GMR
The Original

UNCOVERING KEY GENES LINKED TO DRUG RESISTANCE THROUGH TRANSCRIPTOMIC ANALYSES AND GENOMIC VARIATION IN MDR AND XDR STRAINS OF *MYCOBACTERIUM TUBERCULOSIS*

Suvarna Kapale [1], Joy Hoskeri[2], Prashant Kumar Modi [1*]

[1] Center for Systems Biology and Molecular Medicine, Yenepoya Research Centre, Yenepoya (Deemed to be University), Mangalore -575018, India, Email address: dr.suvarnakapale@gmail.com
[2] Karnataka State Akkamahadevi Women University Vijayapura, Karnataka-586108, Email: joybioinfo@gmail.com

ABSTRACT
The rise of extensively drug-resistant *Mycobacterium tuberculosis* (XDR-MTB) poses a serious global health threat, which emphasizes the need for understanding the molecular mechanisms behind multidrug-resistant *Mycobacterium tuberculosis* (MDR-MTB). This study integrates transcriptome sequencing data to explore genomic variations and transcriptomic profiles in XDR and MDR strains. Using available datasets from NCBI and ENA, quality checks, trimming, and performed comparative genome analysis against a reference genome using the Windows Subsystem for Linux (WSL), and key gene variants unique to MDR and XDR strains were identified using the Galaxy tool. The analysis focused on differentially expressed genes that drive resistance mechanisms. Using WSL and RStudio for transcriptomic analysis, critical regulatory elements, such as non-coding RNAs, were uncovered that significantly influence bacterial persistence and virulence. These regulatory RNAs enhance the bacterium's ability to survive in hostile conditions, enabling drug resistance. Differential gene expression analysis further revealed overexpression of non-coding RNA RVnc0024 (Mcr7) and small regulatory RNA RVnc0036a (MTS2823) in the regulation of virulence pathways, shedding light on the complex mechanisms fueling resistance. These discoveries not only elevate our understanding of drug resistance in *Mycobacterium tuberculosis* but also enable the development of targeted therapeutic strategies. By dissecting these molecular intricacies, this study lays the groundwork for innovative interventions aimed at combating one of the world's most pressing public health crises.

KEYWORDS: Multidrug-resistance *Mycobacterium tuberculosis* (MDR), Extensively drug-resistant *Mycobacterium tuberculosis* (XDR), non-coding RNA RVnc0024 (Mcr7), small regulatory RNA RVnc0036a (MTS2823).

INTRODUCTION

Tuberculosis (TB) is a contagious bacterial infectious disease that mainly targets the lungs, but it can also affect other parts of the body. It is caused by the bacteria Mycobacterium *tuberculosis*. The infection spreads through airborne droplets from the infected person when they cough, sneeze, talk, or sing. Although prolonged, close contact typically enables spread, TB continues to present a serious public health challenge. The human respiratory tract tries to block infection at the tracheal and bronchial linings, but once *M. tuberculosis* bypasses these defenses, it can establish infection in the lungs [1]. According to 2023 data from the World Health Organization, TB remains a global threat: approximately 10.8 million fresh cases were confirmed (range: 10.1-11.7 million), with an incidence rate of 134 per 100,000 population. Of these, multidrug-resistant or rifampicin-resistant TB (MDR/RR-TB) cases numbered 400,000 (5 per 100,000), and TB those diagnosed with HIV reached 662,000 (8.2 per 100,000) [2] , [1]. These figures highlight a serious, ongoing public health crisis. A major factor in TB's persistence is the development of drug resistance, which happens when the bacterium acquires mutations that reduce the efficacy of key anti-TB drugs. Resistance to first-line drugs leads to multidrug-resistant TB (MDR), and further resistance to second-line treatments results in extensively drug-resistant TB (XDR). This evolving pattern underscores the need for continuous genomic research to elucidate the genetic mechanisms underlying drug resistance [1], [3]

MDR TB is defined by resistance to the two principal first-line drugs: isoniazid (INH) and rifampicin (RIF). While second-line treatments remain effective, they are often costly, less effective, and associated with greater side effects [4]. Improper drug use, such as incorrect prescriptions, substandard medications, or premature treatment discontinuation, can promote the emergence of resistance. XDR-TB exhibits resistance to fluoroquinolones and at least one injectable second-line drug, kanamycin, amikacin, or capreomycin, making it particularly difficult to treat [5], [6]. XDR-TB arises through inadequate treatment or direct transmission from an infected individual. First-line drugs, including isoniazid (INH), rifampicin (RIF), ethambutol (EMB), and pyrazinamide (PZA) [7], [8], are the first choice of treatment for normal TB. Second-line drugs are fluoroquinolones, such as ofloxacin

(OFX), levofloxacin (LEV), moxifloxacin (MOX), ciprofloxacin (CIP) , and injectables such as kanamycin (KAN), amikacin (AMK), and capreomycin (CAP) are also prescribed for the treatment of MDR-TB and also in XDR -TB [5], [9]. Other less-effective agents, including ethionamide (ETH), prothionamide (PTH), cycloserine (CS)/terizidone, para-aminosalicylic acid (PAS), and streptomycin (STR), are also available but less commonly used [1] [10].

In this study, we aim to study the genomic and transcriptomic characteristics of drug-resistant *Mycobacterium tuberculosis* strains to gain comprehensive insights into the molecular mechanisms driving antimicrobial resistance. The study involves the acquisition and quality assessment of publicly available transcriptomic datasets, followed by comparative analyses among susceptible (SUS), multidrug-resistant (MDR), and extensively drug-resistant (XDR) strains. Differentially expressed genes (DEGs) will be identified and subjected to pathway enrichment analysis to uncover key biological processes and resistance-associated pathways. Furthermore, variant analysis will be performed across the three strain types to identify specific mutations linked to drug resistance, thereby providing an integrated understanding of the genetic and transcriptomic adaptations that contribute to *M. tuberculosis* persistence and pathogenicity [11].

METHODOLOGY

Data collection

This study utilized data of *Mycobacterium tuberculosis H37Rv* (NCBI RefSeq assembly: GCF_000195955.2; GenBank assembly: GCA_000195955.2) [12]. From a previous study carried out by [13] We used transcriptome sequencing data for *M. tuberculosis* obtained from single tuberculosis (TB) patient, and 12 clinical isolates were collected at different time points during treatment [13]. Drug resistance developed in the patient and was characterized by analyzing serial isolates using single-end RNA sequencing (Bio Project: PRJEB5865; Accession: ERP005297, ERR459293) [13].

The isolates were categorized as follows: Samples 1–3 (ERR459284–ERR459286): Susceptible to all first-line and second-line anti-TB drugs. Samples 4–6 (ERR459287–ERR459289): Multidrug-resistant TB (MDR-TB), resistant to isoniazid (INH), rifampicin (RIF), and streptomycin (STR). Samples 7–9 (ERR459290–ERR459292): Extensively drug-resistant TB (XDR-TB), resistant to INH, RIF, STR, ethionamide (ETH), and fluoroquinolones (FLQ). Samples 10–12 (ERR459293–ERR459295): Continued resistance to INH, RIF, STR, ETH, and FLQ [13].

Transcriptomic analysis

The transcriptomic datasets were obtained from the NCBI repository (Bio Project: PRJEB5865; Accession: ERP005297, ERR459293 [13] and were processed using the Windows Subsystem for Linux (WSL). The downloaded datasets were raw or FastQ format data, which needed to be processed before the datasets could be used in any downstream analyses. The raw transcriptomic dataset quality was assessed using FastQC (version 0.12.1) [14] (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which provided quality metrics such as per base Quality Score, the amount of adapter content, GC content, and determination of quality-related issues in the datasets prior to preprocessing.

The next step was to remove adapter sequences and poor quality bases from the datasets using Cutadapt version 5.0. https://cutadapt.readthedocs.io/en/stable/ [15]; it will trim/clip adapter sequences and basecalls below Q20 or poor quality bases from the datasets. Trimming adjusts the quality of base data generated by DNA sequencing, which improves the quality of the downstream results analyses from datasets.

After Cutadapt processed data files, the quality metrics were visualized in MultiQC (version 1.28) [16], which combined reports generated by FastQC and Cutadapt into one report. The report provides a comprehensive view of the overall quality trend of the datasets and includes how successful the adapter removal was performed as well as the GC content of the datasets. Only if the Quality metrics of the dataset reports produced by MultiQC are greater than or equal to 80% to 90%, the data from that dataset was used for downstream analyses, http://multiqc.info.

RNA STAR (v2.7.11b) http://code.google.com/p/rna-star/. [17], was used to map each processed cutadapt reads to the Reference Genome using an alignment process called "mapping". This required the creation of a reference genome index file prior to the mapping process and aligned output bam files indicating precise alignment positions. All of the aligned transcripts will then be used to produce Count Gene Expression Levels for each gene based on a GTF or a GFF annotation file to generate a count matrix http://www.bioconductor.org [18] that is essential for further analyses on differential gene expression [19].

Differential Gene Expression (DGE) analysis

To perform the Differential Gene Expression (DGE) analysis, RStudio was utilized. We utilized Bioconductor package DESeq2 in RStudia for all analyses [19]. First, the Feature Count Data was imported into RStudio using either read.delim or read.csv. This creates a matrix containing reads per transcript from each sample for each gene. Next, a DESeq2 file was created using both count data and sample metadata (i.e condition/replicate). At this point, the dataset was prepared for both normalization and differential testing, and DESeq functions was run on the dataset (normalization and differential expression). The DESeq function was executed to normalize counts, estimate dispersion, and fit statistical models resulting in log2fold changes, p-values, and adjusted p-values (FDR) for each gene. As a result of the differential expression analysis, log2Fold changes, $p<0.05$, and adjusted $p<0.05$ were extracted from the results(dds), These values help identify significantly differentially expressed genes.

PCA (Principal Components Analysis) was conducted using ggplot2 [20] and ggrepel https://CRAN.R-project.org/package=ggrepel [20, 21] to cluster samples based on their gene expression profile and to expose the degree of variance and to detect outliers and/or batch effects both within and between sample groups. These clusters were represented graphically in volcano plots. Volcano plots highlight gene expressions with statistical significance as well as the ability to highlight changes in expressions of genes within two separate conditions.

Variant analysis

All transcript sets were analyzed using Galaxy https://galaxy-main.usegalaxy.org/ [22] for variation detection. All samples undergoing processing followed the same methodology of using FastQC, Cutadapter and, storing all the data in MultiQC before they were aligned using bwa-mem2 https://github.com/lh3/bwa to the respective reference genomes [23]. The alignment of each trimmed sample produced the alignment data formatted into Bam files that contain where in the reference genome the trimmed reads mapped. It created a pileup file using bcftools mpileup that converted bam files into a formatted pileup file clearly showing the number of reads covering each base on the reference genome and serve as the basis for making variant calls.

Bcftools call allows you to use the bcftools call function for calling variants http://samtools.sourceforge.net [24]. bcftools call creates a VCF file (https://github.com/ekg/vcflib) from a pileup file and generates a list of identified genetic variants within that VCF file along with their locations and quality levels [25]. Subsequentally a Bcftools stats report summarizing the identified variants (number of variants, types of variants and confidence level) was generated so we can make a determination about the quality of variants. The VCF filter is used to further filter identified variants. The VCF filter can use user-defined filters (e.g., quality (≥30) or depth (≥10)) to filter the VCF file. This produces a VCF file that contains only high-confidence variants and likely removes false positives.

RESULTS

Evaluation of quality of raw sequence reads

The quality of the raw sequence reads was evaluated with FastQC, and the results were compiled using MultiQC. The per-base analysis of sequence quality indicated that average Phred quality scores remained high (Q30) across all bases (Figure:1). Most bases had a quality score of >Q30, which means that the accuracy of each base was >99.9%. A decrease in base quality was seen at the first three bases and at the last ~40-51 bases of the read, which is a common feature in Illumina sequencing datasets. The decrease may be due to sequencing artifacts related to Illumina sequencing technology; however, the median base quality scores were all >Q30 and <Q40 indicating that base quality was relatively high throughout the read length and that no substantial loss of base quality occurred as a function of read length.

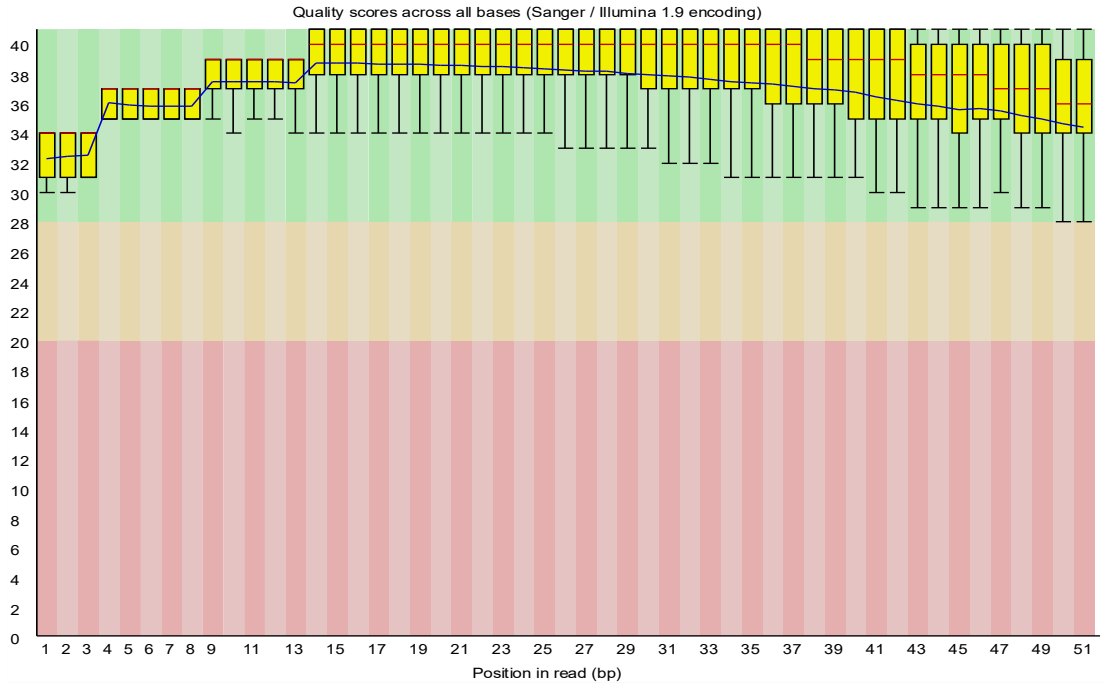

Figure 1: Per-base sequence quality plot. Per base read is represented along x-axis and Phred quality score along y-axis, where most of the bases are showing high quality (>30)

Adapter and quality filtering efficiency

All adapter trimming and quality filtering was performed using Cutadapt. The performance was summarized in figure 2 using MultiQC . Our analysis showed that close to 100% of reads passed the filtering criteria, with very few reads removed due to either length or quality criteria. The results from the analysis also indicate that the quality of the raw sequencing data is high, with minimal adapter contamination, and few low-quality sequences.

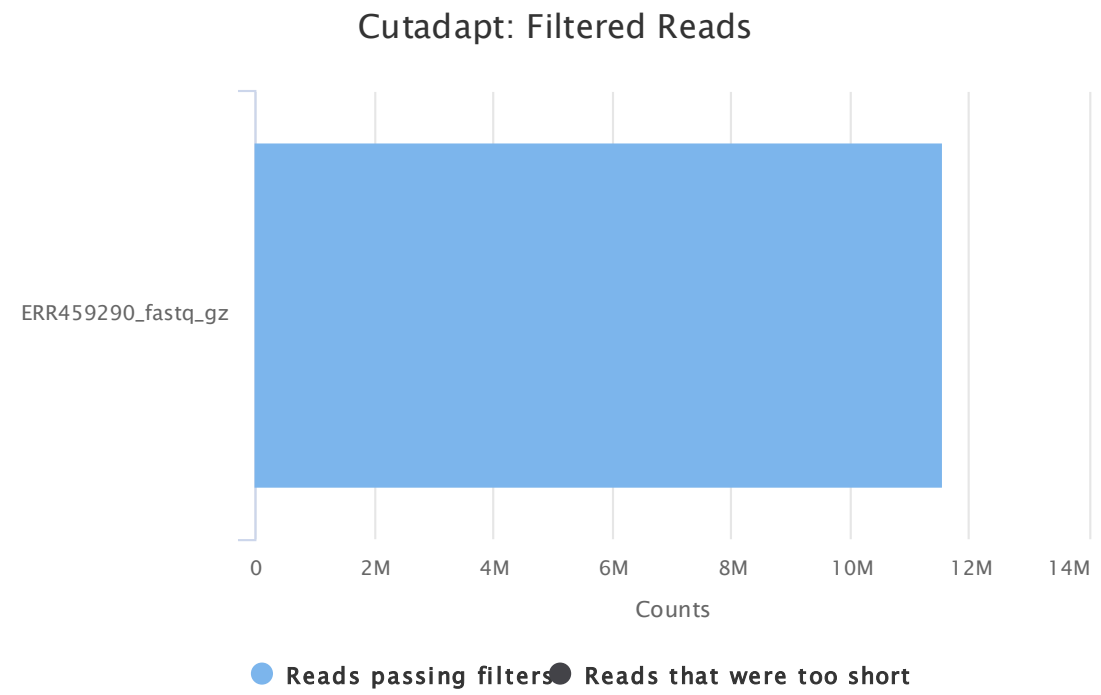

Fig: 2 MultiQC visualization of cutadapted read. Read counts in millions (M) are represented along the X-axis and Sample details on the Y-axis, where blue colour indicating the read successfully passed filtering

Principal component analysis indicates robust clustering and minimal technical variation

Principal component analysis PCA was analysed to probe the overall variation and relationships among the samples on the basis of their expression profiles. Insights from the PCA plot (Figure 3) demonstrate a clear separation between susceptible and drug-resistant strains, primarily along PC1. The susceptible (SUS) samples (Blue) form a distinct cluster on the far-right side (low PC1 values), indicating they exhibit differences from the drug-resistant strains. The MDR, XDR, and XDR1 samples are clustered closely on the left side, indicating greater similarity among the drug-resistant strains. This separation suggests that the principal source of variation in the data is driven by the drug-resistance phenotype.

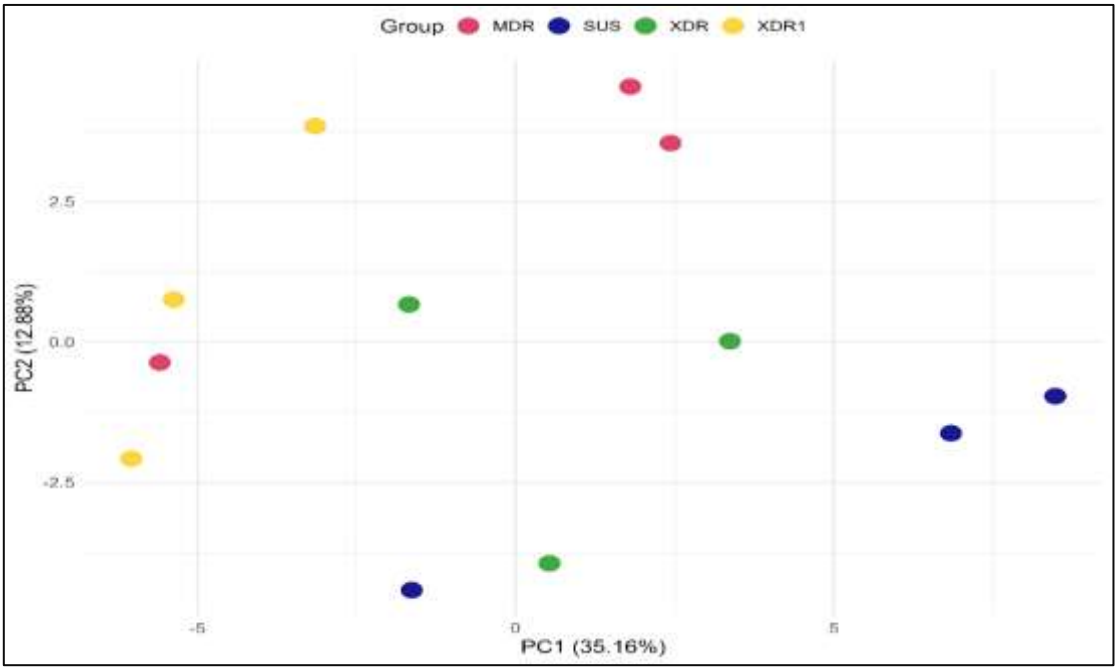

Figure 3: Principal Component Analysis (PCA) plot of transcriptomic data, where the PC1 (35.16% variance) presented along the X-axis and the PC2 (12.88% variance) presented along the Y-axis, indicating the major variation in gene expression pattern across the samples. Each point represents a biological replicate (n = 3), with each group is denoted by similar color: SUS (blue), MDR (red), XDR (green), and XDR1 (yellow).

Differential gene expression analysis between experimental groups

A differential gene expression was analysed and visualised using a volcano plot, from the statistical output file generated during differential gene expression analysis (refer supplementary file-1, file-2, file3, file-4). The x-axis represents the $\log_2$ fold change in expression levels between the two conditions, while the y-axis displays the negative logarithm of the p-value ($-\log_{10}$ p-value), indicating the statistical significance of the observed changes. The genes that did not reach the threshold for significance clustered near the centre of the plot and were considered not significantly differentially expressed. In the plot, upregulated genes are represented by red dots, while downregulated genes are represented by green dots. Analysis revealed a distinct expression pattern in MDR isolates compared to SUS isolates, with insights from Figure 4.

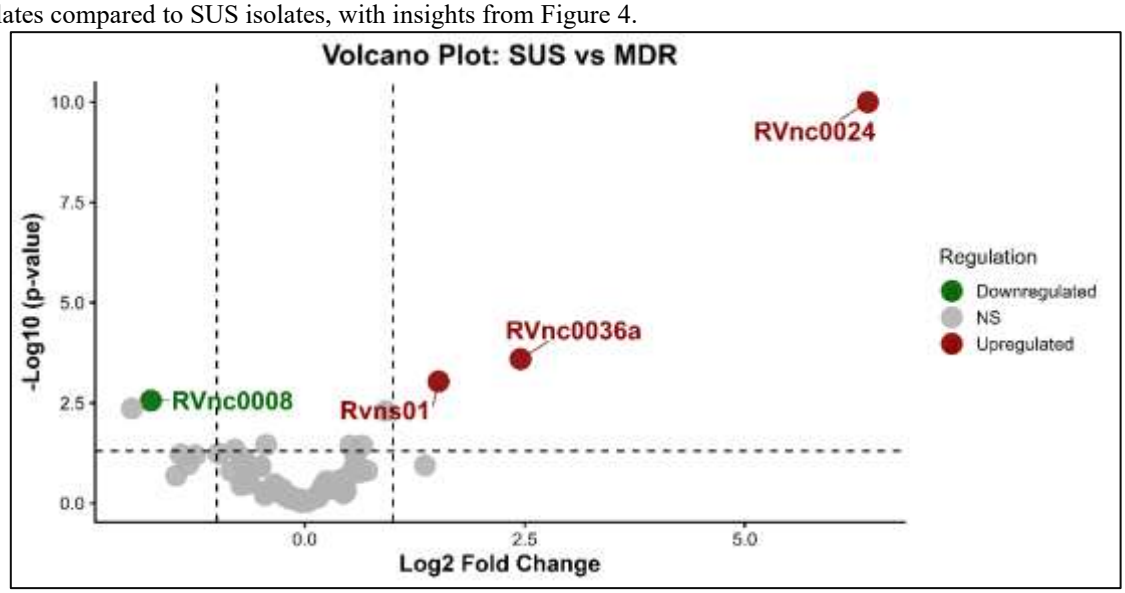

Figure 4: Volcano Plot of SUS vs MDR, log2fold change presented along the X-axis and −log10 (p-value) represents along the Y-axis, indicating the statistical significance of gene expression changes. The genes with significantly increased expression are shown in red, significant decrease in expression shown in green, and non-significant are in grey.

A total of three genes-RVnc0024, RVnc0036a, and RVns01 exhibited notable upregulation in MDR isolates relative to the SUS group. The analysis revealed a distinct expression pattern in XDR isolates compared to SUS isolates, insights from Figure 5.

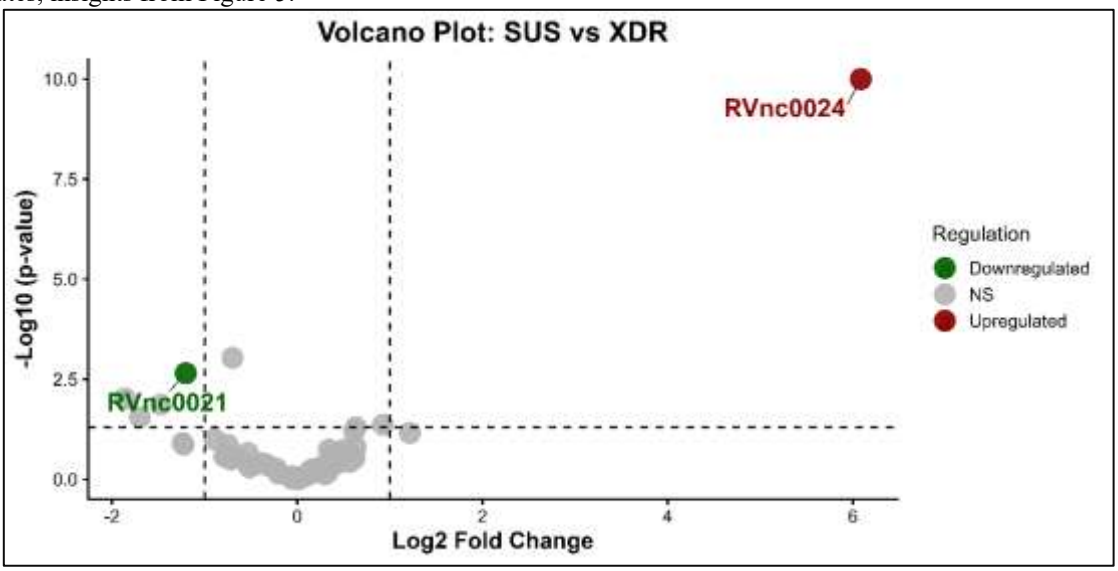

Figure 5: Volcano Plot of SUS vs XDR, log2fold change presented along the X-axis and −log10 (p-value) represents along the Y-axis, indicating the statistical significance of gene expression changes. The genes with significantly increased expression are shown in red, significant decrease in expression shown in green, and non-significant are in grey.

Only one gene-RVnc0024 exhibited notable upregulation in XDR isolates relative to the SUS group. The analysis revealed a distinct expression pattern in XDR1 isolates compared to SUS isolates, insights from Figure 6.

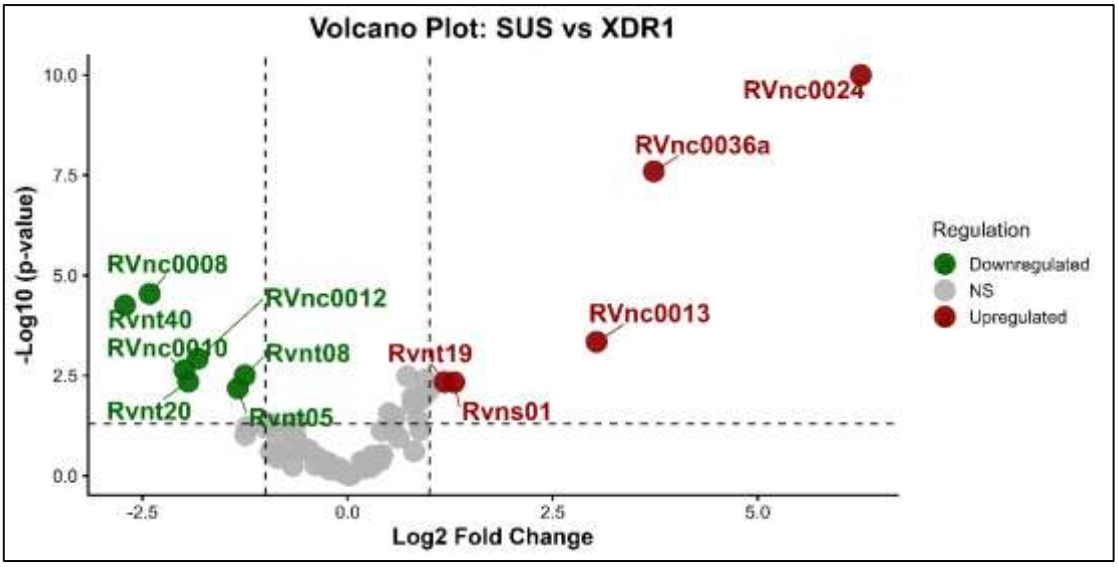

Figure 6: Volcano Plot of SUS vs XDR1, log2fold change presented along the X-axis and −log10 (p-value) represents along the Y-axis, indicating the statistical significance of gene expression changes. The genes with significantly increased expression are shown in red, significant decrease in expression shown in green, and non-significant are in grey.

A total of three genes such as RVnc0024, RVnc0036a, and RVnc0013 exhibited notable upregulation in XDR1 isolates relative to the SUS group. Three statistical tables of differentially expressed genes were generated using the susceptible condition as the reference. The tables include the following parameters: baseMean -average expression across samples; $\log_2$ Fold Change-log-transformed fold difference in expression between two conditions; lfcSE -standard error of the $\log_2$ fold change; stat-test statistic; p-value-unadjusted statistical significance; padj-Adjusted p-value (FDR).

Differentially expressed genes between SUS and MDR: Among the identified genes, as shown in Table 1, RVnc0024 exhibited the highest upregulation in MDR isolates, with a $\log_2$ fold change ($\log_2$FC) of 6.39 and an extremely significant adjusted p-value of 1.27×10^{-71}, indicating strong differential expression. RVnc0036a and Rvns01 were also significantly upregulated in MDR samples, with $\log_2$FC values of 2.45 and 1.52, and adjusted p-values of 0.0082 and 0.0199, respectively (For more detailed results, see the Supplementary file: sus_vs_mdr_results).

Table-1, Differentially expressed genes in MDR condition comparison with SUS

Gene ID	baseMean	Change	lfcSE	stat	pvalue	padj
RVnc0008	25.81594	-1.7472	0.583396	-2.99488	0.002746	0.044615
Rvns01	130507.2	1.519452	0.45851	3.313888	0.00092	0.019935
RVnc0024	716.9444	6.397412	0.352927	18.12675	1.96E-73	1.27E-71
RVnc0036a	609434.4	2.451453	0.670032	3.658712	0.000253	0.008238

Differentially expressed genes between SUS and XDR: Among the identified genes, as shown in Table 2, RVnc0024 exhibited the highest upregulation in XDR isolates, with a $\log_2$ fold change ($\log_2$FC) of 6.08 and an extremely significant adjusted p-value of 8.54×10^{-61}, indicating strong differential expression (For more detailed results, see the Supplementary file: sus_vs_xdr_results).

Table-2. Differentially expressed genes in XDR condition comparison with SUS

Gene ID	baseMean	Log2Fold Change	lfcSE	stat	pvalue	padj
RVnc0021	136.5243	-1.20362	0.393764	-3.05671	0.002238	0.048486
RVnc0024	716.9444	6.08511	0.364381	16.69985	1.31E-62	8.54E-61
Rvnt38	189.3615	-0.70072	0.21148	-3.31341	0.000922	0.029953

Differentially expressed genes between SUS and XDR1:

Among the identified genes, as shown in Table 3, RVnc0024 showed the highest upregulation in MDR isolates, with a $\log_2$ fold change ($\log_2$FC) of 6.25 and an extremely significant adjusted p-value of 2.14×10^{-70}, indicating strong differential expression. RVnc0036a and RVnc0013 were also significantly upregulated in XDR1 samples,

with log_2FC values of 3.73 and 3.04, and adjusted p-values of 8.16×10^{-7} and 0.00588, respectively (For more detailed results, see the Supplementary file: sus_vs_xdr1_results).

Table-3. Differentially expressed genes in XDR1 condition comparison with SUS

Gene ID	Mean	log2Fold Change	lfcSE	stat	pvalue	padj
RVnc0008	25.81594	-2.4127	0.576663	-4.1839	2.87E-05	0.000621
Rvnt04	107.9813	0.727104	0.247389	2.939112	0.003292	0.022962
Rvnt05	15.31575	-1.33719	0.491743	-2.71929	0.006542	0.030374
Rvnt08	92.72049	-1.25246	0.423694	-2.95604	0.003116	0.022962
RVnc0012	12.81844	-1.82691	0.563505	-3.24205	0.001187	0.012856
RVnc0013	2910.658	3.035747	0.865496	3.507524	0.000452	0.00588
Rvnr01	478178.9	0.921129	0.349323	2.6369	0.008367	0.03399
Rvnr02	11363826	0.943925	0.327523	2.882014	0.003951	0.022962
Rvnr03	41536.78	0.992053	0.369718	2.68327	0.007291	0.031593
Rvnt19	48.48385	1.181803	0.416467	2.837687	0.004544	0.022962
Rvnt20	7.414157	-1.94222	0.683307	-2.84238	0.004478	0.022962
RVnc0010	11.71745	-1.98842	0.652253	-3.04854	0.0023	0.021353
Rvns01	130507.2	1.299555	0.458509	2.834309	0.004592	0.022962
RVnc0024	716.9444	6.256999	0.348174	17.9709	3.29E-72	2.14E-70
Rvnt29	71.30882	0.791359	0.311952	2.536798	0.011187	0.042774
Rvnt37	35.03865	0.907099	0.36116	2.511629	0.012018	0.043397
RVnc0036a	609434.4	3.733803	0.670034	5.572558	2.51E-08	8.16E-07
Rvnt40	12.45356	-2.71474	0.673649	-4.0299	5.58E-05	0.000907

Variants in SUS, MDR, XDR, XDR1

The variant analysis illustrates how *Mycobacterium tuberculosis* evolves in response to antibiotic pressure. This analysis highlights significant genetic differences between the reference genome *Mycobacterium tuberculosis* H37Rv and various drug-resistant strains, specifically Susceptible, Multi-Drug Resistant, Extensively Drug-Resistant, and Extensively Drug-Resistant 1, as shown in Table 4. The examination of single nucleotide polymorphisms and insertions/deletions (INDELs) across different strains of *Mycobacterium tuberculosis* demonstrates that genetic diversity plays a crucial role in the progression of drug resistance. The increased number of variants in MDR and XDR strains indicates ongoing adaptation and potential processes that enable these bacteria to survive under drug pressure.

Table-4. Variants in SUS, MDR, XDR, XDR1

Groups	Samples	SNPs	INDELs	sites
SUS	sample1	4096	24	2977324
SUS	sample2	4615	23	2542611
SUS	sample3	3836	18	2029735
MDR	sample4	3578	36	3751967

MDR	sample5	4305	40	3320956
MDR	sample6	4208	36	3345346
XDR	sample7	4069	27	3276770
XDR	sample8	4029	19	2506935
XDR	sample9	4457	22	2632087
XDR1	sample10	4926	78	3740011
XDR1	sample11	4647	94	3859294
XDR1	sample12	4827	75	3735015

DISCUSSION

This study delves into the genomic variations and transcriptomic profiles of extreme drug-resistant *Mycobacterium tuberculosis* strains, aiming to uncover key genes linked to extreme drug resistance. The results illuminate critical regulatory mechanisms and pathways that enhance the adaptability and virulence of *M. tuberculosis*, particularly within the challenging context of drug resistance Our findings suggests that RVnc0024 (Mcr7) is a small non-coding RNA, and RVnc0036 (MTS2823) is a small regulatory RNA that plays a crucial role in the two-component system in MTB [26], [27].

RVnc0024 is a small non-coading RNA, plays a primary role in the post-transcriptional regulation of the TAT secretion system (twin-arginine translocation pathway, which is also known as the two-component system), along with metabolic and secretory pathways by modulating TATc mRNA, where the TAT secretion pathway is essential for the secretion of major and important virulence factors https://www.kegg.jp/pathway/map02020 . Proper regulation of the pathway also affects the release of the immunodominant Ag85 complex and beta-lactamase, which are crucial for the pathogenicity of the bacteria. This RVnc0024 is modulated by PhoP protein. PhoP proteinis a component of PhoPR proteins of the TAT secretion system, which plays a critical role in regulating the expression of genes that are responsible for the virulence of MTB, along with the cell wall synthesis and lipid metabolism of the bacteria. The ability of MTB to adapt to changing environmental conditions, especially within the host, is largely dependent on PhoP regulating RVnc0024, and RVnc0024 regulating the TAT secretion system. There is an interplay between PhoP and RVnc0024. Regulation of RVnc0024 by PhoP highlights the critical link between environmental sensing and its virulence factor secretions in the bacteria [26], [28], [29].

RVnc0036a is a small regulatory RNA that plays a critical role in the persistence and pathogenic potential of drug-resistant strains of MTB. Recent research shows that RVnc0036a is associated with early expression linked to the suppression of pro-inflammatory cytokines such as TNF-alpha and interleukin-6, in addition to a notable decrease in the production of nitric oxide, which are important molecules used by human macrophages to kill the bacteria under stress conditions. So, this finding suggests that RVnc0036a is responsible for the survivance of the bacteria. This also regulates the RecA protein (recombinase A protein), which is critical for DNA repair and maintaining the gene integrity of *Mycobacterium tuberculosis* under stress conditions [27], [30], [31].

Overall, results indicate that both RVnc0024 (Mcr7) a small non-coding RNA-and RVnc0036 (MTS2823) a small regulatory RNA-have an important function in adaptation and persistence of drug-resistant, regulation of virulence, immune evasion and stress response pathways in *Mycobacterium tuberculosis* via two-component system. However, this work is hampered by the absence of experimental validation of the predicted regulatory interactions. Future investigations with functional assays involving greater sample size will confirm how these non-coding RNAs operate and provide insights into their potential utility as drug targets.

CONCLUSION

This research explores the genomic variations and transcriptomic profiles of multiple drug-resistant strains and extensively drug-resistant strains of *Mycobacterium tuberculosis*, with the objective of identifying crucial genes associated with extensive drug resistance. The findings reveal significant regulatory mechanisms and pathways that bolster the adaptability and severity of *M. tuberculosis*, especially concerning drug resistance. At the forefront of these discoveries is the non-coding RNA RVnc0024, which is regulated by the PhoP transcription factor. RVnc0024 is instrumental in the post-transcriptional control of the Tat secretion system, showcasing the bacterium's exceptional capacity to adjust its responses to challenging environments, such as those present within host macrophages. Furthermore, the small regulatory RNA RVnc0036a has emerged as a vital component in the intracellular survival of multidrug-resistant strains. Through its modulation of key cellular processes, RVnc0036a effectively inhibits the secretion of pro-inflammatory cytokines and nitric oxide, enabling *Mycobacterium tuberculosis* to evade the host's immune defences and promote bacterial persistence. Additionally, the recombinase protein RecA which is regulated by RVnc0036a, is crucial for DNA damage repair caused by various antimicrobial agents, thus enhancing genetic adaptation and survival in adverse conditions, solidifying its role in the development of tuberculosis. Overall, over results identify RVnc0024 (Mcr7)and RVnc0036 (MTS2823) as a potential modulator of drug resistance in *M. tuberculosis.* However, our presented work mainly dependents on

bioinformatics analysis of available data, further investigations including wet lab experiments such as qPCR with larger sample size is required to prove RVnc0024 (Mcr7) and RVnc0036 (MTS2823) role in multi-drug resistance in *M. tuberculosis.*

Abbreviation

MTB Mycobacterium tuberculosis
TB Tuberculosis
SUS-TB Susceptible- tuberculosis
MDR-TB Multidrug-resistant tuberculosis
XDR-TB Extensively-resistant tuberculosis
RR-TB Refampicin-resistance tuberculosis
AR Antibiotic resistance
RNA Ribonucleic acid
RNA-seq RNA sequencing
DNA Deoxyribonucleic acid
ncRNA Non-coding RNA
DEGs Differentially expressed genes
DGE Differential gene expression
FDR False discovery rate
PCA Principal component analysis
SNP Single nucleotide polymorphism
INDELs Insertions and deletions
NCBI National Center for Biotechnology Information
ENA European Nucleotide Archive
BWA-MEM Burrows-Wheeler Aligner Maximal Exact Matches
STAR Spliced Transcripts Alignment to a Reference
GTF Gene Transfer Format
GFF General Feature Format
DESeq2 Differential expression analysis tool
INH Isoniazid
RIF Rifampicin
PZA Pyrazinamide
ETH Ethionamide
KAN Kanamycin
AMK Amikacin
CIP Ciprofloxacin
OFX Ofloxacin
LEV Levofloxacin
PAS Para-aminosalicylic acid

Acknowledgements

The authors acknowledge the support of the Department of Biotechnology, Government of India, to the Yenepoya (Deemed to be University) through the project on "Skill Development in Mass Spectrometry-based metabolomics technology BIC" (BT/PR40202/BTIS/137/53/2023).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

All the data related to the study are provided in the manuscript and supplementary data.

Ethics approval

This study did not require ethical approval.

Author Contributions

Prashant Kumar Modi provided conceptual guidance and critically reviewed and edited the manuscript. **Suvarna Kapale** performed the data analysis, prepared the figures, and drafted the manuscript. **Joy Hoskeri** conceived the idea, designed the experiments, and contributed to data interpretation. All authors read and approved the final version of the manuscript.

Competing Interests

The authors have no competing financial or non-financial interests to disclose.

Figure legends:

1) Figure 1: Per-base sequence quality plot. Per base read is represented along x-axis and Phred quality score along y-axis, where most of the bases are showing high quality (>30)
2) Fig: 2 MultiQC visualization of cutadapted read. Read counts in millions (M) are represented along the X-axis and Sample details on the Y-axis, where blue colour indicating the read successfully passed filtering
3) Figure 3: Principal Component Analysis (PCA) plot of transcriptomic data, where the PC1 (35.16% variance) presented along the X-axis and the PC2 (12.88% variance) presented along the Y-axis, indicating the major variation in gene expression pattern across the samples. Each point represents a biological replicate (n = 3), with each group is denoted by similar color: SUS (blue), MDR (red), XDR (green), and XDR1 (yellow).
4) Figure 4: Volcano Plot of SUS vs MDR, log2fold change presented along the X-axis and −log10 (p-value) represents along the Y-axis, indicating the statistical significance of gene expression changes. The genes with significantly increased expression are shown in red, significant decrease in expression shown in green, and non-significant are in grey.
5) Figure 5: Volcano Plot of SUS vs XDR, log2fold change presented along the X-axis and −log10 (p-value) represents along the Y-axis, indicating the statistical significance of gene expression changes. The genes with significantly increased expression are shown in red, significant decrease in expression shown in green, and non-significant are in grey.
6) Figure 6: Volcano Plot of SUS vs XDR1, log2fold change presented along the X-axis and −log10 (p-value) represents along the Y-axis, indicating the statistical significance of gene expression changes. The genes with significantly increased expression are shown in red, significant decrease in expression shown in green, and non-significant are in grey.

REFERENCE

1. Zhi Chen TWaJDea. Decoding the WHO Global Tuberculosis Report 2024: A Critical Analysis of Global and Chinese Key Data. . Zoonoses Vol 5(1) 2025. doi: 10.15212/ZOONOSES-2024-0061.
2. Chen Z, Wang T, Du J, Sun L, Wang G, Ni R, et al. Decoding the WHO global tuberculosis report 2024: a critical analysis of global and Chinese key data. Zoonoses. 2025;5(1):999.
3. Advani J, Verma R, Chatterjee O, Pachouri PK, Upadhyay P, Singh R, et al. Whole Genome Sequencing of Mycobacterium tuberculosis Clinical Isolates From India Reveals Genetic Heterogeneity and Region-Specific Variations That Might Affect Drug Susceptibility. Front Microbiol. 2019;10:309. doi: 10.3389/fmicb.2019.00309.
4. Smith SE, Kurbatova EV, Cavanaugh JS, Cegielski JP. Global isoniazid resistance patterns in rifampin-resistant and rifampin-susceptible tuberculosis. Int J Tuberc Lung Dis. 2012;16(2):203-5. doi: 10.5588/ijtld.11.0445.
5. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, et al. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet. 2006;368(9547):1575-80. doi: 10.1016/S0140-6736(06)69573-1.
6. Hailemariam TG, Ayele A, Gelanew T, Atnafu A, Brennan M, Tilahun M, et al. Integrating genomic, transcriptomic, and phenotypic information to explore drug resistance in Mycobacterium tuberculosis sub-lineage 4.2.2.2. J Appl Microbiol. 2025;136(3). doi: 10.1093/jambio/lxaf063.
7. Perumal R, Naidoo K, Naidoo A, Ramachandran G, Requena-Mendez A, Sekaggya-Wiltshire C, et al. A systematic review and meta-analysis of first-line tuberculosis drug concentrations and treatment outcomes. Int J Tuberc Lung Dis. 2020;24(1):48-64. doi: 10.5588/ijtld.19.0025.
8. Zumla A, Nahid P, Cole ST. Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov. 2013;12(5):388-404. doi: 10.1038/nrd4001.
9. Falzon D, Gandhi N, Migliori GB, Sotgiu G, Cox HS, Holtz TH, et al. Resistance to fluoroquinolones and second-line injectable drugs: impact on multidrug-resistant TB outcomes. Eur Respir J. 2013;42(1):156-68. doi: 10.1183/09031936.00134712.
10. Michael JS, John TJ. Extensively drug-resistant tuberculosis in India: a review. Indian J Med Res. 2012;136(4):599-604.
11. Verma H, Nagar S, Vohra S, Pandey S, Lal D, Negi RK, et al. Genome analyses of 174 strains of Mycobacterium tuberculosis provide insight into the evolution of drug resistance and reveal potential drug targets. Microb Genom. 2021;7(3). doi: 10.1099/mgen.0.000542.
12. Chitale P, Lemenze AD, Fogarty EC, Shah A, Grady C, Odom-Mabey AR, et al. A comprehensive update to the Mycobacterium tuberculosis H37Rv reference genome. Nature Communications. 2022;13(1):7068. doi: 10.1038/s41467-022-34853-x.
13. Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, Caugant DA, et al. Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient. Genome Biol. 2014;15(11):490. doi: 10.1186/s13059-014-0490-3.
14. S. A. FastQC: a quality control tool for high throughput sequence data [Internet]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ 2010.
15. M. M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):10–12. . 2011.
16. Ewels P MM, Lundin S, Käller M. . MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–3048. 2016.

17. Dobin A DC, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21. . 2013.
18. Liao Y SG, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. . Bioinformatics 2014;30(7):923–930. 2014. doi: https://doi.org/10.1093/bioinformatics/btt656.
19. 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.
20. H. W. ggplot2: elegant graphics for data analysis. . New York: Springer-Verlag; 2016 2016. doi: doi:10.1007/978-3-319-24277-4.
21. K. S. ggrepel: automatically position non-overlapping text labels with 'ggplot2' [Internet]. 2024 [cited 2026 Apr 12]. Available from: https://CRAN.R-project.org/package=ggrepel. 2024.
22. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018;46(W1):W537-W44. doi: 10.1093/nar/gky379.
23. H. L. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. 2013. Available from: https://doi.org/10.48550/arXiv.1303.3997. 2013.
24. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987-93. doi: 10.1093/bioinformatics/btr509.
25. https://github.com/ekg/vcflib GEvIAf: https://github.com/ekg/vcflib (2015). Accessed.
26. Namouchi A, Gomez-Munoz M, Frye SA, Moen LV, Rognes T, Tonjum T, et al. The Mycobacterium tuberculosis transcriptional landscape under genotoxic stress. BMC Genomics. 2016;17(1):791. doi: 10.1186/s12864-016-3132-1.
27. Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, et al. The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in Mycobacterium tuberculosis. PLoS Pathog. 2014;10(5):e1004183. doi: 10.1371/journal.ppat.1004183.
28. Ahmed W, Zheng K, Liu ZF. Small Non-Coding RNAs: New Insights in Modulation of Host Immune Response by Intracellular Bacterial Pathogens. Front Immunol. 2016;7:431. doi: 10.3389/fimmu.2016.00431.
29. Palomino JC. Molecular detection, identification and drug resistance detection in Mycobacterium tuberculosis. FEMS Immunol Med Microbiol. 2009;56(2):103-11. doi: 10.1111/j.1574-695X.2009.00555.x.
30. Chen Z, Jiang W, Zhang M, Yu B, Li W, Liu J, et al. Mycobacterium tuberculosis sRNA MTS2823 regulates the growth of the multidrug-resistant strain in macrophages. FEMS Microbiol Lett. 2022;369(1). doi: 10.1093/femsle/fnac106.
31. Nautiyal A, Patil KN, Muniyappa K. Suramin is a potent and selective inhibitor of Mycobacterium tuberculosis RecA protein and the SOS response: RecA as a potential target for antibacterial drug discovery. J Antimicrob Chemother. 2014;69(7):1834-43. doi: 10.1093/jac/dku080.