Table SI. Databases of lncRNAs-related information.

Focus	Databases	Origins	Usage	Advantages	PubMed ID
LncRNAs biogenesis	ChIPBase v2.0	10 species including Human	ChIPBase v2.0 can be used to illustrate the clinically relevant interactions between TFs and lncRNAs	1. Expanding with 20-fold ChIP-seq datasets 2. Providing comprehensive annotations of transcriptional regulation of lncRNAs and PCGs 3. Integrating a large number of RNA-seq and miRNA-seq data for researchers to explore the co-expression associations of TF-lncRNA and TF-PCG	27924033
LncRNAs structure	RMDB version 2.0	-	RMDB version 2.0 offers interactive graphical browsing of lncRNAs structural, thermodynamic and kinetic data at single-nucleotide resolution	1. Providing a new interface for more information about the probed lncRNAs at single-nucleotide resolution 2. Providing the FORNA applet to display chemical mapping reactivity on lncRNAs secondary structures 3. Simplifying the submission process for researchers	30053264
	DMfold	-	DMfold is a single model to predict the secondary structure with pseudoknots of different lncRNAs	1. Firstly combining the deep learning and IBPMP 2. Using auxiliary sequence to help predict lncRNA secondary structure and using deep learning model to automatically extract RNA features	30886627

				3. Only using the target sequence as the input for greatly simplifying the method 4. Breaking through the restriction of parameter insufficiency in traditional computational methods	
LncRNAs subcellular distribution	LncATLAS	Human	LncATLAS is a comprehensive resource of lncRNA localization in human cells based on RNA-seq datasets	1. Containing 6,768 GENCODE-annotated lncRNAs in 15 cell lines 2. Introducing the relative concentration index as a useful measure of localization 3. Using standard GENCODE annotations to ensure its integration with diverse other genomics datasets	28386015
	LncSLdb	Human, mouse, fruit fly	LncSLdb collects qualitative and quantitative subcellular localization information of lncRNAs by manually curating the literature	1. Containing information about the subcellular location of >11,000 lncRNA transcripts from 9,494 genes 2. Classifying lncRNAs into three basic types and three subtypes with evidence of biological experiments	30219837
	lncLocator 2.0	Human	lncLocator 2.0 is a cell-line-specific subcellular localization predictor for lncRNAs	1. Constructing benchmark datasets of lncRNA subcellular localization for 15 cell lines 2. Predicting lncRNA subcellular localization from sequences by an end-to-end deep model	33630066

				3. Finding potential sequence patterns associated with localization compared with other methods	
	RNALocate v2.0	104 species including Human	RNALocate is the most comprehensive map of the subcellular localization of the transcriptome to date	1. Integrating >213,000 RNA subcellular localization entries at 171 locations across 104 species 2. Providing a tool named iLoc-lncRNA for predicting the subcellular localization of lncRNAs	34551440
LncRNAs comprehensive information	LncRNome	Human	LncRNome integrates biologically oriented datasets and resources on lncRNAs and manual annotations with the aim of providing a one-stop solution for annotation information on lncRNAs	1. Providing information about essential hairpin and quadruplex motifs in lncRNAs 2. Containing experimental and prediction datasets on lncRNA-protein interactions	23846593
	LncRNA2Function	Human	LncRNA2Function is the first ontology-driven web interface for researchers to browse lncRNAs associated with a specific functional term, functional terms associated with a specific lncRNA and functional terms associated with a set of human lncRNA genes	Annotating 9,625 human lncRNAs with enriched functions from RNA-seq datasets of 19 human normal tissues	25707511

	LNCat	Human	LNCat is the first analysis of commonalities and differences of 24 publicly lncRNA annotation resources	1. Reviewing 24 available lncRNA annotation resources referring to >205,000 lncRNAs in >50 tissues and cell lines 2. Characterizing annotation resources from different aspects including exon structure 3. Providing a genome browser of lncRNA structures and downloading combinations of lncRNA annotation 4. Supporting the rapid exploration, comparison and integration of lncRNA annotation resources	26944085
	LncVar	H. sapiens, M. musculus, D. rerio, C. elegans, D. melanogaster,	LncVar is a database on genetic variations associated with lnc genes	1. Systematically integrating TF binding sites and m6A modification sites of lncRNAs, and providing comprehensive effects of SNPs on transcription and modification of lncRNAs 2. Collecting putatively translated ORFs in lncRNAs, and identifying synonymous and non-synonymous SNPs in ORFs 3. Collecting eQTLs of lncRNAs from the literature 4. Identifying lncRNAs in CNV regions as prognostic biomarker candidates of tumors, and predicting lncRNA gene fusion events	27605101

		A. thaliana		from RNA-seq data from cell lines	
	lncRNASNP 2	Human, mouse	LncRNASNP2 offers comprehensive information on SNPs and mutations of lncRNAs, and their influences on lncRNA structure and function	1. Containing 859,534 non-coding variants from COSMIC cancer data in lncRNAs, and their effects on lncRNA structure and function 2. Providing 315,234 cancer mutations in lncRNAs from TCGA and their effects 3. Offering lncRNA expression profiling of 20 cancer types in tumor and adjacent samples 4. Expanding lncRNA-associated diseases 5. Optimizing results on lncRNAs structure change induced by variants 6. Reducing false positives in miRNA and lncRNA interaction results	29077939
	LNCediting	Human, rhesus, mouse, fly	LNCediting provides a useful resource for the functional effects of RNA editing in lncRNAs	1. Containing A-to-I editing sites in lncRNAs across four species 2. Predicting the effects of editing sites on lncRNA secondary structures and lncRNA-miRNA interactions	27651464
	LNCipedia 5	Human	LNCipedia 5 provides a complete set of human lncRNAs with better	1. Providing insights of 1,555 human lncRNAs by curating 2,482 lncRNA publications	30371849

			annotations and official HGNC gene names	2. Using official lncRNAs symbols when available 3. Developing an improved filtering pipeline for a higher quality reference lncRNA gene set	
	LncBook	Human	LncBook is a comprehensive collection of human lncRNAs by multi-omics data integration, functional annotation and disease association	1. Containing 270,044 lncRNAs and 1,867 characterized lncRNAs with 3,762 lncRNA-function associations 2. Integrating abundant multi-omics data from expression, methylation, genome variation and lncRNA-miRNA interaction 3. Identifying 3,772 experimentally validated lncRNA-disease associations and 97,998 lncRNAs that are putatively disease-associated	30715521
	AnnoLnc2	Human, mouse	AnnoLnc2 is the most comprehensive annotation tool for novel lncRNAs in humans and mice in 2020	1. Systematically annotating new lncRNA information, including sequences, structure, expression, regulation, genetic association, and evolution 2. Providing a novel module for assessing lncRNAs subcellular localization based on separated nucleus/cytosol profiling data 3. Updating CLIP-seq dataset, which contains 188 human RBPs and 62 mouse RBPs of	32406920

				lncRNAs 4. Providing a fully functional standalone package for large-scale offline analysis	
	LncSEA	Human	LncSEA focuses on accommodating various available resources of human lncRNAs, and performing annotation and enrichment analyses of lncRNA lists submitted by users	1. Supporting >40,000 reference lncRNA sets across 18 categories, and 66 sub-categories, including >50,000 lncRNAs 2. Collecting ChIP-seq, DNase-seq, ATAC-seq, and H3K27ac ChIP-seq data from hundreds of human cell types to identify a large number of lncRNA sets regulated by upstream TFs and DNA regulatory elements 3. Simultaneously performing lncRNA set enrichment analyses associated with upstream regulators and downstream targets of lncRNAs	33045741
	LncExpDB	Human	LncExpDB provides comprehensive expression profiles of lncRNA genes, exploring their expression features and capacities, identifying featured genes with potentially important functions, and building interactions with protein-coding genes across various biological contexts/conditions	1. Collecting the expression profiles of 101,293 human lncRNA genes derived from 1,977 samples of 337 biological conditions across 9 biological contexts; 2. Estimating lncRNA genes expression reliability and capacities, and identifying 25,191 characterized genes, and further obtaining 28,443,865 lncRNA-mRNA interactions	33045751

	LncRBase V.2	Human, rat, the fruit fly, zebrafish, chicken, cow, C. elegans	LncRBase V.2 is a readily available source of tissue-specific expression of lncRNAs corresponding to multiple species.	1. Providing TF binding sites in the lncRNA promoter region 2. Searching subcellular localization of lncRNAs 3. Predicting possible sORFs within lncRNAs; 4. Integrating target molecules and disease association of lncRNA genes 5. Offering distribution of lncRNAs in multiple tissues across species 6. Hosting ClinicLSNP in LncRBase V.2, which has a comprehensive catalog of lncRNAs originating from 561 RNA-seq datasets of breast, ovarian and cervical cancers	33112702
	NONCODE V6	Humans, mice, plants	NONCODE V6 is a comprehensive database of collection and annotation of ncRNAs, particularly lncRNAs in both animals and plants	1. Increasing the number of lncRNAs since 2017; 2. Updating human lncRNA-disease associations, particularly cancer	33196801
	LincSNP 3.0	Human	LincSNP 3.0 aims specifically to store and annotate disease or phenotype-associated variants and RNA editing in human lncRNAs or their regulatory elements	1. Adding more types of variants, including SNPs and somatic mutations in human lncRNAs 2. Expanding regulatory elements including TF binding sites, enhancers and topologically	33219661

				associated domains 3. Collecting more experimentally supported variant-lncRNA-disease /phenotype associations	
	ColorCells	Human, mouse, rat, zebrafish, yeast, fruit fly	ColorCells is a database of comparative analysis of lncRNAs and mRNAs expression, classification and functions in scRNA-seq data	1. Collecting scRNA-seq datasets of 205 studies from six species, and classifying cell clusters based on lncRNAs, mRNAs and all RNAs 2. Displaying results by principal component analysis and T-distributed stochastic neighbor embedding algorithm in two-and three-dimensional space 3. Performing cell-cell similarities estimation by shared nearest neighbor and Pearson correlation methods 4. Predicting the function of lncRNA and protein-coding genes in each group through gene function enrichment analysis	33313674
	exoRBase 2.0	Human	exoRBase 2.0 remains the only online resource available for exploring biotypes of different long RNA, including mRNAs, lncRNAs and circRNAs, in EVs in human normal and cancerous biofluids	1. Containing 905 exLR-seq data from four types of human biofluids that were collected from healthy individuals and patients with 13 types of cancer or other diseases 2. Obtaining annotation information and expression profiles of 19,643 mRNAs, 15,645	34918744

				lncRNAs and 79,084 circRNAs in EVs 3. Covering the enrichment scores of 11,536 Molecular Signatures Database pathways for each sample generated by ssGSEA on exLR expression profiles	
LncRNAs function prediction	SEEKR	Human, mouse	SEEKR is a method to compare sequence content between any group of lncRNAs to quantify nonlinear homology for predicting protein binding, structure and subcellular localization of lncRNAs via k-mer signature	1. Revealing that evolutionarily unrelated lncRNAs can encode similar function through different spatial arrangements of related sequence motifs 2. Demonstrating that k-mer-based classification is a powerful approach to detect recurrent associations between sequence and function in lncRNAs	30224646

LncRNAs, long non-coding RNAs; ChIP, chromatin immunoprecipitation; seq, sequencing; TFs, transcriptional factors; PCG, protein coding gene; miRNA, microRNA; FORNA, force-directed RNA; IBPMP, improved base pair maximization principle; lnc, long non-coding; SNPs, single nucleotide polymorphisms; ORFs, open reading frames; eQTLs, expression quantitative trait locis; CNV, copy number variation; TCGA, The Cancer Genome Atlas; A, adenoine; I, inosine; HGNC, HUGO Gene Nomenclature Committee; CLIP, cross-linking and immunoprecipitation; RBPs, RNA binding proteins; DNase, deoxyribonuclease; ATAC, Assay for Transposase Accessible Chromatin; H3K27ac, acetylation of histone 3 on lysine (K)-27; sORFs, small open reading frames; ncRNAs, non-coding RNAs; scRNA, single-cell RNA; circRNAs, circular RNAs; EVs, extracellular vesicles; exLR, EV Long RNA; ssGSEA, single sample Gene set Enrichment Analysis.

Table SII. Databases of lncRNAs associated with DNAs.

Databases	Origins	Usage	Advantages	PubMed ID
Triplexator	*H. sapiens, M. musculus, D. rerio, D. melanogaster, C. elegans*	Triplexator can predict whether nuclear lncRNAs have the potential to interact with genomic DNA by forming a triplex	1. Identifying potential TFO in lncRNAs single-strand 2. Identifying suitable TTS in DNA double-stranded nucleotide sequences	22550012
LongTarget	Human, mouse	LongTarget can predict lncRNA DNA-binding motif and binding sites	1. Expanding base-pairing rules including canonical and potential non-canonical rules between lncRNAs and DNA 2. Determining the single best TFO sequences on lncRNA in triplexes 3. Determining the best TFO's TTSs in DNA of triplexes in a genome region 4. Evaluating triplex stability	25262155
Lnc2Meth	Human	Lnc2Meth explores the associations of lncRNAs and DNA methylation in various human disorders	1. Abundant DNA methylation profiles for lncRNAs and protein coding genes (PCGs) in diseases 2. Providing prognostic values for each lncRNA transcript from clinical data 3. Identifying differential methylation patterns for lncRNAs and PCGs with external datasets	29069510

LnChrom	Human, mouse	LnChrom can search lncRNA-chromatin interaction, and help to determine the effects of lncRNA-chromatin interaction on targeted genes	LnChrom provides the first comprehensive map of lncRNA interactions with chromatin, and contains 382,743 interactions involving 2,390 lncRNAs and 34,345 target genes	29788225
Fasim-LongTarget	Human, mouse	Fasim-LongTarget is a novel tool for predicting genome-wide lncRNA/DNA binding	1. It is 200-fold faster than LongTarget 2. It is more accurate than TDF It is feasible for genome-wide prediction of lncRNA/DNA binding	35832611

LncRNAs, long non-coding RNAs.

Table SIII. Databases of lncRNAs interaction with RNAs.

Databases	Origins	Usage	Merits	PMID
Starbase v2.0	CLIP-seq of 14 kinds of tumor samples	Starbase v2.0 exhibited the most comprehensive miRNA-lncRNA interaction networks in 2014	1. Identifying RNA-RNA interaction networks from 108 CLIP-seq datasets generated by 37 independent studies systematically 2. Providing the most comprehensive CLIP-seq experimentally supported miRNA-lncRNA interaction networks in 2014 3. Developing ceRNAFunction web servers to predict the function of lncRNAs from miRNA-mediated regulatory networks	24297251
RISE	human, mouse, yeast	RISE is a comprehensive database of RNA interactome from transcriptome-wide and targeted sequencing-based experiments, and several primary databases and publications	1. Containing 328,811 RRIs, while ＞50% of the RRIs are among mRNA and lncRNAs 2. Providing details of the interacting sites 3. Offering extensive annotations for each RRI	29040625
LnCeVar	Human	LnCeVar documents personalized lncRNA–variation–ceRNA events from published literature and high-throughput identification from individual genomics data.	1. Containing >2,000 experimentally supported circulating, drug-resistant and prognosis-related lncRNA biomarkers 2. Integrating expression profiles of ceRNAs and individual clinical information from >10,000 samples from TCGA, COSMIC and 1000 Genomes Project 3. Developing several tools for users to perform following analyses: Searching ceRNAs; identifying dysregulated cancer hallmarks of variation–ceRNA events; performing COX	31617563

			regression analyses and producing survival curves for variation–ceRNA events	
DIANA-LncBase v3	Human, mouse	DIANA-LncBase v3.0 is a reference repository with experimentally supported miRNA targets on non-coding transcripts from publications and analysis of >300 high-throughput data	1. It is the first database to employ a CLIP-seq-guided algorithm and microCLIP framework to analyze 236 AGO-CLIP-Seq libraries and catalog 370,000 miRNA binding events 2. Providing 500,000 entries, corresponding to 240,000 unique tissue and cell type specific miRNA-lncRNA pairs 3. Containing >67,000 experimentally supported target sites and lncRNA expression profiles in different cellular compartments	31732741
RNAInter v4.0	28 species including human	RNAInter v4.0 provides a comprehensive RNA interactome platform from manually text mining, database integration and computational prediction	1. Achieving a redefined confidence scoring system 2. Exhibiting an updated interface that enables fast retrieval and browsing of RRIs 3. Containing >4,700,000 entries collected from the literature and six databases	34718726
LncACTdb 3.0	25 species including human	LncACTdb 3.0 is a comprehensive database of experimentally supported interactions among ceRNA and corresponding personalized networks thus contributing to precision medicine	1. Collecting 5,669 experimentally validated ceRNA interactions in 537 diseases/phenotypes 2. Containing personalized ceRNA interactions and networks for 16,228 patients from 62 datasets in TCGA and GEO 3. Containg subcellular and extracellular vesicle locations of ceRNA 4. Describing >10,000 experimentally supported lncRNA biomarkers associated with tumor diagnosis and therapy	34850125

			5. Containing lncRNA/mRNA/miRNA expression profiles with clinical and pathological information of thousands of patients with cancer	

LncRNAs, long non-coding RNAs; CLIP, cross-linking and immunoprecipitation; seq, sequencing; miRNA, microRNA; RRIs, RNA-RNA interactions; mRNA, messenger RNA; ceRNA, competing endogenous RNA; TCGA, The Cancer Genome Atlas; AGO, argonaute; GEO, Gene Expression Omnibus.

Table SIV. Databases of lncRNAs encoding potential or confirmed interaction with proteins.

Focus	Databases	Origins	Usage	Merits	PubMed ID
Encoding potential of lncRNAs	CPPred	Human	CPPred can provide the isoelectric point, hydropathicity and stability of a predicted peptide translated by RNA transcripts to distinguish coding RNAs and ncRNAs, particularly small coding RNAs and small ncRNAs.	1. Plotting ORF length, Fickett score and Hexamer score in three dimensions for 33,360 coding RNAs, 24,163 ncRNAs and 1,016 small RNA transcripts 2. Firstly proposing that composition, transition and distribution of predicted peptides could be used to distinguish coding RNAs from ncRNAs in eukaryotes 3. Only using FASTA format sequence files as necessary inputs	30753596
	CNIT	11 animal species including humans, and 27 plant species	CNIT can predict protein-coding or non-coding sequences generated by high-throughput sequencing data, and analyze sequence features across species.	1. It is faster and more accurate than the former CNCI 2. It is applicable for species that lack a whole-genome sequence or annotated information 3. It is freely available as an online or offline tool	31147700
	LncPep	39 species including human	LncPep can be used to evaluate the coding potential of lncRNAs, and provide a convenient interface to	1. Containg 10,580,228 peptides that were predicted to be translated from 883,804 lncRNAs 2. Predicting the coding potential of	35141219

			browse information of each lncRNA and its translated peptide	novel lncRNAs transcripts based on user-customized sequences 3. Providing a download option for full datasets of LncPep for users' further exploration	
Interaction with proteins	catRAPID omics v2.0	*Homo sapiens, Mus musculus, Rattus norvegicus, Xenopus tropicalis, Danio rerio, Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae*	catRAPID omics v2.0 web server is a powerful tool for the characterization and classification of RNA-protein interactions	1. Listing integration of the sequence fragmentation approach to identify binding regions in proteins and lncRNAs 2. Providing calculation of lncRNA-binding abilities of proteins 3. Exploiting orthology associations to explore the evolutionary conservation of predicted lncRNA-protein interactions 4. For the 500 most interacting lncRNA-protein pairs, the software calculates lncRNA-protein binding sites to increase the power and resolution of their high-throughput method	34086933
	SFPEL-LPI	-	SFPEL-LPI can predict uncleared interactions between lncRNAs and proteins	Integrating sequence-derived features and similarities with a feature projection ensemble learning frame	30533006

LncRNAs, long non-coding RNAs; ncRNAs, non-coding RNAs; ORF, open reading frame; CNCI, Coding-Non-Coding Index.