Functional & Integrative Genomics (2022) 22:1105–1112
https://doi.org/10.1007/s10142-022-00915-y

REVIEW

Computational approaches and challenges for identification and annotation of non-coding RNAs using RNA-Seq

Kiran Dindhoria[1] · Isha Monga[2] · Amarinder Singh Thind[3,4]

Received: 11 July 2022 / Revised: 4 November 2022 / Accepted: 4 November 2022 / Published online: 21 November 2022

Abstract

Significant innovations in next-generation sequencing techniques and bioinformatics tools have impacted our appreciation and understanding of RNA. Practical RNASeq applications have evolved in conjunction with sequence technology and bioinformatic tool advances. In this review, we explained various computational resources, tools, and bioinformatics analyses advancement for small and large non-coding RNAs. These include non-coding RNAs (ncRNAs) such as piwi, micro, circular, and long ncRNAs. In addition, this article discusses future challenges, single-cell level sequencing for non-coding RNAs, and advantages of using long-read sequencing to annotate lncRNAs.

Keywords Next-generation sequencing · RNAseq applications · Bioinformatics advancements · Transcriptomics · Omics

Introduction

In the recent era, next-generation sequencing (NGS) is a vastly used method to answer various biological research questions at different omics levels such as genomics, epigenomics, transcriptomics, metabolomics, etc. For transcriptomics research, RNA-Seq was a revolution (Mortazavi et al. 2008) (Pan et al. 2008) due to its ability to measure the transcription of 90% of the genomic DNA in eukaryotes and can also be utilized for other variety of analyses (Copy number alterations, TWAS, neoantigen pre) and to reveal complex events (Thind et al. 2021). Most of the transcribed DNA includes RNAs without any coding capacity, commonly known as the non-coding RNAs (ncRNAs). With the evolution of species, the approximate amount of coding genes remains the same while the number of non-coding sequences rises with the increase in organism complexity (Amaral and Mattick 2008). The fact that most ncRNAs are expressed at much lower levels as compared to mRNAs indicates that ncRNAs are primarily playing role in regulation of the gene expression (Geisler and Coller 2013). ncRNAs consist of ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs), long ncRNAs (lncRNAs), small interfering RNAs (siRNAs), microRNAs (miRNA), PIWI-interacting RNAs (PiRNAs), circular RNAs, etc. Studies have shown the role of ncRNAs in the occurrence and regulation of normal physiological processes (Dinger et al. 2008), regulation of gene expression (Holoch and Moazed 2015), and human diseases (Fig. 1). Moreover, genetic and epigenetic deformities in miRNAs or their machinery may cause many diseases (Wang et al. 2013a) and it has applications in forensic science (Rocchi et al. 2020). However, ncRNAs role is mainly studied in cancer (Huarte 2015) and cardiovascular diseases (Fang et al. 2020). The abnormal expression of LncRNA's can cause the development and progression of cancer (Mercer et al. 2009; Hauptman and Glavač 2013). Circular RNAs and piwi RNAs are also known to play role in cardiovascular diseases (Altesha et al. 2019; Zeng et al. 2021). It is essential to recognize the full repertoire of available ncRNAs to understand their regulatory function with respect to normal developmental processes and human diseases. Expression of ncRNA varies among healthy cell types, proven by various studies such as LncRNA expression ability to resolve various cell types using single cell RNAseq (scRNASeq) (Mortazavi

✉ Amarinder Singh Thind
athind@uow.edu.au

[1] Institute of Microbial Technology, Council of Scientific and Industrial Research, Chandigarh, India

[2] Department of Dermatology, Columbia University Irving Medical Center, New York, NY, USA

[3] Graduate school of Medicine, University of Wollongong, Wollongong, Australia

[4] Illawarra Health and Medical Research Institute, Wollongong, Australia

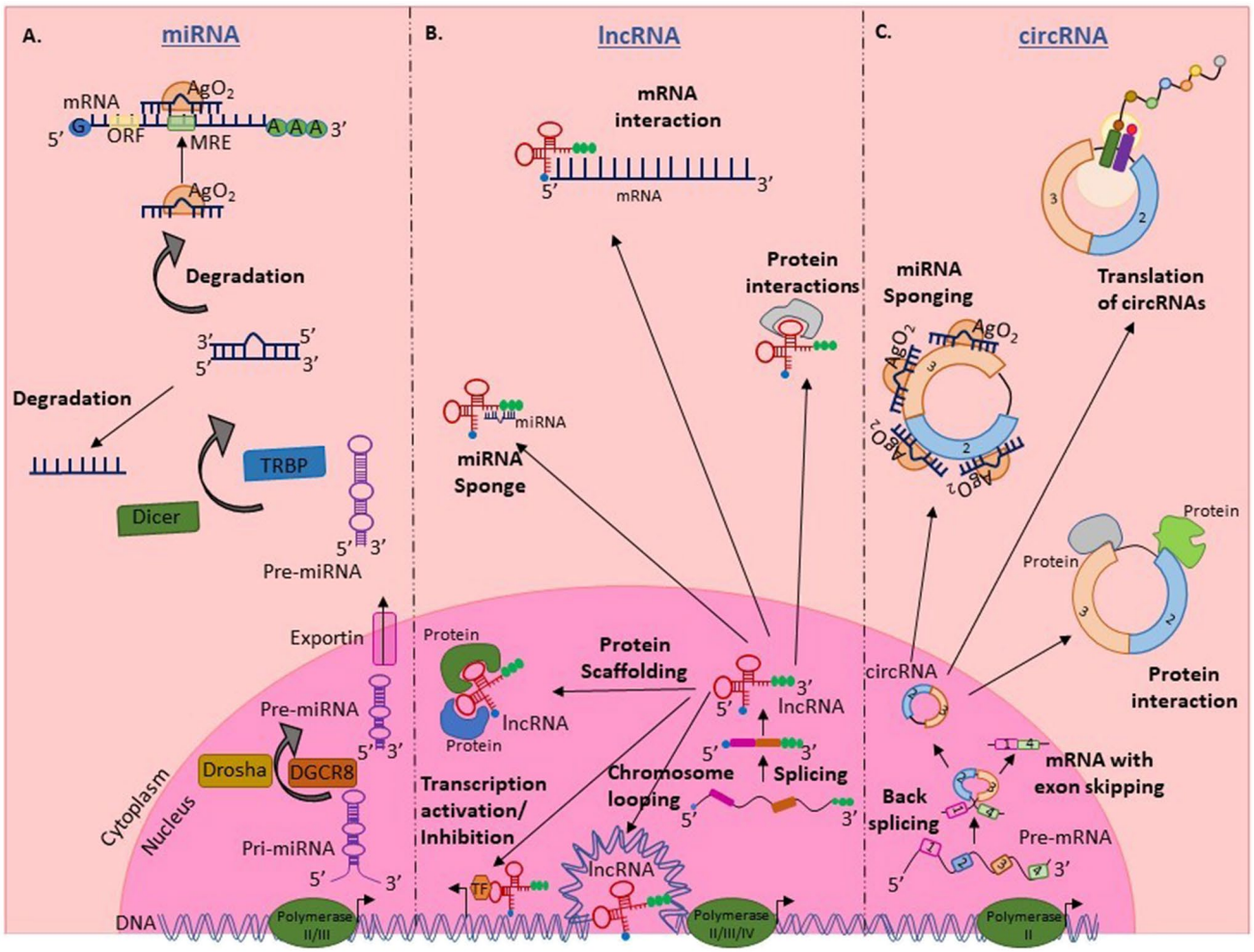

Fig. 1 Schematic representation of biosynthesis of different types of non-coding RNAs and their applications. A miRNAs. B lncRNAs. C circRNAs

et al. 2008). The recognition of total available ncRNAs is difficult to achieve via contemporary molecular biology techniques.

However, third generation sequencing has provided advancement in this field to discover and characterize the role of ncRNA. In the past few years, the submission of RNA-seq datasets eventually increased exponentially in public databases, which can be reutilized for many novel analyses using advanced bioinformatics methods. Here, in this review, we discuss the resources of ncRNASeq data and the advanced tool to analyze ncRNAseq data. Furthermore, we discussed some key challenges that needed to be solved.

Types of non-coding RNAs

There is growing appreciation and understanding of non-coding regions' roles in gene function and expression. Many non-coding RNA (ncRNA) molecules have been reported and attributed to various roles depending upon their length, function, action, location of transcription, etc. Examples of these ncRNA molecules microRNA (miRNA), small-interfering RNAs (siRNAs), small nuclear RNAs (snRNAs), small-nucleolar RNAs (snoRNAs), Piwi-interacting RNAs (piRNAs), long ncRNAs (lncRNAs), and circular ncRNA.

Long non-coding RNA and circular RNA

Among all ncRNA molecules, lncRNAs are the most versatile, critical molecule implicated in diverse gene regulatory processes (Mercer et al. 2009; Rinn and Chang 2012). LncRNAs, in the broadest sense, are defined as a novel class of functional ncRNAs of length > 200 nucleotides (Ma et al. 2013). They play a significant role in regulating gene expression (both at transcriptional/post-transcriptional level), chromatin remodeling, protein localization, etc. (Jarroux et al. 2017). They modulate target gene expression at transcriptional (via RNA–DNA hybrid), post-transcriptional, epigenetic

modifier by interacting with chromatin complexes, and at the level of 3D-genome structure. There are various kinds of lncRNAs molecules based on their length, as proposed by Amarar et al. In lncRNAdb (Amaral et al. 2011). Later, these were classified based on their position, the direction of transcription (relative to protein-coding genes), and their shape such as long intergenic non-coding RNAs (lincRNAs): lncRNA transcribed between the exons; long intronic non-coding RNAs (lncRNAs): lncRNA from the intronic regions; Telomere-associated ncRNAs (TERRAs): lncRNAs from the telomere region; Transcribed-ultraconserved regions (T-UCRs), antisense lncRNAs; and circular RNAs (circRNAs).

CircRNAs are the most recently identified form of lncRNAs and are formed by back splicing. They can be classified into intronic, single exon, multiple exons, and intronic exonic based on which parts are parts of the gene are retained in the closed circular. circRNAs are usually cell-type specific and more stable than linear RNAs because exonucleases cannot easily degrade their closed circular structure. They are excellent competing endogenous RNAs (ceRNAs) acting as a miRNA sponge to compete with the host cell's miRNAome and allow gene expression. A major proportion of lncRNAs reside in intergenic regions of the genome, so they are termed as long intergenic ncRNA (lincRNAs). lincRNAs target their gene through long-range chromatin looping; they do not need cohesion proteins to close the 3D chromatin loops. These studies opened an exciting function of lincRNAs as a linker for 3D-genome structure maintenance. Given these alterations in lncRNAs expression promotes disease states like tumor formation, progression, and metastasis. Increased knowledge of the molecular mechanisms of lncRNAs could provide novel therapeutic targets for treating various diseases like autoimmune disorders, cancers, and viral infections.

Small non-coding RNA analysis

In general RNAseq applications lie in identifying and characterizing two types of small RNAs, i.e., MicroRNA and Piwi-RNA. Complementary DNA library construction protocol is widely used for small non-coding RNAseq, but it generates a bias in the sequencing results, partially due to RNA modifications. RNA modifications interfere with adapter ligation and reverse transcription processes and prevent the detection of sncRNAs bearing these modifications (Shi et al. 2021).

miRNA

MicroRNAs are 20–22 bp long small noncoding RNAs. It acts as a post-transcriptional regulator by interacting with mRNA, lncRNA, and circular RNA molecules. miRNAs also serve as biomarkers in various diseases, including cancer. NGS-based miRNA analyses evolved very fast. For example, the alignment of the short reads of miRNA (20–22 bp) to the reference was a bit challenging, but much work is done so far to solve this issue (Ziemann et al. 2016). Currently, many advanced command-line and user-friendly pipelines are available to carry out miRNA analyses (Aparicio-Puerta, et al. 2019; Riffo-Campos et al. 2016; Chen et al. 2019a; Farrell 2017; Liu et al. 2021). These pipelines can identify novel miRNA, predict miRNA structure, and perform expression quantification and differential expression analyses (Bortolomeazzi et al. 2019). Each of these steps has its importance, e.g., miRNA's secondary structure can have a conformational role in modulating miRNA-mRNA interactions. Expression quantification is required for differential expression of miRNA.

PIWI-RNA

PIWI-interacting RNA (piRNA) is a recently discovered class of small ncRNAs with ~ 19–33 nt in length implicated in gene regulation of transposable elements (TEs) in the germline cells (Cox et al. 1998) and amongst the most abundant small ncRNAs in the germline cells (Wang et al. 2019). As mentioned before, the eukaryotic genomes encode millions of copies of selfish DNA elements like repetitive sequences, transposons, SINE, LINE, etc. (Ünsal and Morgan 1995; Ernst et al. 2017; Jurka 2000; Frith et al. 2005). The TEs' movement plays an important role in genomic evolution by creating novel genes, diverse immune responses viz V(D)J recombination for MHC-alleles. These movements need to be controlled because unrestricted movement creates a threat to genomics integrity, which could cause deadly diseases like cancers, autoimmunity, and genetic disorders (Castañeda et al. 2011). piRNA sequences are loaded onto the germline argonaute (AGO) proteins termed as PIWI (PIWI) proteins and regulate the TEs expression (Castañeda et al. 2011). The PIWI proteins were first identified by their roles in maintaining (Cox et al. 2000) and patterning (Wilson et al. 1996).

The new small RNA-seq technologies and advanced bioinformatics tools have contributed to the piRNA repertoire's growth that helped to improve prediction tools for novel piRNA (Jensen et al. 2020). Experimentally known piRNAs datasets are usually used to train classifiers that predict piRNA sequences from the genome. Basic features used for training the algorithms are nucleotide usage, physicochemical properties, RNA secondary structure, etc. After training, these algorithms' performance is assessed to predict novel piRNA sequences using statistical measures like sensitivity, specificity, and Matthews Correlation coefficient (MCC). Many tools and databases for prRNA published recently have various functions such as piRNA prediction,

identification of novel piRNA, can differentiate transposon-derived piRNAs from non-piRNAs, identification of Piwi-Interacting RNAs, database of piRNA, detection of piRNA-mediated transposon-silencing and discriminate the siRNAs and piRNAs, etc.

Computational tools for identification, annotations, and analysis of ncRNAs

The recent transcriptomics methods and current computational resources in the ncRNA field are helping to improve the classification, annotation, and analysis of ncRNAs, helping the scientific community identify, annotate, store, predict, and analyze ncRNA data (Thind et al. 2022). In Table 1, we summarized various computational tools based on non-coding RNA types and functions.

Future prospects and challenges of non-coding RNAs

Despite advancements in the technology of RNA sequencing, there are several technical challenges in this field that are needed to be addressed. For instance, the expression of ncRNAs is generally restricted to a specific cell lineage and are expressed in lower amount as compared to that of the other genes (Guttman et al. 2010; Iyer et al. 2015). Due to lower expression levels, their exact quantification is very difficult to achieve thus impacting the differential analysis studies (Everaert et al. 2017). In order to obtain the proper quantification via differential expression studies higher sequence coverage is required (e.g., 100–200 million reads using total-RNA-Seq library for deep whole transcriptomics analysis of human RNA-Seq data). Another challenge is in dealing with natural antisense transcripts which are widespread in the class of lncRNAs (Pasmant et al. 2007; Beltran et al. 2008). The antisense transcripts of lncRNAs and miRNAs with overlapping exons on the opposite gene strand are also difficult to count. To deal with such issues various computational methods are developed to correctly identify antisense transcription utilizing the information of location and orientation of splicing sites and poly(A) tails (Lorenzi et al. 2019).

Both the lncRNAs and circRNAs have also been observed to be present in extracellular vesicles (EVs) secreted by diseased cells. The RNA content of these vesicles generally acts as biomarkers for a particular disease (Mohankumar and Patel 2016; Hinger et al. 2018; Li et al. 2020). However, the amount of the RNA present in EVs is very less. Thus, again making it challenging to quantify. Similarly, the detection of ncRNA in a single cell is also difficult due to lower abundance. In order to carry out accurate prediction and quantification, the detection methods like (lnc) RNAs capture sequencing techniques have also been developed. This process involves biotinylated probes for capturing the target (lnc)RNAs, improving the coverage for low-abundant lncRNAs (Kato and Carninci 2020). In normal RNA-seq library preparations, CircRNAs get depleted at poly(A) enrichment step because it lacks a poly(A) tail. However, they are found to retain in rRNA-depleted libraries and libraries treated with RNAse R degraded linear RNAs. The RNase R treatment followed by RT-quantitative PCR (qPCR), is a popular experimental strategy for validating the circRNAs obtained from rRNA-depleted samples thus allowing the targeted confirmation of true positives (Szabo and Salzman 2016).

Table 1 Bioinformatics tools used for the identification, prediction, and annotation of different types of non-coding RNAs

S. No	Type of non-coding RNA	Functions	Bioinformatic tools	References
1	lncRNAs	Annotation	GENCODE, NONCODE, lnciPedia, FEELnc, LncRNA-ID	Kawai et al. 2001; Zhao et al. 2016; Volders et al. 2013; Sun et al. 2015; Han et al. 2019; Liu et al. 2019; Ge et al. 2016; Baek et al. 2018; Baek et al. 2018; Xu et al. 2020; Fan and Zhang 2015; Wucher et al. 2017; Achawanantakun et al. 2015)
		Identification	lncRscan-SVM, LncFinder, LncRNAnet, LncRNA-MFDL	
		Prediction	LPBNI, PredLnc-GFStack	
2	Circular RNAs	Annotation	CircTest, CIRCexplorer, DEBKS	Gao et al. 2018; Cheng et al. 2016; Song et al. 2016; Zhang et al. 2016; Pan and Xiong 2015; Wang and Wang 2019; Zhang et al. 2020; Liu et al. 2021)
		Identification	CIRI2, DCC, UROBORUS, Predcir-cRNA, CIRIquant	
		Prediction	DeepCirCode	
3	miRNA	Annotation	miRViz, miEAA	Agarwal et al. 2015; Kertesz et al. 2007; Betel et al. 2010; Ru et al. 2014; Quillet et al. 2020; Backes et al. 2016; Giroux et al. 2020)
		Identification	miRanda	
		Prediction	Targetscan, PITA, MultiMir, miRabel	
4	Piwi-RNA	Annotation	PIANO, piRNA, RAPID	Liu et al. 2014; Monga and Banerjee 2019; Li et al. 2016; Boucheham et al. 2017; Han et al. 2015; Pogorelcnik et al. 2018; Uhrig and Klein 2019; Karunanithi et al. 2019)
		Identification	sRNAPipe, PingPongPro	
		Prediction	IpRId, PILFER	

Neither CircRNAs nor lncRNAs have a standard naming convention. The naming of lncRNAs is mostly based on their functions, structures, and mechanisms of action(Gong et al. 2021). The same circRNA is called by different names in various circDatabases. For instance, circBase takes into account species and numeric code and other proposed new naming based on genomic coordinates (e.g., chr10:126,970,702|127,127,764), which is also inconsistent since reference genomes are updated periodically, and newly developed databases could use hg38/others instead of hg19/old. In addition, the naming of genomic coordinates can be influenced by the zero/one index formats (chr10:126,970,701|12,712,776 could also be named as chr10:126,970,702|127,127,766). Based on genomic coordinates from UCSC resources, circBank and circAtlas use gene symbols to identify transcriptional units that generate circRNAs; however, there may be discrepancies in the names in these databases due to the non-consistent transcriptional unit defined for a particular gene (e.g., hsa_circAEBP2_003 in circAtlas could be hsa_circAEBP2_001 or hsa_circAEBP2_002 in circBank).

With deep sequencing technologies, both known and novel miRNAs can now be detected at a large scale. As most organisms do not have their genomes completely sequenced, even mapping reads to genomes can be challenging. Although there are several tools available for miRNA profiling, some of them are already mentioned in Table 1. These methods depend on databases consisting of known miRNAs thus the accuracy of predicted novel miRNAs is still questionable. Also, it is observed that many different sequences can be produced from a single miRNA locus. These variable length short sequences may have various 5′ and 3′ ends as compared to that of the miRNA reads stored in public databases. They may possess the regulatory activity and require more of the properly stored information in the form of databases such as already existing databases: YM500 (Cheng et al. 2013) and isomiRex (Sablok et al. 2013). On the other hand, the limitation with piRNAs is the absence of a reliable and efficient method for the detection in tissues other than the germline. Due to the lack of proper databases, the detection and characterization of piRNAs in somatic cells are still difficult. Similar to miRNA isomers, identical piRNA sequences are produced from multiple loci thus adding to the higher complexity and lower precision of the generated data (Geles et al. 2021).

Single-cell and long-read technology for non-coding RNAs sequencing

Single-cell RNA-Seq (ScRNA-seq) is a very recent and transformative technology. With the help of single-cell RNASeq, the role of non-coding RNA in cell specificity (Gawronski and Kim 2017), embryonic development (Fu 2018), and cell reprogramming has been revealed (Luginbühl et al. 2017). It is used to search for the answers which bulk RNA sequencing cannot give, for instance, it helps in the gene expression analysis of an individual cell among the group of cells. The non-coding RNAs (ncRNAs) play an important role in the differentiation of cells by changing the overall genomic program in a small subset of the cells. They are also expressed in lower amounts, transiently expressed, or in association with transcription events involved in regulatory processes. Therefore, they cannot be easily detected by the bulk RNA-seq analysis and require single cell transcriptome sequencing to evaluate their role in a particular type of cell. Traditional approaches for sequencing small RNAs required a huge amount of cell material that limits the possibilities for single-cell analyses.

Recently, various single-cell specific protocols for non-coding RNAs are being developed. CAS-seq and Small-seq are single-cell small RNA sequencing method that enables the capture, sequencing, and molecular counting of small RNAs (Yang et al. 2019; Hagemann-Jensen et al. 2018). Small-seq is a ligation-based approach. Not only sequencing protocols are advanced but also tools specific to single-cell data are evolving, e.g., miReact software infers miRNA activities from single-cell mRNAseq that use motif enrichment analysis to derive miRNA activity estimates from scRNAseq data (Nielsen and Pedersen 2021). With the availability of long-read sequencing technologies, there is an improvement in the current annotations and large-scale initiatives are taken to complete the human lncRNA transcriptome map (Uszczynska-Ratajczak et al. 2018). lncRNAs are probably the most beneficial class of transcripts that would have improved annotation using long-read sequencing technology. Compared to protein-coding genes, long non-coding RNAs (lncRNAs) annotations are poorly characterized due to trade-offs between quality and size, often unappreciated consequences for downstream studies. Furthermore, the impact of short and long-read sequencing on the identification of lncRNAs in humans and plants is documented (Chiquitto et al. 2022) where a significant improvement in annotations of lncRNA in humans is observed using tools such as CPAT (Wang et al. 2013b), RNAmining (Ramos et al. 2021), lncRNAnet (Baek et al. 2018), and LncADeep (Yang et al. 2021). The ScRNA-seq has shown applications in the identification of the role of non-coding RNA in gene regulatory networks (Zhao et al. 2022), cell specificity (Gawronski and Kim 2017), embryonic development (Fu 2018), and cell reprogramming (Luginbühl et al. 2017).

ScRNA-seq is a powerful tool to study the expression and regulation of cell-specific ncRNAs. However, current single-cell sequencing methods are not well optimized, so many limitations and issues exist. For example, a very small amount of the starting material is generally obtained

for scRNA-seq causing lower capture efficiency and higher dropouts, thus leading to the detection of a minority of expressed genes (Hwang et al. 2018). Since the ncRNAs have lower expression so the dropout events may have prominent effects on the analysis. scRNA-seq produces noisier and more complex sequencing data as compared to the bulk RNA-seq data, thus making the computational analysis of the data difficult. The batch effects caused due to slight variations in sample preparations are generally found. Besides, biological variations due to the state of the cell, size, cycle, etc. also affect the transcriptomic analysis. To minimize both the technical and the biological errors, repeated analysis of multiple cells is required. To resolve this, recently, scLVM59 approach was developed to minimize the errors caused by the latent variables (Chen et al. 2019b). In gene regulatory network analysis, many different tools such as SCODE (Matsumoto et al. 2017), SCGRNs (Turki and Taguchi 2020), scGNN (Wang et al. 2021), etc., but none of these have been tested for ncRNAs gene regulatory network mapping. In this regard, multi-omics data integration may be helpful as it cross-validates the regulatory interactions in multiple datasets (Hu et al. 2020). Based on the fact that ncRNAs are emerging players in cell differentiation, interactions, and reprogramming and are less explored as compared to single-cell mRNA and bulk RNA, their investigation in a specific type of cell would provide a new outlook in near future.

Acknowledgements The authors would like to acknowledge the authors of various tools discussed here in the article.

Author contribution K.D, I.M, and A.S.T wrote, edited, and reviewed the original review article; KD prepared the figure and table.

Data availability All the relevant data discussed in the article is provided in the article.

Declarations

Consent for publication Not applicable.

Human and animal ethics Not applicable.

Competing interests The authors declare no competing interests.

References

Achawanantakun R et al (2015) LncRNA-ID: Long non-coding RNA identification using balanced random forests. Bioinform 31(24):3897–3905

Agarwal V et al (2015) Predicting effective microRNA target sites in mammalian mRNAs. Elite 4:e05005

Altesha MA et al (2019) Circular RNA in cardiovascular disease. J Cell Physiol 234(5):5588–5600

Amaral PP, Mattick JS (2008) Noncoding RNA in development. Mamm Genome 19(7):454–492

Amaral PP et al (2011) lncRNAdb: A reference database for long noncoding RNAs. Nucleic Acid Res 39(1):D146–D151

Aparicio-Puerta E et al (2019) sRNAbench and sRNAtoolbox 2019: Intuitive fast small RNA profiling and differential expression. Nucleic Acids Res 47(1):W530–W535

Backes C et al (2016) miEAA: microRNA enrichment analysis and annotation. Nucleic Acids Res 44(W1):W110–W116

Baek J et al (2018) LncRNAnet: Long non-coding RNA identification using deep learning. Bioinform 34(22):3889–3897

Baek J et al (2018) LncRNAnet: Long non-coding RNA identification using deep learning. Bioinform 34(22):3889–3897

Beltran M et al (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition. Genes Dev 22(6):756–769

Betel D et al (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 11(8):1–14

Bortolomeazzi M, Gaffo E, Bortoluzzi S (2019) A survey of software tools for microRNA discovery and characterization using RNA-seq. Brief Bioinform. 20(3):918–930

Boucheham A et al (2017) IpiRId: Integrative approach for piRNA prediction using genomic and epigenomic data. Plos One 12(6):e0179787

Castañeda J et al (2011) piRNAs, transposon silencing, and germline genome integrity. Mutat Res/Fundam Mol Mech Mutagen 714(1–2):95–104

Chen L et al (2019) Trends in the development of miRNA bioinformatics tools. Brief Bioinform 20(5):1836–1852. https://doi.org/10.1093/bib/bby054

Chen G, Ning B, Shi T (2019b) Single-cell RNA-seq technologies and related computational data analysis. Front Genet317

Cheng W-C et al (2013) YM500: A small RNA sequencing (smRNA-seq) database for microRNA research. Nucleic Acids Res 41(D1):D285–D294

Cheng J, Metge F, Dieterich CJB (2016) Specific Identification and Quantification of Circular RNAs from Sequencing Data. Bioinform 32(7):1094–1096

Chiquitto AG et al (2022) Impact of sequencing technologies on long non-coding RNA computational identification. BioRxiv. https://doi.org/10.1101/2022.04.15.488462

Cox DN et al (1998) A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. Genes Dev 12(23):3715–3727

Cox DN, Chao A, Lin HJD (2000) Piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development 127(3):503–514

Dinger ME et al (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 18(9):1433–1445

Ernst C, Odom DT, Kutter C (2017) The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. Nat Commun 8(1):1–10

Everaert C et al (2017) Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Sci Rep 7(1):1–11

Fan XN, Zhang SW (2015) lncRNA-MFDL: Identification of human long non-coding RNAs by fusing multiple features and using deep learning. Mol BioSyst 11(3):892–897

Fang Y et al (2020) Recent advances on the roles of LncRNAs in cardiovascular disease. J Cell Mol Med 24(21):12246–12257

Farrell D (2017) Smallrnaseq: short non coding RNA-seq analysis with Python. Biorxiv :110585. https://doi.org/10.1101/110585

Frith MC, Pheasant M, Mattick JS (2005) The amazing complexity of the human transcriptome. Eur J Hum Genetics 13(8):894–897

Fu Q et al (2018) Single-cell non-coding RNA in embryonic development. Single Cell Biomed :19–32. https://doi.org/10.1007/978-981-13-0502-3_3
Gao Y, Zhang J, Zhao F (2018) Circular RNA identification based on multiple seed matching. Brief Bioinform 19(5):803–810
Gawronski KA, Kim J (2017) Single cell transcriptomics of non-coding RNAs and their cell-specificity. Wiley Interdiscip Rev RNA 8(6):e1433
Ge M et al (2016) A bipartite network-based method for prediction of long non-coding RNA–protein interactions. Genomics Proteomics Bioinformatics 14(1):62–71
Geisler S, Coller J (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 14(11):699–712
Geles K et al (2021) WIND (Workflow for pIRNAs aNd beyonD): a strategy for in-depth analysis of small RNA-seq data. F1000Res 10:1. https://doi.org/10.12688/f1000research.27868.3
Giroux P et al (2020) miRViz: A novel webserver application to visualize and interpret microRNA datasets. Nucleic Acids Res 48(W1):W252–W261
Gong Y et al (2021) Bioinformatics analysis of long non-coding RNA and related diseases: An overview. Front Genet 12:813873. https://doi.org/10.3389/fgene.2021.813873
Guttman M et al (2010) Ab initio reconstruction of cell type–specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28(5):503–510
Hagemann-Jensen M et al (2018) Small-seq for single-cell small-RNA sequencing. Nat Protoc 13(10):2407–2424
Han BW et al (2015) piPipes: A set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-Seq, Degradome-and CAGE-Seq, ChIP-Seq and genomic DNA sequencing. Bioinformatics 31(4):593–595
Han S et al (2019) LncFinder: An integrated platform for long non-coding RNA identification utilizing sequence intrinsic composition, structural information and physicochemical property. Brief Bioinform 20(6):2009–2027
Hauptman N, Glavač D (2013) Long non-coding RNA in cancer. Int J Mol Sci 14(3):4655–4669
Hinger SA et al (2018) Diverse long RNAs are differentially sorted into extracellular vesicles secreted by colorectal cancer cells. Cell Rep 25(3):715–725
Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 16(2):71–84
Hu X et al (2020) Integration of single-cell multi-omics for gene regulatory network inference. Comput Struct Biotechnol J 18:1925–1938
Huarte M (2015) The emerging role of lncRNAs in cancer. Nat Med 21(11):1253–1261
Hwang B, Lee JH, Bang D (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50(8):1–14
Iyer MK et al (2015) The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 47(3):199–208
Jarroux J, Morillon A, Pinskaya M (2017) History, discovery, and classification of lncRNAs. Adv Exp Med Biol 1008:1-46
Jensen S et al (2020) Conserved small nucleotidic elements at the origin of concerted piRNA biogenesis from genes and lncRNAs. Cells 9(6):1491
Jurka J (2000) Repbase update: a database and an electronic journal of repetitive elements. Trend Genet 16(9):418–420
Karunanithi S, Simon M, Schulz MHJP (2019) Automated Analysis of Small RNA Datasets with RAPID. PeerJ 7:e6710
Kato M, Carninci P (2020) Genome-wide technologies to study RNA–chromatin interactions. Noncoding RNA 6(2):20
Kawai J et al (2001) Functional annotation of a full-length mouse cDNA collection. Nature 409(6821):685–689
Kertesz M et al (2007) The role of site accessibility in microRNA target recognition. Nat Genet 39(10):1278–1284
Li D et al (2016) A genetic algorithm-based weighted ensemble method for predicting transposon-derivedd piRNAs. BMC Bioinform 17(1):1–11
Li Z, Zhu X, Huang S (2020) Extracellular vesicle long non-coding RNAs and circular RNAs: Biology, functions and applications in cancer. Cancer Lett 489:111–120
Liu X, Ding J, Gong J (2014) piRNA identification based on motif discovery. Mol BioSyst 10(12):3075–3080
Liu Q et al (2021) Small Noncoding RNA Discovery and Profiling with sRNAtools Based on High-Throughput Sequencing. Brief Bioinform 22(1):463–473
Liu Z et al (2021) DEBKS: A tool to detect differentially expressed circular RNA
Liu S et al (2019) PredLnc-GFStack: A global sequence feature based on a stacked ensemble learning method for predicting lncRNAs from transcripts. Genes (Basel) 10(9):672
Lorenzi L et al (2019) Long noncoding RNA expression profiling in cancer: Challenges and opportunities. Genes Chromosom Cancer 58(4):191–199
Luginbühl J, Sivaraman DM, Shin JW (2017) The essentiality of non-coding RNAs in cell reprogramming. Noncoding RNA Res 2(1):74–82
Ma L, Bajic VB, Zhang Z (2013) On the classification of long non-coding RNAs. RNA Biol 10(6):924–933
Matsumoto H et al (2017) SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics 33(15):2314–2321
Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: Insights into functions. Nat Rev Genet 10(3):155–159
Mohankumar S, Patel T (2016) Extracellular vesicle long noncoding RNA as potential biomarkers of liver cancer. Brief Funct Genomics 15(3):249–256
Monga I, Banerjee I (2019) Computational identification of piRNAs using features based on rna sequence, structure, thermodynamic and physicochemical properties. Curr Genom 20(7):508–518
Mortazavi A et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628
Nielsen MM, Pedersen JS (2021) miRNA activity inferred from single cell mRNA expression. Sci Rep 11(1):1–8
Pan X, Xiong K (2015) PredcircRNA: Computational classification of circular RNA from other long non-coding RNA using hybrid features. Mol Biosyst 11(8):2219–2226
Pan Q et al (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40(12):1413–1415
Pasmant E et al (2007) Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. Can Res 67(8):3963–3969
Pogorelcnik R et al (2018) sRNAPipe: a Galaxy-based pipeline for bioinformatic in-depth exploration of small RNAseq data. Mobile DNA 9(1):1–6
Quillet A et al (2020) Improving Bioinformatics Prediction of microRNA Targets by Ranks Aggregation. Front Genet 10:1330
Ramos TA et al (2021) RNAmining: A machine learning standalone and web server tool for RNA coding potential prediction. F1000Res 10:323. https://doi.org/10.12688/f1000research.52350.2
Riffo-Campos ÁL, Riquelme I, Brebi-Mieville P (2016) Tools for sequence-based miRNA target prediction: What to choose? Int J Mol Sci 17(12):1987
Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166

Rocchi A et al (2020) MicroRNAs: An update of applications in forensic science. Diagnostics 11(1):32
Ru Y et al (2014) The multiMiR R package and database: Integration of microRNA–target interactions along with their disease and drug associations. Nucleic Acids Res 42(17):e133–e133
Sablok G et al (2013) isomiRex: Web-based identification of microRNAs, isomiR variations and differential expression using next-generation sequencing datasets. FEBS Lett 587(16):2629–2634
Shi J et al (2021) PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. Nat Cell Biol 23(4):424–436
Song X et al (2016) Circular RNA profile in gliomas revealed by identification toolUROBORUS. Nucleic Acids Res 44(9):e87–e87
Sun L et al (2015) lncRScan-SVM: A tool for predicting long non-coding RNAs using support vector machine. Plos One 10(10):e0139654
Szabo L, Salzman J (2016) Detecting circular RNAs: Bioinformatic and experimental challenges. Nat Rev Genet 17(11):679–692
Thind AS et al (2021) Demystifying emerging bulk RNA-Seq applications: The application and utility of bioinformatic methodology. Brief Bioinform 22(6):bbab259
Thind AS, Kaur K, Monga I (2022) An overview of databases and tools for lncrna genomics advancing precision medicine. Mach Learn Syst Biol Genomics Health :49–67. https://doi.org/10.1007/978-981-16-5993-5_3
Turki T, Taguchi Y (2020) SCGRNs: Novel supervised inference of single-cell gene regulatory networks of complex diseases. Comput Biol Med 118:103656
Uhrig S, Klein H (2019) PingPongPro: A tool for the detection of piRNA-mediated transposon-silencing in small RNA-Seq data. Bioinform 35(2):335–336
Ünsal K, Morgan GT (1995) A novel group of families of short interspersed repetitive elements (SINEs) inXenopus: Evidence of a specific target site for dna-mediated transposition of inverted-repeat SINEs. J Mol Biol 248(4):812–823
Uszczynska-Ratajczak B et al (2018) Towards a complete map of the human long non-coding RNA transcriptome. Nat Rev Genet 19(9):535–548
Volders PJ et al (2013) LNCipedia: A database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 41(D1):D246–D251
Wang J, Wang LJB (2019) Deep learning of the back-splicing code for circular RNA formation. Bioinform 35(24):5235–5242
Wang Y et al (2013a) The role of miRNA-29 family in cancer. Eur J Cell Biol 92(3):123–128
Wang L et al (2013b) CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res 41(6):e74–e74
Wang J et al (2019) piRBase: A comprehensive database of piRNA sequences. Nucleic Acids Res 47(D1):D175–D180
Wang J et al (2021) scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. Nat Commun 12(1):1–11
Wilson JE, Connell JE, Macdonald PM (1996) aubergine enhances oskar translation in the Drosophila ovary. Development 122(5):1631–1639
Wucher V et al (2017) FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Res 45(8):e57–e57
Xu Y et al (2020) Predicting long non-coding RNAs through feature ensemble learning. BMC Genom 21(13):1–12
Yang Q et al (2019) Single-cell CAS-seq reveals a class of short PIWI-interacting RNAs in human oocytes. Nat Commun 10(1):1–15
Yang C et al (2021) LncADeep performance on full-length transcripts. Nat Mach Intell 3(3):197–198
Zeng Q et al (2021) PIWI-interacting RNAs and PIWI proteins in diabetes and cardiovascular disease: Molecular pathogenesis and role as biomarkers. Clin Chim Acta 518:33–37
Zhang X-O et al (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. Genome Res 26(9):1277–1287
Zhang J et al (2020) Accurate quantification of circular RNAs identifies extensive circular isoform switching events. Nat Commun 11(1):1–14
Zhao Y, Yuan J, Chen R (2016) NONCODEv4: Annotation of noncoding RNAs with emphasis on long noncoding RNAs. Long Non-Coding RNAs. Springer, pp 243–254
Zhao X, Lan Y, Chen D (2022) Exploring long non-coding RNA networks from single cell omics data. Comput Struct Biotechnol J 20:4381–4389. https://doi.org/10.1016/j.csbj.2022.08.003
Ziemann M, Kaspi A, El-Osta AJR (2016) Evaluation of microRNA alignment techniques. RNA 22(8):1120–1138