Long Non-coding RNAs: Mechanisms, Experimental, and Computational Approaches in Identification, Characterization, and Their Biomarker Potential in Cancer

Anshika Chowdhary, Venkata Satagopam* and Reinhard Schneider

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Long non-coding RNAs are diverse class of non-coding RNA molecules >200 base pairs of length having various functions like gene regulation, dosage compensation, epigenetic regulation. Dysregulation and genomic variations of several lncRNAs have been implicated in several diseases. Their tissue and developmental specific expression are contributing factors for them to be viable indicators of physiological states of the cells. Here we present an comprehensive review the molecular mechanisms and functions, state of the art experimental and computational pipelines and challenges involved in the identification and functional annotation of lncRNAs and their prospects as biomarkers. We also illustrate the application of co-expression networks on the TCGA-LIHC dataset for putative functional predictions of lncRNAs having a therapeutic potential in Hepatocellular carcinoma (HCC).

INTRODUCTION

Advancement in Next Generation Sequencing (NGS) technologies and genome wide analysis of gene expression have revealed at least 80% of the human genome is active (Palazzo and Lee, 2015). However, only up to 1.5% of the genome is translated to protein which implicate RNAs to have more diverse roles than an intermediate component as templates in the genetic flow of information from gene to protein. They are categorized into mRNAs which are translated into proteins and non-coding RNAs (ncRNAs) which have little or no coding potential but are involved in transcriptional regulatory mechanisms.

The evolutionary development of an organism is associated with the increase in complexity of the regulatory potential of these ncRNAs which constitute the majority of the transcriptome. Non-coding RNAs are further categorised as short ncRNAs which include microRNAs (miRNAs), small RNA (sRNA), piwi-interacting RNAs (piRNAs), siRNAs, and long non-coding RNAs (lncRNAs) consisting long intergenic non-coding RNAs (lincRNAs), circular RNAs (circRNAs), and competitive endogenous RNAs (CeRNAs) (Hombach and Kretz, 2016). These RNAs have known to have functions involved in cellular functions like mRNA translation, alternative splicing events, RNA editing and also regulatory mechanisms like RNA silencing involving miRNA and mRNA interference via siRNA (Mattick and Makunin, 2006). LncRNAs have emerged as a latest class of RNA molecules which are more diverse than short ncRNAs having complex gene regulatory

functions in the cells. In this article we present and review the various biological characteristics and mechanisms of lncRNAs in transcriptional regulation and the latest development in experimental and computational methods for their identification, annotation and putative function prediction.

There are more than 30,000 lncRNAs in humans available in the GENCODE (Harrow et al., 2012), and more and more new lncRNAs are being discovered overtime. Long non-coding RNAs are typically longer than 200 nucleotides of length and sometimes have similar features to that of protein-coding genes, such as a 5' cap, exons and poly A tail and are spliced post-transcriptionally, but don't possess functional open reading frames and cannot be translated to functional proteins. (Fang and Fullwood, 2016). Their varied molecular properties enable them to function in various methods regulating gene expression at various stages of cellular development (Hanahan and Weinberg, 2000).

LncRNAs are also not stable in comparison to mRNAs, localized mainly across the nucleus and cytoplasm and also not conserved across species, transcribed mostly by RNA polymerase II and exhibit tissue specific expression. However, high conservation patterns have been observed in the exonic regions and promoters regions of the lncRNA. Recently, it has been discovered that some lncRNAs can in fact translate to small peptide chains which could have significant biological functions (Hubé and Francastel, 2018; Li and Liu, 2019).

One way to classify lncRNAs is based on the genomic locations from where they are transcribed relative to protein coding genomic regions: (1) lincRNAs: long intergenic non-coding RNAs which are transcribed from the intergenic regions between the protein coding genes; (2) Sense lncRNAs: transcribed from the sense strand of the protein coding genes and may overlap with a part or the entire sequence of a protein coding gene; (3) Antisense lncRNAs: transcribed from the antisense strand of the protein coding genes which may overlap of exons, only from the intronic region and overlapping the entire gene in the antisense strand. (Ma et al., 2013); (4) Intronic lncRNAs: transcribed from the intronic regions between the exomes of a gene. (5) Bidirectional lncRNAs: transcribed from both sense and antisense directions of TSS (Hanahan and Weinberg, 2000; He et al., 2014, 2017).

FUNCTIONS AND MECHANISMS OF LONG NON-CODING RNAS

The elucidation of the mechanisms of long non-coding RNAs is mostly based on empirical evidence of the subcellular localization, developmental stage of the cell and tissue specific expression. The function of lncRNAs can be stratified into four types of molecular mechanisms described and illustrated (**Figure 1**; Zanella, 2021) below.

Signals
Transcriptional regulation aided by lncRNA where they function as signals are brought by various factors like developmental stages, organismal stress, re-programming of cells and state of the cell at a particular space and time in response to the environment
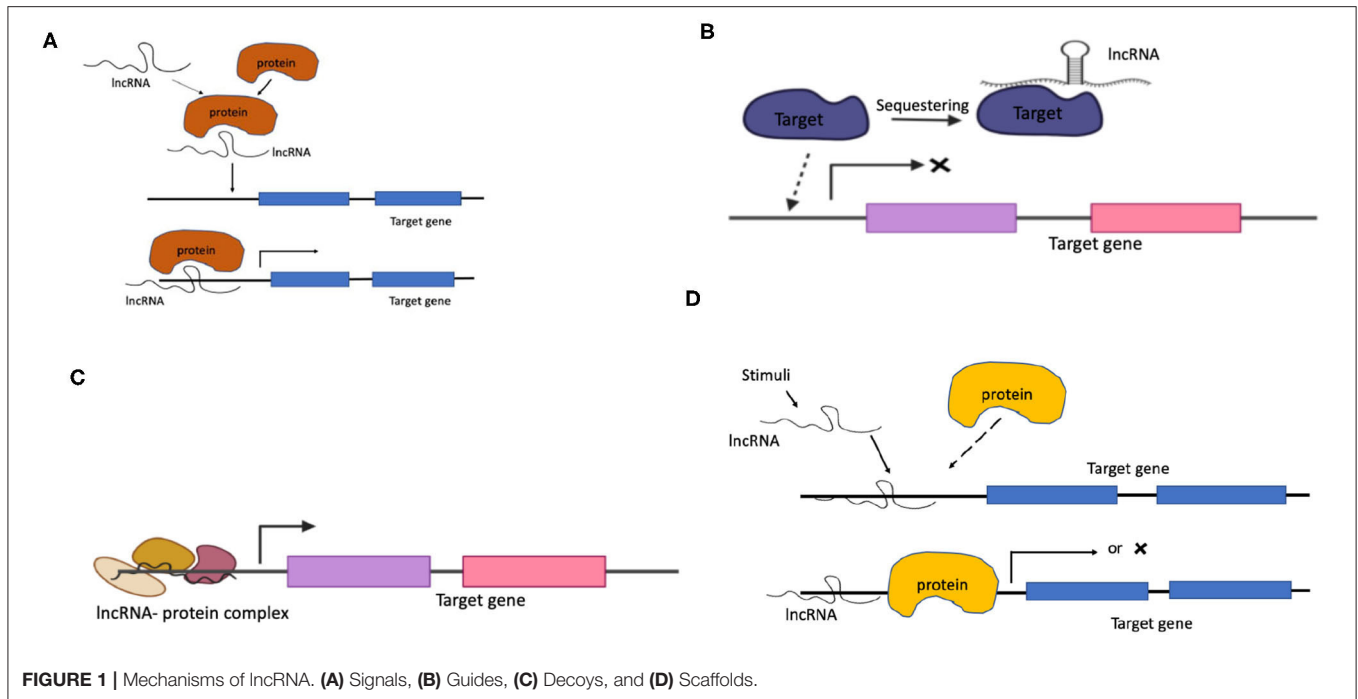
and their expression could be a phenotypical indicator of these states (Wang and Chang, 2011). A prominent example is the chromatin regulation for dosage compensation in females in X-chromosome inactivation (XCI) (Engreitz et al., 2013; Wasko et al., 2019). The mechanism includes expression of XIST lncRNA from one of the X chromosome which coats itself leading to its silencing, which is also aided by the accumulation of the lncRNA Jpx. The antisense transcript of XIST, TSIX represses the activity of XIST in the other chromosome rendering it to be active (Starmer and Magnuson, 2009; Wang and Chang, 2011; Carmona et al., 2018). Another example of epigenetic re-programming that takes place in plants mediated by lncRNAs is to switch between vegetative to reproductive state. In *Arabidopsis thaliana* with the decrease in temperature for an extended period of time during winter COOLAIR is expressed and accumulated in large amounts which represses the expression of the FLOWERING LOCUS C (FLC). This is gene mediated by the PRC2 complex which when expressed normally in winter stops flowering in the plant. So, gradually upon the approach of spring and warmer temperatures COOLAIR enables vernalization of plants (Swiezewski et al., 2009; Tian et al., 2010; Heo and Sung, 2011; Wang and Chang, 2011).

Guides
As guides lncRNAs bind to proteins and direct them to specific sites, also leading to expression or silencing of the target genomic regions. This essentially involves recruitment of chromatin modifying enzymes which alter the chromatin state with the formation of complex structures with RNA-RNA, RNA-DNA, RNA-DNA-effector proteins. For instance, XIST transcription has also known to be induced by recruiting the Polycomb Repressive Complex 2 (PRC2) by RepA RNA. Additionally XIST also interacts with a matrix protein hnRNP U for its accumulation at the chromosome (Wang and Chang, 2011). Some other examples of lncRNAs acting as signals and guides include COLDAIR, HOTTIP, HOTAIR, ROR and some PRC2-bound RNAs (Rinn et al., 2007; Loewer et al., 2010; Wang et al., 2011; Kim et al., 2017).

Decoys
LncRNAs can also regulate transcription by acting as endogenous target mimics (eTMs) where they bind to intermediary regulatory proteins, RNA, DNA molecules and sequester them away from their respective target site. These otherwise known as competitive endogenous RNA (ceRNA) act as *sponges* generating a "sponge effect" by base pairing with target molecules which include transcription factors, miRNAs, chromatin modifiers (Wang and Chang, 2011) among others at their active sites and render them to be unavailable for interaction for their target molecules. An example of such activity is that of the lncRNA transcribed at the minor promoter of the DHFR gene which pairs and forms a complex with the DNA at the promoter region of the same gene. The complex inhibits formation of the preinitiation complex and also interacts with transcription factor IIB (TFIIB) which was also further confirmed by siRNA knockdown of the lncRNA (Martianov et al., 2007; Wang and Chang, 2011). MALAT1 (Tripathi et al., 2010), TERRA (Redon et al., 2010),

FIGURE 1 | Mechanisms of lncRNA. **(A)** Signals, **(B)** Guides, **(C)** Decoys, and **(D)** Scaffolds.

Gas5 (Kino et al., 2010) are also examples that exhibit the 'sponge'/sequestering mechanism. ceRNA mechanism has been extensively studied with several computational algorithms and repositories also being developed in order to identify and store potential and experimentally verified targets of lncRNA (listed in **Table 1**). However, verification of their mechanism have to be contended with transcriptional levels of miRNA and lncRNA to be sufficient enough for them to function as competitive endogenous RNAs (Denzler et al., 2014, 2016; Zhang et al., 2019).

Scaffolds

LncRNAs serve as structural supports where other effector proteins and DNA/RNA molecules bind to form a functional complex and are then directed to appropriate localization of the complex for its function. Gene repression by HOTAIR forming a complex with the polycomb complex PRC2 for methylation at H3K27 (Rinn et al., 2007; Wang and Chang, 2011) and also forming a complex with LSD1, CoREST and REST (Wang and Chang, 2011) exhibits this mechanism. TERC also assembles the telomerase complex and mediates reverse transcriptase activity by binding with telomere targeting proteins (Balas and Johnson, 2018). The lncRNAs ANRIL (Yap et al., 2010; Kotake et al., 2011), SRP(Signal Recognition component), LINP1(LncRNA In Nonhomologous End Joining Pathway 1) (Sakthianandeswaren et al., 2016) are also found to have similar mechanisms.

IDENTIFICATION AND ANNOTATION

Experimental Approaches

Widely used experimental approaches to identify and annotate lncRNAs include Microarray, RNAseq, SAGE, CAGE among others with customized adaptations to identify and annotate lncRNAs based on their molecular characteristics as described in the following sections and listed in **Table 2**.

Adaptations in Microarray Technology

Probesets in conventional microarray platforms do not have lncRNAs annotations and not suitable for identifying and measuring lncRNA levels. Some of the mRNAs from these previous microarrays that have been correctly identified as lncRNAs have been re-annotated and their expression levels have been re-analyzed accordingly (Michelhaugh et al., 2011; Ma et al., 2012). ArrayStar Human LncRNA microarrays (V4.0) has been designed to profile both lncRNA and mRNA on the same array with 40,173 lncRNAs with 7,506 gold standard lncRNAs, 20,730 mRNAs among 60,903 distinct probes (Shi and Shang, 2016). As the expression of lncRNAs indicates the relative physiological state of a cell, differential expression between samples at different conditions can provide us information to understand the regulatory lncRNAs at these conditions. (Zhang et al., 2017) identified novel circulating lncRNAs: TINCR, CCAT2, AOC4P, BANCR, and LINC00857 which are differentially expressed in gastric cancer patients and be detected from the plasma of patients and hence function as biomarkers. Similarly, it was found that the lncRNA ENST00000551152 was upregulated and the lncRNA TCO.NS_00001368 was downregulated in cervical cancer cell lines (Huang et al., 2018) in a study by Huang et. al using Agilent DNA microarray. Whole-genome tiling arrays are used for the sequenced regions which are not annotated for lncRNA isolation and identification. (Lund et al., 2014) used this in their experimental design where they used tiled probes from chr8: 127,640,000–129,120,000 at locus 8q24 to analyze prostate tissue from prostate cancer patients.

TABLE 1 | Databases and computational pipelines predicting lncRNAs functioning as ceRNAs.

Databases/Computational pipeline	Description	References
DIANA-LncBase v3	Database dedicated to cataloging miRNA and lncRNA interactions, includes ceRNABase	Karagkouni et al., 2020
StarBase v2.0	RNA-RNA and protein-RNA interactions from CLIP-Seq experiments predicting ceRNA function	Li et al., 2014
spongeScan	predicts miRNA target sites in lncRNAs	Furió-Tarí et al., 2016
lnCeDB	stores lncRNAs acting as ceRNAs with targets from StarBase and TargetScan Grimson et al., 2007	Das et al., 2014
LncCeRBase	lncRNA-miRNA-mRNA interactions collected from literature	Pian et al., 2018
Linc2GO	predicts linc RNAs functions using miRNA and mRNA interactions based ceRNA hypothesis	Liu et al., 2013

TABLE 2 | Experimental approaches in lncRNA profiling.

Experimental approaches	Features
RNA-seq	Identifies on novel lncRNA transcripts
Microarray	Reannotations of existing microarrays
	Arrays specifically designed for lncRNAs
Tiling arrays	Ability to profile transcriptome for specific regions(whole) in the genome.
SAGE	Accurate quantification and novel transcript identification
CAGE	Identification of transcription start points
PARE, degradome-seq	Used in RNA degradome analysis
GRO-seq	Measures nascent RNA regulating gene transcription
RIP, CLIP	LncRNA-protein interaction identification
TIF-seq	Identification of isoforms of lncRNA
Selective 2'-hydroxyl acylation by primer extension (SHAPE)	LncRNA structure prediction
PARS	LncRNA structure prediction *in vitro*
FragSeq	Transcript structure prediction from RNA fragments
nextPARS	Adaptation of PARS to Illumina technology

RNA-Seq Technologies

RNA-seq is the most prevalent technique used to identify and annotate novel long non-coding transcripts that are less abundant including the isoforms of lncRNAs. RNA-seq offers a broad spectrum of transcript identification with novel transcripts detection and *de novo* assembly as probes are not required in order to hybridize and capture transcripts from samples. Modifications in the RNA-seq pipeline facilitate identification of specific type of lncRNAs, for instance strand-specific RNA-seq allows labeling of *origin of strand* information on the transcripts which allows sense/antisense lncRNA segregation and identification (Mills et al., 2013; Liu et al., 2019).

Wang et al. identified 2895 novel lncRNA in endometrial tissue of pigs; of which 301 were differentially expressed and functionally annotated to be involved in several biological pathways including immune system process and other cellular process of which TCONS_01729386 and TCONS_01325501 have a major functions in embryo pre-implantation (Liu et al., 2017). Functional attributes of lncRNA are validated with qRT-PCR

experimental pipelines in which siRNA, GAPmers are designed to knockdown the lncRNA and the resulting change in gene expression is analyzed to identify its effector genes/molecules. However, in order for *in vitro* studies to correlate with vivo studies several contributing factors involved in the knockdown of lncRNA and its effect on resulting varying gene expression need to be considered. Features of the lncRNA to consider while design of the knockdown strategy is the sub-cellular localization of the lncRNA, along with the developmental stage of the cells. Lennox et al. were able to decipher that nuclear lncRNAs were knocked down at higher levels using antisense strands and cytoplasmic lncRNAs were better knocked down using RNAi (Lennox and Behlke, 2016). In a recent study by Nicola Amod et al. a MALAT1-targeting 16mer LNA gapmeR g#5 showed significant anti-tumor activity in humanized murine model. Inference from transcriptome analysis showed proteasome expression was repressed by g#5 and was instead enriched increased *in vivo* in MALAT1 murine model patients (Amodio et al., 2018). RNA CaptureSeq (Mercer et al., 2011), another derivative of RNA-seq involves tiling arrays prepared for specific target regions of the genome. cDNAs against these regions are hybridized and sequenced. This method supports the identification of novel unannotated lncRNAs along with high fold coverage.

SAGE, CAGE

Serial Analysis of Gene Expression (SAGE) (Velculescu et al., 1995) and Cap Analysis of Gene Expression (CAGE) are based on short sequences tags which are complementary to a given RNA of interest (Kashi et al., 2016). In SAGE these cDNA tags are biotinylated, captured on streptavidin beads (Wang and Chekanova, 2019). They are further ligated and later PCR amplified followed by concatenation and sequencing by mapping to reference genes. This method like RNA-seq facilitates discovery to novel transcripts and enables accurate measurement of expression levels of lncRNAs but has a drawback of small cDNA sequences mapping to multiple genes in the reference genome. Gibb et al. analyzed 272 SAGE libraries normal(26) and cancer(19) tissues from human which elucidated the tissue specific and aberrant expression lncRNAs in cancer tissues implicating them in disease development (Gibb et al., 2011). In a study by Jia et al. (2018) SAGE datasets of OPL(Oral premalignant lesions) from GEO were analyzed to identify 10 differentially expressed lncRNAs among with the lncRNA NEAT1 was the highly expressed in OPL. NEAT1 has been

also implicated in lung cancer metastasis and hepatocellular carcinoma (Dong et al., 2018).

Cap analysis gene expression (CAGE), was a development upon SAGE to over come its drawbacks where cDNA tags can be generated from the 5′ end of the RNA of interest. The cap structure of the transcripts are biotinylated in the CAP-trapper method followed by cDNA tag generation, cleaving by restriction enzymes, PCR, ligation and cloning of tags and mapping to reference genome (Shiraki et al., 2003). CAGE allows the expression analyzes at promoter regions but is restricted only to capped RNAs. CAGE method has better throughput with the use of sequence tags and is also cheap in comparison to cDNA library (Shiraki et al., 2003). Hon et al. (2017) collated 27,919 human lncRNAs from 1,829 datasets from CAGE and other methods in the FANTOM5 project. HeliScopeCAGE (Kanamori-Katayama et al., 2011) nanoCAGE (Poulain et al., 2017) CAGEscan (Bertin et al., 2017), DeepCAGE (Valen et al., 2009) are also protocols based on the CAGE technology for profiling the mammalian transcriptome.

Other Approaches

Parallel analysis of RNA-ends (PARE) (German et al., 2008), genome-wide mapping of uncapped transcripts (GMUCT) (Gregory et al., 2008), degradome-seq are among other techniques developed to map transcripts that are not stable and get degraded i.e., they act as templates for other non-coding RNAs like miRNA. RNA-seq measures transcripts at equilibrium conditions where as on the other hand Gro-seq (Global run-on sequencing) is able to sequence nascent RNA. This has revealed genome wide view of the transcripts by measuring half life of transcripts at various time points. RNA-seq and GRO-seq analyzes have revealed that divergent transcription occurs at the promoter regions of protein-coding genes (Kashi et al., 2016). 5′-bromo-uridine immunoprecipitation chase—deep sequencing analysis (bric-seq) method involves labeling of transcripts with 5′-bromo-uridine (BrU) which are isolated at sequential time intervals and recovered by immunopurification followed by RT-qPCR (Tani et al., 2012; Kashi et al., 2016). TIF-seq, an approach developed by Pelechano et al. (2013), jointly sequences both 5' and 3' ends of RNA molecules enabling characterization isoform heterogeneity of RNA molecules.

Other than perturbation by silencing of lncRNAs by RNA interference as mentioned in above section, functional characterization of lncRNA also involves methods like RNA centric purification methods when the RNA is pulled down exogenously based on *in vitro* affinity capture methods or endogenously under native or ultraviolet (UV) cross-linking conditions (Cipriano and Ballarino, 2018). On the other hand protein centric purification involves immunoprecipitation of lncRNAs and their target proteins with specific antibodies. RNA immunoprecipitation (RIP) is used to functionally characterize the lncRNA by purifying RNAs associated with target proteins. Cross-linking immunoprecipitation (CLIP), combination of CLIP with high-throughput sequencing (HITS-CLIP or CLIP-seq) and Photo Activatable Ribonucleotide-enhanced (PAR-CLIP) (Spitzer et al., 2014) been developed to analyze interactions of RNA binding proteins but these methods carry disadvantages

like loss of cDNAs and de-crosslinking along with being expensive (Barra and Leucci, 2017). Chromatin isolation by RNA purification (ChIRP) has been used to identify lncRNAs and their interactions with chromatin during gene regulation (Chu et al., 2011; Kashi et al., 2016). Further more, techniques have been developed to probe the RNA structures, such as Selective 2′ -hydroxyl acylation by primer extension (SHAPE) [67], parallel analysis of RNA structure (PARS) (Kertesz et al., 2010) and FragSeq (Underwood et al., 2010) which can provide an extensive evidence on mode of action and interactions with other regulatory molecules (Guo et al., 2016). More recently, Saus et al. described nextPARS an adaptation to PARS technique on the Illumina's sequencing technology where parallel execution of highly specific enzymatic digestion of single an double stranded genomic regions make the "capable of tagging both all the bases in single and double-stranded conformation at a genome-wide scale" making it cost effective with better throughput (Saus et al., 2018). CRISPRlnc, containing manually curated and validated 2184 CRISPR/Cas9 sgRNAs for 335 lncRNAs from different species, (Chen et al., 2019) was developed by Chen et al. which would further help design CRISPR/Cas9 experiments to investigate lncRNAs functions.

Computational Approaches

Novel Computational tools and pipelines are quintessential in combination with novel experimental techniques to identify putative transcripts as lncRNAs and further elucidate their functional roles involving interactions with other DNA, RNA and proteins. Computational pipelines to process NGS data are modified for the annotation of putative lncRNAs from novel transcripts. For the genome wide identification of lncRNA transcripts from data sets generated by the most widely RNA-seq techniques for novel lncRNA identification typically involves the following steps: alignment of reads from the experiment to the target regions in reference genome. This is followed by transcripts assembly and isoform identification and scoring the transcripts for protein coding potential (Coding Potential Calculator) (Jalali et al., 2015) and also include attributes like presence of open reading frames, poly-A tails and exonic regions and strand information into consideration. Standard programs like HISAT2, (Trapnell et al., 2009), STAR (Dobin and Gingeras, 2015) are used for mapping and StringTie (Ghosh and Chan, 2016), Scripture (Schoenbeck, 2016) for assembly. After transcripts of length >200 bp are filtered out, other types of transcripts such as tRNA, rRNA, snoRNA, miRNA, siRNA etc are searched in different databases and removed. Following this, based on their homology scores using programs like BLAST, BLAT the candidate lncRNAs are annotated with information from lncRNA databases. Sequence alignment and similarity search methods such as BLASTX and HMMER3 (Eddy, 2009) search against data repositories like UniProt, PDB and filter RNA transcripts which have similar homologous domains and can be translated to proteins (Gish and States, 1993; Eddy, 2011; Jalali et al., 2015). On comparing the performance of various alignment methods (Zheng et al., 2019) Kallisto or Salmon in combination with full transcriptome annotation performed best

for lncRNA detection on both un-stranded and stranded RNA-Seq datasets.

ORF is also among the features which help categorization of novel transcripts as lncRNAs; for example ORF length predicted by EMBOSS tools (Itaya et al., 2013) (getORF). ORFs of length greater than 100 codons categorised as mRNAs are filtered out as coding transcripts but it is not a definite threshold with certain exceptions like XIST, H19 among others which having ORFs longer than 100 amino acids (Dinger et al., 2008; Jalali et al., 2015).

Another approach is use of machine learning based tools developed on SVM, logistic regression models use sequence features to compute the protein coding potential which predict the transcript to be a lncRNA/mRNA. ORF, conservation of the exonic regions of the transcript, nucleotide composition, sequence motif and codon usage are inclusive feature vectors from the transcript sequences to train the models. In order to compute transcript's coding potential two methods have been developed CPC (Coding Potential Calculator) (Altschul et al., 1997; Kong et al., 2007; Ma et al., 2012) based on SVM models with sequence features and the comparative genomics features and ii) A later faster version CPC2 that can be for novel transcripts of organism which have improper genome assembly and poorly annotated (Kang et al., 2017). CONC (for coding and non coding) (Liu et al., 2006) also trains SVM models based on a comprehensive set of RNA features like the peptide length and composition, secondary structure, compositional entropy among others to classify transcripts as lncRNAs and mRNAs. Lu et al. have further integrated quantitative properties like a GC content, conservation patterns, level of expression which is lower of lncRNAs in comparision to mRNAs to predict lncRNAs in *C. elegans* in their machine learning model (Lu et al., 2011; Ma et al., 2012). The pipeline employed by Sun et al. lncRScan-SVM (Sun et al., 2015), which after a standard processing of RNASeq transcripts identifies transcripts as lncRNAs by a SVM model trained on GTF positive and negative samples. iSeeRNA is also a similar tool that identifies putative lincRNAs by on SVM based classifier (Sun et al., 2013). COME, a coding potential calculator, developed by Hu et al. (2017) integrated multiple features from both sequences and experiments like poly(A) enrichment, methylation taken from RNA-seq data sets had more accuracy over transcripts of different lengths. In the COME method, an index for the whole genome splitting it into bins of 100-nucleotide(nt) on which the feature vectors were generated and subsequently a balanced random forest (BRF) was trained.

Attempts to functionally characterize novel lncRNAs by computational methods have been challenging. In the case of protein-coding genes a putative function is assigned to transcripts based on their similarity with already characterized proteins (de Hoon et al., 2015); as they have highly conserved regions across species which is not the same with lncRNAs. Their tissue specificity and low abundance along with varied mechanisms involved with various other biological molecules further add to the complexity of modeling their functionality *in-silico*.

Co-expression Evidence Analysis and Network Inference

Data analysis of microarrays and tiling experiments include identification of differential expressed transcripts followed by network analysis based on co-expression patterns. To infer the putative function of a lncRNA 'guilt by association' algorithm has been developed based on the co-expression patterns of lncRNA and protein coding genes (PCGS) which suggest their functional relatedness and regulatory relationships. The tissue and condition specific expression, subcellular localization are distinctive attributes of lncRNA expression which are combined with differential expression to infer putative functions and target proteins interactions of the lncRNA and their role in disease development (Li et al., 2016; Gao et al., 2019a). The correlation scores between expression profiles of lncRNAs and PCGs at a given condition/tissue/time series are calculated which represents a network by a transformed correlation-adjacency matrix. From these networks, clusters of co-expressed lncRNAs and mRNAs are identified. The functional regulation of lncRNAs are annotated based on the functional enrichment of the PCGs in the clusters with which it is co-expressed.

Co-lncRNA is one such tool/database developed by Wu et al. (2016) where they were able to analyze lncRNA-mRNA co-expression patterns, consistent with previous established related lncRNA-mRNAs like HOTAIR, BRCA2, MMP9 and MMP11 and also novel lncRNA RP11-118E18 validated by TANRIC. Such network based clustering approaches have also been further extended to include other non-coding RNAs and regulatory proteins like miRNAs to predict more specific mechanisms like cis-regulatory relationships where whole transcriptomic data is analyzed (Signal et al., 2016).

Several studies have been done to understand the pathogenesis of complex diseases from available data of lncRNA and their interacting proteins (Sumathipala et al., 2019). The approaches consist of Machine learning (ML) based models trained over expression profiles to extract patterns from which lncRNA functionality and disease associations are predicted, random walk based models on networks representing the similar expression patterns or a combination of both. (Chen and Yan, 2013) included disease information into identify lncRNA disease associations from lncRNA expression levels by developing a semi-supervised learning model Laplacian Regularized Least Squares for LncRNA Disease Association (LRLSLDA).

Chen et al. further developed novel lncRNA functional similarity calculation models (LNCSIM) by associating the semantic similarity between lncRNA and disease groups (Chen and Yan, 2013; Chen, 2015a,b). Guo et al. (2019) developed LDASR to identify lncRNA-disease associations where Guassian profile similarities and neural network for dimensional reduction and finally rotating forests were used to predict disease associations (Guo et al., 2019). DislncRF also uses random forest models trained over lncRNA-disease associated protein coding genes in order to score the association of lncRNA for a particular disease (Pan et al., 2019). Liao et al. developed a method called GrwLDA which is based on global network

random walk model in order to predict lncRNA and their associated diseases (Gu et al., 2017). Xuan et al. (2019) also recently proposed a tool graph convolutional network and convolutional neural network (GCNLDA) to explore network and come up with lncRNA-disease candidate pairs. Bipartite Network inference (LPBNI), a computational pipeline developed by Ge et al. (2016) used two-step propagation in the bipartite network to rank target proteins for lncRNAs; BPLLDA developed to predict lncRNA-disease links from a network of heterogenous lncRNAs and associated diseases based on their node interaction paths (Xiao et al., 2018). TPGLDA also had been developed to predict lncRNA-disease from lncRNA-disease-gene tripartite graph constructed base on was developed by Ding et al. (2018) where they could predict lncRNAs like GAS5, UCA1, implicated in lung, hepatocellular, ovarian cancer (Ding et al., 2018). The above mentioned tools are all based on network propagation and inference. Recently, a similar network diffusion algorithm called LION was developed to infer key candidate lncRNAs (Sumathipala et al., 2019) by Sumathipala et al. with better prediction results for cardiovascular diseases and cancer. Another recent approach IDHI-MIRW by Integrating Diverse Heterogeneous Information(IDHI) with positive pointwise Mutual Information and Random Walk(MIRW) was also proposed by Fan et al. (2019) which integrates lncRNA-miRNA/protein and expression profiles along with disease ontology information.

Conservation and Structure Prediction

Although the conservation scores of lncRNA molecules are lower than mRNA, when used within awareness of biological context including information about potential interactions with other RNA, DNA, proteins, can decipher evidences to categorise novel transcripts to lncRNAs. Algorithms like BLAST (Altschul et al., 1990), ClustalW (Thompson et al., 2002), MAFFT (Katoh et al., 2009), ConSurf (Glaser et al., 2003), MUSCLE (Edgar, 2004) among others perform multiple sequence alignment. Furthermore tools like RNAz 2.0 (Gruber et al., 2010), Evofold (Pedersen et al., 2006) can predict conserved RNA structures from multiple sequence alignment. RNAstructure (Reuter and Mathews, 2010), GTFold (Swenson et al., 2012), CentroidFold (Sato et al., 2009), RNAfold (Denman, 1993), Mfold (Zuker, 2003), CentroidHomfold-LAST (Hamada et al., 2011), and Seqfold (Ouyang et al., 2013), FARNA (Alam et al., 2017), iFoldRNA (Sharma et al., 2008) are among the tools to predict RNA secondary and tertiary structures, respectively from primary sequence. The RNA-RNA interaction prediction methods mainly employ alignment algorithms, comparative (homology) methods and *in silico* energy calculations (Umu and Gardner, 2017). Minimum Free Energy based methods are based on computation of the minimum free energy of the RNA-RNA molecules taking the inter- and/or intra molecular base-pairing into account. On the other hand, as perceivable, alignment and homology based methods include algorithms using tools for multiple sequence alignment and seed match-extension.

IntaRNA (Mann et al., 2017), RNAhybrid (Krüger and Rehmsmeier, 2006), Pairfold (Andronescu et al., 2003), RNAplex (Tafer and Hofacker, 2008), RIsearch (Wenzel et al., 2012), RIblast (Fukunaga and Hamada, 2017), Bindigo (Hodas and

Aalberts, 2004), and GUUGle (Gerlach and Giegerich, 2006) are some examples of tools used to predict RNA-RNA interactions. These are also integrated in pipelines to predict lncRNA-RNA interactions in humans. For instance, (Terai et al., 2016), developed a pipeline using RACCESS (Kiryu et al., 2011) to extract accessible regions from RNA molecules followed by masking tandem repeats using TanTan (Frith, 2011) and finding seed match using LAST and then calculate the interaction energy between two RNA molecules using IntaRNA and finally predict the joint secondary structure (RactIP) (Kato et al., 2010) to predict lncRNA-mRNA interactions (Szcześniak and Makałowska, 2016) proposed a similarity based method to predict RNA-RNA interactions using LAST (Kiełbasa et al., 2011), miRanda (Betel et al., 2010) tools in some pipelines. Similarly, RNA-protein interactions are also be predicted from sequence based methods which use physiochemical properties of amino/nucleic acids in tools like lncPRO (Lu et al., 2013) and catRAPID (Bellucci et al., 2011). Along with these sequence features, secondary structures of RNA are incl in tools like RPI-Pred (Suresh et al., 2015). PARIS (Lu et al., 2016), SPLASH (Aw et al., 2016), LIGR-seq (Sharma et al., 2016), and MARIO (Nguyen et al., 2016) to identify RNA-RNA interactions based on proximity ligation *in vivo* (Fukunaga and Hamada, 2017).

LNCRNA DATABASES

The publicly available datasets from RNA-seq and microarray experiments have led to rapid increase of annotated lncRNAs with dedicated databases for lncRNA and their molecular and disease associations. Many pipelines and tools have been benchmarked from the data available from these knowledge bases. NONCODEv5, the largest database for noncoding RNAs (majorly lncRNAs) contains 548,640 lncRNA transcripts from several model organisms (Fang et al., 2018), of which 96,308 lncRNA genes are from humans. The data has been curated from published literature and annotated with information from public resources like RefSeq, Ensembl, GenBank, lncRNAdb, lncipedia. The FANTOM (Functional ANnoTation Of the Mammalian genome) consortium led by RIKEN has systematically investigated and annotated about 27,919 human lncRNA genes across 1829 samples in the FANTOM database (FANTOM5) (Abugessaisa et al., 2017). Some of the databases provide experimentally validated and/or computationally predicted interactions of lncRNAs with other RNA and proteins. Analysis of data from RNA-seq and microarray experiments on disease cell lines have also helped in discovery of the roles lncRNA in disease mechanisms which have been recorded in disease-association databases. For instance LNCipedia provides lncRNA from humans with experimental and putative annotations along with miRNA-lncRNAs associations (Volders et al., 2013). Similarly, lncRNAdb is repository for functionally annotated lncRNAs along with TF-lncRNA associations. LncRNome, a lncRNA database for human complied form GENCODE has lncRNAs with annotations of their biomolecular interactions and disease associations. LncATLAS provides information on lncRNA localization in cells from RNA-sequencing data, from GENCODE (Mas-Ponte et al., 2017), lnc2CAncer has 1,488 entries of lncRNAs from

experimentally supported validations which are associated with cancer (Ning et al., 2016). **Table 3** contains a list of databases and their references.

CASE STUDY: CO-EXPRESSION NETWORK ANALYSIS IDENTIFYING PRO-INFLAMMATORY LNCRNAS IMPLICATED IN HCC

Cancer is caused by continuous accumulation of unfavourable genetic alterations that cause deregulation of genetic networks and cellular pathways ultimately (Huarte, 2015) leading to unceasing growth of cells and tissue. The mechanisms of these dysregulations are complex, involving altered gene expressions and molecular interactions which are yet to be discovered comprehensively; thus leading to the necessity to analyze the anomalies at all omics levels. In fact, LncRNAs are diversely associated in most of the hall marks of cancer. Many of the studies on cancer associated lncRNAs have mainly analyzed expression profile variations of lncRNA in cancer vs. healthy tissue and its effects on deregulated pathways and identification their regulatory targets. Also, approaches to identify RNA folding and stable complexes to evaluate lncRNA functions have depicted that genetic alterations like SNPs can also majorly impact the RNA structure and eventually their function with changes in active/binding sites of lncRNAs (Wan et al., 2014; Schmitt and Chang, 2016). Chronic inflammation has known be a vital in

TABLE 3 | Overview of databases of LncRNAs.

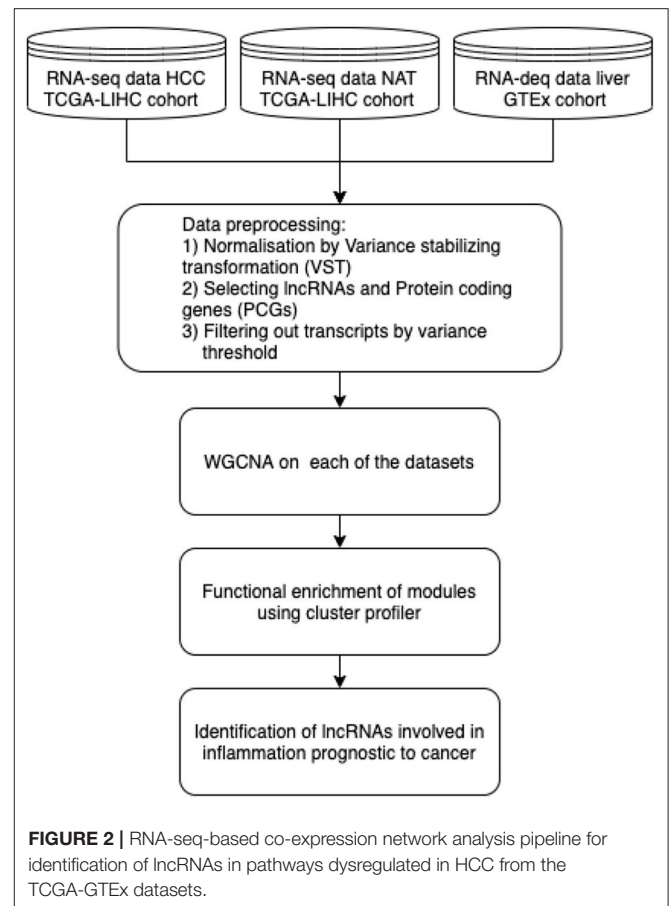| Database	Description	References
NONCODEV5	Knowledge base for ncRNAs	Fang et al., 2018
LNCipedia	lncRNA with secondary structure prediction, protein coding potential and microRNA binding sites	Volders et al., 2013
lncRNAdb v2.0	Manually curated lncRNAs from literature	Quek et al., 2015
LncATLAS	lncRNA annotated with subcellular localisatiom	Mas-Ponte et al., 2017
lncRNAdisease 2.0	Experimentally supported lncRNA disease association and molecular targets	Bao et al., 2019
LncRBase	lncRNA with information about their subtypes and interactions	Chakraborty et al., 2014
lncRNome	lncRNA with interactions with other RNAs	Bhartiya et al., 2013
GreeNC v1.1.2	Database for plant lncRNAs	Paytuvi Gallart et al., 2016
Lnc2Cancer v2.0	Manually curated database with experimentally supported lncRNA-cancer associations	Gao et al., 2019b
EVLncRNAs	Manually curated database with validated with low-throughput experiments	Zhou et al., 2018
ChIPBase v2.0	lncRNAs and other ncRNA from ChIP seq data	Zhou et al., 2017
DIANA-LncBase v3	Database dedicated to cataloging miRNA and lncRNA interactions	Karagkouni et al., 2020
LNCediting	Information of lncRNA editing, its impact and interactions with miRNAs	Gong et al., 2017
TCLA	Cancer LncRNome Atlas: lncRNAs predicted from TCGA datasets	Yan et al., 2015
MNDR v2.0	Experimental and predicted ncRNA-disease associations	Cui et al., 2018
lncRNASNP2	lncRNA variants and their disease associations	Miao et al., 2018
Lnc2Meth	Manually curated database of regulatory relationships between long non-coding RNAs and DNA methylation associated with human disease	Zhi et al., 2018
DES-ncRNA	Database of human miRNA and lncRNA from literature	Salhi et al., 2017
LincSNP2.0	disease associated SNPs with lncRNAs	Ning et al., 2017
LncVar	lncRNAs with associated genetic variations	Chen et al., 2017
deepBase v2.0	ncRNA database from deep sequencing data	Zheng et al., 2016
C-It-loci	Tissue specific transcriptome data (protein coding genes and ncRNA)	Weirick et al., 2015
LncRNA2Target v2.0	lncRNA and lncRNA-to-target genes after lncRNA knockdown and over expression	Cheng et al., 2019
LncTarD	Manually curated database of lncRNAs and target regulations	Zhao et al., 2020
CRlncRNA	Cancer related lncRNAs along with associations and interactions	Wang et al., 2018
lncRNAKB	Cancer related lncRNAs along with associations and interactions	Seifuddin et al., 2020
Cancer LncRNA Census (CLC)	lncRNAs from GENCODE involved in cancer	Carlevaro-Fita et al., 2020

TABLE 4 | Details of datasets used in the case study.

Dataset	Project	Number of samples	Number of modules	Data source
RNA-Seq expression data HCC	TCGA-LIHC	372	27	Tomczak et al., 2015
RNA-Seq expression data NAT	TCGA-LIHC	50	76	Tomczak et al., 2015
RNA-Seq expression from liver	GTEx	208	43	Lonsdale et al., 2013

cancer progression in case of Hepatocellular carcinoma(HCC). Some of the pathways known to be chronically upregulated causing hepatoma cell profileration include JAK/STAT signalling, NF-Kappa B signalling, PI3K/AKT/mTOR pathway, WNT pathway, and MAPK pathway (Chen et al., 2018; Yang et al., 2019). In order to investigate the application of co-expression network based on the "guilt by association" principle analysis of RNA-seq data, we applied the Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008) on the following datasets: The RNA-Seq dataset from The Cancer Genome Atlas (Tomczak et al., 2015) Liver Hepatocellular Carcinoma (TCGA-LIHC) project and the GTEx dataset (Lonsdale et al., 2013) (**Table 4**) samples to identify the pathways dysregulated in HCC with regards to chronic inflammation in HCC progression. The steps in the pipeline are illustrated in **Figure 2**. The datasets were collected and analyzed using the TCGAbiolinks, WGCNA packages in R.
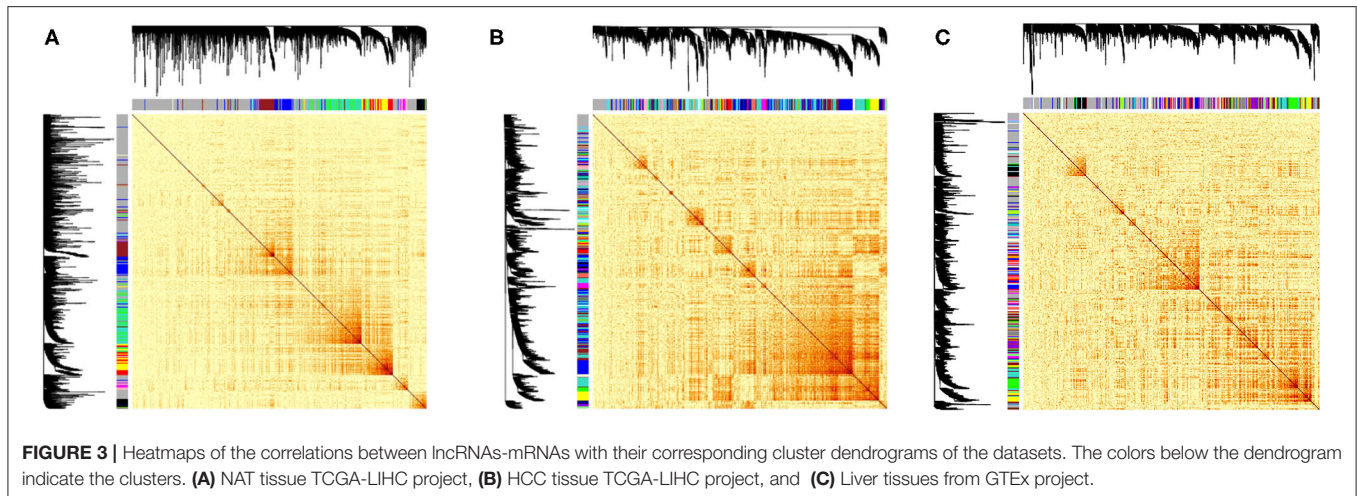
WGCNA analysis consists of the following steps: correlations across the normalized expression values of the samples are computed and raised to a soft threshold power based on the scale free topology criterion generating an adjacency matrix representing the co-expression network. This is followed by hierarchical clustering is used to identify clusters of co-expressed lncRNAs and protein coding genes among the network, each of which is labeled with a color/number. Co-expression Network using WGCNA was generated across all the 3 datasets and modules obtained in each case were enriched for functional process by cluster profiler. The modules which were identified for pathways dysregulated in case of HCC were selected and the lncRNAs which were highly connected, i.e., being significant for each module were identified for having bio-marker prognostic potential.

For HCC, NAT and GTEx profiles 27, 76 and 43 modules were identified, respectively from the hierarchical clustering with the cut height being selected 0.99, 0.98, 0.98 (**Figure 3**), respectively. These includes all the PCGs and lncRNAs transcripts. Each module was labeled with a color allocated by the WGCNA function and were enriched for KEGG pathways with threshold $p < 0.05$. The *red*, *yellow* in TCGA-HCC dataset and *turquoise*, *green* modules in TCGA-NAT dataset were enriched for the pathways involved in inflammation including JAK-STAT signaling pathway, cytokine-cytokine receptor interaction, NF-kappa B signaling pathway, T cell receptor signaling pathway among others contributing in inflammatory response. The network properties of all the networks were calculated based on which the transcripts in these modules were sorted according to their connectivity. The top highly connected lncRNAs(top 10) putatively having important regulatory mechanisms in these modules were selected for having biomarker potential in regards to chronic inflammation both in the tumour and its is surrounding micro environment proceeding to NAT. The common lncRNAs among the both phenotypes across these modules were *PCED1B-AS1, TRG-AS1, MIR155HG, MIAT, LINC00996*. MIAT has been known to be implicated in several cancers such as breast cancer, gastrointestinal cancer and NSCLC and also its silencing has known to inhibit cell proliferation and

FIGURE 2 | RNA-seq-based co-expression network analysis pipeline for identification of lncRNAs in pathways dysregulated in HCC from the TCGA-GTEx datasets.

tumorogenesis in HCC (Zhao et al., 2019). In a recent study by Peng et al. (2020) it has been postulated that *MIAT* regulates the expression of JAK2 among other genes and has an important role in controlling the tumour microenvironment in HCC.

LINC00996 has also been known to have regulatory mechanism in the JAK-STAT signalling pathway in colorectal cancer in a study by Ge et al. (2018). These pathways are dysregulated in the case of HCC as seen in the clusters from the TCGA datasets (HCC and NAT) but not the GTEx dataset. This provides us with corroboration pointing that NAT is subjected to an inflammatory environment prompted by the malignant tissue. This is similar to micro tumour environment with higher proliferation rate than a healthy hepatocyte. Identification of these modules and lncRNAs provides extended empirical evidence of lncRNA regulation in inflammation and pertaining to cancer progression. This analysis provides support to the "guilt by association" hypothesis of co-expression of lncRNAs with the genes involving in similar functions. However, few of the lncRNAs like MEG3, MALAT1, H19, UCA1 which have been studied for their implications in HCC didn't show an expression in the GTEx greater the variance threshold and could not be characterized in the co-expression networks while comparing to the TCGA datasets. This could be attributed to the batch effects of the RNA-Seq experiments across the GTeX and TCGA projects

FIGURE 3 | Heatmaps of the correlations between lncRNAs-mRNAs with their corresponding cluster dendrograms of the datasets. The colors below the dendrogram indicate the clusters. **(A)** NAT tissue TCGA-LIHC project, **(B)** HCC tissue TCGA-LIHC project, and **(C)** Liver tissues from GTEx project.

which can be addressed and corrected while pre-processing the raw reads together from all the datasets. The understanding of such complex networks in which dysregulation of lncRNAs occurs impacting cancer progression and metastasis, which also being tissue specific can set lncRNAs to become excellent biomarkers in cancer therapy (Schmitt and Chang, 2016).

CONCLUDING REMARKS

The recent discovery of the lncRNAs in the non-coding genome has led to a paradigm shift in the understanding of the mechanism of information flow from the genetic code and the genotype-phenotype map. But, as discussed, the mechanisms in which lncRNAs functions are very complex involving interactions with various molecules from other 'omic' levels. Advancements in RNA technologies have helped to elucidate some of the diverse mechanisms of lncRNAs but the regulatory potential of the majority of these noncoding genes have yet to be discovered. Differential co-expression of lncRNAs, RNA secondary structure and sequence analysis and prediction, ML based approaches in computational pipelines have aided in the identification and characterization of lncRNAs from RNA-seq experiments. This has to be supported by experimental validations and clarifications on cis-trans regulatory processes. Genome wide transcriptome profiling has identified several lncRNAs which have significant roles in

diseases like cancer exhibiting cell- and/or tissue/tumor-specific expression and hence can be excellent candidate targets for therapy. It has been demonstrated that silencing of certain disease associated lncRNAs exhibited tumor suppression. In summary, a comprehensive knowledge of lncRNAs shall provide researchers insights into genotype-phenotype distinction and genetic disorders leading to more effective therapeutic strategies for diseases and with emergence of new experimental designs and computational pipelines we can advance our understanding of the transcriptome.

AUTHOR CONTRIBUTIONS

VS and RS: original idea for the manuscript, contributed to design and conceptualization of the study, and supervision. AC: literature, data analysis, and writing initial draft. VS: writing, review, and editing. RS: project administration and funding acquisition. All authors also critically reviewed, wrote, and approved the final version.

FUNDING

REFERENCES

Abugessaisa, I., Noguchi, S., Carninci, P., and Kasukawa, T. (2017). The FANTOM5 Computation Ecosystem: Genomic Information Hub for Promoters and Active Enhancers. *Methods Mol. Biol.* 1611, 199–217. doi: 10.1007/978-1-4939-7015-5_15

Alam, T., Uludag, M., Essack, M., Salhi, A., Ashoor, H., Hanks, J. B., et al. (2017). FARNA: knowledgebase of inferred functions of non-coding RNA transcripts. *Nucleic Acids Res.* 45, 2838–2848. doi: 10.1093/nar/gkw973

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Amodio, N., Stamato, M. A., Juli, G., Morelli, E., Fulciniti, M., Manzoni, M., et al. (2018). Drugging the lncRNA MALAT1 via LNA gapmeR ASO inhibits gene expression of proteasome subunits and triggers anti-multiple myeloma activity. *Leukemia.* 32, 1948–1957. doi: 10.1038/s41375-018-0067-3

Andronescu, M., Aguirre-Hernández, R., Condon, A., and Hoos, H. H. (2003). RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Res.* 31, 3416–3422. doi: 10.1093/nar/gkg612

Aw, J. G., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K. L., et al. (2016). *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol Cell* 62, 603–617. doi: 10.1016/j.molcel.2016.04.028

Balas, M. M., and Johnson, A. M. (2018). Exploring the mechanisms behind long noncoding RNAs and cancer. *Noncoding RNA Res.* 3, 108–117. doi: 10.1016/j.ncrna.2018.03.001

Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*. 47, D1034–D1037. doi: 10.1093/nar/gky905

Barra, J., and Leucci, E. (2017). Probing long non-coding RNA-protein interactions. *Front. Mol. Biosci.* 4:45. doi: 10.3389/fmolb.2017.00045

Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G. G. (2011). Predicting protein associations with long noncoding RNAs. *Nat. Methods* 8, 444–445. doi: 10.1038/nmeth.1611

Bertin, N., Mendez, M., Hasegawa, A., Lizio, M., Abugessaisa, I., Severin, J., et al. (2017). Linking FANTOM5 CAGE peaks to annotations with CAGEscan. *Sci. Data* 4:170147. doi: 10.1038/sdata.2017.147

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11, R90. doi: 10.1186/gb-2010-11-8-r90

Bhartiya, D., Pal, K., Ghosh, S., Kapoor, S., Jalali, S., Panwar, B., et al. (2013). lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database (Oxford)* 2013:bat034. doi: 10.1093/database/bat034

Carlevaro-Fita, J., Lanzós, A., Feuerbach, L., Hong, C., Mas-Ponte, D., Pedersen, J. S., et al. (2020). Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun Biol* 3, 56. doi: 10.1038/s42003-019-0741-7

Carmona, S., Lin, B., Chou, T., Arroyo, K., and Sun, S. (2018). LncRNA Jpx induces Xist expression in mice using both trans and cis mechanisms. *PLoS Genet.* 14:e1007378. doi: 10.1371/journal.pgen.1007378

Chakraborty, S., Deb, A., Maji, R. K., Saha, S., and Ghosh, Z. (2014). LncRBase: an enriched resource for lncRNA information. *PLoS ONE* 9:e108010. doi: 10.1371/journal.pone.0108010

Chen, H.-J., Hu, M.-H., Xu, F.-G., Xu, H.-J., She, J.-J., and Xia, H.-P. (2018). Understanding the inflammation-cancer transformation in the development of primary liver cancer. *Hepatoma Res.* 4:29. doi: 10.20517/2394-5079.2018.18

Chen, W., Zhang, G., Li, J., Zhang, X., Huang, S., Xiang, S., et al. (2019). CRISPRlnc: a manually curated database of validated sgRNAs for lncRNAs. *Nucleic Acids Res.* 47, D63–D68. doi: 10.1093/nar/gky904

Chen, X. (2015a). KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Sci. Rep.* 5:16840. doi: 10.1038/srep16840

Chen, X. (2015b). Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. *Sci. Rep.* 5:13186. doi: 10.1038/srep13186

Chen, X., Hao, Y., Cui, Y., Fan, Z., He, S., Luo, J., et al. (2017). LncVar: a database of genetic variation associated with long non-coding genes. *Bioinformatics* 33, 112–118. doi: 10.1093/bioinformatics/btw581

Chen, X., and Yan, G. Y. (2013). Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* 29, 2617–2624. doi: 10.1093/bioinformatics/btt426

Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2019). LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* 47, D140–D144. doi: 10.1093/nar/gky1051

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027

Cipriano, A., and Ballarino, M. (2018). The ever-evolving concept of the gene: the use of RNA/protein experimental techniques to understand genome functions. *Front. Mol. Biosci.* 5:20. doi: 10.3389/fmolb.2018.00020

Cui, T., Zhang, L., Huang, Y., Yi, Y., Tan, P., Zhao, Y., et al. (2018). MNDR v2.0: an updated resource of ncRNA-disease associations in mammals. *Nucleic Acids Res.* 46, D371–D374. doi: 10.1093/nar/gkx1025

Das, S., Ghosal, S., Sen, R., and Chakrabarti, J. (2014). lnCeDB: database of human long noncoding RNA acting as competing endogenous RNA. *PLoS ONE* 9:e98965. doi: 10.1371/journal.pone.0098965

de Hoon, M., Shin, J. W., and Carninci, P. (2015). Paradigm shifts in genomics through the FANTOM projects. *Mamm. Genome* 26, 391–402. doi: 10.1007/s00335-015-9593-8

Denman, R. B. (1993). Using RNAFOLD to predict the activity of small catalytic RNAs. *Biotechniques* 15, 1090–1095.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D. P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell* 54, 766–776. doi: 10.1016/j.molcel.2014.03.045

Denzler, R., McGeary, S. E., Title, A. C., Agarwal, V., Bartel, D. P., and Stoffel, M. (2016). Impact of MicroRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. *Mol. Cell* 64, 565–579. doi: 10.1016/j.molcel.2016.09.027

Ding, L., Wang, M., Sun, D., and Li, A. (2018). TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph. *Sci. Rep.* 8, 1065. doi: 10.1038/s41598-018-19357-3

Dinger, M. E., Pang, K. C., Mercer, T. R., and Mattick, J. S. (2008). Differentiating protein-coding and non-coding RNA: challenges and ambiguities. *PLoS Comput. Biol.* 4:e1000176. doi: 10.1371/journal.pcbi.1000176

Dobin, A., and Gingeras, T. R. (2015). Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* 51:1–11. doi: 10.1002/0471250953.bi1114s51

Dong, P., Xiong, Y., Yue, J., Hanley, S. J. B., Kobayashi, N., Todo, Y., et al. (2018). Long non-coding RNA NEAT1: a novel target for diagnosis and therapy in human tumors. *Front.Genet.* 9:471. doi: 10.3389/fgene.2018.00471

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211. doi: 10.1142/9781848165632_0019

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341:1237973. doi: 10.1126/science.1237973

Fan, X. N., Zhang, S. W., Zhang, S. Y., Zhu, K., and Lu, S. (2019). Prediction of lncRNA-disease associations by integrating diverse heterogeneous information sources with RWR algorithm and positive pointwise mutual information. *BMC Bioinformatics* 20:87. doi: 10.1186/s12859-019-2675-y

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., et al. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 46, D308–D314. doi: 10.1093/nar/gkx1107

Fang, Y., and Fullwood, M. J. (2016). Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics Proteomics Bioinformatics* 14, 42–54. doi: 10.1016/j.gpb.2015.09.006

Frith, M. C. (2011). A new repeat-masking method enables specific detection of homologous sequences. *Nucleic Acids Res.* 39, e23. doi: 10.1093/nar/gkq1212

Fukunaga, T., and Hamada, M. (2017). RIblast: an ultrafast RNA-RNA interaction prediction system based on a seed-and-extension approach. *Bioinformatics* 33, 2666–2674. doi: 10.1093/bioinformatics/btx287

Furió-Tarí, P., Tarazona, S., Gabaldón, T., Enright, A. J., and Conesa, A. (2016). spongeScan: A web for detecting microRNA binding elements in lncRNA sequences. *Nucleic Acids Res* 44, 176–180. doi: 10.1093/nar/gkw443

Gao, C., Zhao, D., Zhao, Q., Dong, D., Mu, L., Zhao, X., et al. (2019a). Microarray profiling and co-expression network analysis of lncRNAs and mRNAs in ovarian cancer. *Cell Death Discov.* 5:93. doi: 10.1038/s41420-019-0173-7

Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., et al. (2019b). Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.* 47, D1028–D1033. doi: 10.1093/nar/gky1096

Ge, H., Yan, Y., Wu, D., Huang, Y., and Tian, F. (2018). Potential role of LINC00996 in colorectal cancer: A study based on data mining and bioinformatics. *Onco Targets Ther.* 11, 4845–4855. doi: 10.2147/OTT.S173225

Ge, M., Li, A., and Wang, M. (2016). A bipartite network-based method for prediction of long non-coding RNA-protein interactions. *Genomics Proteomics Bioinformatics* 14, 62–71. doi: 10.1016/j.gpb.2016.01.004

Gerlach, W., and Giegerich, R. (2006). GUUGle: a utility for fast exact matching under RNA complementary rules including G-U base pairing. *Bioinformatics* 22, 762–764. doi: 10.1093/bioinformatics/btk041

German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* 26, 941–946. doi: 10.1038/nbt1417

Ghosh, S., and Chan, C. K. (2016). Analysis of RNA-Seq data using TopHat and cufflinks. *Methods Mol. Biol.* 1374, 339–361. doi: 10.1007/978-1-4939-3167-5_18

Gibb, E. A., Vucic, E. A., Enfield, K. S., Stewart, G. L., Lonergan, K. M., Kennett, J. Y., et al. (2011). Human cancer long non-coding RNA transcriptomes. *PLoS ONE* 6:e25915. doi: 10.1371/journal.pone.0025915

Gish, W., and States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat. Genet.* 3, 266–272. doi: 10.1038/ng0393-266

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., et al. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 19, 163–164. doi: 10.1093/bioinformatics/19.1.163

Gong, J., Liu, C., Liu, W., Xiang, Y., Diao, L., Guo, A. Y., et al. (2017). LNCediting: a database for functional effects of RNA editing in lncRNAs. *Nucleic Acids Res.* 45, D79–D84. doi: 10.1093/nar/gkw835

Gregory, B. D., O'Malley, R. C., Lister, R., Urich, M. A., Tonti-Filippini, J., Chen, H., et al. (2008). A link between RNA metabolism and silencing affecting Arabidopsis development. *Dev. Cell* 14, 854–866. doi: 10.1016/j.devcel.2008.04.005

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell.* 27, 91–105. doi: 10.1016/j.molcel.2007.06.017

Gruber, A. R., Findeiß, S., Washietl, S., Hofacker, I. L., and Stadler, P. F. (2010). RNAz 2.0: improved noncoding RNA detection. *Pac Symp Biocomput pages* 69–79. doi: 10.1142/9789814295291_0009

Gu, C., Liao, B., Li, X., Cai, L., Li, Z., Li, K., et al. (2017). Global network random walk for predicting potential human lncRNA-disease associations. *Sci. Rep.* 7, 12442. doi: 10.1038/s41598-017-12763-z

Guo, X., Gao, L., Wang, Y., Chiu, D. K., Wang, T., and Deng, Y. (2016). Advances in long noncoding RNAs: identification, structure prediction and function annotation. *Brief Funct. Genomics* 15, 38–46. doi: 10.1093/bfgp/elv022

Guo, Z. H., You, Z. H., Wang, Y. B., Yi, H. C., and Chen, Z. H. (2019). A learning-based method for LncRNA-disease association identification combing similarity information and rotation forest. *iScience* 19, 786–795. doi: 10.1016/j.isci.2019.08.030

Hamada, M., Yamada, K., Sato, K., Frith, M. C., and Asai, K. (2011). CentroidHomfold-LAST: accurate prediction of RNA secondary structure using automatically collected homologous sequences. *Nucleic Acids Res.* 39, W100–106. doi: 10.1093/nar/gkr290

Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 100, 57–70. doi: 10.1016/S0092-8674(00)81683-9

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111

He, X., Ou, C., Xiao, Y., Han, Q., Li, H., and Zhou, S. (2017). LncRNAs: key players and novel insights into diabetes mellitus. *Oncotarget* 8, 71325–71341. doi: 10.18632/oncotarget.19921

He, Y., Meng, X. M., Huang, C., Wu, B. M., Zhang, L., Lv, X. W., et al. (2014). Long noncoding RNAs: Novel insights into hepatocelluar carcinoma. *Cancer Lett.* 344, 20–27. doi: 10.1016/j.canlet.2013.10.021

Heo, J. B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331, 76–79. doi: 10.1126/science.1197349

Hodas, N. O., and Aalberts, D. P. (2004). Efficient computation of optimal oligo-RNA binding. *Nucleic Acids Res.* 32, 6636–6642. doi: 10.1093/nar/gkh1008

Hombach, S., and Kretz, M. (2016). Non-coding RNAs: classification, biology and functioning. *Adv. Exp. Med. Biol.* 937, 3–17. doi: 10.1007/978-3-319-42059-2_1

Hon, C. C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J., Gough, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* 543, 199–204. doi: 10.1038/nature21374

Hu, L., Xu, Z., Hu, B., and Lu, Z. J. (2017). COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Res.* 45, e2. doi: 10.1093/nar/gkw798

Huang, J., Liu, T., Shang, C., Zhao, Y., Wang, W., Liang, Y., et al. (2018). Identification of lncRNAs by microarray analysis reveals the potential role of lncRNAs in cervical cancer pathogenesis. *Oncol. Lett.* 15, 5584–5592. doi: 10.3892/ol.2018.8037

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* 21, 1253–1261. doi: 10.1038/nm.3981

Hubé, F., and Francastel, C. (2018). Coding and Non-coding RNAs, the Frontier Has Never Been So Blurred. *Front. Genet.* 9:140. doi: 10.3389/fgene.2018.00140

Itaya, H., Oshita, K., Arakawa, K., and Tomita, M. (2013). GEMBASSY: an EMBOSS associated software package for comprehensive genome analyses. *Source Code Biol. Med.* 8:17. doi: 10.1186/1751-0473-8-17

Jalali, S., Kapoor, S., Sivadas, A., Bhartiya, D., and Scaria, V. (2015). Computational approaches towards understanding human long non-coding RNA biology. *Bioinformatics* 31, 2241–2251. doi: 10.1093/bioinformatics/btv148

Jia, H., Wang, X., and Sun, Z. (2018). Exploring the molecular pathogenesis and biomarkers of high risk oral premalignant lesions on the basis of long noncoding RNA expression profiling by serial analysis of gene expression. *Eur. J. Cancer Prev.* 27, 370–378. doi: 10.1097/CEJ.0000000000000346

Kanamori-Katayama, M., Itoh, M., Kawaji, H., Lassmann, T., Katayama, S., Kojima, M., et al. (2011). Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* 21, 1150–1159. doi: 10.1101/gr.115469.110

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Karagkouni, D., Paraskevopoulou, M. D., Tastsoglou, S., Skoufos, G., Karavangeli, A., Pierros, V., et al. (2020). DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Res.* 48, D101–D110. doi: 10.1093/nar/gkz1036

Kashi, K., Henderson, L., Bonetti, A., and Carninci, P. (2016). Discovery and functional analysis of lncRNAs: methodologies to investigate an uncharacterized transcriptome. *Biochim. Biophys. Acta* 1859, 3–15. doi: 10.1016/j.bbagrm.2015.10.010

Kato, Y., Sato, K., Hamada, M., Watanabe, Y., Asai, K., and Akutsu, T. (2010). RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming. *Bioinformatics* 26, i460–466. doi: 10.1093/bioinformatics/btq372

Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 537, 39–64. doi: 10.1007/978-1-59745-251-9_3

Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., et al. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107. doi: 10.1038/nature09322

Kiełbasa, S. M., Wan, R., Sato, K., Horton, P., and Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Res.* 21, 487–493. doi: 10.1101/gr.113985.110

Kim, D. H., Xi, Y., and Sung, S. (2017). Modular function of long noncoding RNA, COLDAIR in the vernalization response. *PLoS Genet.* 13:e1006939. doi: 10.1371/journal.pgen.1006939

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal* 3, ra8. doi: 10.1126/scisignal.2000568

Kiryu, H., Terai, G., Imamura, O., Yoneyama, H., Suzuki, K., and Asai, K. (2011). A detailed investigation of accessibilities around target sites of siRNAs and miRNAs. *Bioinformatics* 27, 1788–1797. doi: 10.1093/bioinformatics/btr276

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, W345–W349. doi: 10.1093/nar/gkm391

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., et al. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to

and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956–1962. doi: 10.1038/onc.2010.568

Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* 34, W451–W454. doi: 10.1093/nar/gkl243

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Lennox, K. A., and Behlke, M. A. (2016). Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* 44, 863–877. doi: 10.1093/nar/gkv1206

Li, J., and Liu, C. (2019). Coding or Noncoding, the Converging Concepts of RNAs. *Front Genet* 10:496. doi: 10.3389/fgene.2019.00496

Li, J., Xu, Y., Xu, J., Wang, J., and Wu, L. (2016). Dynamic co-expression network analysis of lncRNAs and mRNAs associated with venous congestion. *Mol. Med. Rep.* 14, 2045–2051. doi: 10.3892/mmr.2016.5480

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97. doi: 10.1093/nar/gkt1248

Liu, J., Gough, J., and Rost, B. (2006). Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet.* 2:e29. doi: 10.1371/journal.pgen.0020029

Liu, K., Yan, Z., Li, Y., and Sun, Z. (2013). Linc2GO: a human LincRNA function annotation resource based on ceRNA hypothesis. *Bioinformatics* 29, 2221–2222. doi: 10.1093/bioinformatics/btt361

Liu, X., Ma, Y., Yin, K., Li, W., Chen, W., Zhang, Y., et al. (2019). Long non-coding and coding RNA profiling using strand-specific RNA-seq in human hypertrophic cardiomyopathy. *Sci. Data* 6, 90. doi: 10.1038/s41597-019-0094-6

Liu, Y., Sun, Y., Li, Y., Bai, H., Xue, F., Xu, S., et al. (2017). Analyses of Long Non-Coding RNA and mRNA profiling using RNA sequencing in chicken testis with extreme sperm motility. *Sci. Rep.* 7, 9055. doi: 10.1038/s41598-017-08738-9

Loewer, S., Cabili, M. N., Guttman, M., Loh, Y. H., Thomas, K., Park, I. H., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* 42, 1113–1117. doi: 10.1038/ng.710

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45, 580–585. doi: 10.1038/ng.2653

Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X., et al. (2013). Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics* 14:651. doi: 10.1186/1471-2164-14-651

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., et al. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279. doi: 10.1016/j.cell.2016.04.028

Lu, Z. J., Yip, K. Y., Wang, G., Shou, C., Hillier, L. W., Khurana, E., et al. (2011). Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome Res.* 21, 276–285. doi: 10.1101/gr.110189.110

Lund, S. H., Gudbjartsson, D. F., Rafnar, T., Sigurdsson, A., Gudjonsson, S. A., Gudmundsson, J., et al. (2014). A method for detecting long non-coding RNAs with tiled RNA expression microarrays. *PLoS ONE* 9:e99899. doi: 10.1371/journal.pone.0099899

Ma, H., Hao, Y., Dong, X., Gong, Q., Chen, J., Zhang, J., et al. (2012). Molecular mechanisms and function prediction of long noncoding RNA. *Sci. World J.* 2012:541786. doi: 10.1100/2012/541786

Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biol.* 10, 925–933. doi: 10.4161/rna.24604

Mann, M., Wright, P. R., and Backofen, R. (2017). IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.* 45, W435–W439. doi: 10.1093/nar/gkx279

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445, 666–670. doi: 10.1038/nature05519

Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Hermoso Pulido, T., Guigo, R., and Johnson, R. (2017). LncATLAS database for subcellular localization of long noncoding RNAs. *RNA* 23, 1080–1087. doi: 10.1261/rna.060814.117

Mattick, J. S., and Makunin, I. V. (2006). Non-coding RNA. *Hum. Mol. Genet.* 15, 17–29. doi: 10.1093/hmg/ddl046

Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddeloh, J. A., et al. (2011). Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* 30, 99–104. doi: 10.1038/nbt.2024

Miao, Y. R., Liu, W., Zhang, Q., and Guo, A. Y. (2018). lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic Acids Res.* 46, D276–D280. doi: 10.1093/nar/gkx1004

Michelhaugh, S. K., Lipovich, L., Blythe, J., Jia, H., Kapatos, G., and Bannon, M. J. (2011). Mining Affymetrix microarray data for long non-coding RNAs: altered expression in the nucleus accumbens of heroin abusers. *J. Neurochem.* 116, 459–466. doi: 10.1111/j.1471-4159.2010.07126.x

Mills, J. D., Kawahara, Y., and Janitz, M. (2013). Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. *Curr. Genomics* 14, 173–181. doi: 10.2174/1389202911314030003

Nguyen, T. C., Cao, X., Yu, P., Xiao, S., Lu, J., Biase, F. H., et al. (2016). Mapping RNA-RNA interactome and RNA structure *in vivo* by MARIO. *Nat. Commun.* 7:12023. doi: 10.1038/ncomms12023

Ning, S., Yue, M., Wang, P., Liu, Y., Zhi, H., Zhang, Y., et al. (2017). LincSNP 2.0: an updated database for linking disease-associated SNPs to human long non-coding RNAs and their TFBSs. *Nucleic Acids Res.* 4, :D74–D78. doi: 10.1093/nar/gkw945

Ning, S., Zhang, J., Wang, P., Zhi, H., Wang, J., Liu, Y., et al. (2016). Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res.* 44, D980–D985. doi: 10.1093/nar/gkv1094

Ouyang, Z., Snyder, M. P., and Chang, H. Y. (2013). SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* 23, 377–387. doi: 10.1101/gr.138545.112

Palazzo, A. F., and Lee, E. S. (2015). Non-coding RNA: what is functional and what is junk? *Front. Genet.* 6:2. doi: 10.3389/fgene.2015.00002

Pan, X., Jensen, L. J., and Gorodkin, J. (2019). Inferring disease-associated long non-coding RNAs using genome-wide tissue expression profiles. *Bioinformatics* 35, 1494–1502. doi: 10.1093/bioinformatics/bty859

Paytuvi Gallart, A., Hermoso Pulido, A., Anzar Martinez de Lagran, I., Sanseverino, W., and Aiese Cigliano, R. (2016). GREENC: a Wiki-based database of plant lncRNAs. *Nucleic Acids Res.* 44, D1161–D1166. doi: 10.1093/nar/gkv1215

Pedersen, J. S., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E. S., et al. (2006). Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol.* 2:e33. doi: 10.1371/journal.pcbi.0020033

Pelechano, V., Wei, W., and Steinmetz, L. M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 497, 127–131. doi: 10.1038/nature12121

Peng, L., Chen, Y., Ou, Q., Wang, X., and Tang, N. (2020). LncRNA MIAT correlates with immune infiltrates and drug reactions in hepatocellular carcinoma. *Int. Immunopharmacol.* 89(Pt A):107071. doi: 10.1016/j.intimp.2020.107071

Pian, C., Zhang, G., Tu, T., Ma, X., and Li, F. (2018). LncCeRBase: a database of experimentally validated human competing endogenous long non-coding RNAs. *Database (Oxford)* 2018:bay061. doi: 10.1093/database/bay061

Poulain, S., Kato, S., Arnaud, O., Morlighem, J., Suzuki, M., Plessy, C., et al. (2017). NanoCAGE: a method for the analysis of coding and noncoding 5'-capped transcriptomes. *Methods Mol. Biol.* 1543, 57–109. doi: 10.1007/978-1-4939-6716-2_4

Quek, X. C., Thomson, D. W., Maag, J. L., Bartonicek, N., Signal, B., Clark, M. B., et al. (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 43, D168–D173. doi: 10.1093/nar/gku988

Redon, S., Reichenbach, P., and Lingner, J. (2010). The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res* 38, 5797–5806. doi: 10.1093/nar/gkq296

Reuter, J. S., and Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11:129. doi: 10.1186/1471-2105-11-129

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323. doi: 10.1016/j.cell.2007.05.022

Sakthianandeswaren, A., Liu, S., and Sieber, O. M. (2016). Long noncoding RNA LINP1: scaffolding non-homologous end joining. *Cell Death Discov.* 2:16059. doi: 10.1038/cddiscovery.2016.59

Salhi, A., Essack, M., Alam, T., Bajic, V. P., Ma, L., Radovanovic, A., et al. (2017). DES-ncRNA: A knowledgebase for exploring information about human micro and long noncoding RNAs based on literature-mining. *RNA Biol.* 14, 963–971. doi: 10.1080/15476286.2017.1312243

Sato, K., Hamada, M., Asai, K., and Mituyama, T. (2009). CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic Acids Res.* 37, W277–W280. doi: 10.1093/nar/gkp367

Saus, E., Willis, J. R., Pryszcz, L. P., Hafez, A., Llorens, C., Himmelbauer, H., et al. (2018). nextPARS: parallel probing of RNA structures in Illumina. *RNA* 24, 609–619. doi: 10.1261/rna.063073.117

Schmitt, A. M., and Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. *Cancer Cell* 29, 452–463. doi: 10.1016/j.ccell.2016.03.010

Schoenbeck, S. L. (2016). GUIDELINES FOR APPROPRIATELY USING scripture at the bedside. *J. Christ. Nurs.* 33, 108–111. doi: 10.1097/CNJ.0000000000000260

Seifuddin, F., Singh, K., Suresh, A., Judy, J. T., Chen, Y. C., Chaitankar, V., et al. (2020). lncRNAKB a knowledgebase of tissue-specific functional annotation and trait association of long noncoding RNA. *Sci. Data* 7, 326. doi: 10.1038/s41597-020-00659-z

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B. J. (2016). Global mapping of human RNA-RNA interactions. *Mol. Cell* 62, 618–626. doi: 10.1016/j.molcel.2016.04.030

Sharma, S., Ding, F., and Dokholyan, N. V. (2008). iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* 24, 1951–1952. doi: 10.1093/bioinformatics/btn328

Shi, Y., and Shang, J. (2016). Long noncoding RNA expression profiling using arraystar LncRNA microarrays. *Methods Mol. Biol.* 1402, 43–61. doi: 10.1007/978-1-4939-3378-5_6

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15776–15781. doi: 10.1073/pnas.2136655100

Signal, B., Gloss, B. S., and Dinger, M. E. (2016). Computational approaches for functional prediction and characterisation of long noncoding RNAs. *Trends Genet.* 32, 620–637. doi: 10.1016/j.tig.2016.08.004

Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., Wardle, G., et al. (2014). PAR-CLIP (Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Meth. Enzymol.* 539, 113–161.

Starmer, J., and Magnuson, T. (2009). A new model for random X chromosome inactivation. *Development* 136, 1–10. doi: 10.1242/dev.025908

Sumathipala, M., Maiorino, E., Weiss, S. T., and Sharma, A. (2019). Network diffusion approach to predict LncRNA disease associations using multi-type biological networks: LION. *Front. Physiol.* 10:888. doi: 10.3389/fphys.2019.00888

Sun, K., Chen, X., Jiang, P., Song, X., Wang, H., and Sun, H. (2013). iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. *BMC Genomics* (14 Suppl.) 2:S7. doi: 10.1186/1471-2164-14-S2-S7

Sun, L., Liu, H., Zhang, L., and Meng, J. (2015). lncRScan-SVM: a tool for predicting long non-coding RNAs using support vector machine. *PLoS ONE* 10:e0139654. doi: 10.1371/journal.pone.0139654

Suresh, V., Liu, L., Adjeroh, D., and Zhou, X. (2015). RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res.* 43, 1370–1379. doi: 10.1093/nar/gkv020

Swenson, M. S., Anderson, J., Ash, A., Gaurav, P., Sükösd, Z., Bader, D. A., et al. (2012). GTfold: enabling parallel RNA secondary structure prediction on multi-core desktops. *BMC Res. Notes* 5:341. doi: 10.1186/1756-0500-5-341

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* 462, 799–802. doi: 10.1038/nature08618

Szcześniak, M. W., and Makałowska, I. (2016). lncRNA-RNA interactions across the human transcriptome. *PLoS ONE* 11:e0150353. doi: 10.1371/journal.pone.0150353

Tafer, H., and Hofacker, I. L. (2008). RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics* 24, 2657–2663. doi: 10.1093/bioinformatics/btn193

Tani, H., Mizutani, R., Salam, K. A., Tano, K., Ijiri, K., Wakamatsu, A., et al. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* 22, 947–956. doi: 10.1101/gr.130559.111

Terai, G., Iwakiri, J., Kameda, T., Hamada, M., and Asai, K. (2016). Comprehensive prediction of lncRNA-RNA interactions in human transcriptome. *BMC Genomics* (17 Suppl.):12. doi: 10.1186/s12864-015-2307-5

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* Chapter 2: Unit 2.3. doi: 10.1002/0471250953.bi0203s00

Tian, D., Sun, S., and Lee, J. T. (2010). The long noncoding RNA Jpx, is a molecular switch for X chromosome inactivation. *Cell* 143, 390–403. doi: 10.1016/j.cell.2010.09.049

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol. (Pozn)* 19, 68–77. doi: 10.5114/wo.2014.47136

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938. doi: 10.1016/j.molcel.2010.08.011

Umu, S. U., and Gardner, P. P. (2017). A comprehensive benchmark of RNA-RNA interaction prediction tools for all domains of life. *Bioinformatics* 33, 988–996. doi: 10.1093/bioinformatics/btw728

Underwood, J. G., Uzilov, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., et al. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7, 995–1001. doi: 10.1038/nmeth.1529

Valen, E., Pascarella, G., Chalk, A., Maeda, N., Kojima, M., Kawazu, C., et al. (2009). Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res.* 19, 255–265. doi: 10.1101/gr.084541.108

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* 270, 484–487. doi: 10.1126/science.270.5235.484

Volders, P. J., Helsens, K., Wang, X., Menten, B., Martens, L., Gevaert, K., et al. (2013). LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* 41, D246–D251. doi: 10.1093/nar/gks915

Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709. doi: 10.1038/nature12946

Wang, H. V., and Chekanova, J. A. (2019). An Overview of methodologies in studying lncRNAs in the high-throughput era: when acronyms ATTACK! *Methods Mol. Biol.* 1933, 1–30. doi: 10.1007/978-1-4939-9045-0_1

Wang, J., Zhang, X., Chen, W., Li, J., and Liu, C. (2018). CRlncRNA: a manually curated database of cancer-related long non-coding RNAs with experimental proof of functions on clinicopathological and molecular features. *BMC Med. Genomics* 11(Suppl. 6):114. doi: 10.1186/s12920-018-0430-2

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–124. doi: 10.1038/nature09819

Wasko, U., Zheng, Z., and Bhatnagar, S. (2019). Visualization of xist long noncoding RNA with a fluorescent CRISPR/Cas9 system. *Methods Mol. Biol.* 1870, 41–50. doi: 10.1007/978-1-4939-8808-2_3

Weirick, T., John, D., Dimmeler, S., and Uchida, S. (2015). C-It-Loci: a knowledge database for tissue-enriched loci. *Bioinformatics* 31, 3537–3543. doi: 10.1093/bioinformatics/btv410

Wenzel, A., Akbasli, E., and Gorodkin, J. (2012). RIsearch: fast RNA-RNA interaction search using a simplified nearest-neighbor energy model. *Bioinformatics* 28, 2738–2746. doi: 10.1093/bioinformatics/bts519

Wu, W., Wagner, E. K., Hao, Y., Rao, X., Dai, H., Han, J., et al. (2016). Tissue-specific co-expression of long non-coding and coding RNAs associated with breast cancer. *Sci. Rep.* 6:32731. doi: 10.1038/srep32731

Xiao, X., Zhu, W., Liao, B., Xu, J., Gu, C., Ji, B., et al. (2018). BPLLDA: predicting lncRNA-disease associations based on simple paths with limited lengths in a heterogeneous network. *Front. Genet.* 9:411. doi: 10.3389/fgene.2018.00411

Xuan, P., Pan, S., Zhang, T., Liu, Y., and Sun, H. (2019). Graph convolutional network and convolutional neural network based method for predicting lncrna-disease associations. *Cells* 8, 1012. doi: 10.3390/cells8091012

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S. D., et al. (2015). Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell* 28, 529–540. doi: 10.1016/j.ccell.2015.09.006

Yang, Y. M., Kim, S. Y., and Seki, E. (2019). Inflammation and liver cancer: molecular mechanisms and therapeutic targets. *Semin Liver Dis.* 39, 26–42. doi: 10.1055/s-0038-1676806

Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., et al. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* 38, 662–674. doi: 10.1016/j.molcel.2010.03.021

Zanella, C. (2021). How do I Cite BioRender? Available online at: https://help.Biorender.com/en/articles/3619405-how-do-i-cite-biorender

Zhang, K., Shi, H., Xi, H., Wu, X., Cui, J., Gao, Y., et al. (2017). Genome-wide lncRNA microarray profiling identifies novel circulating lncRNAs for detection of gastric cancer. *Theranostics* 7, 213–227. doi: 10.7150/thno.16044

Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., et al. (2019). Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *Int. J. Mol. Sci.* 20:5573. doi: 10.3390/ijms20225573

Zhao, H., Shi, J., Zhang, Y., Xie, A., Yu, L., Zhang, C., et al. (2020). LncTarD: a manually-curated database of experimentally-supported functional lncRNA-target regulations in human diseases. *Nucleic Acids Res.* 48, D118–D126. doi: 10.1093/nar/gkz985

Zhao, L., Hu, K., Cao, J., Wang, P., Li, J., Zeng, K., et al. (2019). lncRNA miat functions as a ceRNA to upregulate sirt1 by sponging miR-22-3p in HCC cellular senescence. *Aging (Albany NY)* 11, 7098–7122. doi: 10.18632/aging.102240

Zheng, H., Brennan, K., Hernaez, M., and Gevaert, O. (2019). Benchmark of long non-coding RNA quantification for RNA sequencing of cancer samples. *Gigascience* 8:giz145. doi: 10.1093/gigascience/giz145

Zheng, L. L., Li, J. H., Wu, J., Sun, W. J., Liu, S., Wang, Z. L., et al. (2016). deepBase v2.0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res.* 44, 196–202. doi: 10.1093/nar/gkv1273

Zhi, H., Li, X., Wang, P., Gao, Y., Gao, B., Zhou, D., et al. (2018). Lnc2Meth: a manually curated database of regulatory relationships between long non-coding RNAs and DNA methylation associated with human disease. *Nucleic Acids Res.* 46, D133–D138. doi: 10.1093/nar/gkx985

Zhou, B., Zhao, H., Yu, J., Guo, C., Dou, X., Song, F., et al. (2018). EVLncRNAs: a manually curated database for long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.* 46, D100–D105. doi: 10.1093/nar/gkx677

Zhou, K. R., Liu, S., Sun, W. J., Zheng, L. L., Zhou, H., Yang, J. H., et al. (2017). ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.* 45, D43–D50. doi: 10.1093/nar/gkw965

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595