



# mtTB: A Web-Based R/Shiny App for Pulmonary Tuberculosis Screening

Zhougui Ling<sup>\*†</sup>, Shuangping Huang<sup>†</sup>, Zhongwei Wen, Zhenming Tang, Ying Huang, Ni Wei, Mei Liu and Jinyan Wu

Department of Pulmonary and Critical Care Medicine, The Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, China

## OPEN ACCESS

### Edited by:

Alexandre Dias Tavares Costa,  
Carlos Chagas Institute (CCI), Brazil

### Reviewed by:

Shaoqiu Chen,  
University of Hawaii at Mānoa,  
United States

Kaifeng Guo,  
Fudan University, China

Meilin Wei,  
Second Affiliated Hospital of  
Nanchang University, China

### \*Correspondence:

Zhougui Ling  
lzg228@163.com;  
lzhougui@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Clinical Microbiology,  
a section of the journal  
Frontiers in Cellular and  
Infection Microbiology

Received: 07 January 2022

Accepted: 18 February 2022

Published: 18 March 2022

### Citation:

Ling Z, Huang S, Wen Z, Tang Z,  
Huang Y, Wei N, Liu M and Wu J  
(2022) mtTB: A Web-Based  
R/Shiny App for Pulmonary  
Tuberculosis Screening.  
Front. Cell. Infect. Microbiol. 12:850279.  
doi: 10.3389/fcimb.2022.850279

Pulmonary tuberculosis caused by *Mycobacterium tuberculosis* remains a global issue. However, the diagnosis of active pulmonary tuberculosis (TB) remains a challenge in the clinic. Small non-coding RNAs are potential diagnostic biomarkers for pulmonary tuberculosis. However, the current normalization methods are not stable and usually fail to reliably detect differentially expressed sncRNAs. To identify reliable biomarkers for pulmonary tuberculosis screening, we utilized the ratio-based method on the newly discovered mitochondria-derived small RNAs in human peripheral blood mononuclear cells. The prediction model of seven mtRNA biomarkers noteworthily enables the discrimination between pulmonary tuberculosis patients and controls in discovery (AUC = 0.906, 23 patients) and independent validation cohort (AUC = 0.968, 20 patients). Moreover, we present mtTB (<https://tuberculosis.shinyapps.io/mtTB/>), a novel R Graphical User Interface (GUI) that provides reliable biomarkers for the feasibility of blood-based screening, and produce a more accurate tool for pulmonary tuberculosis diagnosis in real clinical practice.

**Keywords:** TB, peripheral blood, mitochondria-derived small RNAs, ratio-based method, biomarkers

## INTRODUCTION

Pulmonary tuberculosis (TB) is a chronic pulmonary infectious disease caused by *Mycobacterium tuberculosis* (Mtb) and is the second most predominant infectious disease across the world (Bando-Campos et al., 2019). Current diagnostic tools comprise smear microscopy, microbiological culture, and molecular detection by Xpert MTB/RIF (Xpert) or Xpert MTB/RIF Ultra (Ultra). However, there are additional shortcomings of each approach, such as the insufficient sensitivity of microscopy, the time delay for culture, the high cost of molecular tests, and false-positive Ultra results (Turner et al., 2020).

Mitochondria are critical organelles for maintaining cell energy metabolism and play an important role in the development and progression of lung cancer (Roberts and Thomas, 2013). The human mitochondrial DNA (mtDNA) encodes 37 genes including 2 rRNAs, 22 tRNAs, and 13 protein-coding genes (Larriba et al., 2018). Moreover, approximately 12% of the unique small RNAs identified were encoded in the mitochondrial genome (Riggs and Podrabsky, 2017; Hirose et al., 2019). Recent studies reveal different types of sncRNAs that are associated with the mitochondrial genome, and these sncRNAs generated from the mitochondrial DNA were proposed to regulate and communicate with various pathways that interact with the nuclear genome (Larriba et al., 2018).

Therefore, mitochondrial-derived RNAs (mtRNAs) could play an important role in pathophysiological processes and infectious diseases. Since various sncRNAs, such as miRNA, snoRNA, and piRNA, are widely studied in the diagnosis of TB, no clear research has been given for mitochondria-derived small RNAs (mtRNAs) (Wang et al., 2011; de Araujo et al., 2019). Moreover, most studies of sncRNA normalization methods are based on synthetic external spiked-in controls or published endogenous miRNA controls. However, those references are too labile to use directly in sncRNA studies. The ratio-based method provides a solution for the difficult normalization problem for sncRNA data to identify reliable biomarkers to reach the real clinical application (Deng et al., 2019). Here, we aim to develop a new reliable model to predict TB patients based on peripheral blood mtRNAs and the ratio-based method.

## METHODS

### Datasets

Discovery and the independent validation set were downloaded from the Gene Expression Omnibus (GEO) repository (GSE148861, GSE148862). Each small RNA-seq was aligned using SPORTS1.1 software to extract mtRNA expression levels (Shi et al., 2018). At first, all miRNA-seq FASTQ files removed adapter sequences from raw reads using nf-core/smrnaseq software (Ewels et al., 2020). The trimmed sequence reads were aligned to the mitotRNAdb database using the STAR algorithm (Jühling et al., 2009). Raw counts from mapped reads were obtained using the htseq-count script from the HTSeq tools (Anders et al., 2015). Missing values were imputed by MetImp 1.2 (Wei et al., 2018).

## Ratio-Based Normalization Method and Shiny App

To stabilize the mtRNA expression profile, we performed the ratio-based method to the mtRNAs (Deng et al., 2019). The mtRNA paired ratios were calculated according to the equation:  $\text{Ratio}(\text{mtRNA1\_to\_mtRNA2}) = \text{mtRNA1}/\text{mtRNA2}$ . All results are displayed as the mean  $\pm$  SEM. Differentially expressed (DE) mtRNA analysis was performed on the discovery group derived from GSE148861 using unpaired Student's t-tests. mtTB uses Shiny's reactivity with built-in R functions from packages for prediction model analysis and best subset selection including "survminer", "shiny", "precrec", "glmnet", and "randomForest". Statistical significance was assigned as  $p < 0.05$ .

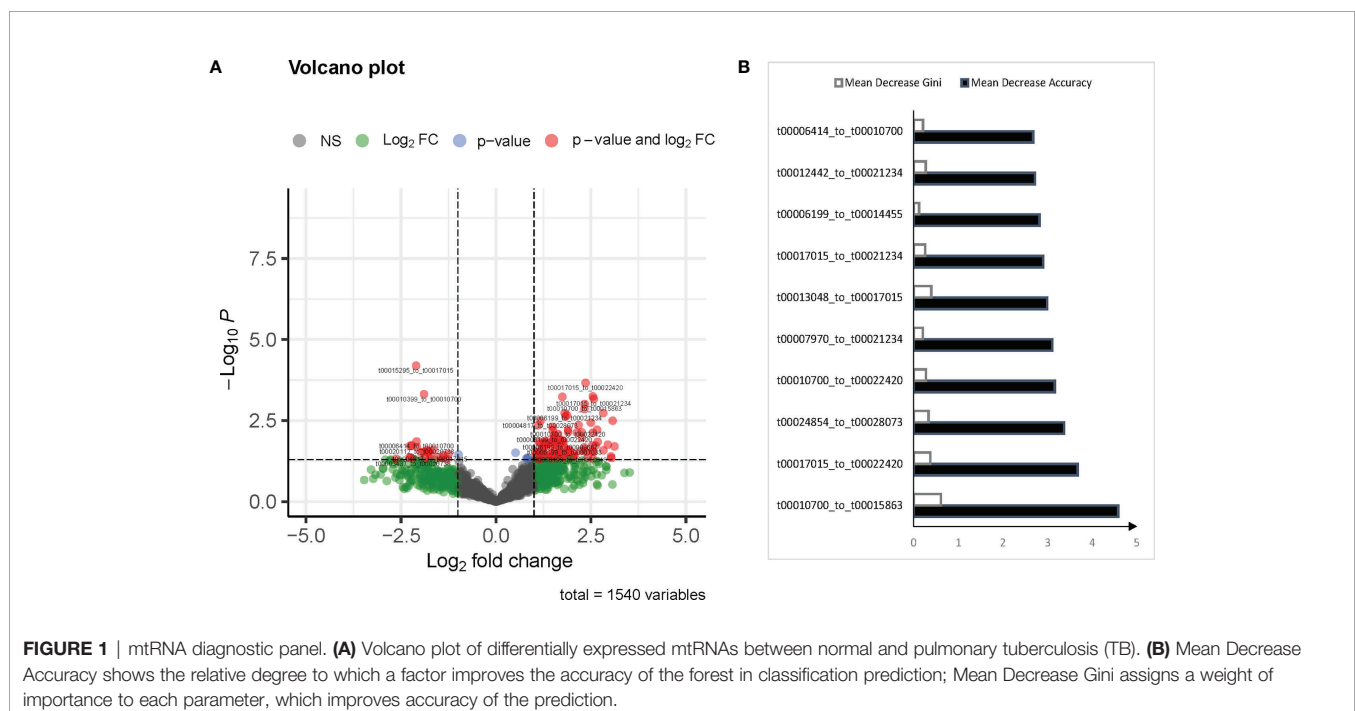
## Prediction Model Construction

Differentially expressed (DE) ( $p$  value  $< 0.01$ ) and  $|\text{Log}_2(\text{fold change})| > 1$  paired mtRNAs were enrolled using the randomForest prediction model. The Mean decrease of accuracy and mean decrease Gini of each paired mtRNA were calculated by the randomForest model. Feature selection based on the overlap of the top 10 Mean decrease of accuracy and mean decrease Gini (Figure 1B). The final prediction model was built by selected features.

## RESULTS

### Patient Cohorts and the Molecular Signature Composed of mtRNAs

The clinical characteristics of the two TB cohorts are summarized in Supplementary Table 1. Our study included



43 cases, composed of 18 patients with TB and 25 healthy controls. In the discovery cohort, the average age of the TB group was  $35 \pm 1.1$ , and the control group was  $52.7 \pm 1.0$ . In the validation dataset, the average age of the TB group was  $38.1 \pm 2.8$ , and the control group was  $45 \pm 1.85$ . In total, we identified 9 mtRNA species in human peripheral blood samples (**Supplementary Table 2**). There are 7 types of mtRNAs by grouping mtRNA species into subcategories according to their parent tRNA types (i.e., mt-tRNA-GAA, mt-tRNA-Ser-GCT\_5\_end). The mtRNA sequences length ranged from 15 to 32 nt with an average length of 19.6 nt.

## Dysregulated mtRNAs in TB and Prediction Model Construction

According to the inclusion criteria described in *Methods*, 127 mtRNA pairs were significantly different by Student t-test (**Figure 1A** and **Supplementary Table S3**). The random forest (RF) algorithm was performed to select the most effective variables from 127 mtRNA pairs to construct prediction models. According to the RF mean decrease of accuracy and RF mean decrease GINI (**Figure 1B**), seven mtRNA pairs were selected, including upregulated t00013048\_to\_t00017015 and downregulated t00010700\_to\_t00015863, t00010700\_to\_t00022420, t00012442\_to\_t00021234, t00017015\_to\_t00021234, t00017015\_to\_t00022420, and t00024854\_to\_t00028073 in TB samples (**Figure 2B**). Out-of-bag (OOB) estimations were used to assess the predicted error. We evaluated the model performance by a receiver operating characteristic curve (ROC curve) and Precision-Recall curve (PR curve). The area under the receiver operating characteristic (ROC) curve (AUC) was

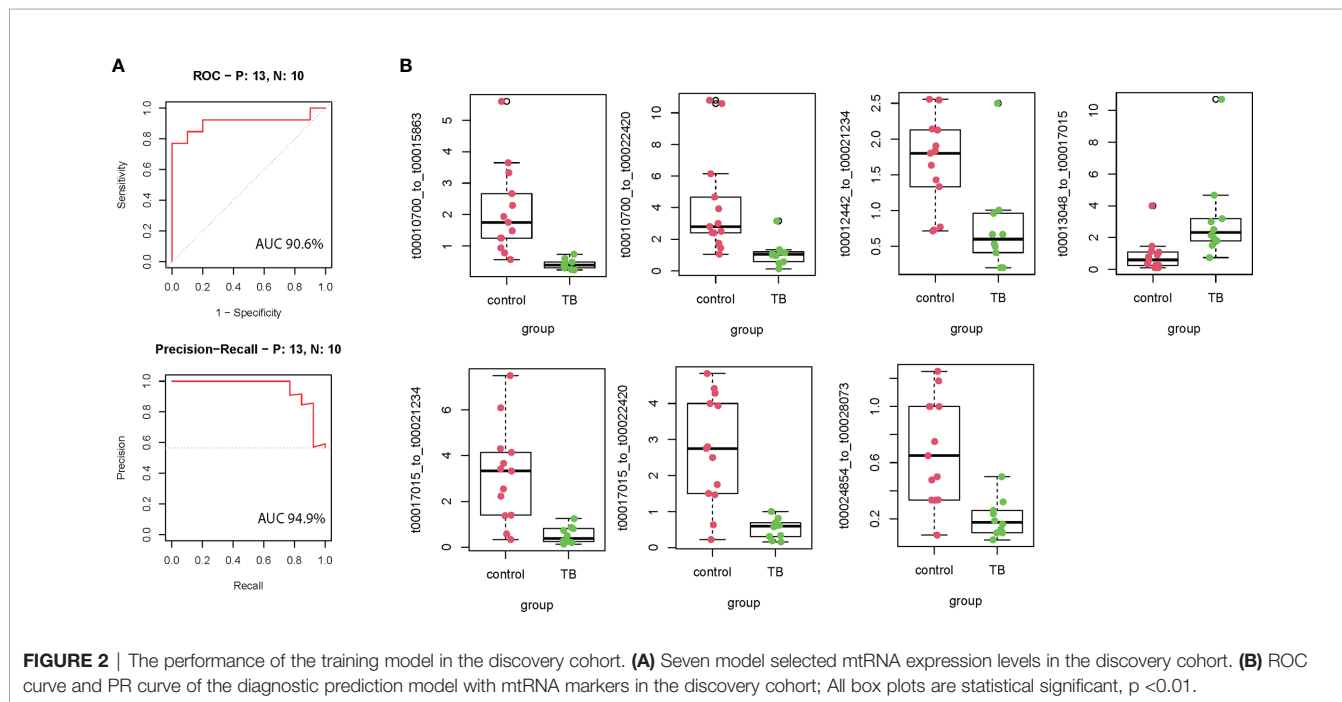
0.906 between TB and control subjects and 0.949 (AUC) in the PR curve (**Figure 2A**).

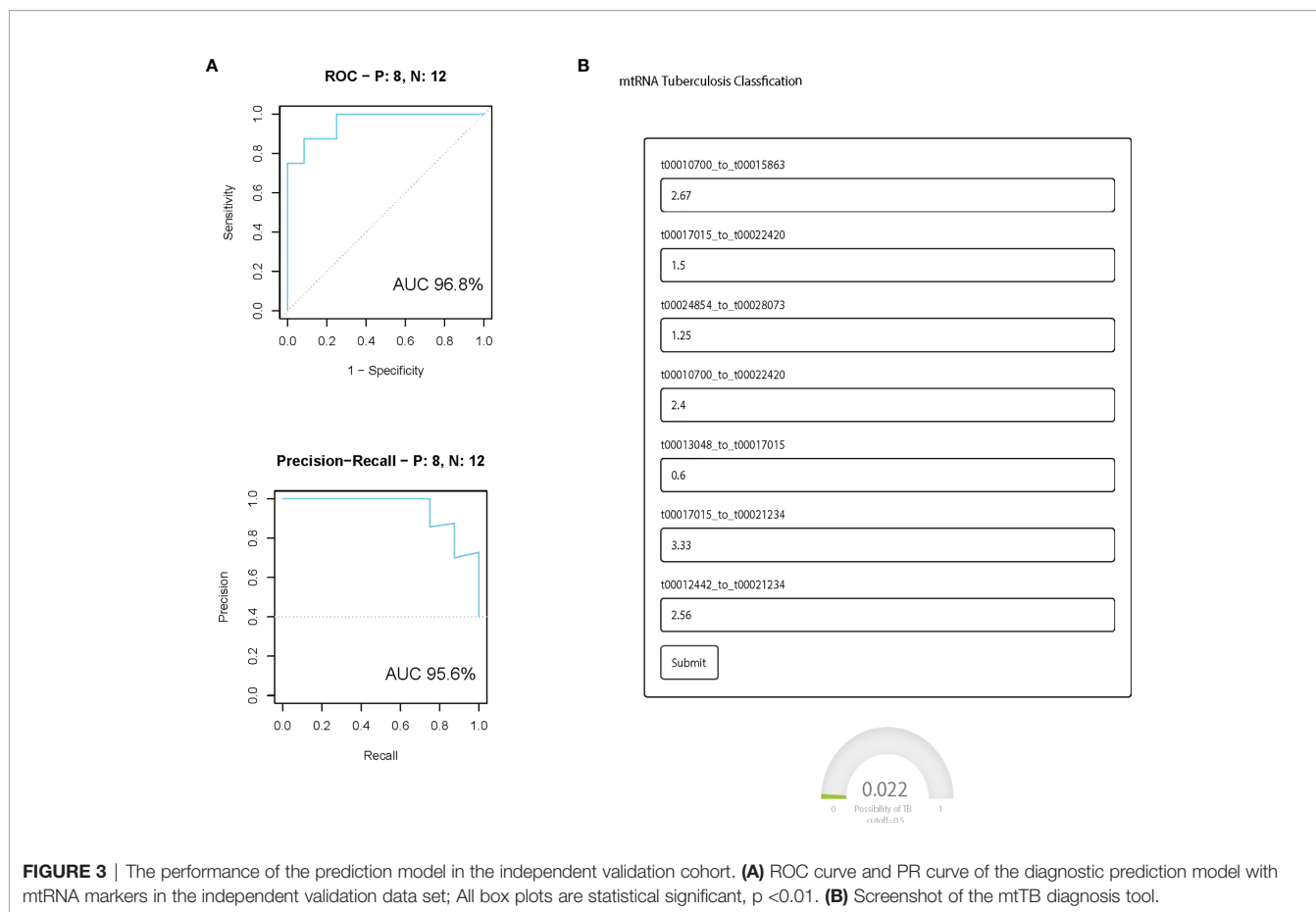
## The Prediction Model in the Independent Validation Cohort and mtTB

We further evaluated the prediction model in the independent validation cohort. The boxplot shows eight selected mtRNA expression levels in the validation dataset (**Figure 3A**). There is significant variation remaining, except for the t00016493\_to\_t00024522, which is marginal for the validation cohort. For the prediction model, the AUC was 0.968 between TB and non-TB cases and 95.6 (AUC) in the PR curve which infers the strong classification power for TB screening. At the same time, we have developed a user-friendly webpage where doctors only need to input the mtRNA pairs to get the probability of TB diagnosis (**Figure 3B**).

## DISCUSSION

Globally, *M. tuberculosis* drastically affects not only TB patients but also asymptomatic undiagnosed subjects in the community. Fast and precise diagnosis is critical for the control of TB spread and sufficient antimicrobial therapy. Although there are multiple methods in the clinical diagnosis of pulmonary TB, such as sputum smear which provides rapid results and is widely used in clinical laboratories, this traditional method shows a low positive rate of 20% to 30%. Moreover, the gold standard of pulmonary TB diagnosis requires a long incubation time (4–8 weeks) (McNerney et al., 2012). In the early stage of TB infection, one





unmet challenge in TB diagnosis is to accurately differentiate other lung diseases from TB with similar clinical symptoms and radiological features.

Circulating small non-coding RNAs have been broadly explored as novel and non-invasive diagnostic and prognostic biomarkers. Many studies have shown that circulating miRNAs serve as potential biomarkers for the detection of TB. However, the performance of miRNA-based TB diagnostic signatures is limited (Zhang et al., 2013; Latorre et al., 2015). Interestingly, our mtRNA signature, derived from the PBMC non-canonical sncRNAs, shows superiority over the miRNA-based signature in diagnosing pulmonary tuberculosis (Pedersen et al., 2019). Compared with miRNA, non-canonical small RNAs such as tsRNAs exhibit a surprising complexity and variability in their sequence (Shi et al., 2019). Moreover, their extraordinary performance in cancer diagnosis and prognosis may be due to the additional complex coating of non-canonical small non-coding RNAs (Gu et al., 2020; Zhu et al., 2021; Zuo et al., 2021).

An affordable, reproducible, and non-invasive method for predicting the severity of TB is required to support longitudinal management and clinical decision-making. In this study, we aimed to develop blood-based screening to improve the sensitivity and specificity of classifications between normal and TB patients. To our knowledge, this is the first time that machine

learning algorithms have been used to diagnose TB by mtRNA on the clinical system. Furthermore, this algorithm has been implemented into a user-friendly Shiny app, an R package that makes it easy to build interactive web apps straight from R, to support further independent investigations of its clinical practice (Sievert, 2020). Previous miRNA-based TB diagnostic tools were either inaccurate or difficult to use (Zhou et al., 2016; Sampath et al., 2021). This shiny app only needs to input the expression ratio of mtRNA, and the result can be obtained quickly after clicking submit, which greatly reduces the user's time.

However, the model was established for the diagnosis of TB and the questions that require further investigation still remain. First, the mtRNA-based signature functions are unknown. Second, large-scale, multicenter case-control studies are warranted to validate our results and identify the signature. Third, since we obtained the sequences of mtRNA signature, all candidate genes need to be validated by quantitative PCR.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

The Ethics Committee of Bengbu Medical College approved this study, with written informed consent obtained from all subjects, which conformed to the standard indicated by the Declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

ZL and SH had the design and launched the study. ZL and SH processed the statistical data analyses, and all authors revised the manuscript and approved the version for publication. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Anders, S., Pyl, P. T., and Huber, W. (2015). Htseq—A Python Framework to Work With High-Throughput Sequencing Data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Bando-Campos, G., Juárez-López, D., Román-González, S. A., Castillo-Rodal, A. I., Olvera, C., López-Vidal, Y., et al. (2019). Recombinant O-Mannosylated Protein Production (Psts-1) From Mycobacterium Tuberculosis in Pichia Pastoris (Komagataella Phaffii) as a Tool to Study Tuberculosis Infection. *Microb. Cell Fact.* 18, 1–19. doi: 10.1186/s12934-019-1059-3
- de Araujo, L. S., Ribeiro-Alves, M., Leal-Calvo, T., Leung, J., Durán, V., Samir, M., et al. (2019). Reprogramming of Small Noncoding RNA Populations in Peripheral Blood Reveals Host Biomarkers for Latent and Active Mycobacterium Tuberculosis Infection. *mBio* 10, e01037–e01019. doi: 10.1128/mBio.01037-19
- Deng, Y., Zhu, Y., Wang, H., Khadka, V. S., Hu, L., Ai, J., et al. (2019). Ratio-Based Method to Identify True Biomarkers by Normalizing Circulating Nerna Sequencing and Quantitative PCR Data. *Anal. Chem.* 91, 6746–6753. doi: 10.1021/acs.analchem.9b00821
- Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., et al. (2020). The NF-Core Framework for Community-Curated Bioinformatics Pipelines. *Nat. Biotechnol.* 38, 276–278. doi: 10.1038/s41587-020-0439-x
- Gu, W., Shi, J., Liu, H., Zhang, X., Zhou, J. J., Li, M., et al. (2020). Peripheral Blood Non-Canonical Small Non-Coding RNAs as Novel Biomarkers in Lung Cancer. *Mol. Cancer* 19, 1–6. doi: 10.1186/s12943-020-01280-9
- Hirose, M., Küstner, A., Schilf, P., Tietjen, A. K., Jöhren, O., Huebbe, P., et al. (2019). A Natural Mtdna Polymorphism in Complex III Is a Modifier of Healthspan in Mice. *Int. J. Mol. Sci.* 20, 2359. doi: 10.3390/ijms20092359
- Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., and Pütz, J. (2009). Trnadb 2009: Compilation of tRNA Sequences and tRNA Genes. *Nucleic Acids Res.* 37, D159–D162. doi: 10.1093/nar/gkn772
- Larriba, E., Rial, E., and Del Mazo, J. (2018). The Landscape of Mitochondrial Small Non-Coding RNAs in the Pgc's of Male Mice, Spermatogonia, Gametes and in Zygotes. *BMC Genomics* 19, 1–12. doi: 10.1186/s12864-018-5020-3
- Latorre, I., Leidinger, P., Backes, C., Domínguez, J., de Souza-Galvão, M. L., Maldonado, J., et al. (2015). A Novel Whole-Blood MiRNA Signature for a Rapid Diagnosis of Pulmonary Tuberculosis. *Eur. Respir. J.* 45, 1173–1176. doi: 10.1183/09031936.00221514
- McNerney, R., Mauerer, M., Abubakar, I., Marais, B., Mchugh, T. D., Ford, N., et al. (2012). Tuberculosis Diagnostics and Biomarkers: Needs, Challenges, Recent Advances, and Opportunities. *J. Infect. Dis.* 205, S147–S158. doi: 10.1093/infdis/jir860
- Pedersen, J. L., Bokil, N. J., and Saunders, B. M. (2019). Developing New TB Biomarkers, Are MiRNA the Answer? *Tuberculosis* 118, 101860. doi: 10.1016/j.tube.2019.101860
- Riggs, C. L., and Podrabsky, J. E. (2018). Mitochondria-Derived Small Non-Coding RNAs in Extreme Anoxia Tolerance. *FASEB J.* 31, 1080–1082. doi: 10.1096/fasebj.31.1\_supplement.1080.2
- Roberts, E. R., and Thomas, K. J. (2013). The Role of Mitochondria in the Development and Progression of Lung Cancer. *Comput. Struct. Biotechnol. J.* 6, e201303019. doi: 10.5936/csbt.201303019
- Sampath, P., Periyasamy, K. M., Ranganathan, U. D., and Bethunaickan, R. (2021). Monocyte and Macrophage MiRNA: Potent Biomarker and Target for Host-Directed Therapy for Tuberculosis. *Front. Immunol.* 12. doi: 10.3389/fimmu.2021.667206
- Shi, J., Ko, E.-A., Sanders, K. M., Chen, Q., and Zhou, T. (2018). Proteomics, and Bioinformatics, SPORTS1.0: A Tool for Annotating and Profiling Non-Coding Rnas Optimized for rRNA-and tRNA-Derived Small Rnas. *Genomics Proteomics Bioinf.* 16, 144–151. doi: 10.1016/j.gpb.2018.04.004
- Shi, J., Zhang, Y., Zhou, T., and Chen, Q. (2019). tsRNAs: The Swiss Army Knife for Translational Regulation. *Trends Biochem. Sci.* 44, 185–189. doi: 10.1016/j.tibs.2018.09.007
- Sievert, C. (2020). *Interactive Web-Based Data Visualization With R, Plotly, and Shiny* (CRC Press).
- Turner, C. T., Gupta, R. K., Tsaliki, E., Roe, J. K., Mondal, P., Nyawo, G. R., et al. (2020). Blood Transcriptional Biomarkers for Active Pulmonary Tuberculosis in a High-Burden Setting: A Prospective, Observational, Diagnostic Accuracy Study. *Lancet Respir. Med.* 8, 407–419. doi: 10.1016/S2213-2600(19)30469-2
- Wang, C., Yang, S., Sun, G., Tang, X., Lu, S., Neyrolles, O., et al. (2011). Comparative Mirna Expression Profiles in Individuals With Latent and Active Tuberculosis. *PLoS One* 6, e25832. doi: 10.1371/journal.pone.0025832
- Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., et al. (2018). Missing Value Imputation Approach for Mass Spectrometry-Based Metabolomics Data. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-017-19120-0
- Zhang, X., Guo, J., Fan, S., Li, Y., Wei, L., Yang, X., et al. (2013). Screening and Identification of Six Serum Micronas as Novel Potential Combination Biomarkers for Pulmonary Tuberculosis Diagnosis. *PLoS One* 8, e81076. doi: 10.1371/journal.pone.0081076
- Zhou, M., Yu, G., Yang, X., Zhu, C., Zhang, Z., and X., Zhan. (2016). Circulating Micronas as Biomarkers for the Early Diagnosis of Childhood Tuberculosis Infection. *Mol. Med. Rep.* 13, 4620–4626. doi: 10.3892/mmr.2016.5097
- Zhu, Y., Chen, S., Ling, Z., Winnicki, A., Xu, L., Xu, S., et al. (2021). Comprehensive Analysis of a tRNA-Derived Small RNA in Colorectal Cancer. *Front. Oncol.* doi: 10.3389/fonc.2021.701440
- Zuo, Y., Chen, S., Yan, L., Hu, L., Bowler, S., Zitello, E., et al. (2021). Development of a tRNA-Derived Small RNA Diagnostic and Prognostic Signature in Liver Cancer. *Genes Dis.* 9 (2), 393–400.

## FUNDING

This work was supported in part by funding from the Key Research and Development Program of Guangxi Zhuang Autonomous Region (No. AB16380152), in part from the Key Research and Development Program of Liuzhou (2018BJ10509) and in part from the ‘139’ Incubation Program for high-level medical talents in Guangxi.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.850279/full#supplementary-material>

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ling, Huang, Wen, Tang, Huang, Wei, Liu and Wu. 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.