International Journal Of Latest Technology In Engineering & Management (IJLTEM)
Volume 9- Issue 5 {September-October 2024} {Page: 28-32}

Comprehensive Overview of Bioinformatics Tools and Databases for Small and Long Non-Coding RNAs in the Human Genome

Dr. Vinay Kumar Singh
Information Officer, Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh 221005

ABSTRACT: *Non-coding RNAs (ncRNAs), including small non-coding RNAs (sncRNAs) and long non-coding RNAs (lncRNAs), are key regulators of gene expression, epigenetic modifications, and cellular processes within the human genome. The rapid advancements in next-generation sequencing (NGS) technologies have accelerated the development of numerous bioinformatics tools and databases tailored for the discovery, characterization, and functional analysis of ncRNAs. This manuscript provides a comprehensive review of the most widely used bioinformatics tools, such as BLAST, RNAfold, miRDeep, CNCI, lncRNApred, and integrative platforms like miRNet, ShinyGO, and iDEP. These tools play pivotal roles in RNA structure prediction, miRNA discovery, coding versus non-coding RNA differentiation, and gene enrichment analysis. Understanding the utility and application of these tools allows researchers to gain deeper insights into the biological roles of ncRNAs, particularly in the context of gene regulation, RNA interactions, and disease mechanisms. This review emphasizes the growing significance of these computational resources in the realm of genomic research, especially for decoding the complexities of non-coding RNA biology.*

Keywords: *Non-coding RNA, small non-coding RNA, long non-coding RNA, bioinformatics tools, RNA databases, RNA structure prediction, miRNA, lncRNA, gene enrichment, RNA networks*

INTRODUCTION

The human genome, with approximately 3.2 billion base pairs, is a vast repository of genetic information, much of which does not code for proteins. Despite early assumptions that non-protein-coding regions were functionally insignificant, recent research has highlighted the crucial roles of non-coding RNAs (ncRNAs) in regulating various biological processes. ncRNAs, including small non-coding RNAs (sncRNAs) such as microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and small interfering RNAs (siRNAs), along with long non-coding RNAs (lncRNAs), are involved in regulating gene expression, RNA splicing, chromatin remodeling, and other essential cellular functions.

While sncRNAs, such as miRNAs, generally exert their functions through post-transcriptional gene silencing, lncRNAs, which are typically over 200 nucleotides long, modulate gene expression via chromatin modification, transcriptional interference, or post-transcriptional mechanisms. The complexity and diversity of ncRNAs necessitate the development of specialized computational tools for their identification, annotation, and functional prediction. This review aims to provide a comprehensive overview of the key bioinformatics tools and databases used for studying small and long non-coding RNAs in the human genome.

METHODOLOGY

This review is based on an extensive examination of scientific literature focusing on bioinformatics tools specifically designed for analyzing ncRNAs. Each tool discussed has been selected for its significance in peer-reviewed studies and its applicability to specific RNA-related analyses, including RNA structure prediction, sequence alignment, coding vs. non-coding RNA differentiation, miRNA discovery, and gene set enrichment analysis. The databases covered in this manuscript were chosen for their comprehensive collections of ncRNA sequences, targets, and interactions, making them essential resources for researchers in this field.

DATABASES & TOOLS

The study of small non-coding RNAs (sncRNAs) such as microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and small interfering RNAs (siRNAs) has been greatly facilitated by several specialized databases and tools. Among the most notable databases for miRNAs, miRBase provides comprehensive annotations of miRNA sequences, while TarBase curates experimentally validated miRNA targets. piRBase and piRNAdb are key resources for piRNAs, offering

information on sequences and functional annotations. For siRNAs, siRNAdb is a central resource. Additional databases like deepBase cover both miRNAs and piRNAs, providing expression and interaction data. mirTarBase focuses on validated miRNA-mRNA interactions, while RepTar and snoRNABase provide data on miRNA target sites and small nucleolar RNAs (snoRNAs), respectively.

In terms of bioinformatics tools, several platforms have been developed to analyze small ncRNAs. miRDeep is essential for identifying novel miRNAs from deep sequencing data. BLAST is widely used for sequence alignment of ncRNAs, while RNAfold predicts RNA secondary structure. Tools like sRNAbench and miRanalyzer assist in analyzing small RNA-seq data, particularly for miRNAs, piRNAs, and siRNAs. Additionally, IntaRNA predicts miRNA-mRNA interactions, and TargetScan is used for miRNA target prediction.

For long non-coding RNAs (lncRNAs), specialized databases such as LncBase catalog lncRNA interactions with miRNAs, while NONCODE serves as a repository for a wide range of non-coding RNAs, including lncRNAs. Databases like Lnc2Cancer focus on lncRNAs involved in cancer, and lncRNADisease provides information on lncRNAs related to human diseases. Other notable resources include deepBase, LncATLAS, and LNCipedia, which provide expression, localization, and annotation data for lncRNAs.

Tools for lncRNA research have also advanced, with platforms like CNCI and FEELnc distinguishing coding from non-coding RNA sequences. CPC2 predicts the coding potential of transcripts, while tools like LncTar and LncADeep predict RNA-RNA interactions and lncRNA classifications using machine learning algorithms. PLEK and lncFinder are crucial for identifying lncRNAs based on sequence characteristics, while LncRNApred employs machine learning models to classify lncRNAs. Table 1 summarizing few important databases and tools related to small non-coding RNAs (sncRNAs) and long non-coding RNAs (lncRNAs), along with references.

Table 1: Databases and tools related to small non-coding RNAs (sncRNAs) and long non-coding RNAs

Database	**Description**	**RNA Types**
miRBase	A comprehensive database of microRNA sequences and annotations.	miRNAs
piRBase	A database for piwi-interacting RNA sequences.	piRNAs
siRNAdb	A resource for small interfering RNAs derived from endogenous and exogenous sources.	siRNAs
LncBase	A database of lncRNA interactions with miRNAs.	lncRNAs, miRNAs
NONCODE	A repository of non-coding RNAs including miRNAs, piRNAs, lncRNAs, and other sncRNAs.	lncRNAs, miRNAs, piRNAs, siRNAs
TarBase	A curated database for experimentally validated microRNA targets.	miRNAs
RNAcentral	A comprehensive resource for all non-coding RNA sequences.	lncRNAs, miRNAs, piRNAs, siRNAs
piRNAdb	A database for functional annotations of piRNAs.	piRNAs
starBase	A platform for decoding miRNA-ceRNA interactions, including lncRNAs.	lncRNAs, miRNAs
snoRNA-LBME-db	A specialized database for small nucleolar RNAs (snoRNAs).	snoRNAs
PLncDB	A database dedicated to plant lncRNAs.	lncRNAs
deepBase	A resource for small RNAs and lncRNAs, with expression and interaction data.	lncRNAs, miRNAs, piRNAs, siRNAs
sRNAtoolbox	A suite of tools for analysis of small RNA-seq data.	miRNAs, siRNAs, piRNAs
lncRNASNP	A database of lncRNA-related SNPs and their potential effects.	lncRNAs
Rfam	A database of RNA families, including non-coding RNAs.	lncRNAs, miRNAs, snoRNAs

Tools related to small non-coding RNAs and long non-coding RNAs

1. BLAST (Basic Local Alignment Search Tool)
BLAST is a widely used sequence alignment tool that compares query RNA sequences against known databases to identify homologous sequences. Its versatility allows it to be applied across various ncRNA research areas, including the identification of novel ncRNAs based on sequence similarity. BLAST has been instrumental in discovering novel miRNAs, lncRNAs, and other ncRNAs by aligning unknown sequences with reference databases.

2. RNAfold (RNA Secondary Structure Prediction)
RNAfold predicts the secondary structure of RNA sequences based on thermodynamic parameters. By minimizing the free energy (MFE) of the RNA molecule, RNAfold can generate highly accurate predictions of hairpin structures, stem-loops, and other RNA configurations. Widely used in the study of miRNAs and lncRNAs, RNAfold provides insight into the structural features that influence RNA function.

3. miRDeep (miRNA Discovery Tool)
The miRDeep identifies novel miRNAs from deep sequencing data by analyzing the RNA sequence's structural features and mapping reads to precursor miRNA hairpins. The miRDeep has been essential for discovering novel miRNAs from high-throughput sequencing data, particularly in species where miRNA annotations are incomplete.

4. CNCI (Coding-Non-Coding Index)
The CNCI distinguishes between coding and non-coding RNA sequences based on intrinsic sequence features without relying on known annotations, making it useful for identifying novel lncRNAs. The CNCI is especially valuable for transcriptome-wide studies where distinguishing between protein-coding genes and ncRNAs is crucial for understanding gene regulatory networks.

5. lncRNApred (lncRNA Prediction)
The lncRNApred employs machine learning models to predict lncRNAs based on features such as sequence composition, k-mers, and secondary structure. It is widely used to classify and predict novel lncRNAs from RNA-seq data, aiding in understanding the diverse functional roles of lncRNAs in gene regulation.

6. miRNet (miRNA-Centric Network Visual Analytics Platform)
Function: miRNet is an integrative platform for analyzing and visualizing miRNA-target interactions, offering comprehensive miRNA-centric functional and network analyses. The miRNet supports network-based studies of miRNAs in various biological contexts, including cancer and developmental biology.

7. ShinyGO (Gene Enrichment Analysis Tool)
The ShinyGO provides interactive visualization for gene set enrichment analysis, linking miRNAs and lncRNAs to their target genes and associated pathways. It has become a popular tool for researchers seeking to interpret high-throughput RNA-seq data in terms of biological pathways and gene regulatory networks.

8. iDEP (Integrated Differential Expression and Pathway Analysis)
The iDEP integrates RNA-seq differential expression analysis with pathway enrichment tools, facilitating the exploration of functional changes in gene expression. It allows researchers to comprehensively explore RNA-seq data by linking differential expression to functional pathways and biological processes.

DISCUSSION

The availability of diverse bioinformatics tools has transformed the landscape of ncRNA research by offering comprehensive solutions for identifying, analyzing, and interpreting RNA species that do not code for proteins. BLAST and RNAfold provide foundational insights into RNA sequences and their structures, aiding in the functional characterization of ncRNAs. miRDeep and CNCI facilitate the discovery of novel miRNAs and lncRNAs, while advanced platforms such as miRNet and ShinyGO enable the functional exploration of RNA interactions and regulatory networks.

While these tools have significantly enhanced our understanding of ncRNAs, several challenges remain, particularly in characterizing the full spectrum of lncRNA functions due to their inherent diversity and context-dependent roles. Future

research will undoubtedly benefit from continued improvements in sequencing technologies, machine learning algorithms, and integrative computational platforms. These advancements will further illuminate the multifaceted roles of ncRNAs in gene regulation, disease mechanisms, and therapeutic development.

CONCLUSION

Bioinformatics tools and databases have become indispensable in ncRNA research, providing powerful resources for the discovery, classification, and functional analysis of these crucial genomic elements. The tools reviewed in this manuscript offer a wide range of capabilities for studying small and long ncRNAs, enabling researchers to uncover new insights into gene regulatory networks, RNA interactions, and disease pathways. As the field of ncRNA research continues to grow, the development of even more sophisticated tools will play a vital role in unlocking the mysteries of the non-coding genome.

REFERENCE

[1] *Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2*

[2] *Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., & Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. Nucleic acids research, 47(D1), D1034–D1037. https://doi.org/10.1093/nar/gky905*

[3] *Bartel D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116(2), 281–297. https://doi.org/10.1016/s0092-8674(04)00045-5*

[4] *Cai, Y., Yu, X., Hu, S., & Yu, J. (2009). A brief review on the mechanisms of miRNA regulation. Genomics, proteomics & bioinformatics, 7(4), 147–154. https://doi.org/10.1016/S1672-0229(08)60044-3*

[5] *Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., Huang, W. C., Sun, T. H., Tu, S. J., Lee, W. H., Chiew, M. Y., Tai, C. S., Wei, T. Y., Tsai, T. R., Huang, H. T., Wang, C. Y., Wu, H. Y., Ho, S. Y., Chen, P. R., Chuang, C. H., ... Huang, H. D. (2018). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic acids research, 46(D1), D296–D302. https://doi.org/10.1093/nar/gkx1067*

[6] *Elefant, N., Berger, A., Shein, H., Hofree, M., Margalit, H., & Altuvia, Y. (2011). RepTar: a database of predicted cellular targets of host and viral miRNAs. Nucleic acids research, 39(Database issue), D188–D194. https://doi.org/10.1093/nar/gkq1233*

[7] *Esteller M. (2011). Non-coding RNAs in human disease. Nature reviews. Genetics, 12(12), 861–874. https://doi.org/10.1038/nrg3074*

[8] *Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic acids research, 40(1), 37–52. https://doi.org/10.1093/nar/gkr688*

[9] *Griffiths-Jones, S., Saini, H. K., van Dongen, S., & Enright, A. J. (2008). miRBase: tools for microRNA genomics. Nucleic acids research, 36(Database issue), D154–D158. https://doi.org/10.1093/nar/gkm952*

[10] *Hackenberg, M., Sturm, M., Langenberger, D., Falcón-Pérez, J. M., & Aransay, A. M. (2009). miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic acids research, 37(Web Server issue), W68–W76. https://doi.org/10.1093/nar/gkp347*

[11] *Hofacker I. L. (2003). Vienna RNA secondary structure server. Nucleic acids research, 31(13), 3429–3431. https://doi.org/10.1093/nar/gkg599*

[12] *Wang, J., Zhang, P., Lu, Y., Li, Y., Zheng, Y., Kan, Y., Chen, R., & He, S. (2019). piRBase: a comprehensive database of piRNA sequences. Nucleic acids research, 47(D1), D175–D180. https://doi.org/10.1093/nar/gky1043*

[13] *Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., & Petrov, A. I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic acids research, 46(D1), D335–D342. https://doi.org/10.1093/nar/gkx1038*

[14] *Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., & Gao, G. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic acids research, 45(W1), W12–W16. https://doi.org/10.1093/nar/gkx428*

[15] *Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., Papadimitriou, D., Kavakiotis, I., Maniou, S., Skoufos, G., Vergoulis, T., Dalamagas, T., & Hatzigeorgiou, A. G. (2018). DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. Nucleic acids research, 46(D1), D239–D245. https://doi.org/10.1093/nar/gkx1141*

[16] *Mazière, P., & Enright, A. J. (2007). Prediction of microRNA targets. Drug discovery today, 12(11-12), 452–458. https://doi.org/10.1016/j.drudis.2007.04.002*

[17] *Oulas, A., Karathanasis, N., Louloupi, A., Pavlopoulos, G. A., Poirazi, P., Kalantidis, K., & Iliopoulos, I. (2015). Prediction of miRNA targets. Methods in molecular biology (Clifton, N.J.), 1269, 207–229. https://doi.org/10.1007/978-1-4939-2291-8_13*

[18] *Li, J. H., Liu, S., Zhou, H., Qu, L. H., & 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 research, 42(Database issue), D92–D97. https://doi.org/10.1093/nar/gkt1248*

[19] *Li, J., Ma, W., Zeng, P., Wang, J., Geng, B., Yang, J., & Cui, Q. (2015). LncTar: a tool for predicting the RNA targets of long noncoding RNAs. Briefings in bioinformatics, 16(5), 806–812. https://doi.org/10.1093/bib/bbu048*

[20] *Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Hermoso Pulido, T., Guigo, R., & Johnson, R. (2017). LncATLAS database for subcellular localization of long noncoding RNAs. RNA (New York, N.Y.), 23(7), 1080–1087. https://doi.org/10.1261/rna.060814.117*

[21] *Miao, Y. R., Liu, W., Zhang, Q., & Guo, A. Y. (2018). lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. Nucleic acids research, 46(D1), D276–D280. https://doi.org/10.1093/nar/gkx1004*

[22] *Paraskevopoulou, M. D., Vlachos, I. S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., Zagganas, K., Tsanakas, P., Floros, E., Dalamagas, T., & Hatzigeorgiou, A. G. (2016). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic acids research, 44(D1), D231–D238. https://doi.org/10.1093/nar/gkv1270*

[23] *Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. Annual review of biochemistry, 81, 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902*

[24] *Rueda, A., Barturen, G., Lebrón, R., Gómez-Martín, C., Alganza, Á., Oliver, J. L., & Hackenberg, M. (2015). sRNAtoolbox: an integrated collection of small RNA research tools. Nucleic acids research, 43(W1), W467–W473. https://doi.org/10.1093/nar/gkv555*

[25] *Chen, X., Sun, Y. Z., Guan, N. N., Qu, J., Huang, Z. A., Zhu, Z. X., & Li, J. Q. (2019). Computational models for lncRNA function prediction and functional similarity calculation. Briefings in functional genomics, 18(1), 58–82. https://doi.org/10.1093/bfgp/ely031*

[26] *Volders, P. J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J., & Mestdagh, P. (2015). An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic acids research, 43(Database issue), D174–D180. https://doi.org/10.1093/nar/gku1060*

[27] *Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., Jagannathan, V., Cadieu, E., David, A., Lohi, H., Cirera, S., Fredholm, M., Botherel, N., Leegwater, P. A. J., Le Béguec, C., Fieten, H., Johnson, J., Alföldi, J., André, C., Lindblad-Toh, K., ... Derrien, T. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic acids research, 45(8), e57. https://doi.org/10.1093/nar/gkw1306*

[28] *Yang, J. H., Shao, P., Zhou, H., Chen, Y. Q., & Qu, L. H. (2010). deepBase: a database for deeply annotating and mining deep sequencing data. Nucleic acids research, 38(Database issue), D123–D130. https://doi.org/10.1093/nar/gkp943*

[29] *Zhao, J., Song, X., & Wang, K. (2016). lncScore: alignment-free identification of long noncoding RNA from assembled novel transcripts. Scientific reports, 6, 34838. https://doi.org/10.1038/srep34838*

[30] *Zhao, L., Wang, J., Li, Y., Song, T., Wu, Y., Fang, S., Bu, D., Li, H., Sun, L., Pei, D., Zheng, Y., Huang, J., Xu, M., Chen, R., Zhao, Y., & He, S. (2021). NONCODEV6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. Nucleic acids research, 49(D1), D165–D171. https://doi.org/10.1093/nar/gkaa1046*

[31] *Jin, J., Liu, J., Wang, H., Wong, L., & Chua, N. H. (2013). PLncDB: plant long non-coding RNA database. Bioinformatics (Oxford, England), 29(8), 1068–1071. https://doi.org/10.1093/bioinformatics/btt107*

[32] *Lestrade, L., & Weber, M. J. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic acids research, 34(Database issue), D158–D162. https://doi.org/10.1093/nar/gkj002*

[33] *Ning, S., Zhang, J., Wang, P., Zhi, H., Wang, J., Liu, Y., Gao, Y., Guo, M., Yue, M., Wang, L., & Li, X. (2016). Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. Nucleic acids research, 44(D1), D980–D985. https://doi.org/10.1093/nar/gkv1094*

[34] *Fan, Y., Siklenka, K., Arora, S. K., Ribeiro, P., Kimmins, S., & Xia, J. (2016). miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. Nucleic acids research, 44(W1), W135–W141. https://doi.org/10.1093/nar/gkw288*

[35] *Ge, S. X., Son, E. W., & Yao, R. (2018). iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC bioinformatics, 19(1), 534. https://doi.org/10.1186/s12859-018-2486-6*

[36] *Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11), 2498–2504. https://doi.org/10.1101/gr.1239303*