Article

Can miRNA indicate risk of illness after continuous exposure to *M. tuberculosis*?

Cleonardo Augusto Silva [1,2,†], Arthur Ribeiro-dos-Santos [1,†], Wanderson Gonçalves e Gonçalves [1,2], Pablo Pinto [1,2], Rafael P. Pantoja [1], Tatiana Vinasco-Sandoval [4], André M. Ribeiro-dos-Santos [1], Mara H. Hutz [3], Amanda F. Vidal [1], Gilderlanio S. de Araújo [1,‡], Ândrea Ribeiro-dos-Santos [1,2,‡,*] and Sidney Santos [1,2,‡]

[1] PPG em Genética e Biologia Molecular, Laboratório de Genética Humana e Médica, Universidade Federal do Pará, Belém, 66075-110, Brasil
[2] PPG em Oncologia e Ciências Médicas, Núcleo de Pesquisas em Oncologia, Universidade Federal do Pará, Belém, 66073-005, Brasil
[3] Departamento de Genética, Universidade Federal do Rio Grande do Sul, Brasil
[4] Commisariat à l'Énergie Atomique et aux Energies Alternatives (CEA)
* Correspondence: akelyufpa@gmail.com
† These authors contributed equally as First Authors
‡ These authors contributed equally as Senior Authors

Simple Summary: Tuberculosis, one of the oldest infectious diseases of humanity, is the leading cause of mortality from a single infectious agent and is among the top 10 causes of death worldwide. Despite that, few studies focus on regulatory elements such as small non-coding RNAs in tuberculosis.This pilot work applied Next Generation Sequencing techniques to evaluate global miRNA expression profile of patients with active tuberculosis, their respective healthy physicians, which are in constant danger of infection, and a group of external healthy controls. Also, we observe miRNA-gene interactions affected by exposure to the bacteria.Our findings indicate a list of miRNAs that could be used as potential biomarkers, which could improve treatment strategies at early stages, pathways interactions modified due to differential miRNA expression profile, highlighting possible discoveries related to the immune system's response to tuberculosis, and finally, a warning alert to developing new strategies to avoid long-term exposure by healthy physicians, considering how closely related their miRNA profile were to tuberculosis patients using current protocols.

Abstract: Molecular studies regarding regulatory elements such as small ncRNAs and their mechanisms are poorly understood in infectious diseases. Tuberculosis is one of the oldest infectious diseases of humanity, and it is still a challenge to prevent and treat it. The control of the infection as well as its diagnosis are still complex, and treatments used are linked to several side effects. This study aimed to investigate miRNA's expression profile to identify possible biomarkers for tuberculosis. We applied NGS techniques to investigate miRNA's global expression profile from blood samples of infected patients with tuberculosis, their respective healthy physicians, and external healthy individuals as controls. Samples from 22 individuals run through a differential expression, target genes, gene set enrichment, and miRNA-gene network analysis. We observed 153 altered miRNAs, among which, only three DEmiRNAs (hsa-let-7g-5p, hsa-miR-486-3p and hsa-miR-4732-5p) were found between the investigated patients and their respective physicians. These DEmiRNAs are suggested to play an important role in granuloma regulation and their immune physiopathology. Our results propose that miRNAs may be involved in immune modulation, regulating the repertoire of genes expressed in the immune system's cells. Our findings encourage the application of miRNAs as potential biomarkers for tuberculosis.

Keywords: miRNA, Tuberculosis, differential expression analysis

 doi:10.20944/preprints202101.0195.v1

1. Introduction

Tuberculosis, a disease caused by *Mycobacterium tuberculosis* infection, is the leading cause of mortality from a single infectious agent and is among the top 10 causes of death worldwide, accounting for about 1.3 million deaths in 2017, and 10.4 million new cases of active tuberculosis in the world for the same year [1,2]. A third of the world population carries latent *Mycobacterium* and as consequence the potential evolution to active tuberculosis [3]. HIV co-infection and age at first exposure are factors that contribute to the progression of the active form of the disease but explain only a fraction of this conversion threshold [4]. Disease progression may occur if host cell-mediated immunity fails, as suggested by the fact that an increased risk of latent tuberculosis reactivation is posed by anti-tumor necrosis factor therapy [5], which causes immune suppression and considerably high active tuberculosis prevalence among HIV-infected patients with T-cell immunity deficiency [5]. Identifying active tuberculosis high-risk groups would allow strategies for a more effective prophylactic treatment and would prevent the evolution of the disease to highly infectious symptomatic stages, avoiding transmission [6].

Gene regulation is an essential mechanism for both maintenance and normal functionality of the cell. Changes in gene expression may impact cellular state and even lead to complex and infectious diseases [7]. Non-coding RNAs (ncRNAs) are some of the main molecular elements that act in gene regulation. Among ncRNAs, miRNAs (small fragments of 18-22 nucleotides) are able to regulate the expression of genes at post-transcriptional level by binding to the 3'-UTR of messenger RNAs (mRNAs) and inhibiting their translation or inducing their degradation [7,8].

Considering their impact in cells, miRNAs have the potential to be used as biomarkers for diagnosis, response to treatment and therapeutic interventions [7]. Thus, miRNAs may regulate host response to tuberculosis, potentially being critical for the establishment of the infection, however, the dynamics of the miRNA expression and the implications for immune response in the lungs are unknown [9].

Some studies also provide insight into the role of differentially expressed miRNAs (DEmiRNAs) in tuberculosis, such as discrimination of active tuberculosis and healthy controls [6,10,11], and the regulation of pathogen-host interactions [12–14]. However, there are few studies on the expression profile of miRNAs in tuberculosis using Next Generation Sequencing (NGS) techniques, which offers higher accuracy and sensitivity identification of miRNA sequences, being able to identify new and known miRNAs [15].

Risk prediction based on molecular biomarkers of individuals who will develop active tuberculosis, within a population exposed to *Mycobacterium tuberculosis*, is extremely important. [13]. Therefore, the objective of the present study is to apply NGS technology to evaluate the global miRNA expression profile (miRnome) of patients with active tuberculosis, their respective healthy physicians and a group of external healthy controls, aiming to screening and consolidate potential biomarkers of this infectious disease.

2. Results and Discussion

In tuberculosis pathogenesis, host cellular immune response determines whether an infection becomes latent tuberculosis infection or progresses to active infectious or extrapulmonary tuberculosis [16]. It has been established that both adaptive and innate immune response is significantly regulated by miRNAs. miRNAs control the differentiation of B cells, antibody generation, T cell development [17] and function and innate immune cell activation [18]. Under mycobacterial infection, components of inflammatory and immune pathways are regulated by miRNAs [19,20].

A recent study of miRNAs in Tuberculosis [21] reviewed a list of regulatory functions of host's inflammatory and immune response signaling pathways regulated by miRNAs, the listed

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

regulatory functions were: i) Apoptosis; ii) Cytokine production; iii) Nitric oxide suppression; iv) T cell proliferation; v) Inhibition of antimicrobial peptides, and vi) Autophagy. Lipid metabolism, a regulatory function associated with tuberculosis studied by Ouimet et al. (2016)[22] can also be regulated by miRNAs.

2.1. RNA-seq's miRNA data overview

In order to identify potential biomarkers, miRNA sequence expression of 22 whole blood samples was analyzed. Samples were divided into three groups: External Control (n=7); Hospital Control (n=6), which are physicians responsible for the tuberculosis treatment, and Tuberculosis patients (n=9).

RNA-seq data showed an average number of mapped reads per sample of 3,224 - ranging from 0 to 925,920 raw read counts. We identified inside the list of 2,576 known miRNAs a total of 210 expressed miRNAs (miRNAs that reached a total read count $\geq$ 10 in at least one sample), among which, hsa-miR-486-5p, hsa-miR-92a-3p, and hsa-miR-16-5p were the most abundant in all samples, representing about 82% of the total read count, these miRNAs were later removed from the analysis to avoid statistical bias.

Regarding miRNA expression profile, we found a clear contrast between the miRNA expression distribution in External Control group from both Hospital Control and Tuberculosis groups, which showed a higher number of miRNAs with low expression level (Figure 1A and 1B). Contrary to our initial belief, Hospital Control had even more miRNAs with low expression than Tuberculosis patients. External Control group, as indicated by Figure 1C, had a higher expression profile compared to the other groups and presented a very low number of miRNAs with low expression value. High peaks of miRNAs expression are represented for each group and sample in Figure1D and Figure 1E, respectively.

miRNA expression's profile data were submitted to statistical analysis process to identify miRNAs with significant expression level differences among each pair of groups to further investigate possible biomarkers and genetic processes involved in the infectious process of the disease.

2.2. Differential expression analysis

Differential expression analysis process involves comparing only two groups (a case and a control classes) to find statistically relevant differences between these groups' miRNA expression profiles. Besides the three possible two-elements combinations of our three initial groups, two new comparisons were introduced to the study, one which united Hospital Control and External Control inside a new Control group and compared it to Tuberculosis patients (case class) and the last comparison, which merged Tuberculosis patients and Hospital Control groups in a Medical Samples group (case class) to be compared to the External Control group.

Among these five comparisons, two had perhaps the most interesting results and had to be discussed in this paper: i) Medical Samples versus External Control, and ii) Tuberculosis patients versus Hospital Control. Posterior, these results were used to perform both target genes analysis and gene set enrichment analysis.

2.2.1. Medical Samples *vs* External Control

This comparison was chosen to be discussed for two reasons: i) initial groups involved, and ii) differentially expressed miRNAs (DEmiRNAs) results. Medical Samples versus External Control comparison contained all initial groups (Tuberculosis patients, Hospital Control and External Control) inside of it. This enabled us to obtain an overview of DEmiRNAs among these groups and to observe how closely related or distant they were among themselves.

DEmiRNAs expression data analysis obtained from EdgeR showed difference in expression levels of miRNAs among the three initial groups (Figure 2A). This analysis used Kendall's hierarchical clustering correlation between the miRNA expression in each sample and identified a clear contrast between External Control and Medical Samples groups. This contrast is presented in two clusters of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

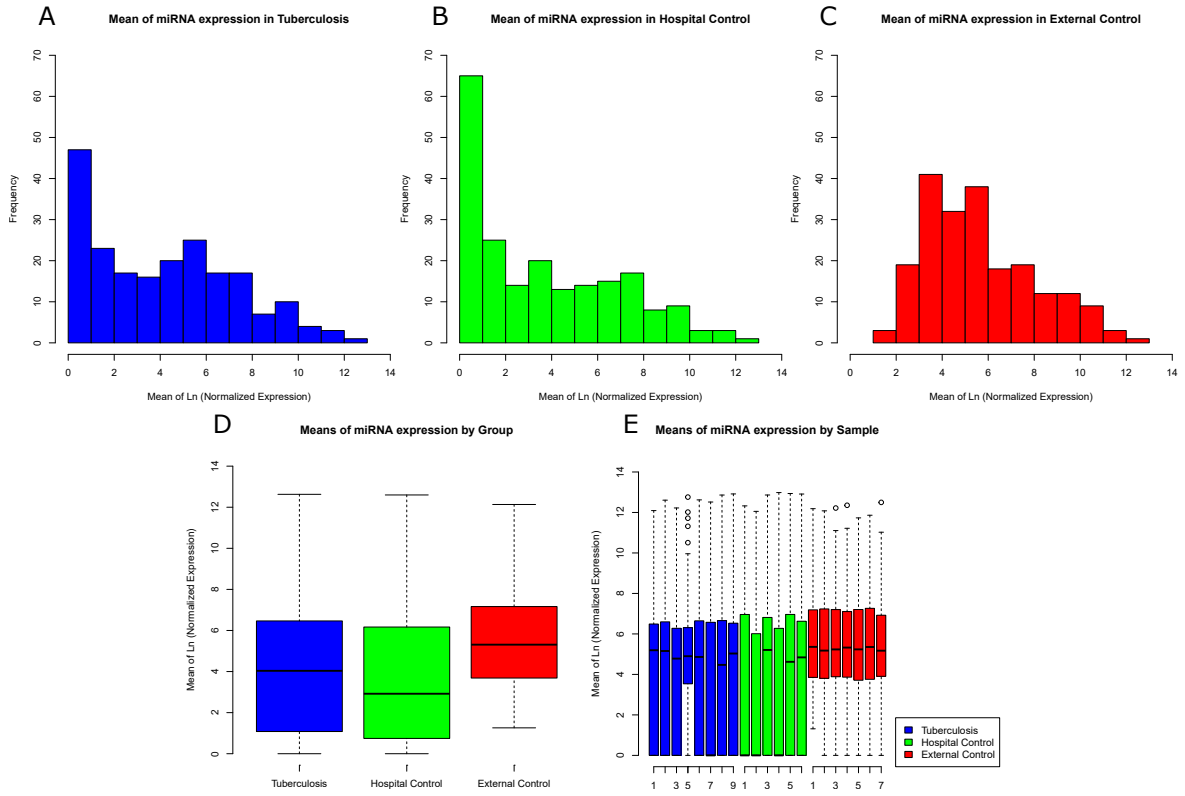

Figure 1. Frequency distrution of miRNAs' reads among the groups: A) Normalized miRNA's Ln mean frequency distribution in Tuberculosis samples; B) Normalized miRNA's Ln mean frequency distribution in Hospital Control samples; C) Normalized miRNA's Ln mean frequency distribution in External Control samples; D) Normalized miRNAs' Ln mean expression by groups; and E) Normalized miRNAs' Ln mean expression by sample.

DEmiRNAs. In the first cluster, External Control group presented a higher expression level, in the second, External Control presented a lower expression level. However, this comparison could not sort Tuberculosis patients from Hospital Control.

A total of 130 DEmiRNAs were found in the differential expression analysis results obtained by the comparison between Medical Samples and External Control groups (Figure 2B). This was the second highest number of DEmiRNAs found in all five comparisons. While Tuberculosis patients versus External Control comparison contained a higher number of DEmiRNAs (135 DEmiRNAs), 122 DEmiRNAs were shared between both comparisons.

Principal Component Analysis (PCA) was performed to verify how closely related the three initial groups were (Figure 2C) regarding their miRNA expression profile. The results showed that External Control group could be completely isolated while Tuberculosis patients could not be separated from within Hospital Control group. This, along with Heatmap results, were our first insights that Hospital Control group was closer to Tuberculosis patients in a miRNA interaction level than it was to External Control.

2.2.2. Tuberculosis patients *vs* Hospital Control

Probably this is the most important comparison performed. Hospital Control group's members had a high contact with Tuberculosis patients, thus, they have been constantly exposed to *Mycobacterium tuberculosis* bacterias. This high exposure promotes changes in the organism, including miRNA expression profile. However, a biological mechanisms is keeping Hospital Control group from acquiring active tuberculosis. This differential expression analysis' results could potentially identify which miRNAs are regulating these mechanisms and use them as biomarkers for the disease.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

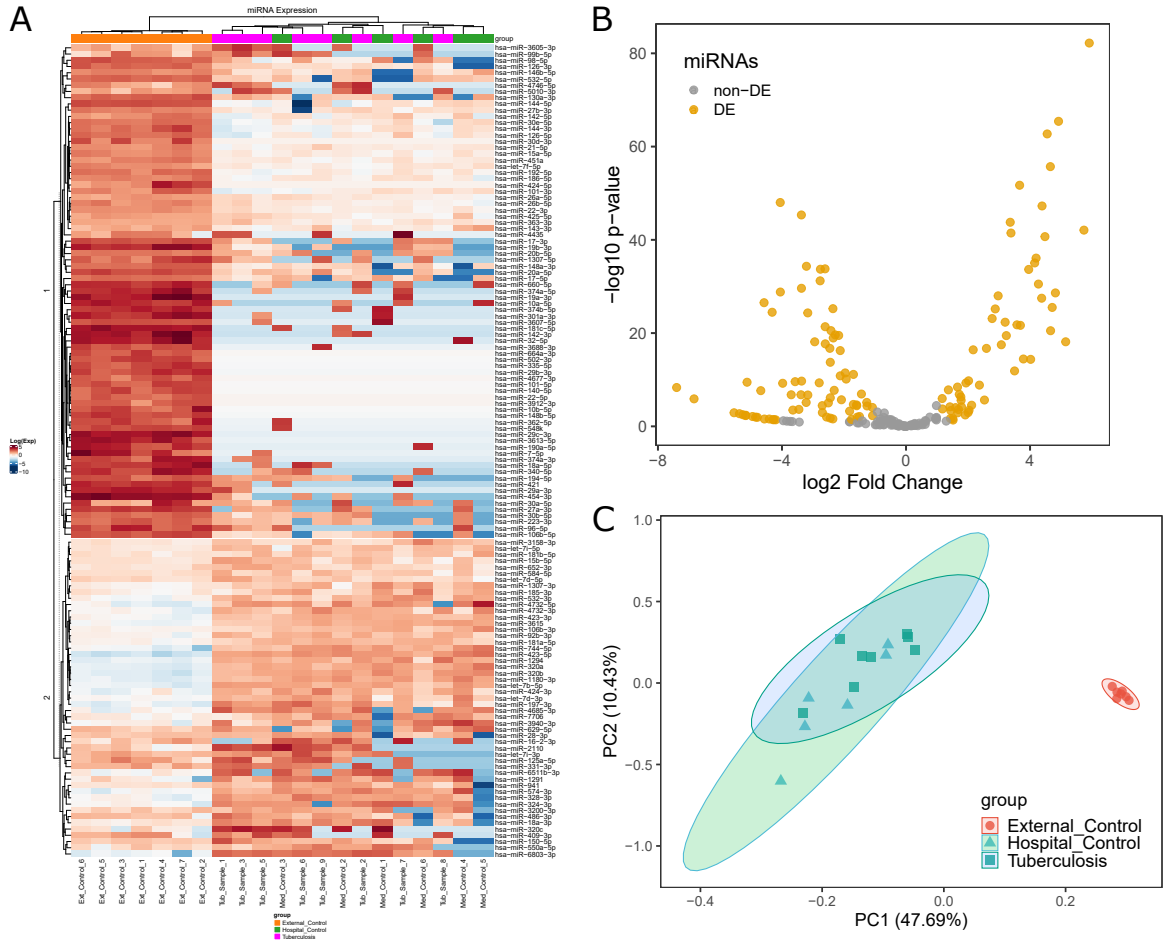

Figure 2. miRNAs profile's differential analysis between Medical Samples and External Control: A) Difference in expression levels of miRNAs between the three initial groups (EC: orange; HC: pink; TP: green) indicated by a Heatmap; B) Volcano plot highlighting 130 DEmiRNAs in Medical Samples versus External Control; C) PCA plot highlighting the homogeneity and separation of the three initial groups based on miRNA expression.

The results from edgeR's statistical differential analysis found only three DEmiRNAs (*hsa-miR-4732-5p*, *hsa-miR-486-3p* and *hsa-let-7g-5p*) (see Figure 3B). The first two with higher expression in Hospital Control and the third one with higher expression in Tuberculosis patients. This was the lowest number of miRNAs found in any of the performed comparisons, indicating a lack of DEmiRNAs involved in the process of keeping the organism safe. Both Heatmap (Figure 3A) and PCA (3C) analysis does not reveal a clear distinction between the two groups. Results from this comparison were alarming, both the number of DEmiRNAs found and distance of the two groups were unexpectedly low, indicating that Hospital Control group's members were been exposed to a higher risk level than previously expected.

2.3. Gene set enrichment analysis

To perform this analysis, DEmiRNAs found in the previous analysis were submitted to both target gene analysis and later to functional enrichment analysis using Reactome platform. Data from Medical Samples versus External Control comparison run through a new quality pipeline to boost target genes analysis. Enrichment analysis results from both comparisons indicated that the DEmiRNAs may interact with cell proliferation inhibition and/or inflammatory response processes, which are some of the affected biological processes expected in tuberculosis disease. Enriched pathways were selected

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

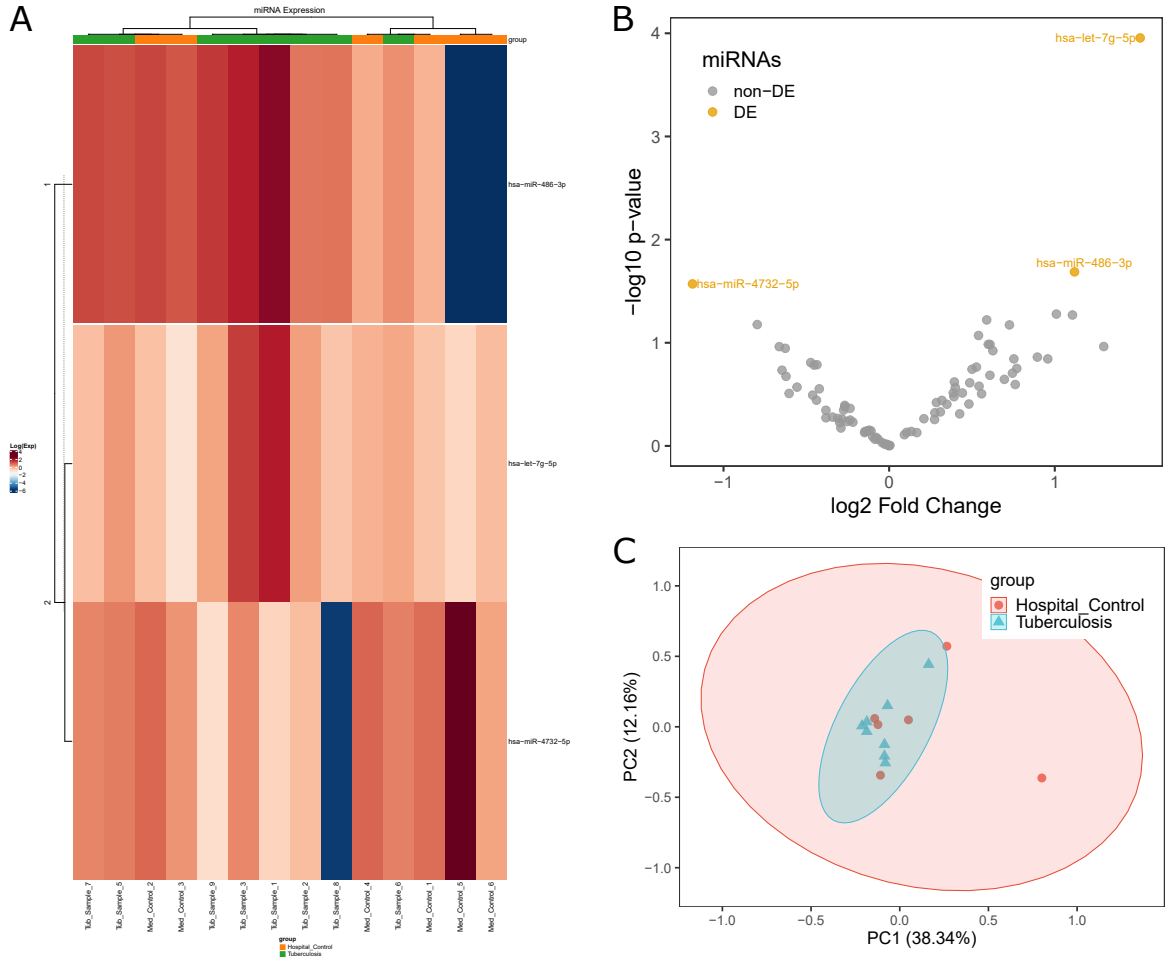

Figure 3. miRNAs profile's differential analysis between Tuberculosis patients and Hospital Control samples: A) Difference in expression levels of miRNAs between the two groups (HC: orange; TP: green) indicated by a Heatmap; B) Volcano plot highlighting three miRNAs with differential expression; C) PCA plot highlighting the homogeneity and separation of the groups based on miRNA expression.

and observed according to the regulatory functions related to tuberculosis pathogenicity that could potentially be regulated by miRNAs.

Inflammatory response and apoptosis inhibition process have been reported by Chai et al. (2019) through targeting of FOXO1 [23]. Interleukin-6's pro-inflammatory properties were observed by Scheller et al.(2011) [24]. Bi et al. (2019) showed the suppression of cell proliferation related to the inhibition of Egr1/TGF-β/Smad pathway by a MiRNA-181a-5p [25]. Spizzo et al. (2010) indicated a TP53 dependent proapoptotic regulatory loop [26]. And finally, Sabir et al. (2018) indicated in their study regulation of TLR and TNF by miRNAs [21]. Therefore, in line with the aforementioned studies, we have also investigated and identified in our analyzes enrichment for pathways related to FOXO, IL-6, TGF-β, TP53, TLR and TNF.

2.3.1. Target genes analysis in Medical Samples *vs* External Control

Only DEmiRNAs with $|log2FoldChange|$ ($|FC|$) > 2 were kept for this analysis, resulting in a total of 97 DEmiRNAs. Considering that miRNAs with positive FC values could potentially regulate different mechanisms of the human body compared to the miRNAs with negative FC values, these 97 DEmiRNAs were separated into 2 new groups, downregulated and upregulated. The downregulated group presented 59 DEmiRNAs, while the upregulated group presented 38.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

While performing target genes investigation to both groups, it was noticed that among the downregulated group, 54 of 59 DEmiRNAs presented experimental interaction validation with 948 different genes, while the upregulated group had 24 validated interactions with 221 genes, resulting in a total of 1051 different genes to be analyzed. To verify which pathways were enriched by each group of DEmiRNAs, genes found in one of the groups were submitted to a gene set enrichment analysis using Reactome, this resulted in 2 gene set analyses performed for this comparison.

Gene enrichment analysis for genes associated with upregulated miRNAs resulted in 1161 associated pathways in Reactome, among which, 346 had a significant association (FDR value <= 0.05), while results regarding downregulated miRNAs found 1732 Reactome pathways (465 significant). Pathways related to tuberculosis were selected and their results were presented in Table 1 and Table 2, which contains the pathway, number of genes interacting with it, and FDR value for the pathway, for upregulated and downregulated groups.

Pathway	Entities found	FDR Value
Cytokine signaling in immune system	82	6,2E-14
Apoptosis	23	1,55E-8
FOXO-mediated transcription of cell cycle genes	8	8,98E-6
TP53 regulates transcription of cell death genes	11	9,55E-5
Signaling by TGF-β Receptor complex	9	1,87E-3
Transcriptional regulation of pluripotent stem cells	6	4,48E-3
FcϵRI signaling	10	1,04E-1
Interleukin-6 signaling	2	1,34E-1
Adaptive immune system	30	1,64E-1
Innate immune system	37	1,64E-1
Regulation of TLR by endogenous ligand	2	1,64E-1
Signaling by B Cell Receptor	7	1,64E-1
Autophagy	6	1,95E-1
Metabolism of nitric oxide: NOS3 activation and regulation	1	6,17E-1
TNF signaling	1	6,97E-1
FcγR dependent phagocytosis	3	8,29E-1
Antimicrobial peptides	1	9,44E-1
ADORA2B mediated anti-inflammatory cytokines production	1	9,76E-1
Metabolism of lipids	19	9,99E-1

Table 1. Gene set enrichment analysis results for genes associated with upregulated miRNAs (Medical Samples *vs* External Control analysis).

2.3.2. Target genes analysis in Tuberculosis patients *vs* Hospital Control

Considering the low number of DEmiRNAs for this comparison (see Figure 3), the three DEmiRNAs (*hsa-miR-4732-5p*,*hsa-miR-486-3p* and *hsa-let-7g-5p*) were not divided in upregulated and downregulated groups and target genes analysis was performed. These DEmiRNAs were verified to have, respectively, interaction with 78, 202, 341 genes and interacting with a total of 599 different genes. Interestingly, these DEmiRNAs show an interaction overlap, three genes, *CDKN1A*, *ZC3HAV1L* and *TUBB2A*, are common targets to the three DEmiRNAs.

Studies shows that *CDKN1A* may be involved in p53/TP53 mediated inhibition of cellular proliferation in response to DNA damage [27,28], while *TUBB2A* encodes β-tubulin. Both genes are closely related to the inhibition of cellular proliferation and might be related to the formation and maturation of granulomas.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

Pathway	Entities found	FDR Value
Cytokine Signaling in Immune system	278	2,11E-14
FOXO-mediated transcription of cell cycle genes	23	6,02E-13
Signaling by TGF-β Receptor complex	34	3,67E-10
Transcriptional regulation of pluripotent stem cells	21	9,86E-8
Apoptosis	44	1,39E-6
TP53 regulates transcription of cell death genes	23	2,07E-6
Interleukin-6 signaling	10	1,23E-4
FcϵRI signaling	35	2,17E-2
TNF signaling	11	2,55E-2
Regulation of TLR by endogenous ligand	8	2,85E-2
Autophagy	25	4,44E-2
Signaling by B Cell Receptor	25	1,19E-1
Innate immune system	137	2,33E-1
FcγR dependent phagocytosis	22	2,6E-1
Adaptive immune system	97	3,6E-1
Metabolism of nitric oxide: NOS3 activation and regulation	3	7,34E-1
ADORA2B mediated anti-inflammatory cytokines production	9	9,59E-1
Antimicrobial peptides	4	9,97E-1
Metabolism of lipids	69	1

Table 2. Gene set enrichment analysis results for genes associated with downregulated miRNAs (Medical Samples *vs* External Control analysis).

2.4. Network analysis of target genes

For this analysis, three different complex networks were built, two for interactions between miRNA-genes in Medical Samples versus External Control comparison and one for interactions in Tuberculosis patients versus Hospital Control comparison. These complex networks were built regarding experimentally interactions catalogued in miRTArBase.

Figure 4A is the graphical representation of the miRNA-gene network for genes regulated by downregulated miRNAs in Medical Samples versus External Control comparison, while Figure 4B represents the miRNA-gene network for genes regulated by upregulated miRNAs. We observed that the network regarding downregulated miRNAs in tuberculosis was much bigger and had a higher number of interactions per miRNA compared to the upregulated network, indicating that there are more genetic interactions been regulated in External Control groups than in both Tuberculosis patients and Hospital Control groups. A total of 118 genes were regulated in both networks as indicated by the Venn diagram in Figure 4C.

For Tuberculosis patients versus Hospital Control comparison, we used as input for ncRNA-network tool a total of 599 different genes. Figure 5A represents the miRNA-gene network for this comparison. Among the three DEmiRNAs an interaction overlap of three genes, *CDKN1A*, *ZC3HAV1L* and *TUBB2A* was found (Figure 5B). Interestingly, *hsa-let-7g-5p*, which had a higher expression in Tuberculosis patients, also had the highest number of regulated genes in this comparison.

3. Materials and Methods

3.1. Sample collection

The present study consisted of 22 samples distributed in three groups (Tuberculosis – 8 samples; Hospital Controls – 6 samples; and External Control – 7 samples). Tuberculosis patients group was obtained from a set of patients with active pulmonary tuberculosis in regular treatment using standard tuberculosis treatment protocols in Brazil (MS, 2011). Hospital Control was composed by health

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

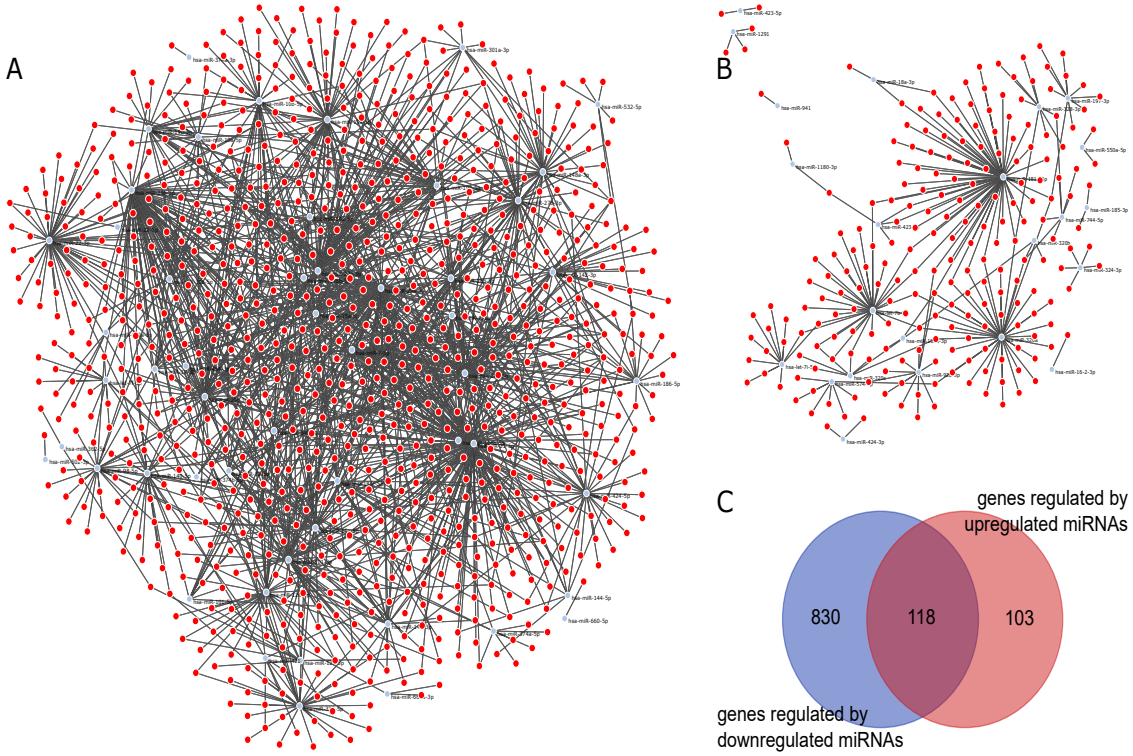

Figure 4. miRNA-gene networks for MS *vs* EC comparison generated by ncRNA-network tool: A) Graphical representation of the miRNA-gene network of downregulated miRNAs in MS *vs* EC comparison. B) Graphical representation of the miRNA-gene network of downregulated miRNAs in MS *vs* EC comparison. C) Venn-diagram of genes regulated by downregulated miRNAs and upregulated miRNAs.

professionals without prior history of tuberculosis and had constant contact with patients with active tuberculosis. External Control (EC) Group was comprised by volunteers without tuberculosis and with no contact with Tuberculosis patients in treatment. Tuberculosis patients samples were obtained from patients attending the Pulmonary Tuberculosis Clinic of the João Barros Barreto Hospital and Hospital Control samples from health professional of the same clinic, both were retrieved from May 17th-22nd, 2017. Biological material was collected in a 3 mL tube containing RNA later and stored until RNA extraction. All Tuberculosis patients' samples were collected directly by their attending physician.

All research procedures were carried out in accordance with the Declaration of Helsinki and the Nuremberg Code, following Research Standards Involving Human Beings (Res. CNS 196/96) of the National Health Council, respecting the ethical standards and the rights of patients. The project was approved by the Human Research Ethics Committee of the Federal University of Pará, under HUJBB's protocol number 350507. The data were collected after the research was explained and the patients signed an informed consent form.

3.2. RNA extraction and quantification

Peripheral blood samples (5 mL) were collected using Tempus Blood RNA Tube (Thermo Fisher Scientific, US) and stored at -20 ∘C until extraction. Total RNA was extracted using MagMAX RNA Isolation Kit (Thermo Fisher Scientific, US) and quantified with NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, US). Agilent RNA ScreenTape assay and 2200 TapeStation Instrument (Agilent Technologies, US) were used to detect and ensure RNA integrity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

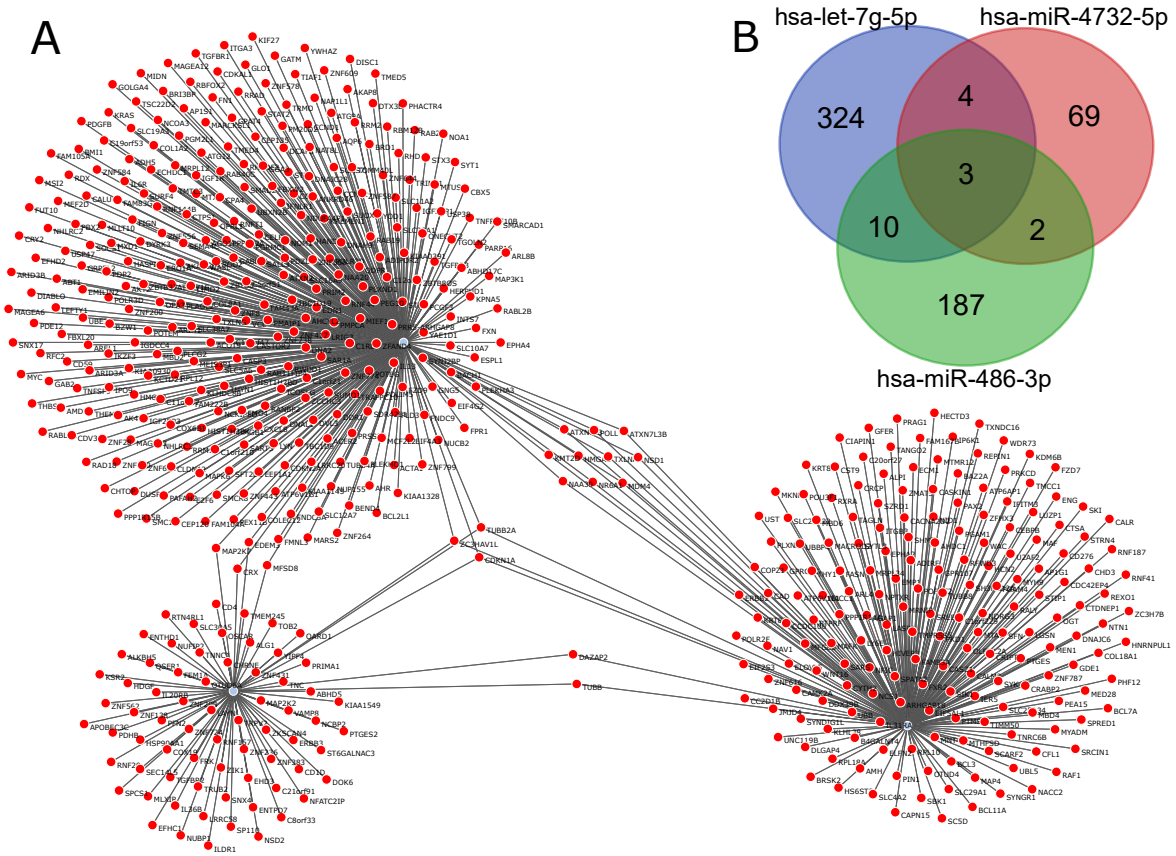

Figure 5. miRNA-gene network for TP *vs* HC comparison generated by ncRNA-network tool: A) miRNA-gene network for the three DEmiRNAs (*hsa-miR-4732-5p*, *hsa-miR-486-3p* and *hsa-let-7g-5p*); B) Venn diagram indicating intersections between genes for each miRNA.

3.3. Library construction and sequencing

For small RNA-Seq, 1 (one) μg of total RNA per sample was used for library preparation using TruSeq Small RNA Sample Prep Kits (Illumina, San Diego, CA, USA). Size-distribution was measured with the DNA ScreenTape assay on a 2200 TapeStation system (AgilentTechnologies, US). A total library pool of 4 nM was sequenced using a MiSeq Reagent Kit v3 150 cycle on a MiSeq System (Illumina, San Diego, CA, USA).

3.4. miRNA quantification and normalization

Biological data obtained from the 22 sequenced samples underwent a quality control pipeline to remove adapters used during the sequencing, as well as to trim and filter the obtained sequences. This quality control process was performed using Trimmomatic software version 0.36, "ILLUMINACLIP" (with a custom adapters list), "LEADING:10" and "TRAILING:10" (both cut bases with quality lower than 10 at the beginning or end of the read respectively), "SLIDINGWINDOW:3:22" (performs a sliding window approach with a 3 bases window size, cutting bases once the average quality within the window falls below 22) and "MINLEN:16" (remove reads with less than 16 bases) parameters were used. Sequences were then aligned with the human genome (HG19) using STAR software. Result files, generated as ".sam" format by STAR, were manipulated using samtools and converted to ".bam" files.

miRNA's expression quantification was performed with HTSeq software, using the human genome annotation file (".gff") and "type" parameter set to "miRNA". Data from alignment process were properly classified before being submitted to a new quality control, which kept miRNAs with at least 10 total read count in at least 1 sample and removed samples with a total read count lower than

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2021 doi:10.20944/preprints202101.0195.v1

1000. This procedure resulted in the removal of 1 sample from the analysis (sample belonging to TB group), this new data served as the raw data for the differential expression analysis.

Two different types of normalization were employed, differential expression analysis was performed using EdgeR package function “calcNormFactors"(which utilizes TMM’s normalization method as package default). For other analysis such as histograms, boxplots and principal component analysis, raw data was normalised by counts per million (CPM), a measurement of reads abundance used to compare the expression of miRNA in different samples or libraries sizes.

3.5. Differential analysis

Data exploratory analysis was performed with R version 3.5.0, Rstudio (v1.1) and shell script. Since the samples were divided into 3 different groups, a total of 5 differential expression analysis were performed comparing: i) Tuberculosis patients versus Controls (both Hospital and External Control groups); ii) Tuberculosis patients versus Hospital Control; iii) Tuberculosis patients versus External Control; iv) Hospital Control versus External Control; and v) Medical Samples (formed by both Tuberculosis patients and Hospital Control groups) versus External Control. Differential expression analysis was performed with edgeR program package, using FDR’s adjustment method. miRNAs with adjusted p-value < 0.05 and $|FC| > 1$ were isolated an considered as differentially expressed.

3.6. Modeling a miRNA-gene networks and analysis

Interactions between miRNAs and their target genes are extremely complex, so it is necessary to implement computational methods to allow their better understanding. Thus, network modeling is a valuable approach to measure and visualize interactions between different components of regulatory networks. We modeled a network of miRNA-gene interaction based on public data of miRTarBase [29]. miRTarBase have catalogued 300k of interactions between miRNAs and genes, which are validated by different types of experimentally studies with microarray data, western blot, report assays and next-generation sequencing.

Our miRNA-network was modeled as a bipartite graph G = (V, U, E), in which G is a graph comprised by two distinct sets of regulatory elements: U, the set of miRNAs and V, the set of genes. Interactions between miRNA and genes were defined if there are two or more experimentally validation studies based on evidences catalogued in miRTarBase.

For the miRNA-gene network we computed the number of interactions (degree of a node) as a centrality index for both regulatory elements, genes and miRNAs. After differential expression analysis we classified DEmiRNAs on the network as upregulated and downregulated miRNAS to investigate patterns of interactions between both distinct groups.

The network were constructed and graphically represented with NetworkX implemented in Python 3 and an in-house tool available at www.lgmh.ufpa.br/ncrnas-network.

4. Conclusions

Almost all differential expression analysis resulted in a high number of DEmiRNAs found (at least 90 DEmiRNAs per comparison), indicating a large set of DEmiRNA, except for differential expression analysis between Tuberculosis patients versus Hospital Control group, which resulted in only three DEmiRNAs, this concerned our group, considering it indicates that only a small group of miRNAs were potentially keeping the physicians from acquiring the active form of tuberculosis. The provided results indicate that, despite not having the active form of the disease, the physicians had a very similar miRNA expression level to tuberculosis patients. It is possible that this proximity is a result of the treatment provided before taking the tuberculosis samples (since all Tuberculosis patients’ samples could only be collected after the treatment had already started), however, despite this possibility, the difference between these two group and External Control’s miRNA expression level is still high.

Differential analysis results from all the comparisons provided a total of 153 different DEmiRNAs to be further studied to provide a better insight of the mechanisms involved in the response to the disease.

5. Study limitations

In this current work, we highlight some limitations: i) a small sample number that could hide important DEmiRNAs; and ii) patients involved in the study were in different stages of treatment, some in acute phase of anti-tuberculosis multidrug therapy and others in maintenance phase. However, this is the first work to bring data related to health professionals in daily contact with high bacillary loads of tuberculosis, which could indicate possible biomarkers of better diagnostic accuracy, besides proposing a new mechanism of labor monitoring of these professionals regarding the risk of manifesting the disease.

Author Contributions: CAS, ARS and SS designed research; CAS enrolled patients, performed, and registered clinical diagnosis; PP, TV-S, AMRS and AFV performed research; ArthurRS, RPP, AMRS and GSA analyzed the data; CAS, ArthurRS, WGG, MHH and GSA wrote the article; CAS, ArthurRS, WGG, PP, RPP, TV-S, AMRS, MHH, AFV, GSA, ARS and SS agree with manuscript results and conclusions.

Funding: We acknowledge the support of CNPq (306815/2018-3 grant for Ândrea Ribeiro-dos-Santos; and 305258/2013-3 grant for Sidney Santos), CAPES PROAMAZONIA (88887.200498/2018-00), CAPES Biocomputacional – Rede PGPH (3381/2013), PROPESP/UFPA. The funders had no role on the manuscript data or concept.

Acknowledgments: The authors are grateful to the entire team of professionals from the Hospital João de Barros Barreto of Universidade Federal do Pará (UFPA) and the tuberculosis patients, without which it would not be possible to make this paper.

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

References

1. Xu, Z.; Xia, A.; Li, X.; Zhu, Z.; Shen, Y.; Jin, S.; Lan, T.; Xie, Y.; Wu, H.; Meng, C.; others. Rapid loss of early antigen-presenting activity of lymph node dendritic cells against Ag85A protein following Mycobacterium bovis BCG infection. *BMC immunology* **2018**, *19*, 19.
2. Organization, W.H.; others. *Global tuberculosis report 2018*; World Health Organization, 2018.
3. Organization, W.H. *Global tuberculosis control: surveillance, planning, financing: WHO report 2008*; Vol. 393, World Health Organization, 2008.
4. Marais, B.J.; Gie, R.P.; Schaaf, H.S.; Starke, J.R.; Hesseling, A.C.; Donald, P.R.; Beyers, N. A proposed radiological classification of childhood intra-thoracic tuberculosis. *Pediatric radiology* **2004**, *34*, 886–894.
5. Fu, Y.; Yi, Z.; Wu, X.; Li, J.; Xu, F. Circulating microRNAs in patients with active pulmonary tuberculosis. *Journal of clinical microbiology* **2011**, *49*, 4246–4251.
6. Ndzi, E.N.; Nkenfou, C.N.; Mekue, L.M.; Zentilin, L.; Tamgue, O.; Pefura, E.W.Y.; Kuiaté, J.R.; Giacca, M.; Ndjolo, A. MicroRNA hsa-miR-29a-3p is a plasma biomarker for the differential diagnosis and monitoring of tuberculosis. *Tuberculosis* **2019**, *114*, 69–76.
7. van Rensburg, I.; du Toit, L.; Walzl, G.; du Plessis, N.; Loxton, A. Decreased neutrophil–associated miRNA and increased B-cell associated miRNA expression during tuberculosis. *Gene* **2018**, *655*, 35–41.
8. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *cell* **2009**, *136*, 215–233.
9. Malardo, T.; Gardinassi, L.G.; Moreira, B.P.; Padilha, É.; Lorenzi, J.C.C.; Soares, L.S.; Gembre, A.F.; Fontoura, I.C.; de Almeida, L.P.; de Miranda Santos, I.K.F.; others. MicroRNA expression signatures in lungs of mice infected with Mycobacterium tuberculosis. *Tuberculosis* **2016**, *101*, 151–159.
10. Cheng, H.; Li, H.; Feng, Y.; Zhang, Z. Correlation analysis between SNPs in microRNA-machinery genes and tuberculosis susceptibility in the Chinese Uygur population. *Medicine* **2018**, *97*.
11. Wang, C.; Yang, S.; Sun, G.; Tang, X.; Lu, S.; Neyrolles, O.; Gao, Q. Comparative miRNA expression profiles in individuals with latent and active tuberculosis. *PloS one* **2011**, *6*, e25832.
12. von Both, U.; Berk, M.; Agapow, P.M.; Wright, J.D.; Git, A.; Hamilton, M.S.; Goldgof, G.; Siddiqui, N.; Bellos, E.; Wright, V.J.; others. Mycobacterium tuberculosis exploits a molecular off switch of the immune system for intracellular survival. *Scientific reports* **2018**, *8*, 661.

13. Duffy, F.J.; Thompson, E.; Downing, K.; Suliman, S.; Mayanja-Kizza, H.; Boom, W.H.; Thiel, B.; Weiner III, J.; Kaufmann, S.H.; Dover, D.; others. A serum circulating miRNA signature for short-term risk of progression to active tuberculosis among household contacts. *Frontiers in immunology* **2018**, *9*, 661.
14. Rothchild, A.C.; Sissons, J.R.; Shafiani, S.; Plaisier, C.; Min, D.; Mai, D.; Gilchrist, M.; Peschon, J.; Larson, R.P.; Bergthaler, A.; others. MiR-155–regulated molecular network orchestrates cell fate in the innate and adaptive immune response to Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences* **2016**, *113*, E6172–E6181.
15. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends in genetics* **2008**, *24*, 133–141.
16. Smith, I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews* **2003**, *16*, 463–496.
17. Singh, Y.; Kaul, V.; Mehra, A.; Chatterjee, S.; Tousif, S.; Dwivedi, V.P.; Suar, M.; Van Kaer, L.; Bishai, W.R.; Das, G. Mycobacterium tuberculosis controls microRNA-99b (miR-99b) expression in infected murine dendritic cells to modulate host immunity. *Journal of Biological Chemistry* **2013**, *288*, 5056–5061.
18. Belver, L.; Papavasiliou, F.N.; Ramiro, A.R. MicroRNA control of lymphocyte differentiation and function. *Current opinion in immunology* **2011**, *23*, 368–373.
19. Liu, Y.; Wang, X.; Jiang, J.; Cao, Z.; Yang, B.; Cheng, X. Modulation of T cell cytokine production by miR-144* with elevated expression in patients with pulmonary tuberculosis. *Molecular immunology* **2011**, *48*, 1084–1090.
20. Chatterjee, S.; Dwivedi, V.P.; Singh, Y.; Siddiqui, I.; Sharma, P.; Van Kaer, L.; Chattopadhyay, D.; Das, G. Early secreted antigen ESAT-6 of Mycobacterium tuberculosis promotes protective T helper 17 cell responses in a toll-like receptor-2-dependent manner. *PLoS pathogens* **2011**, *7*, e1002378.
21. Sabir, N.; Hussain, T.; Shah, S.Z.A.; Peramo, A.; Zhao, D.; Zhou, X. miRNAs in tuberculosis: new avenues for diagnosis and host-directed therapy. *Frontiers in microbiology* **2018**, *9*, 602.
22. Ouimet, M.; Koster, S.; Sakowski, E.; Ramkhelawon, B.; Van Solingen, C.; Oldebeken, S.; Karunakaran, D.; Portal-Celhay, C.; Sheedy, F.J.; Ray, T.D.; others. Mycobacterium tuberculosis induces the miR-33 locus to reprogram autophagy and host lipid metabolism. *Nature immunology* **2016**, *17*, 677–686.
23. Chai, X.; Si, H.; Song, J.; Chong, Y.; Wang, J.; Zhao, G. miR-486-5p Inhibits Inflammatory Response, Matrix Degradation and Apoptosis of Nucleus Pulposus Cells through Directly Targeting FOXO1 in Intervertebral Disc Degeneration. *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology* **2019**, *52*, 109–118.
24. Scheller, J.; Chalaris, A.; Schmidt-Arras, D.; Rose-John, S. The pro-and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **2011**, *1813*, 878–888.
25. Bi, J.g.; Zheng, J.f.; Li, Q.; Bao, S.y.; Yu, X.f.; Xu, P.; Liao, C.x. MicroRNA-181a-5p suppresses cell proliferation by targeting Egr1 and inhibiting Egr1/TGF-β/Smad pathway in hepatocellular carcinoma. *The international journal of biochemistry & cell biology* **2019**, *106*, 107–116.
26. Spizzo, R.; Nicoloso, M.; Lupini, L.; Lu, Y.; Fogarty, J.; Rossi, S.; Zagatti, B.; Fabbri, M.; Veronese, A.; Liu, X.; others. miR-145 participates with TP53 in a death-promoting regulatory loop and targets estrogen receptor-α in human breast cancer cells. *Cell Death & Differentiation* **2010**, *17*, 246–254.
27. Broude, E.V.; Demidenko, Z.N.; Vivo, C.; Swift, M.E.; Davis, B.M.; Blagosklonny, M.V.; Roninson, I.B. p21 (CDKN1A) is a negative regulator of p53 stability. *Cell Cycle* **2007**, *6*, 1467–1470.
28. Löhr, K.; Möritz, C.; Contente, A.; Dobbelstein, M. p21/CDKN1A mediates negative regulation of transcription by p53. *Journal of Biological Chemistry* **2003**, *278*, 32507–32516.
29. Chou, C.H.; Shrestha, S.; Yang, C.D.; Chang, N.W.; Lin, Y.L.; Liao, K.W.; Huang, W.C.; Sun, T.H.; Tu, S.J.; Lee, W.H.; others. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic acids research* **2017**, *46*, D296–D302.