BioTechniques

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ibtn20

RNA extraction and RNA-sequencing method for transcriptomic analysis of *Mycobacterium tuberculosis*

Morgan R. Hiebert, Meenu K. Sharma, Alwyn Go, Christine Bonner, Vanessa Laminman, Morag Graham & Hafid Soualhine

To cite this article: Morgan R. Hiebert, Meenu K. Sharma, Alwyn Go, Christine Bonner, Vanessa Laminman, Morag Graham & Hafid Soualhine (16 Feb 2025): RNA extraction and RNA-sequencing method for transcriptomic analysis of *Mycobacterium tuberculosis*, BioTechniques, DOI: 10.1080/07366205.2025.2457887

To link to this article: https://doi.org/10.1080/07366205.2025.2457887

© 2025 Copyright of the Crown in Canada. Public Health Agency of Canada. Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 16 Feb 2025.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ibtn20

BIOTECHNIQUES
https://doi.org/10.1080/07366205.2025.2457887

Taylor & Francis
Taylor & Francis Group

REPORT

OPEN ACCESS Check for updates

RNA extraction and RNA-sequencing method for transcriptomic analysis of *Mycobacterium tuberculosis*

Morgan R. Hiebert[a,b], Meenu K. Sharma[a,b], Alwyn Go[c], Christine Bonner[c], Vanessa Laminman[c], Morag Graham[b,c] and Hafid Soualhine[a,b]

[a]National Reference Centre for Mycobacteriology, National Microbiology Laboratory Branch, Public Health Agency of Canada, Winnipeg, Canada; [b]Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Canada; [c]Genomics Core Facility, National Microbiology Laboratory Branch, Public Health Agency of Canada, Winnipeg, Canada

ABSTRACT

RNA-sequencing (RNA-seq) technologies have advanced exponentially in recent years, however, the application of RNA-seq to *Mycobacterium tuberculosis* remains limited. We present a wet-lab and computational protocol for RNA-seq based transcriptomics that was tested on 12 replicates each of 11 clinical isolates of *M. tuberculosis* ($n=132$) grown *in vitro* with and without pyrazinamide exposure. This RNA extraction method uses low-volume cultures, mechanical lysis, TRIzol™ phase separation, and column-based purification to produce high yields of pure, intact RNA followed by rRNA depletion and cDNA library preparation. The detection of unique transcripts was optimized at a sequencing depth of 15 million reads. This method detected differential RNA expression in experimental sets with and without pyrazinamide exposure, demonstrating that the method is suitable for RNA-seq applications.

MULTIDISCIPLINARY ABSTRACT

We propose a complete wet-lab and computational protocol for RNA-sequencing of *Mycobacterium tuberculosis* including RNA extraction, rRNA depletion, cDNA library preparation, high-throughput sequencing, and differential expression analysis. This research contributes to the literature by providing a start-to-finish methodology for RNA-sequencing of *M. tuberculosis*.

METHOD SUMMARY

We describe a complete wet-lab and computational workflow for RNA-sequencing based transcriptomics of *M. tuberculosis* grown as low-volume cultures *in vitro*, including: (1) RNA extraction, (2) rRNA depletion, (3) cDNA library preparation, (4) high-throughput sequencing, and (5) differential expression analysis.

ARTICLE HISTORY

Received 23 August 2024
Accepted 20 January 2025

KEYWORDS

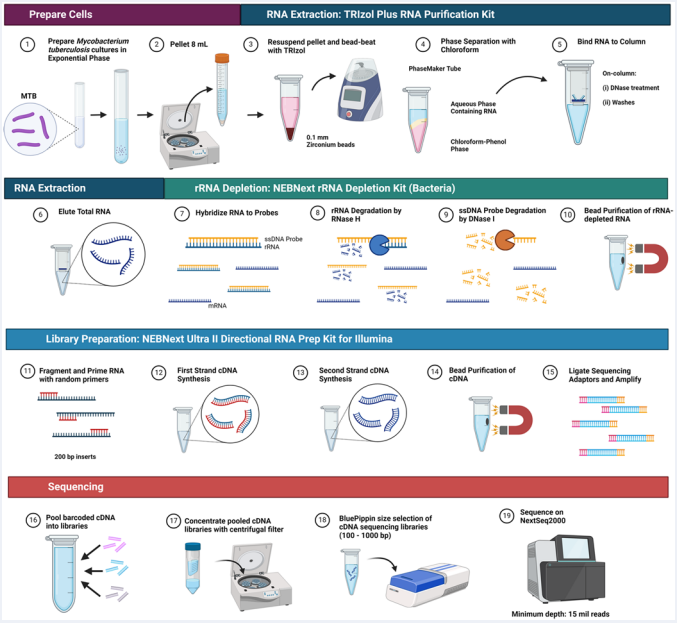
Mycobacterium tuberculosis; next-generation sequencing; RNA extraction; RNA-sequencing; transcriptomics

CONTACT Hafid Soualhine hafid.soualhine@phac-aspc.gc.ca National Reference Centre for Mycobacteriology, National Microbiology Laboratory Branch, Public Health Agency of Canada, Winnipeg, Canada

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07366205.2025.2457887.

© 2025 Copyright of the Crown in Canada. Public Health Agency of Canada. Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

GRAPHICAL ABSTRACT



Created with Biorender.com

1. Introduction

Since the complete genome of *Mycobacterium tuberculosis* was first described in 1998 [1], numerous studies have described methods for mycobacterial RNA isolation [2–10] and transcriptomics [9–17]. Effective methods for RNA extraction consider the stringent biosafety requirements for manipulating this Risk Group 3 (RG-3) organism [18] as well as technical challenges posed by the rigid mycobacterial cell wall [19, 20], requiring harsh lysis conditions that have historically limited the quality and quantity of isolated RNA. Most commonly, studies describe a phenol-chloroform bi-phasic extraction paired with mechanical lysis [2], however, column-based purification kits have been shown to improve mycobacterial RNA yield [4, 5, 7, 8]. Mycobacterial transcriptomics has been described using various technologies, including microarrays [11–13] and reverse-transcriptase quantitative PCR (RT-qPCR) [9, 14]. Despite recent advancements in high-throughput transcriptomic technologies, the implementation of RNA-sequencing (RNA-seq) for *M. tuberculosis* has been limited by high cost, technical challenges, and bioinformatics requirements for data analyses [10, 15–17, 21].

Here, we describe a completely experimental and computational workflow for RNA-seq based transcriptomics of *M. tuberculosis* grown in low-volume cultures *in vitro*, including: (1) RNA extraction, (2) rRNA depletion, (3) cDNA library preparation, (4) high-throughput sequencing, and (5) differential expression analysis. Specifically, *M. tuberculosis* was grown *in vitro* using liquid MGIT™ media, and RNA was extracted using the TRIzol™ Plus RNA Purification Kit and Phasemaker™ Tubes Complete System with modifications to the manufacturer's instructions [22]. Following RNA isolation, we performed rRNA depletion and cDNA library preparation using the NEBNext® rRNA Depletion Kit (Bacteria) and NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® before sequencing on the Illumina NextSeq 2000 platform [23]. By describing a start-to-finish RNA-seq workflow from RNA extraction to computational differential expression analysis, we aim to improve the accessibility of RNA-seq based transcriptomics for *M. tuberculosis* research queries. A detailed, step-by-step protocol is freely available on protocols.io (DOI: dx.doi.org/10.17504/protocols.io.3byl49dqrgo5/v1).

2. Materials and methods

2.1. Materials

2.2.1. Supplies

(1) 0.9 McFarland Standard (E1039; Thermo Fisher Scientific, MA, USA)
(2) 1.5% Agarose dye-free Gel Cassette with internal standards, 250 bp – 1.5 kb (BDF1510; Sage Science, MA, USA)
(3) Agilent RNA 6000 Nano Kit (5067-1511; Agilent Technologies, CA, USA)
(4) Amicon® Ultra-0.5 Centrifugal Filter Units (UFC505096; MilliporeSigma, MA, USA)
(5) BD BACTEC™ MGIT™ 960 PZA Kit (245128; Becton Dickinson, NJ, USA)
(6) BD BACTEC™ MGIT™ 960 PZA Medium (245115; Becton Dickinson)
(7) Chloroform (34854; MilliporeSigma)
(8) Ethanol (P016EA95; Commercial Alcohols Incorporated, ON, CA)
(9) Falcon™ 15 mL Conical Centrifuge tubes (05-527-90; Thermo Fisher Scientific)
(10) Genomic DNA ScreenTape and Reagents (5067-5366; 5067-5365; Agilent Technologies)
(11) Microcentrifuge Tubes: 2.0 mL (05-408-141; Thermo Fisher Scientific)
(12) Microcentrifuge Tubes with Screw Caps: 2 mL (02-681-375; Thermo Fisher Scientific)
(13) NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs) (E6440S; New England Biolabs, MA, USA)
(14) NEBNext® rRNA Depletion Kit (Bacteria) with RNA Sample Purification Beads (E7860X; New England Biolabs)
(15) NEBNext® Ultra™ II Directional RNA Library Prep kit with Sample Purification Beads (E7765L; New England Biolabs)
(16) NextSeq™ 1000/2000 P2 300M Reagents (600 cycles) (20075295; Illumina, CA, USA)
(17) PureLink™ DNase Set (12185010; Thermo Fisher Scientific)
(18) Qubit™ dsDNA High Sensitivity Assay kit (Q32854; Thermo Fisher Scientific)
(19) Qubit™ RNA High Sensitivity Assay kit (Q32852; Thermo Fisher Scientific)
(20) RNaseZap™ RNase Decontamination Solution (AM9780; Invitrogen, MA, USA)
(21) Triple-Pure™ 0.1 mm zirconium beads (D1132-01TP; Benchmark Scientific, NJ, USA)
(22) TRIzol™ Plus RNA Purification Kit and Phasemaker™ Tubes Complete System (A33254; Thermo Fisher Scientific)
(23) UltraPure™ DNase/RNase-free Distilled Water (10977035; Invitrogen)

2.2.2. Equipment

(24) 4200 TapeStation System (G2991BA; Agilent Technologies)
(25) Agilent 2100 Bioanalyzer System (G2938A; Agilent Technologies)
(26) BD BACTEC™ MGIT™ 960 automated mycobacterial detection system (445885; Becton Dickinson)
(27) BluePippin Size Selection System (BLU0001; Sage Science)
(28) Fastprep-24™ Classic bead beating grinder and lysis system (116004500; MP Biomedicals, CA, USA)
(29) Labnet Orbit™ 300 Multipurpose Digital Shaker (S2030-300-B; Mandel Scientific Company Inc, ON, CA).
(30) MiniSpin® Mini Centrifuge (022620100; Eppendorf, Hamburg, Germany)
(31) NanoDrop™ 2000 spectrophotometer (ND-2000; Thermo Fisher Scientific)
(32) NEBNext® Magnetic Separation Rack (S1515S; New England Biolabs)
(33) NextSeq 2000 System (Illumina)
(34) Qubit™ 4.0 Fluorometer (Q33238, Invitrogen)
(35) Sorvall™ ST 16R Centrifuge (75004240; Thermo Fisher Scientific)
(36) Thermocycler
(37) VWR® Analog Vortex Mixer (10153-838; Avantor, PA, USA)

3. Methods

3.1. Study isolates

We assessed the RNA-seq workflow by evaluating the differential expression of *M. tuberculosis* isolates in response to pyrazinamide (PZA) drug exposure. Clinical isolates of *M. tuberculosis* belonging to a

geographically distinct tuberculosis outbreak ($n = 11$) were selected from the National Reference Center for Mycobacteriology culture collection. The attenuated *M. tuberculosis* reference strain H37Ra ATCC 25177 was used for protocol optimization.

3.2. Culture conditions

In a biosafety level 3 (BSL-3) laboratory, liquid cultures were grown in BD BACTEC™ MGIT™ PZA Medium prepared with BACTEC™ MGIT™ 960 PZA Supplement and incubated at 37 °C in an orbital shaker for 2–4 weeks. The exponential phase was estimated by comparing culture turbidity to a 0.9 McFarland Standard. Following, cultures for each isolate were treated in triplicate with PZA at a concentration of 100 µg/mL prepared in sterile de-ionized water for one hour, four hours, and 24 hours prior to RNA extraction in addition to untreated controls, resulting in 12 samples prepared for RNA-seq for each isolate.

3.3. RNA extraction

RNA extraction was performed in a BSL-3 laboratory using the TRIzol™ Plus RNA Purification Kit and Phasemaker™ Tubes Complete System, with modifications to the manufacturer's instructions, and in accordance with biosafety guidelines [18, 22, 23]. Because the protocol steps (i.e., centrifugation) and preparation time may affect the transcriptional profile of *M. tuberculosis* prior to the addition of TRIzol, all steps were carried out on ice or at 4 °C in a temperature-controlled centrifuge whenever possible. Additionally, samples from all experimental sets were processed under identical conditions to ensure consistency and mitigate any possible effect on differential expression results.

M. tuberculosis cultures (8 mL) in exponential phase were placed on ice then pelleted by centrifugation (4,000 × g, 20 min, 4 °C). Keeping samples on ice, pellets were washed twice with RNase/DNase-free distilled water, transferred to a screw-capped tube containing 200 µL of 0.1 mm zirconium beads, and centrifuged (7,000 × g, 10 min, 4 °C) to remove the supernatant. The pellet was resuspended in 1 mL TRIzol™, vortexed to mix, and incubated at room temperature for five minutes. This suspension underwent bead-beating in a FastPrep-24™ tissue homogenizer for three cycles (4.5 m/s for 1 min; 1 min rest on ice between cycles). The combined mechanical and chemical lysis serves to disrupt the mycobacterial cell wall, solubilize cell components, and denature proteins. Samples were centrifuged (12,000 × g, 5 min, 4 °C) to pellet cell debris and ≤ 950 µL supernatant was transferred to Phasemaker™ tubes. Chloroform phase separation and column-based RNA purification with PureLink™ on-column degradation of genomic DNA was performed according to the manufacturer's instructions [22]. The resultant 99 µL of eluted total RNA in DNase/RNase-free distilled water was quantified on a Qubit™ 4.0 fluorometer with an RNA High Sensitivity Assay kit. Purity was assessed on a NanoDrop™ 2000 spectrophotometer. Quality was evaluated by RNA Integrity Number (RIN) using the Agilent 2100 Bioanalyzer System and RNA 6000 Nano kit.

3.4. Ribosomal RNA depletion

RNA samples that met the requirements for integrity (RIN > 7) and quantity (10 ng to 1 µg in 11 µL) proceeded with rRNA depletion. Depletion of rRNA was performed using the manufacturer's protocol for library preparation of intact RNA using NEBNext® rRNA Depletion Kit (Bacteria) [23]. Briefly, 11 µL of total RNA in DNase/RNase-free distilled water was used as input for the probe hybridization reaction in which single-stranded DNA (ssDNA) probes bind rRNA in the total RNA sample. RNase H digestion was used to degrade the rRNA bound to the ssDNA capture probes, and then a DNase I digestion reaction was used to remove the ssDNA probes. The rRNA-depleted RNA was purified using NEBNext® RNA Sample Purification Beads and two washes with 80% ethanol before proceeding to cDNA library preparation.

3.5. cDNA library preparation and sequencing

From the rRNA-depleted RNA, cDNA libraries were prepared using the manufacturer's protocol for the NEBNext® Ultra™ II Directional RNA Library Prep Kit [23]. Briefly, priming and fragmentation of intact RNA (RIN > 7) was

performed by incubating RNA samples at 94 °C for 15 minutes with random primers then transferring the tube to ice. First and second-strand cDNA synthesis, end-prep, and adaptor ligation were performed according to the manufacturer's instructions [23]. Unique dual index primer pairs supplied in the NEBNext® Multiplex Oligos for Illumina® kit were used for PCR amplification consisting of initial denaturation at 98 °C for 30 seconds followed by 12 cycles of (98 °C, 10 seconds; 65 °C, 75 seconds), and final extension at 65 °C for 5 minutes. Indexed cDNA was purified according to the manufacturer's instructions using NEBNext® Sample Purification beads [23]. The cDNA was quantified on a Qubit™ 4.0 fluorometer with a dsDNA high-sensitivity assay kit.

To optimize sequencing depth for the detection of unique transcripts, we performed rarefaction analysis for control (untreated) *M. tuberculosis* H37Ra and after four hours of pyrazinamide exposure. Specifically, libraries from a single RNA sample were pooled to achieve various sequencing depths ranging from 5 to 25 million reads on a NextSeq 2000 using NextSeq 1000/2000 P2 (600 cycles) reagents. The number of unique transcripts detected at each sequencing depth was used to determine the optimal sequencing depth for subsequent experimental sets.

Library pooling for experimental sets was performed to achieve a minimum of 15 million paired-end reads per sample on a NextSeq 2000 employing NextSeq 1000/2000 P2 (600 cycles) reagents. Specifically, 1 ng from each of the 20 samples was pooled to yield library concentrations greater than 750 pM. The pooled libraries were concentrated using an Amicon® Ultra-0.5 centrifugal filter unit and size-selected at a range of 150–1000 bp using a BluePippin. Libraries were quantified on a Qubit™ 4.0 fluorometer with a dsDNA high-sensitivity assay kit. Library size and quality were assessed using the Agilent 4200 TapeStation System with the Genomic DNA ScreenTape Analysis kit before sequencing on an Illumina NextSeq 2000 platform at a loading concentration of 700 pM to achieve a minimum of 15 million reads per sample.

3.6. Sequence analysis

The sequence data was uploaded to the Integrated Rapid Infectious Disease Analysis (IRIDA) platform for organization and management [24]. Pre-trimming read quality was assessed with FastQC in Galaxy (version 0.72) [25]. Trimmomatic in Galaxy was employed for adaptor trimming (ILLUMINACLIP:Custom:2:30:10:8), quality trimming (SLIDINGWINDOW:4:20), and filtering reads below 20 nucleotides in length (MINLEN:20) [26]. Post-trimming read quality was assessed with FastQC (version 0.72) [25]. A threshold Phred score of 20 was used to ensure high sequence quality. Kraken2 (version 2.2) [27] was used to detect the presence of contaminating cDNA. Reads were aligned to the *M. tuberculosis* H37Rv reference genome (accession NC_000962.3) using Bowtie2 (version 2.3.4.3) [28]. Alignment quality was assessed using QualiMap BamQC [29]. Read counts per gene were obtained using featureCounts (version 1.6.0.2) [30] with an annotated *M. tuberculosis* H37Rv reference genome (accession AL_123456.3).

3.7. Differential expression analysis

Normalization of read counts and differential expression analysis was performed using DESeq2 (version 1.42.1) [31]. The Wald test with Benjamini-Hochberg correction for multiple testing was used to compare gene expression between PZA exposure conditions (one hour, four hours, and 24 hours) and untreated controls. A *post hoc* adjusted *P*-value less than 0.05 was considered statistically significant. Packages ggplot2 (version 3.5.0) [32] and EnhancedVolcano (version 1.20.0) [33] were used for visualization.

4. Results and discussion

4.1. Modified extraction protocol yields high-quality RNA from M. tuberculosis

The distribution of RNA yields among 132 extractions from clinical isolates of *M. tuberculosis* is shown in Figure 1A. The mean RNA yield was 17.85 ng/µL, ranging from 1.02 ng/µL to 38.8 ng/µL. Variation in the concentration of extracted RNA was found between replicates of a single isolate (standard deviation ±5.19 ng/µL) and between replicates of the same treatment condition (standard deviation ±8.78 ng/µL). RNA purity was assessed using the $A_{260\,nm}$: $A_{280\,nm}$ ratio [34] and averaged 2.21 for all extractions, indicating pure RNA. The quality and intactness of RNA were assessed by RIN [35]. The average RIN was 8.96, showing that intact

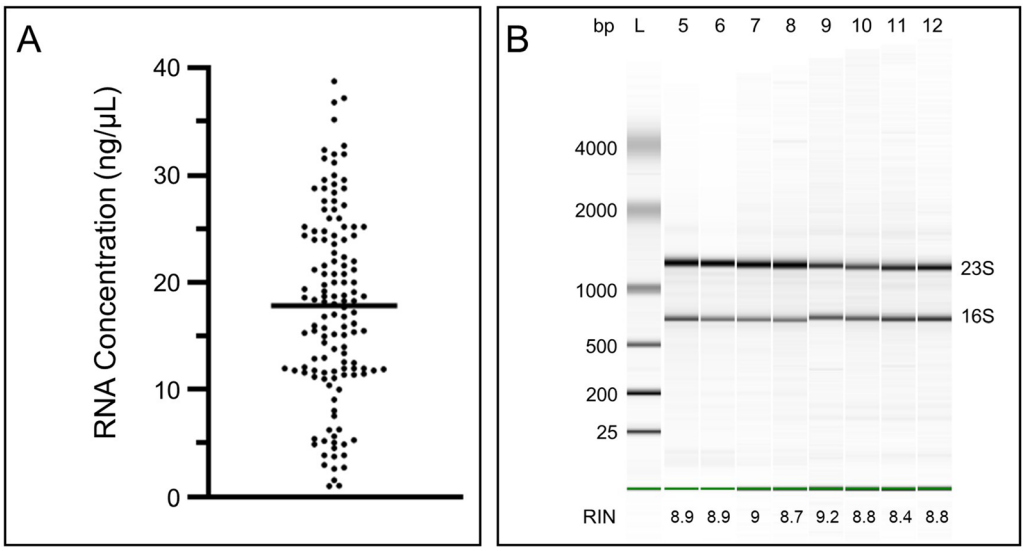


Figure 1. Quantity and quality of RNA extracted from *M. tuberculosis* clinical isolates. (A) Concentrations of RNA extracted from *M. tuberculosis* clinical isolates ($n = 132$). The horizontal line represents the mean RNA concentration. The figure was generated in GraphPad Prism 10. (B) Bioanalyzer Gel Image of RNA extracted from clinical isolates of *M. tuberculosis*. Ladder (L) depicted with base pair (bp) increments. Numbered lanes show 16S and 23S rRNA bands in extracted RNA from representative *M. tuberculosis* cultures. RNA Integrity Number (RIN) is indicated below each lane. Gel and RIN generated by Prokaryote total RNA Nano kit on Agilent 2100 Bioanalyzer for 11 representative RNA extractions from *M. tuberculosis* cultures.

RNA was consistently isolated. A representative subset of RNA extractions is depicted in Figure 1B. These results show that this protocol yields high-quality RNA that surpasses the minimum input requirement of 10 ng RNA in 11 µL for downstream cDNA library preparation in RNA-seq applications [23].

Methods to extract RNA from mycobacteria have been previously described as requiring large culture volumes that are challenging to handle due to biosafety considerations [3, 5, 14, 18]. Here, we describe a method that uses low-volume MGIT™ culture tubes (8 mL) that are commonly used for culturing and phenotypic drug susceptibility testing of *M. tuberculosis* isolates. Others have described that RNA extracted from MGIT™ cultures show degradation and low yields [7]; in contrast, the protocol developed in the present study produced high quantity and quality RNA when culture tubes were agitated in an orbital shaker. We also found that decreasing the diameter of zirconium beads from 0.5 mm to 0.1 mm increased the RNA yield by 30-fold while achieving RNA quality greater than 8.0 RIN under identical culture conditions (Supplementary Table 1). Thus, we present an effective protocol for intact RNA isolation from *M. tuberculosis* grown in low-volume cultures *in vitro* that is suitable for downstream RNA-seq applications in transcriptomics.

4.2. RNA-seq library preparation and sequencing optimized for detection of unique *M. tuberculosis* transcripts

Our RNA-seq workflow for *M. tuberculosis* transcriptomics used the NEBNext® rRNA depletion (bacteria) kit with 10 ng of input RNA in 11 µL and the NEBNext® Ultra™ II Directional RNA Library Prep Kit with Sample Purification Beads. Figure 2A shows the distribution of cDNA library concentrations produced from 132 RNA extractions from clinical isolates of *M. tuberculosis*. The mean concentration of cDNA libraries was 6.28 ng/µL ($n = 132$). Libraries from 20 samples were pooled equimolarly (10 ng each) and the average pool concentration was 3.73 ng/µL. Figure 2B shows Tapestation gel images of three representative pooled libraries of cDNA with a DNA integrity number (DIN) of 1.0, indicating that the pooled libraries comprise sufficiently fragmented inserts for high-throughput sequencing on the NextSeq 2000.

The efficiency of rRNA depletion was determined by comparing the proportion of total non-normalized cDNA reads that align to the rRNA genes in the *M. tuberculosis* reference genome: *rrs* (16S), *rrl* (23S), and *rrf* (5S). rRNA represents 80–95% of the total RNA extracted from *M. tuberculosis* [16, 36]. Here, the mean proportion of total transcripts aligning to rRNA was 0.048% (5S), 5.08% (16S), and 45.66% (23S). These results show efficient depletion of 5S and 16S rRNA and partial depletion of 23S rRNA.

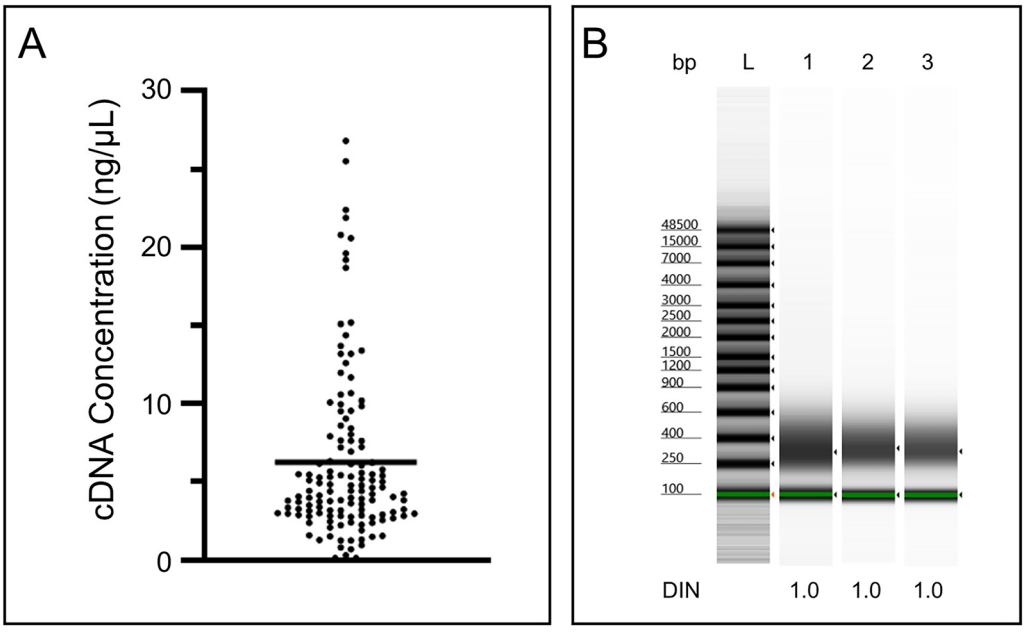


Figure 2. Quantity and quality of cDNA and RNA-seq libraries for *M. tuberculosis* clinical isolates. (A) Concentrations of cDNA generated by library preparation of RNA extracted from cultures of *M. tuberculosis* clinical isolates. The horizontal line represents the mean. $n = 132$. The figure was generated in GraphPad Prism 10. (B) Tapestation Gel Image of representative pooled cDNA sequencing libraries extracted from clinical isolates of *M. tuberculosis*. Ladder (L) depicted with base pair (bp) increments. DNA Integrity Number (DIN) is indicated below each lane. Gel and DIN generated by DNA ScreenTape Analysis kit on the 4200 TapeStation System for three representative cDNA library pools.

To determine the appropriate sequencing depth to optimize the number of unique transcripts detected, we performed rarefaction analysis with *M. tuberculosis* H37Ra in which libraries from a single RNA sample were pooled to achieve varying sequencing depths. The number of unique transcripts detected at each sequencing depth is shown in Figure 3. In general, the number of detectable unique transcripts increased when the sequencing depth was increased from 5 million to 15 million reads and plateaued after a sequencing depth of 15 million reads. This finding was reproducible in both control (no PZA exposure) and PZA exposure conditions. Thus, we ascertained that 15 million reads were optimal for detecting unique transcripts for *M. tuberculosis* RNA-seq applications.

While there remains a gap in the literature regarding RNA-seq of *M. tuberculosis*, two existing methods similarly described the use of rRNA depletion before library preparation [16, 17]. One study proposing dual RNA-seq of *M. tuberculosis* and infected host cells used the NEBNext® rRNA Depletion Kit (Bacteria) in combination with the NEBNext® Depletion Kit (Human/Mouse/Rat) before cDNA library preparation using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina™ [16]. Another study used a different approach by using siTOOLs Biotech's riboPOOL and the NEBNext® Ultra™ II RNA Library Prep Kit for Illumina™ [17]. These methods recommended a sequencing depth of 1 million reads [16] and 10 million reads [17] whereas our rarefaction analysis showed that the discovery of unique bacterial transcripts increased at a depth of 15 million reads.

4.3. Differential expression of pyrazinamide-sensitive M. tuberculosis isolates is well-captured by the current method

The suitability of the present RNA-seq wet-lab workflow for downstream differential expression applications was assessed by evaluating the response of *M. tuberculosis* cultures to PZA drug exposure. Figure 4 depicts the differential expression of PZA-sensitive *M. tuberculosis* cultures in response to PZA exposure (one hour, four hours, and 24 hours) compared to drug-free controls. Specifically, Figure 4A shows volcano plots of the fold-change in the expression of each gene at each post-exposure time-point relative to the drug-free control. In general, these plots show that the number of differentially expressed genes and the magnitude of fold-change increased with PZA exposure time. Figure 4B shows a heatmap of the top 50 significantly differentially expressed genes (P-value < 0.05). Specifically, genes with the highest $\log_2$ fold-change in any condition were selected, and

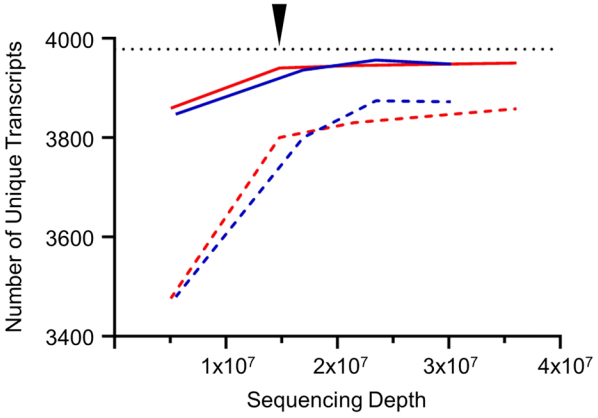


Figure 3. A Rarefaction Curve of Unique Transcripts Detected in *M. tuberculosis* H37Ra. The number of unique transcripts observed at an increasing RNA sequencing depth is shown for *M. tuberculosis* H37Ra without pyrazinamide exposure (blue) and four hours of pyrazinamide exposure (red) where the number of unique genes with counts greater than zero (solid line) and with counts greater than ten (dotted line) is shown. The black dotted line shows the total number of known genes in *M. tuberculosis*, $n = 3978$. The black arrow indicates a sequencing depth of 15 million reads. The figure was generated in GraphPad Prism 10.

the heatmap shows the $\log_2$ fold-change for that gene in all post-exposure time points. Like Figure 4A, this figure shows that the magnitude of fold-change for each gene increases with PZA exposure time. These results show that differential expression between the experimental conditions was well-captured by the present wet lab and RNA-seq analysis workflows. Together, this work shows that the proposed method is suitable for investigating differential expression in downstream RNA-seq analysis applications.

5. Conclusions

This work demonstrates an effective method for RNA isolation, cDNA library preparation, high-throughput sequencing, and differential expression analysis for RNA-seq of *M. tuberculosis*.

6. Future perspective

While high-throughput transcriptomic technologies have undergone exponential advancements in recent years, the application of RNA-seq to *M. tuberculosis* has remained limited. Our efficient and effective method of RNA isolation and cDNA library preparation improves the accessibility of RNA-seq applications for *M. tuberculosis*. Harnessing this method, priority research areas for *M. tuberculosis* including mechanisms of antimicrobial resistance may be studied with a higher resolution compared to older methods such as microarray or RT-qPCR. Future work on the co-purification of proteins may be developed using the present extraction protocol, which would enable simultaneous transcriptomic and proteomic profiling.

7. Article highlights

7.1. Background

- High-throughput transcriptomic technologies have undergone exponential advancements in recent years, however, the application of RNA-sequencing (RNA-seq) to *M. tuberculosis* has remained limited.

7.2. Experimental

- Liquid cultures of clinical *M. tuberculosis* isolates ($n = 11$) grown in low-volume MGIT™ culture tubes with 12 replicates prepared for RNA extraction ($n = 132$).
- RNA extraction from *M. tuberculosis* liquid cultures consisted of mechanical lysis, TRIzol™ phase separation, and column-based purification.

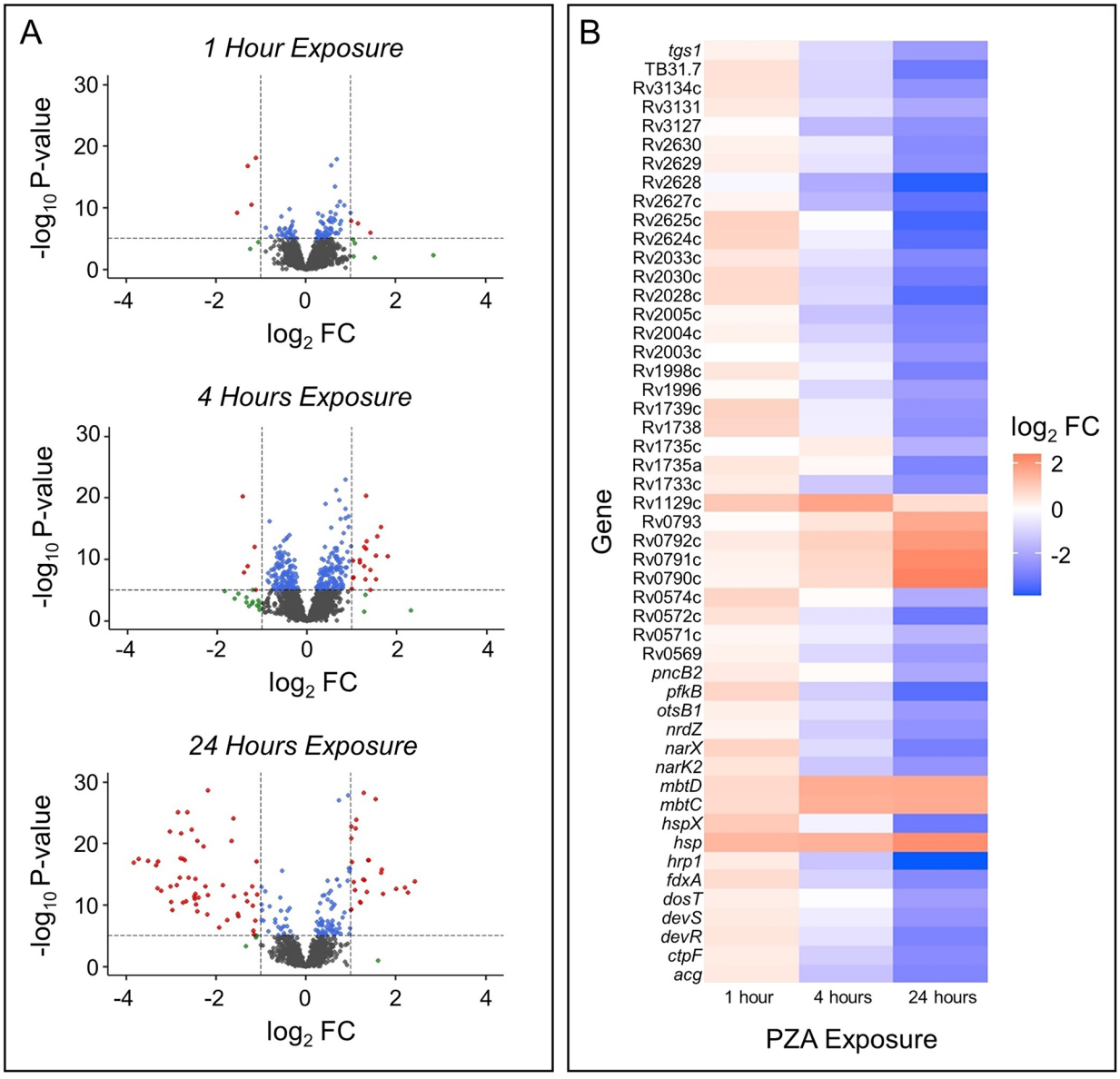


Figure 4. Differential expression of pyrazinamide-sensitive *M. tuberculosis* isolates following pyrazinamide exposure compared to untreated controls. (A) Volcano plot showing differential expression of each gene at one, four, and 24 hours following drug exposure at 100 µg/mL pyrazinamide. Dotted lines represent thresholds for significant differential expression, including a *P*-value of 0.05 (horizontal) and a $\log_2$ fold-change greater than one in magnitude (vertical). (B) Heatmap of the top 50 significantly differentially expressed genes. Genes with the top 50 highest $\log_2$ fold-change in any condition are shown at all post-exposure time points. FC, fold-change. Statistical analysis was performed using DESeq2 (v1.42.1) employing with Wald test with Benjamini-Hochberg correction. Volcano plots were created using the EnhancedVolcano package (v1.20.0) and the heatmap was generated using R stats ggplot2 (v3.5.0). Three biological replicates were performed for each condition for each isolate.

- The NEBNext® rRNA depletion (bacteria) kit and the NEBNext® Ultra™ II Directional RNA Library Prep Kit with Sample Purification Beads were used to prepare cDNA libraries before high-throughput sequencing using an Illumina NextSeq 2000.

7.3 Results and discussion

- The described protocol produced high yields of pure, intact RNA with a mean concentration of 17.85 ng/µL.
- Isolated RNA was successfully used for RNA-seq library preparation and high-throughput sequencing for transcriptomics.

- We determined that 15 million reads were optimal for detecting unique transcripts for *M. tuberculosis* RNA-seq applications.
- The suitability of the present RNA-seq wet-lab workflow for downstream differential expression applications was assessed by evaluating the response of *M. tuberculosis* isolates to pyrazinamide drug exposure.

7.4. Conclusion

- This work demonstrates an effective method for RNA isolation, cDNA library preparation, high-throughput sequencing, and differential transcript expression analysis for RNA-seq of *M. tuberculosis*.

Acknowledgements

We express our gratitude to Dr. Richard LeDuc and Dr. Denice Bay for their valuable consultation for this project. We would also like to acknowledge the National Microbiology Laboratory Branch Genomics Core Facility and the National Reference Centre for Mycobacteriology for their facilities and technical assistance with this project.

Author contributions

Conceptualization: Meenu K. Sharma, Hafid Soualhine. Methodology: Morgan R. Hiebert, Alwyn Go, Christine Bonner, Vanessa Laminman, Hafid Soualhine, Meenu K. Sharma, Morag Graham. Software: Morgan R. Hiebert. Investigation: Morgan R. Hiebert. Resources: Meenu K. Sharma, Hafid Soualhine. Writing (original draft): Morgan R. Hiebert. Writing (review and editing): Morgan R. Hiebert, Meenu K. Sharma. Sharma, Hafid Soualhine, Morag Graham, Alwyn Go, Christine Bonner, Vanessa Laminman. Supervision: Meenu K. Sharma, Hafid Soualhine. Project administration: Meenu K. Sharma, Hafid Soualhine. Funding acquisition: Hafid Soualhine.

Disclosure statement

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was used in the production of this manuscript.

Funding

This study was funded by the Public Health Agency of Canada, the Genomics Research and Development Initiative (GRDI-07, HS), Research Manitoba (MH studentship), and the Manitoba Lung Association (MH studentship). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability statement

The authors declare that the data supporting the findings of this study are available within the article and from the corresponding author upon request. The protocol is freely available on protocols.io (DOI: dx.doi.org/10.17504/protocols.io.3byl49dqrgo5/v1).

References

Papers of special note have been highlighted as either of interest (*) or of considerable interest (**) to readers.

[1] Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393(6685):537–544. doi: 10.1038/31159
[2] Mangan JA, Sole KM, Mitchison DA, et al. An effective method of RNA extraction from bacteria refractory to disruption, including mycobacteria. Nucleic Acids Res. 1997;25(3):675–676. doi: 10.1093/nar/25.3.675
[3] Mahenthiralingam E. Extraction of RNA from mycobacteria. Methods Mol Biol. 1998;101:65–75. doi: 10.1385/0-89603-471-2:65
[4] Dietrich G, Schaible UE, Diehl KD, et al. Isolation of RNA from mycobacteria grown under *in vitro* and *in vivo* conditions. FEMS Microbiol Lett. 2000;186(2):177–180. doi: 10.1111/j.1574-6968.2000.tb09100.x
[5] Rustad TR, Roberts DM, Liao RP, et al. Isolation of mycobacterial RNA. Methods Mol Biol. 2009;465:13–21. doi: 10.1007/978-1-59745-207-6_2

[6] Alli AOT, Mangan JA, Butcher PD, et al. Optimization of RNA Extraction in *Mycobacterium tuberculosis* for studying intracellular gene expression. AJCEM. 2009;10(2):64–79. doi: 10.4314/ajcem.v10i2.7508
[7] Oh TS, Kang HY, Nam YS, et al. An effective method of RNA extraction from *Mycobacterium tuberculosis*. Ann Clin Microbiol. 2016;19(1):20–23. doi: 10.5145/ACM.2021.19.1.20
[8] Mvubu N, Salig A, Moopanar K, et al. A quick, easy and efficient protocol for extracting high-quality RNA from *Mycobacterium tuberculosis* using a spin column commercial kit. BMC Res Notes. 2023;16(1):145. doi: 10.1186/s13104-023-06424-w
[9] Bera JH, Raj ALS, Kumar H, et al. Effective RNA isolation method for gram-positive and acid-fast bacteria: metamorphosed from conventional RNA isolation techniques. Arch Microbiol. 2024;206(9):369. doi: 10.1007/s00203-024-04077-2
[10] Olmo-Fontánez AM, Allué-Guardia A, Garcia-Vilanova A, et al. Impact of the elderly lung mucosa on *Mycobacterium tuberculosis* transcriptional adaptation during infection of alveolar epithelial cells. Microbiol Spectr. 2024;12(12):e0179024. doi: 10.1128/spectrum.01790-24
[11] Poonawala H, Zhang Y, Kuchibhotla S, et al. Transcriptomic responses to antibiotic exposure in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2024;68(5):e0118523. doi: 10.1128/aac.01185-23
[12] Cox RA, Garcia MJ. Adaptation of mycobacteria to growth conditions: a theoretical analysis of changes in gene expression revealed by microarrays. PLOS One. 2013;8(4):e59883. doi: 10.1371/journal.pone.0059883
[13] Waddell SJ, Laing K, Senner C, et al. Microarray analysis of defined *Mycobacterium tuberculosis* populations using RNA amplification strategies. BMC Genomics. 2008;9(1):94. doi: 10.1186/1471-2164-9-94
[14] Khan H, Paul P, Goar H, et al. *Mycobacterium tuberculosis* PhoP integrates stress response to intracellular survival by regulating cAMP level. Elife. 2024;13:RP92136. doi: 10.7554/eLife.92136
[15] Ko EM, Min J, Kim H, et al. Molecular characteristics of drug-susceptible *Mycobacterium tuberculosis* clinical isolates based on treatment duration. Osong Public Health Res Perspect. 2024;15(5):385–394. doi: 10.24171/j.phrp.2024.0101
[16] Pisu D, Huang L, Rin Lee BN, et al. Dual RNA-sequencing of *Mycobacterium tuberculosis*-infected cells from a murine infection model. STAR Protoc. 2020;1(3):100123. doi: 10.1016/j.xpro.2020.100123
[17] Ma S, Jones RMJr, Gleason NS, et al. Experimental and computational workflow for RNA sequencing in *Mycobacterium tuberculosis*: from total RNA to differentially expressed genes. Methods Mol Biol. 2021;2314:481–512. doi: 10.1007/978-1-0716-1460-0_21
[18] Public Health Agency of Canada (PHAC). Biosafety Directive for *Mycobacterium tuberculosis* Complex (MTBC). Canada: PHAC; 2017.
[19] Alderwick LJ, Harrison J, Lloyd GS, et al. The mycobacterial cell wall-peptidoglycan and arabinogalactan. Cold Spring Harb Perspect Med. 2015;5(8):a021113. doi: 10.1101/cshperspect.a021113
[20] Batt SM, Burke CE, Moorey AR, et al. Antibiotics and resistance: the two-sided coin of the mycobacterial cell wall. Cell Surf. 2020;6:100044. doi: 10.1016/j.tcsw.2020.100044
[21] Martínez-Pérez A, Estévez O, González-Fernández Á. Contribution and future of high-throughput transcriptomics in battling tuberculosis. Front Microbiol. 2022;13:835620. doi: 10.3389/fmicb.2022.835620
[22] Invitrogen, ThermoFisher Scientific. User guide: TRIzol plus RNA purification kit and phasemaker tubes complete system. Waltham (MA): Invitrogen ThermoFisher Scientific; 2016.

****The manufacturer's instructions for the TRIzol Plus RNA Purification Kit and Phasemaker Tubes Complete System that were modified to generate the protocol for mycobacterial RNA extraction detailed herein.**

[23] New England Biolabs. Protocol for library preparation of intact RNA using NEBNext rRNA depletion kit (Bacteria) (NEB #E7850, NEB #E7860) and NEBNext Ultra II Directional RNA library prep kit for Illumina (NEB #E7760, NEB #E7765). Ipswich (MA): New England Biolabs.

**** The manufacturer's instructions for the NEBNext rRNA Depletion (bacteria) and NEBNext Directional RNA Library Prep Kit for Illumina that were used to generate the protocol for mycobacterial RNA-sequencing detailed herein.**

[24] Matthews TC, Bristow FR, Griffiths EJ, et al. The integrated rapid infectious disease analysis (IRIDA) platform. bioRxiv Pre-Print. 2018;381830. doi: 10.1101/381830.
[25] FastQC A. Quality control tool for high throughput sequence data. Cambridge (UK): Babraham Institute.
[26] Bolger AM, Lohse M, Usadel B. Genome analysis trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi: 10.1093/bioinformatics/btu170
[27] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257. doi: 10.1186/s13059-019-1891-0
[28] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359. doi: 10.1038/nmeth.1923
[29] Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292–294. doi: 10.1093/bioinformatics/btv566
[30] Liao Y, Smyth GK, Shi W. Sequence analysis feature counts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–930. doi: 10.1093/bioinformatics/btt656
[31] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8
[32] Wickham H. ggplot2: elegant graphics for data analysis, 2nd edition. New York (NY): Springer; 2016.

[33] Blighe K, Rana S, Lewis M. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling [R package]. Version 1.18.0; 2024.
[34] Glasel J. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. Biotechniques. 1995;18(1):62–63.
[35] Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006;7(1):3. doi: 10.1186/1471-2199-7-3
[36] Verma A, Sampla AK, Tyagi JS. *Mycobacterium tuberculosis* rrn promoters: differential usage and growth rate-dependent control. J Bacteriol. 1999;181(14):4326–4333. doi: 10.1128/JB.181.14.4326-4333.1999