



## OPEN ACCESS

## EDITED AND REVIEWED BY

Fangqing Zhao,  
Beijing Institutes of Life Science (CAS),  
China

## \*CORRESPONDENCE

Pingjian Ding,  
✉ pxd210@case.edu  
Min Zeng,  
✉ zengmin@csu.edu.cn  
Rui Yin,  
✉ ruiyin@ufl.edu

RECEIVED 31 July 2023

ACCEPTED 14 August 2023

PUBLISHED 22 August 2023

## CITATION

Ding P, Zeng M and Yin R (2023), Editorial:  
Computational methods to analyze RNA  
data for human diseases.  
*Front. Genet.* 14:1270334.  
doi: 10.3389/fgene.2023.1270334

## COPYRIGHT

© 2023 Ding, Zeng and Yin. This is an  
open-access article distributed under the  
terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Editorial: Computational methods to analyze RNA data for human diseases

Pingjian Ding<sup>1\*</sup>, Min Zeng<sup>2\*</sup> and Rui Yin<sup>3\*</sup>

<sup>1</sup>Center for Artificial Intelligence in Drug Discovery, School of Medicine, Case Western Reserve University, Cleveland, OH, United States, <sup>2</sup>School of Computer Science and Engineering, Central South University, Changsha, China, <sup>3</sup>Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States

## KEYWORDS

RNA, miRNA, LncRNA, RNA virus, ceRNA network, machine learning/statistics, human disease

## Editorial on the Research Topic

### Computational methods to analyze RNA data for human diseases

RNA, as a type of nucleic acid, forms one of the four fundamental macromolecules crucial for all known life forms. Unlike DNA (Deoxyribonucleic Acid), which typically serves as the primary genetic material in cells, many viruses use RNA as their genetic material. RNA viruses are known for their ability to mutate rapidly, and the emergence of novel strains and variants (Yin et al., 2020) is potentially responsible for a wide range of diseases, leading to epidemics or pandemics such as swine-origin flu pandemic (Yin et al., 2018) and COVID-19 (V'kovski et al., 2021; Yin et al., 2018; Ding and Xu, 2023). In addition, RNA plays critical roles in various biological processes, including gene expression, protein synthesis (Frye et al., 2018). Understanding the mechanisms and roles of RNA in disease pathogenesis and progression is crucial for advancing our knowledge of human biology and developing optimized therapeutic strategies to combat RNA-related diseases. Computational approaches like machine learning and statistics, have captured much attention in this field due to increasingly available diverse RNA datasets (Yin et al., 2022; Li et al., 2023; Yin et al., 2023). This Research Topic of Frontiers in Genetics features a Research Topic of the latest advances in applying and developing various kinds of computational methods to analyze RNA data towards non-coding RNAs (e.g., miRNA, lncRNA) and RNA viruses (e.g., influenza, coronavirus).

The ncRNAs are crucial for regulating gene expression at both the transcriptional and posttranscriptional levels within the transcriptome, without encoding proteins (Winkle et al., 2021). In particular, miRNAs are a type of small, single-stranded noncoding RNAs, about 19–25 nucleotides long, that have highly conserved sequences and can regulate gene expression at the post-transcriptional level. Through extensive research on miRNA in the context of development and disease, it has emerged as a compelling target for innovative therapeutic approaches (Shen et al., 2020a; Shen et al., 2020b; Li Peng et al., 2022). In this Research Topic, Luo et al. presented a comprehensive perspective of recent progress in miRNA-targeted therapeutics employing machine learning techniques. In addition to discussing resources and preprocessing of pharmacogenomic data, they also presented the main machine learning algorithms employed in identifying miRNA-disease associations. Given the limitations of current methods in constructing negative sample sets, Wei et al.

introduced a clustering-based sampling approach called CSMDA to predict miRNA-disease associations. This method aims to address the Research Topic associated with negative sample selection in the context of miRNA-disease association prediction. Under a five-fold cross-validation, CSMDA computed an impressive Area Under the Curve (AUC) of 0.9610. Additionally, through validation with the dbDEMC database, it was confirmed that all predicted miRNAs, except hsa-mir-34c, were associated with colon cancer.

LncRNAs are a subset of ncRNAs characterized by their length, which exceeds 200 nucleotides. They have important functions in controlling gene expression at various levels, such as translational, transcriptional, and epigenetic processes (Qin et al., 2020). LncRNAs are crucial in controlling genes and proteins related to a range of human diseases like cancer (Xiao et al., 2018), digestive system Research Topic, and heart problems. Their role in disease regulation is well-established and holds promise for future therapies. Yao et al. proposed a computational model called GCHIRFLDA, which utilizes geometric complement heterogeneous information and random forest to predict lncRNA-disease associations. Under five-fold cross-validation, GCHIRFLDA achieved impressive performance metrics with an AUC of 0.9897 and an AUPR of 0.7040. The study demonstrated that 18 of the predicted lncRNAs were validated through records present in databases or published literature. Meanwhile, the presence of inherent sparsity in known heterogeneous bio-data poses a challenge for computational methods aiming to enhance the accuracy of prediction. Thus, Zhang et al. explored a novel multiple mechanisms to discover underlying lncRNA-disease associations (MM-LDA). By integrating the graph attention network (GAT) and inductive matrix completion (IMC), this approach boosts the prediction accuracy. Firstly, a multiple-operator aggregation was created as part of the n-heads attention mechanism in the GAT. Then, IMC was incorporated into the improved node feature, and subsequently, the LDA network underwent a reconstruction to address the cold start problem caused by insufficient data in either whole rows or columns of a known association matrix. Under 5-fold cross-validation, an AUC of 0.9395 and an AUPR of 0.8057 were computed. The results from MM-LDA suggested a potential link between HOTAIR and HTTAS and gastric cancer.

In recent years, there has been the proposal of a hypothesis about competing endogenous RNA (ceRNA) network (Salmena et al., 2011). Under this hypothesis, lncRNAs possess the capability to function as endogenous molecular sponges for miRNAs, indirectly regulating the expression of messenger RNAs (mRNAs). The intricate nature of the lncRNA-miRNA-mRNA network makes their dysregulation closely linked to the progression and onset of various human diseases. For example, Ye et al. (2019) discovered that the lncRNA MIAT increases the expression of CD47 by acting as a sponge for miR-149-5p, leading to the inhibition of efferocytosis in advanced atherosclerosis. Yang et al. (2021) conducted a study uncovering the role of lncRNA XIST as a ceRNA, promoting atherosclerosis by upregulating TLR4 expression through the mediation of miR-599. Additionally, they identified several putative ceRNA networks, including those associated with implantation failure (Feng et al., 2018), polycystic ovary syndrome (Ma et al., 2021), and epithelial ovarian cancer (Zhao et al., 2019). Chen et al. employed the CIBERSORT algorithm to investigate the potential ceRNA-related mechanism of Peripheral arterial occlusive disease (PAOD) and to identify the associated patterns of immune cell infiltration. They

developed an immune-related core ceRNA network that offered valuable insights into the molecular mechanisms underlying Peripheral Arterial Occlusive Disease (PAOD). This network consisting of CREB1, LINC00221, miR-20b-5p, and miR-17-5p, along with the infiltrating immune cells, specifically M1 macrophages and monocytes. Luo et al. introduced a lncRNA-mRNA network based on POI (POILMN) to identify essential lncRNAs. This research yielded a Research Topic of 288 differentially expressed mRNAs and 244 differentially expressed lncRNA. Ultimately, Through the application of topological analysis, POILMN identified four intersecting lncRNAs based on two centralities, namely, degree and betweenness.

CircRNA is a class of ncRNAs that forms a covalently closed loop structures (Li et al., 2020; Xiao et al., 2020; Peng et al., 2022; Peng et al., 2023). CircRNA molecules have been observed or artificially synthesized in various organisms, including mammals (Xu and Zhang, 2021) and viruses (Tan and Lim, 2021). The interactions between miRNAs and circRNAs have been demonstrated to modify gene expression and play a regulatory role in diseases. Therefore, He et al. introduced a novel approach called GCNCMI, which utilizes a graph convolutional neural (GCN) network to uncover latent associations between miRNAs and circRNAs. GCNCMI initially examines the underlying connections between neighboring nodes in the GCN network. Afterward, it iteratively spreads this connection information across the graph convolutional layers. Lastly, the embeddings produced by each layer were combined to output the ultimate prediction results. GCNCMI achieved an AUC of 0.9312 and an AUPR of 0.9412. The results from GCNCMI showed that 8 interactions involving hsa-miR-149-5p and 7 interactions involving hsa-miR-622 were validated.

Additionally, mitochondrial dysfunction could be among the molecular mechanisms implicated in obstructive sleep apnea (OSA) and its concurrent conditions. Despite several studies reporting the involvement of various proteins and miRNAs in OSA (Targa et al., 2020; Pinilla et al., 2021), the impact of OSA on genes and pathways, particularly concerning mitochondrial dysfunction, remains largely unexplored. In a previous study by Li et al. (2017), differentially expressed miRNAs were reported in OSA, but their specific association with mitochondrial dysfunction was not established. Liu et al. developed a novel diagnostic model consisting of a four-gene signature related to mitochondrial dysfunction. Using gene expression related to mitochondrial dysfunction, all samples were categorized into two clusters, with an additional subdivision of three clusters identified specifically among the samples with OSA. In the OSA samples compared to control samples, Significant differences were noted in the levels of M0 and M1 macrophages as well as plasma cells. Additionally, within the clusters associated with mitochondrial dysfunction in OSA samples, various immune cell types, particularly T cells, showed significant differences.

Although multiple databases offer information on virus-host protein interactions, they often lack detailed information about strain-specific virulence factors or the specific protein domains implicated in the interactions (Yin et al., 2017; Yin et al., 2021). Several databases may have incomplete representation coverage of influenza strains of influenza strains due to the challenge of sifting through extensive literature to gather comprehensive information. No existing database has provided complete records of strain-

specific protein-protein interactions for all types of Influenza A viruses. In particular, Ng et al. presented an innovative network that predicts domain-domain interactions between proteins from the mouse host and influenza A virus (IAV). By incorporating vital virulence details like lethal dose, this network facilitates a methodical exploration of disease factors. They created a network of interacting protein domains from both mouse and viral proteins, representing them as nodes and using weighted edges to show their interactions.

In summary, this Research Topic centers on the recent progress in utilizing and refining diverse computational methods, including machine learning and statistical techniques, to analyze RNA data related to RNA viruses and non-coding RNA. As a result, these analyses have delved into the biological disease mechanisms and aided in the understanding of human diseases, leading to improved preventive measures, diagnoses, and treatments.

## Author contributions

PD: Conceptualization, Formal Analysis, Writing—original draft, Writing—review and editing. MZ: Conceptualization, Formal Analysis, Writing—original draft, Writing—review and editing. RY: Conceptualization, Funding acquisition, Writing—original draft, Writing—review and editing.

## References

- Ding, P., and Xu, R. (2023). Causal association of COVID-19 with brain structure changes: findings from a non-overlapping 2-sample mendelian randomization study. medRxiv 2023.07.16.23292735.
- Feng, C., Shen, J. M., Lv, P. P., Jin, M., Wang, L. Q., Rao, J. P., et al. (2018). Construction of implantation failure related lncRNA-mRNA network and identification of lncRNA biomarkers for predicting endometrial receptivity. *Int. J. Biol. Sci.* 14, 1361–1377. doi:10.7150/ijbs.25081
- Frye, M., Harada, B. T., Behm, M., and He, C. (2018). RNA modifications modulate gene expression during development. *Science* 361, 1346–1349. doi:10.1126/science.aau1646
- Li, G., Luo, J., Wang, D., Liang, C., Xiao, Q., Ding, P., et al. (2020). Potential circRNA-disease association prediction using DeepWalk and network consistency projection. *J. Biomed. Inf.* 112, 103624. doi:10.1016/j.jbi.2020.103624
- Li, K., Wei, P., Qin, Y., and Wei, Y. (2017). MicroRNA expression profiling and bioinformatics analysis of dysregulated microRNAs in obstructive sleep apnea patients. *Medicine* 96, e7917. doi:10.1097/MD.00000000000007917
- Li, M., Zhao, B., Yin, R., Lu, C., Guo, F., and Zeng, M. J. (2023). GraphLncLoc: long non-coding RNA subcellular localization prediction using graph convolutional networks based on sequence to graph transformation. *Briefings Bioinforma.* 24, bbac565. doi:10.1093/bib/bbac565
- Li Peng, Y. T., Huang, L., Yang, L., Fu, X., and Chen, X. (2022). Daestb: inferring associations of small molecule-miRNA via a scalable tree boosting model based on deep autoencoder. *Briefings Bioinforma.* 23, bbac478. doi:10.1093/bib/bbac478
- Ma, Y., Ma, L., Cao, Y., and Zhai, J. (2021). Construction of a ceRNA-based lncRNA-mRNA network to identify functional lncRNAs in polycystic ovarian syndrome. *Aging (Albany NY)* 13, 8481–8496. doi:10.18632/aging.202659
- Peng, L., Yang, C., Chen, Y., and Liu, W. (2023). Predicting CircRNA-disease associations via feature convolution learning with heterogeneous graph attention network. *IEEE J. Biomed. Health Inf.* 27, 3072–3082. doi:10.1109/JBHI.2023.3260863
- Peng, L., Yang, C., Huang, L., Chen, X., Fu, X., and Liu, W. (2022). Rnmflp: predicting circRNA-disease associations based on robust nonnegative matrix factorization and label propagation. *Briefings Bioinforma.* 23, bbac155. doi:10.1093/bib/bbac155
- Pinilla, L., Barbe, F., and De Gonzalo-Calvo, D. J. (2021). MicroRNAs to guide medical decision-making in obstructive sleep apnea: A review. *Sleep. Med. Rev.* 59, 101458. doi:10.1016/j.smrv.2021.101458
- Qin, T., Li, J., and Zhang, K. Q. (2020). Structure, regulation, and function of linear and circular long non-coding RNAs. *Front. Genet.* 11, 150. doi:10.3389/fgene.2020.00150
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* 146, 353–358. doi:10.1016/j.cell.2011.07.014
- Shen, C., Luo, J., Lai, Z., and Ding, P. (2020a). Multiview joint learning-based method for identifying small-molecule-associated miRNAs by integrating pharmacological, genomics, and network knowledge. *J. Chem. Inf. Model.* 60, 4085–4097. doi:10.1021/acs.jcim.0c00244
- Shen, C., Luo, J., Ouyang, W., Ding, P., and Wu, H. (2020b). Identification of small molecule-miRNA associations with graph regularization techniques in heterogeneous networks. *J. Chem. Inf. Model.* 60, 6709–6721. doi:10.1021/acs.jcim.0c00975
- Tan, K. E., and Lim, Y. (2021). Viruses join the circular RNA world. *FEBS J.* 288, 4488–4502. doi:10.1111/febs.15639
- Targa, A., Dakterzada, F., Benítez, I., De Gonzalo-Calvo, D., Moncusí-Moix, A., López, R., et al. (2020). Circulating MicroRNA profile associated with obstructive sleep apnea in alzheimer's disease. *Mol. Neurobiol.* 57, 4363–4372. doi:10.1007/s12035-020-02031-z
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* 19, 155–170. doi:10.1038/s41579-020-00468-6
- Winkle, M., El-Daly, S. M., Fabbri, M., and Calin, G. (2021). Noncoding RNA therapeutics—challenges and potential solutions. *Nat. Rev. Drug Discov.* 20, 629–651. doi:10.1038/s41573-021-00219-z
- Xiao, Q., Luo, J., Liang, C., Li, G., Cai, J., Ding, P., et al. (2018). Identifying lncRNA and mRNA co-expression modules from matched expression data in ovarian cancer. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 17, 623–634. doi:10.1109/TCBB.2018.2864129
- Xiao, Q., Yu, H., Zhong, J., Liang, C., Li, G., Ding, P., et al. (2020). An *in-silico* method with graph-based multi-label learning for large-scale prediction of circRNA-disease associations. *Genomics* 112, 3407–3415. doi:10.1016/j.ygeno.2020.06.017
- Xu, C., and Zhang, J. (2021). Mammalian circular RNAs result largely from splicing errors. *Cell Rep.* 36, 109439. doi:10.1016/j.celrep.2021.109439
- Yang, K., Xue, Y., and Gao, X. (2021). lncRNA XIST promotes atherosclerosis by regulating miR-599/TLR4 axis. *Inflammation* 44, 965–973. doi:10.1007/s10753-020-01391-x
- Ye, Z. M., Yang, S., Xia, Y. P., Hu, R. T., Chen, S., Li, B. W., et al. (2019). lncRNA MIAT sponges miR-149-5p to inhibit efferocytosis in advanced atherosclerosis through CD47 upregulation. *Cell death Dis.* 10, 138. doi:10.1038/s41419-019-1409-4

## Funding

This study was partially supported by grants from Centers for Disease Control and Prevention (1U18DP006512), National Institute of Environmental Health Sciences (R21ES032762) and the NIH National Center for Advancing Translational Sciences (UL1TR001427).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Yin, R., Luo, Z., Zhuang, P., Lin, Z., and Kwoh, C. (2021). VirPreNet: a weighted ensemble convolutional neural network for the virulence prediction of influenza A virus using all eight segments. *Bioinformatics* 37, 737–743. doi:10.1093/bioinformatics/btaa901

Yin, R., Luo, Z., Zhuang, P., Zeng, M., Li, M., Lin, Z., et al. (2023). ViPal: a framework for virulence prediction of influenza viruses with prior viral knowledge using genomic sequences. *J. Biomed. Inf.* 142, 104388. doi:10.1016/j.jbi.2023.104388

Yin, R., Luusua, E., Dabrowski, J., Zhang, Y., and Kwoh, C. (2020). Tempel: time-series mutation prediction of influenza A viruses via attention-based recurrent neural networks. *Bioinformatics* 36, 2697–2704. doi:10.1093/bioinformatics/btaa050

Yin, R., Tran, V. H., Zhou, X., Zheng, J., and Kwoh, C. (2018). Predicting antigenic variants of H1N1 influenza virus based on epidemics and pandemics using a stacking model. *PLoS one* 13, e0207777. doi:10.1371/journal.pone.0207777

Yin, R., Zhou, X., Ivan, F. X., Zheng, J., Chow, V. T., and Kwoh, C. K. (2017). “Identification of potential critical virulent sites based on hemagglutinin of influenza A virus in past pandemic strains,” in Proceedings of the 6th International Conference on Bioinformatics and Biomedical Science, Singapore, June 22 - 24, 2017, 30–36.

Yin, R., Zhu, X., Zeng, M., Wu, P., Li, M., and Kwoh, C. (2022). A framework for predicting variable-length epitopes of human-adapted viruses using machine learning methods. *Briefings Bioinforma.* 23, bbac281. doi:10.1093/bib/bbac281

Zhao, X., Tang, D. Y., Zuo, X., Zhang, T. D., and Wang, C. (2019). Identification of lncRNA–miRNA–mRNA regulatory network associated with epithelial ovarian cancer cisplatin-resistant. *J. Cell. physiology* 234, 19886–19894. doi:10.1002/jcp.28587