

Article

Metatranscriptomic Analysis of Human Lung Metagenomes from Patients with Lung Cancer

Ya-Sian Chang ^{1,2,3}, Ming-Hung Hsu ², Siang-Jyun Tu ², Ju-Chen Yen ¹, Ya-Ting Lee ¹, Hsin-Yuan Fang ⁴ and Jan-Gowth Chang ^{1,2,3,5,*}

- ¹ Epigenome Research Center, China Medical University Hospital, Taichung 404, Taiwan; t25074@mail.cmuh.org.tw (Y.-S.C.); t24399@mail.cmuh.org.tw (J.-C.Y.); t23701@mail.cmuh.org.tw (Y.-T.L.)
² Center for Precision Medicine, China Medical University Hospital, Taichung 404, Taiwan; t35492@mail.cmuh.org.tw (M.-H.H.); t34752@mail.cmuh.org.tw (S.-J.T.); d6781@mail.cmuh.org.tw (J.-G.C.)
³ School of Medicine, China Medical University, Taichung 404, Taiwan
⁴ Department of Thoracic Surgery, China Medical University Hospital, Taichung 404, Taiwan; d17573@mail.cmuh.org.tw
⁵ Department of Bioinformatics and Medical Engineering, Asia University, Taichung 413, Taiwan
* Correspondence: d6781@mail.cmuh.org.tw; Tel.: +886-4-22052121 (ext. 2008)

Citation: Chang, Y.-S.; Hsu, M.-H.; Tu, S.-J.; Yen, J.-C.; Lee, Y.-T.; Fang, H.-Y.; Chang, J.-G.

Metatranscriptomic Analysis of Human Lung Metagenomes from Patients with Lung Cancer. *Genes* **2021**, *12*, 1458. <https://doi.org/10.3390/genes12091458>

Academic Editor: Sudheer Kumar Gara

Received: 30 July 2021

Accepted: 16 September 2021

Published: 21 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: This study was designed to characterize the microbiomes of the lung tissues of lung cancer patients. RNA-sequencing was performed on lung tumor samples from 49 patients with lung cancer. Metatranscriptomics data were analyzed using SAMSA2 and Kraken2 software. 16S rRNA sequencing was also performed. The heterogeneous cellular landscape and immune repertoires of the lung samples were examined using xCell and TRUST4, respectively. We found that nine bacteria were significantly enriched in the lung tissues of cancer patients, and associated with reduced overall survival (OS). We also found that subjects with mutations in the epidermal growth factor receptor gene were less likely to experience the presence of *Pseudomonas aeruginosa*. We found that the presence of CD8+ T-cells, CD4+ naive T-cells, dendritic cells, and CD4+ central memory T cells were associated with a good prognosis, while the presence of pro B-cells was associated with a poor prognosis. Furthermore, high clone numbers were associated with a high ImmuneScore for all immune receptor repertoires. Clone numbers and diversity were significantly higher in unrepresented subjects compared to presented subjects. Our results provide insight into the microbiota of human lung cancer, and how its composition is linked to the tumor immune microenvironment, immune receptor repertoires, and OS.

Keywords: metatranscriptomic; lung cancer; microenvironments; T cell receptor repertoires; B cell receptor repertoires

1. Introduction

Lung cancer is the most common cancers worldwide, and its incidence is rapidly increasing in Taiwan. Globally, there are 1.8 million lung cancer diagnoses and 1.6 million deaths per year [1]. Chemical carcinogens, chronic inflammation, bacterial and viral infections, periodontal diseases, and various other factors promote lung cancer development. Several pathogenic microorganisms are associated with lung cancer including *Haemophilus influenzae*, *Acidovorax*, *Klebsiella*, *Moraxella catarrhalis*, *Mycobacterium tuberculosis* and *Granulicatella adiacens* [2]. Despite recent advances in targeted receptor therapies and immunotherapies, the five-year survival rate of lung cancer remains low. Major reasons for the poor prognosis include late diagnosis and resistance to standard chemotherapy [3].

Over the last decade, a number of studies have investigated the microbiome of cancer patients, including patients with oral [4], lung [5], stomach, esophageal [6], pancreatic [7],

intestinal [8], and prostate [9] cancers. Straussman et al. performed 16S rRNA sequencing on 1526 samples collected from patients at nine medical centers with seven cancer types; they could distinguish between various cancer types based on the microbial DNA present in the samples [10]. In the same year, Knight et al. analyzed the microbial reads in whole-genome and whole-transcriptome sequencing data from The Cancer Genome Atlas for 33 cancer types, and found associations between different cancer types and specific microbiota [11].

The first study of lung tissue microbiota was conducted in 2016 [12], and revealed that the genus *Thermus* is significantly more abundant in advanced-stage cancer patients (stage IIIB and IV). Furthermore, *Acidovorax*, *Cyanobacteria*, *Streptococcus* and *Prevotella* are enriched in lung cancer tissues [5,13,14]. Another recent study explored the relationship between the microbiome of resected lung tissue and lung cancer prognosis [15], and revealed that greater diversity of normal tissue was associated with poorer recurrence-free and disease-free survival. Recently, the lung microbiota of bronchoalveolar lavage fluid and non-malignant, peritumoral and tumor tissue samples, from 18 non-small cell lung cancer (NSCLC) patients, [16] were examined; bronchoalveolar lavage fluid was found to harbor unique microbiota, but few differences were found among the other tissues. In the three types of lung tissue samples, *Pseudomonas*, *Clostridium*, *Kocuria*, *Acinetobacter*, and *Sphingomonas* were the five most abundant genera. All of the studies discussed above used 16S rRNA sequencing to examine human lung tissue microbiota. However, metatranscriptomic analysis has not yet been performed.

Advances in next generation sequencing technologies, including 16S rRNA sequencing and metagenomic technologies, have enabled the study of the microbial gene repertoire. RNA-sequencing (RNA-seq) has facilitated the detection of microorganisms in a diverse range of microbiomes. Moreover, metatranscriptomics has been applied to study various environments, including human hosts, plants, soils, and aquatic environments [17]. Researchers have applied metatranscriptomics to investigate the interactions between microbes and human host [18], and to explore associations of microbes with disease progression [19] and severity [20]. However, this technique has not yet been applied to human lung cancer tissues.

In this study, we conducted a metatranscriptomics pilot study using cancer tissues collected from 49 patients with NSCLC. Eight samples were paired tumor and non-disease samples. Using RNA-seq, we explored the associations between the tumor metatranscriptome, tumor immune microenvironment, and immune receptor repertoire.

2. Materials and Methods

2.1. Sample Population

Tissue specimens were obtained from 49 Taiwanese patients with lung cancer who underwent surgical resection between May 2007 and April 2014 at China Medical University Hospital. Forty-nine lung tumors were examined, including forty adenocarcinomas and nine squamous cell carcinomas. Surgically resected specimens were grossly dissected and preserved immediately in liquid nitrogen following surgery. The present study was approved by the Institutional Review Board of the China Medical University Hospital (CMUH106-REC1-053).

2.2. RNA-Seq

Total RNA was extracted from the clinical tissue samples using a NucleoSpin® RNA Kit (Macherey-Nagel GmbH, Düren, Germany) according to the manufacturer's instructions. The quality, quantity, and integrity of the extracted RNA were evaluated using a NanoDrop1000 spectrophotometer and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Samples with RNA integrity >6.0 were used for RNA-seq. An mRNA-focused, barcoded library was generated using a TruSeq Stranded mRNA Library Preparation Kit (Illumina, San Diego, CA, USA). The libraries were sequenced using the

Illumina Nova Seq 6000 platform (Illumina), using 2×151 bp paired-end sequencing flow cells according to the manufacturer's instructions.

2.3. RNA-Seq Data Analysis

The RNA-seq data were analyzed as described previously [21]. In brief, data quality control at the Q20 level was performed using Trimmomatic [22], read alignment to the GRCh38 human genome was conducted using HISAT2 [23], expression was quantified using GENCODE v22 (excluding including mitochondrial genes), and transcripts were normalized into transcripts per million (TPM) using StringTie [24].

2.4. Cell Enrichment Analysis

xCell [25] was used to examine the enrichment of various immune cells in the tumors, and to calculate an immune score from the TPM expression matrix.

2.5. Detection and Analysis of Immune Receptor Repertoires

To characterize the immune receptor repertoires from the RNA-seq data, we used TRUST4 (v1.0.2) [26] software, which was applied to download the IMGT Repertoire reference [27], perform *de novo* assembly, and annotate and count consensus assemblies of T cell receptors (TCRs) and B cell receptor (BCRs). The clonal diversity of TCRs and BCRs was calculated using VDJtools (v1.2.1) [28].

2.6. Metatranscriptomic Analysis

SAMSA2 [29] and Kraken2 [30] were used independently to analyze metatranscriptomic data. The first step in the SAMSA2 pipeline (v2.2.0) was to merge paired-end reads using PEAR (v0.9.6) [31], and then to remove bacteria rRNA reads using SortMeRNA (v2.1) [32]. Then, DIAMOND (v0.9.36) was used to annotate the transcriptome using the National Center for Biotechnology Information (NCBI) [33] Reference Sequence (RefSeq) database [34]. Sequences with >96% similarity were selected for annotation. Kraken2 (--kmer-len = 120, v2.1.1) [30] was also applied to analyze the metatranscriptomic data after removing rRNA reads. The NCBI RefSeq database was used to annotate transcripts against human, virus, archaea, bacterial and fungal genomes (database version: July, 2020).

2.7. 16S rRNA Sequencing

DNA was extracted from the clinical tissue samples using a DNeasy Blood and Tissue Mini kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Extracted DNA was quantified using Qubit (Life Technologies, Grand Island, NY, USA). The V4 region of the 16S rRNA gene was PCR-amplified using the 515F/806R primer pair [35], which contain Nextera adapter sequences (Illumina) at their 5'-ends, for library preparation (515F: 5'-GTGCCAGCMGCCGCGGTAA-3' and 806R: 5'-GGAC-TACHVGGGTWTCTAAT-3'). PCR was performed using a KAPA HiFi HotStart Ready-Mix PCR Kit (Roche, Cape Town, South Africa). The PCR program consisted of 30 s at 98 °C followed by 30 cycles of 10 s at 98 °C, 30 s at 60 °C, and 30 s at 72 °C, with a final amplification step of 5 min at 72 °C. PCR products were purified using the AMPure XP Beads (Beckman Coulter, Indianapolis, IN, USA) and quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). A Second round of PCR using Illumina dual-index oligos was performed using the KAPA HiFi HotStart ReadyMix PCR Kit (Roche, Basel, Switzerland), as follows: 95 °C for 3 min, eight cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, and a final amplification step at 72 °C for 5 min. Samples were pooled and purified using AMPure XP Beads (Beckman Coulter, Brea, CA, USA). Sequencing was performed on the Illumina MiSeq instrument using the MiSeq reagent kit v3 (600 cycles).

2.8. Bioinformatic Analysis of 16S rRNA Sequencing Data

Sequencing data were analyzed using Illumina local run manager software. In brief, the index reads were demultiplexed, FASTQ files were generated, and the reads were classified against the Greengenes 16S rRNA gene database (version gg_13_5) [36], which achieved up to species-level sensitivity.

2.9. Statistical Analysis

Differences between groups were compared using the Mann-Whitney test. The log-rank test and Cox proportional hazards regression model was used to compare differences in overall survival (OS) between groups. Relationships between bacterial infection status and gene mutation status were determined using Chi-square test or Fisher's exact test. All statistical analyses were calculated using SciPy (v1.2.1) package [37], lifelines (v0.22.3) package [38], GraphPad Prism 8.0.2, or SPSS 22.0. A *p*-value of less than 0.05 was considered statistically significant.

3. Results

3.1. Microorganisms Identified in Lung Cancer

A bacterial infection was considered to be present if the bacterial species was detected by both the SAMSA2 and Kraken2 tools, the bacterium was detected by SAMSA2 and 16S rRNA sequencing, or the bacterium was detected by Kraken2 and 16S rRNA sequencing. Using these criteria, 435 taxonomic groups were detected (Table S1). We applied a log-rank test of OS to evaluate the prognostic impact of each bacterial species, and found that 63 taxonomic groups were associated with survival in lung cancer patients (Table S1). We subjected *Brevundimonas diminuta* (*n* = 3), *Acinetobacter radioresistens* (*n* = 5), *Enterobacter cloacae* (*n* = 3), *Mycobacterium chelonae* (*n* = 3), *Mycobacterium franklinii* (*n* = 5) and *Staphylococcus sp.* (*n* = 3) to further analysis (Figure 1).

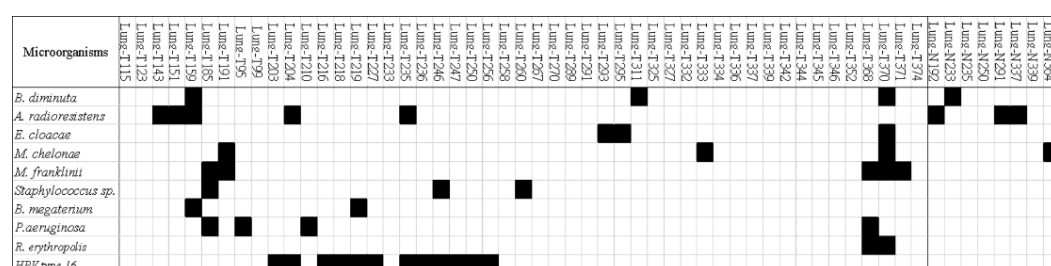


Figure 1. Microorganism composition of the lungs revealed by metatranscriptomic sequencing.

Moreover, we conducted survival analysis of patients with bacterial presences detected by SAMSA2 and Kraken2 analyses of RNA-seq and 16S rRNA sequencing data. Fifty taxonomic groups were detected, of which seven were associated with survival in lung cancer patients (Table S2). We subjected *Bacillus megaterium* (*n* = 2), *P. aeruginosa* (*n* = 4) and *Rhodococcus erythropolis* (*n* = 2) to further analysis (Figure 1).

A virus, human papillomavirus (HPV) type 16, was identified in the metatranscriptome using SAMSA2 and Kraken2 (Figure 1). However, HPV type 16 was not associated with the OS of lung cancer patients.

3.2. Cell Types Associated with Survival in NSCLC

Figure 2 shows the cellular heterogeneity of each sample based on the RNA-seq data. We see that xCell estimates of immune and stromal cell types can be used to cluster two different lung cancer subtypes. Of the 64 cell types identified, 34 were immune cells. We found that five immune cell types were significantly associated with patient survival: CD8+ T-cells, CD4+ naive T-cells, dendritic cells (DC) and CD4+ central memory T-cell (CD4+ Tcm) were associated with high OS (Figure 3A), while pro B-cells enrichment was

predictive of adverse survival outcomes (Figure 3B). Table S3 showed the hazard ratios of the impacts of tumor immune cells on lung cancer. The CD8+ T-cells, CD4+ naive T-cells, DC and CD4+ Tcm maintained the significance after adjustment for stage, grade, gender and histologic cell type ($p = 0.004$, 0.046 , 0.04 and 0.035 , respectively) (Table S3). After adjustment, the pro B-cells showed borderline significance ($p = 0.058$).

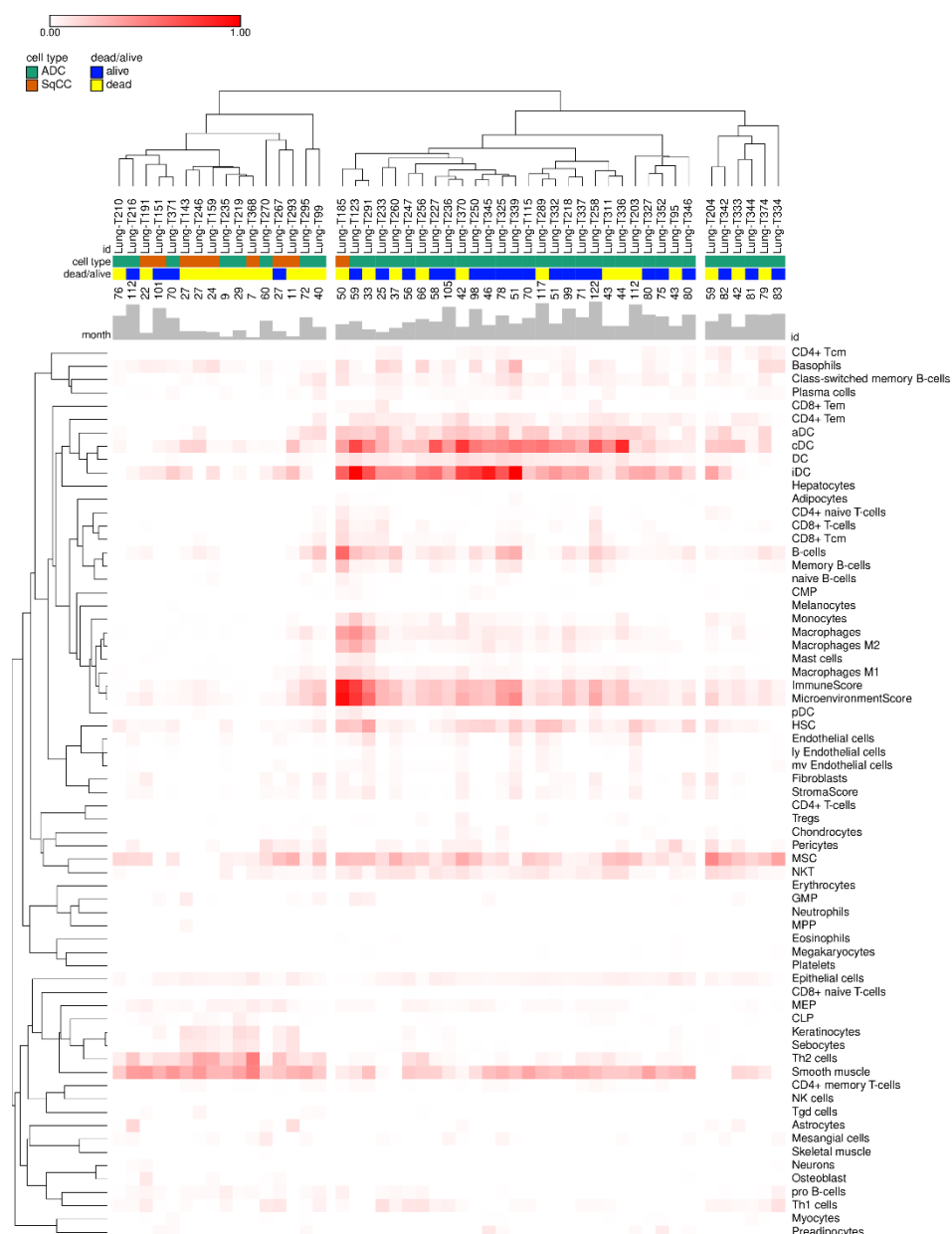


Figure 2. Heatmap showing the 64 cell types found across all samples. Columns are sample id and annotated tissue type with adenocarcinoma (ADC, green) and squamous cell carcinoma (SqCC, orange), survival status with dead (yellow) and alive (blue), survival/follow-up time in month with bar plot. Rows are cell/summary score categories from xCell. Values in heatmap plot are xCell score of each category and samples, and the color map is from white (0) to red (1). Hierarchical clustering was performed with pearson correlation for sample distance calculation and linkage method is “average” on web tool (<https://software.broadinstitute.org/morpheus/>, accessed on 16 August 2021). The three clusters were generated from manually determined distance cut-off.

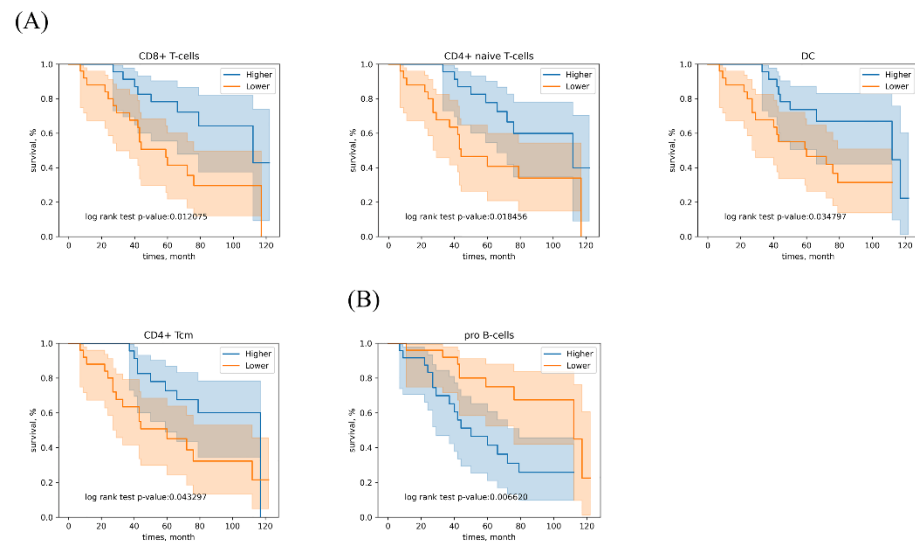


Figure 3. Kaplan-Meier plot showing overall survival (OS) according to immune cell types. (A) The group with high levels of CD8+ T-cells, CD4+ naive T-cells, DC and CD4+ Tcm exhibited significantly higher OS according to the log-rank test. (B) The group with high levels of pro B-cells showed significantly reduced OS.

3.3. Compositional Changes in Tumor Immune Cell Populations

We identified 13 immune cell subtypes, and ImmuneScore and MicroenvironmentScore differed significantly between presented and unpresented subjects (Table S4). In the multivariate analysis which incorporated independence prognostic factors of 13 immune cell subtypes, ImmuneScore and MicroenvironmentScore, we found that there is a significant association between presented subjects and poor survival ($p = 0.001$) (Figure 4).

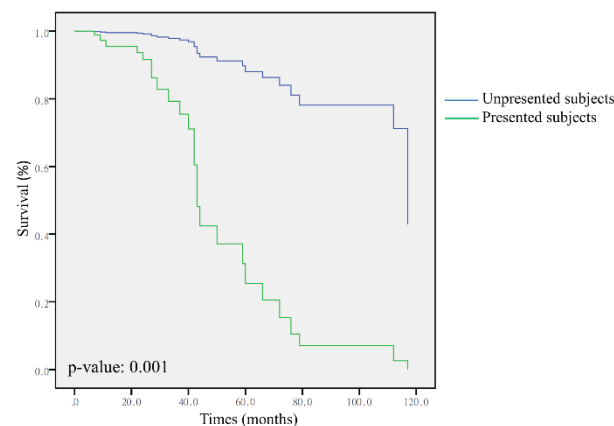


Figure 4. Multivariate survival analysis according to presence or non-presence bacteria in patients with lung cancer.

B. diminuta presence was associated with CD4+ T-cells and regulatory T-cells (Tregs) enrichment (Table S5). *A. radioresistens* presence was associated with relatively low level of activated DC (aDC), CD4+ effector memory T-cells (CD4+ Tem), CD8+ central memory T-cells (CD8+ Tcm), macrophages, macrophages M2, memory B-cells and plasma cells, as well as a low ImmuneScore and MicroenvironmentScore (Table S5). *E. cloacae* presence was linked to relatively high levels of macrophages M1 and Tregs, and relatively low levels of CD4+ naive T-cells (Table S5). Presence with *M. chelonae* and *M. franklinii* was associated with relatively high levels of CD4+ T-cells (Table S5). *Staphylococcus sp.* presence

correlated with relatively high levels of basophils, eosinophils, naive B-cells, natural killer (NK) cells, and pro B-cells (Table S5).

B. megaterium presence was associated with relatively high levels of type 2 T-helper (Th2) cells and relatively low levels of aDC, natural killer T (NKT) cells and plasma cells (Table S5). *R. erythropolis* presence was linked to relatively high levels of CD4+ T-cells (Table S5). *P. aeruginosa* and HPV type 16 presences were not significantly associated with the enrichment of any specific tumor immune cells.

We further analyzed whether the positive effect of presence of bacterial on the survival is related to their positive effect on levels of tumor immune cells or it is an independent effect. The *B. diminuta*, *M. chelonae* and *M. franklinii* maintained the significance after adjustment for tumor immune cells ($p = 0.031$, 0.006 and 0.031 , respectively) (Table S6). After adjustment, the *A. radioresistens*, *E. cloacae*, *Staphylococcus sp.*, *B. megaterium* and *R. erythropolis* showed no significance of the relationship.

3.4. TCR and BCR Repertoires in NSCLC

We systematically analyzed TCR and BCR repertoires in RNA-seq data from 48 lung cancer tissues. One case has not passed the TRUST4 criteria. Detailed information about the TCR (α , β , γ and δ) and BCR (IgL, IgK and IgH) repertoires is included in Table S7. Multivariate analysis, which included tumor stage, grade, histologic cell type and gender as independent prognostic factors, revealed that the clone numbers of TCR α , significantly correlated with OS ($p = 0.044$) (Table 1). For all TCR and BCR repertoires, the ImmuneScore was significantly higher in groups with higher clone numbers (Figure 5A–G).

Table 1. Multivariate analysis (Cox regression) of independent prognostic factors in patients with lung cancer.

Variables		Hazard Ratio	95% CI	p-Value
TCR α clone number	$\leq$ Median	1	0.101–0.972	0.044
	>Median	0.314		
Stage	I	1	0.242–2.946	0.023
	II	0.844		0.791
	III	2.471		0.204
	IV	59.625		0.003
Grade	1	1	0.044–5.930	0.523
	2	0.512		0.592
	3	0.334		0.429
	4	3.269		0.568
Gender	M	1	0.578–6.2	0.289
	F	1.907		
Histologic cell type	ADC	1	8.743–513.175	0.000
	SqCC	66.984		

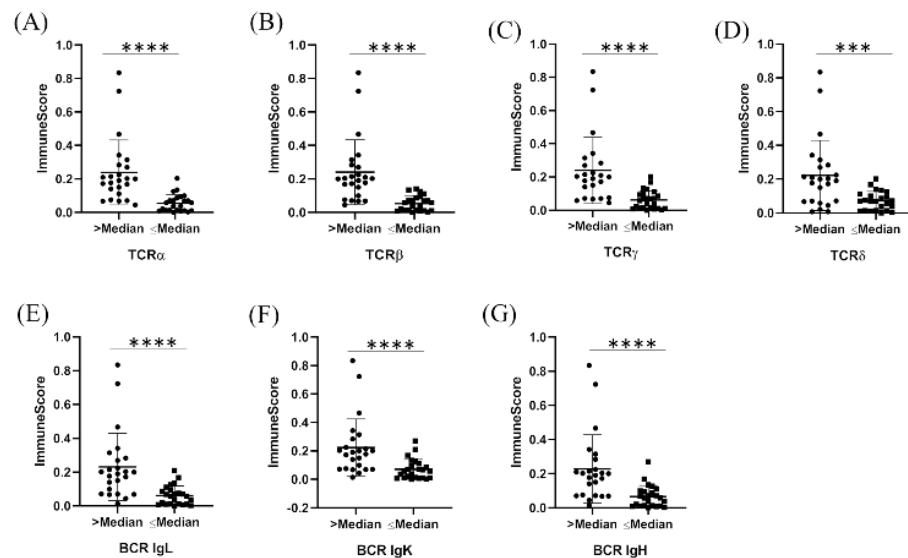


Figure 5. Correlation between clone number of immune receptor repertoires and the Immunoscore, calculated with xCell. (A) TCR α , (B) TCR β , (C) TCR γ , (D) TCR δ , (E) BCR IgL, (F) BCR IgK, (G) BCR IgH. *** $p < 0.001$, **** $p < 0.0001$.

3.5. Compositional Differences between TCR and BCR Repertoires

Next, we analyzed the relationships of bacterial presence with the number of unique immune receptor clones and clonal diversity in 48 subjects. We found that patients with bacterial presences had fewer unique TCR and BCR clones compared to unrepresented subjects (Figure 6A). In addition, the clonal diversity of T and B cells was significantly higher in unrepresented subjects (Figure 6B). However, we did not find any significant associations between HPV type 16 presence and immune receptor repertoires.

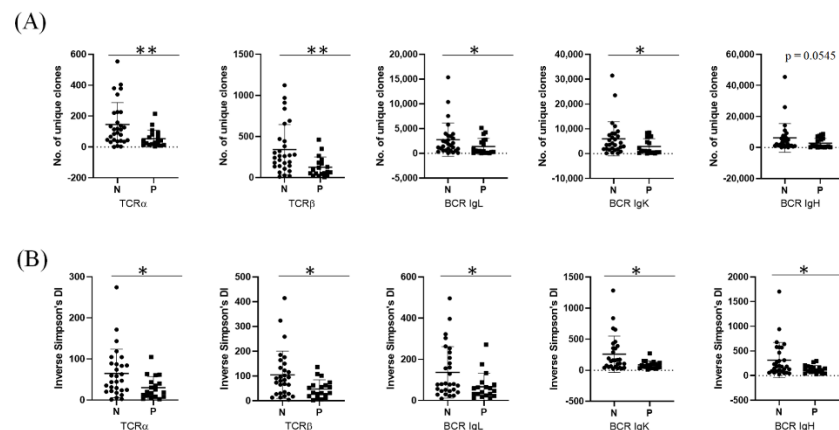


Figure 6. Comparison of the immune receptor repertoires between presented and unrepresented subjects. (A) Number of unique clone and (B) clonal diversity. * $p < 0.05$, ** $p < 0.01$.

3.6. Associations between the Metatranscriptome and Genetic Alterations

The most prevalent mutations in lung tumors are found in the epidermal growth factor receptor (*EGFR*), *KRAS*, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit α and *TP53* genes. Therefore, we investigated the association between the lung metatranscriptome and host genetic mutations. We found a higher rate of mutation in the *EGFR* gene among patients not presented with *P. aeruginosa*, exhibiting borderline statistical significance ($p = 0.050$) (Table S8). However, we found no other associations between bacterial species and gene mutations.

4. Discussion

This is the first study to employ RNA-seq to investigate the metatranscriptome of human lung cancer. *B. diminuta*, *A. radioresistens*, *E. cloacae*, *M. chelonae*, *M. franklinii*, *Staphylococcus* sp., *B. megaterium*, *P. aeruginosa*, and *R. erythropolis* were enriched in lung cancer tissues and significantly associated with prognosis. Most of the bacterial species identified in this study (except for *M. franklinii* and *B. megaterium*) were consistent with those detected in a recent study, in which 16S rRNA sequencing was applied to identify species present in lung tissues [10]. Using xCell, we estimated the relative proportions of immune cells, and investigated their relationships with the metatranscriptome. We also investigated immune receptor repertoires, and their associations with species identified in the metatranscriptome. Moreover, we found that *EGFR* mutations may protect against *P. aeruginosa* presences.

Bacillus sp. produce anti-cancer and anti-proliferative biomolecules [39]. *B. megaterium* forms large spores, which distinguish it from other *Bacillus* sp. Moreover, *B. megaterium* strain SAmt17 produces extracellular polymeric substances that suppress the expression of hepatocyte carcinoma G2 cells, and strain ATCC 13368 produces four cytotoxic compounds suppressing human melanoma cells. Infections caused by *B. megaterium* are rare; only five cases (eye, skin, brain, pleuritis with pleural effusion [40], and soft tissue infections [41]) have been described in the literature so far. In the present study, we detected *B. megaterium* in two lung tumor tissue samples. Eight normal tissue samples were not detected. Our results suggest that *B. megaterium* presence may play a role in lung cancer carcinogenesis. The tumor microenvironments of *B. megaterium*-positive patients tended to have high levels Th2 cells and low levels of aDC, NKT and plasma cells.

The clinically isolated human pathogen *M. franklinii* is a mycobacterial species and member of the *M. chelonae*-*M. abscessus* complex. Clinical laboratories typically diagnose *M. franklinii* infections by partially sequencing *rpoB*, *hsp65*, *sodA* and internal transcribed spacer DNA targets, or by complete 16S rRNA gene sequencing analysis in conjunction with assessment of cefoxitin and minocycline susceptibility [42]. Most *M. franklinii* isolates have been collected from the respiratory specimens of patients with underlying pulmonary diseases. In the present study, we detected *M. franklinii* in five lung cancer tissue samples. Eight normal tissue samples were not detected, which may play a role in lung carcinogenesis. Furthermore, high levels of CD4⁺ T-cells in the tumor microenvironment were associated with *M. franklinii* presences.

The immune responses of cancer patients influence prognosis and survival. In this study, the presence of certain bacteria of lung cancer tissues correlated with poor prognoses and short survival times. Furthermore, the immune responses of presented patients were disrupted, reflected in higher proportions of immune-suppressing cells and decreased TCR and BCR diversity, which may have increased the likelihood of bacterial presence. Our results suggest that presences in cancer tissues likely occurred due to the weak immune systems of the patients, which increased their susceptibility to bacterial presence.

A limitation of this study was the small number of cases, which resulted in a lack of significance for some parameters, and could have led to biased and unclear conclusions. Furthermore, the depth of the RNA-seq metatranscriptomics data was low, which may have resulted in false-negative results. There were also some limitations to the analysis tools applied in this study. However, to overcome these, we performed multiple analyses using different tools to ensure the reliability of the results.

5. Conclusions

In this study, we employed RNA-seq to investigate the metatranscriptome of lung cancer patients. Nine bacteria were significantly associated with reduced OS. The presence of two bacterial species, *B. megaterium* and *M. franklinii*, may play an important role in lung tumor carcinogenesis. We also found a correlation between metatranscriptomic

changes in expression profiles and tumor immune cell enrichment, as well as immune receptor repertoires.

Supplementary Materials: The following are available online at www.mdpi.com/article/10.3390/genes12091458/s1. Table S1: Log-rank test of bacterial of lung microbiome revealed by both the SAMSA2 and Kraken2 tools, the bacterium was detected by SAMSA2 and 16S rRNA sequencing, or the bacterium was detected by Kraken2 and 16S rRNA sequencing; Table S2: Log-rank test of bacterial of lung microbiome revealed by SAMSA2 and Kraken2 analyses of RNA-seq and 16S rRNA sequencing; Table S3: Hazard ratio of tumor immune cells for the lung cancer; Table S4: Compositional differences in tumor immune cells between patients with one of nine bacteria presence and non-bacteria presence groups; Table S5: Compositional differences in tumor immune cells between patients with bacteria presence and non-bacteria presence groups; Table S6: Hazard ratio of presence of bacterial species for the lung cancer; Table S7: Statistical characteristics of the immune receptor repertoires data in 48 lung cancer patients; Table S8: Correlation of *EGFR*, *PIK3CA*, *KRAS*, and *TP53* mutations with microbial presence.

Author Contributions: Conceptualization, J.-G.C.; data curation, M.-H.H., S.-J.T., J.-C.Y. and Y.-T.L.; formal analysis, Y.-S.C.; funding acquisition, J.-G.C.; investigation, Y.-S.C.; methodology, Y.-S.C., M.-H.H., S.-J.T., J.-C.Y. and Y.-T.L.; software, M.-H.H. and S.-J.T.; supervision, J.-G.C.; visualization, H.-Y.F.; writing—review and editing, H.-Y.F. and J.-G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China Medical University Hospital, grant number DMR-104-076.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of China Medical University Hospital (CMUH106-REC1-053).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The RNA-Seq data from this study was submitted to the NCBI Sequence read Archive (SRA) under BioProject accession nos. PRJNA698419.

Acknowledgments: The data analyzed in this study were partially provided by the China Medical University Hospital Cancer Registry.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2015**, *136*, E359–E386, doi:10.1002/ijc.29210.
2. Kovaleva, O.V.; Romashin, D.; Zborovskaya, I.B.; Davydov, M.M.; Shogenov, M.S.; Gratchev, A. Human Lung Microbiome on the Way to Cancer. *J. Immunol. Res.* **2019**, *2019*, 1–6, doi:10.1155/2019/1394191.
3. Sugimura, H.; Yang, P. Long-term survivorship in lung cancer: A review. *Chest* **2006**, *129*, 1088–1097.
4. Wang, H.; Funchain, P.; Bebek, G.; Altemus, J.; Zhang, H.; Niazi, F.; Peterson, C.; Lee, W.T.; Burkey, B.B.; Eng, C. Microbiomic differences in tumor and paired-normal tissue in head and neck squamous cell carcinomas. *Genome Med.* **2017**, *9*, 1–10, doi:10.1186/s13073-017-0405-5.
5. Greathouse, K.L.; White, J.R.; Vargas, A.; Bliskovsky, V.V.; Beck, J.A.; Von Muhlinen, N.; Polley, E.C.; Bowman, E.D.; Khan, M.A.; Robles, A.; et al. Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol.* **2018**, *19*, 1–16, doi:10.1186/s13059-018-1501-6.
6. Huang, J.Q.; Zheng, G.F.; Sumanac, K.; Irvine, E.; Hunt, R.H. Meta-analysis of the relationship between cagA seropositivity and gastric cancer. *Gastroenterology* **2003**, *125*, 1636–1644, doi:10.1053/j.gastro.2003.08.033.
7. Wang, C.; Li, J. Pathogenic Microorganisms and Pancreatic Cancer. *Gastrointest Tumors* **2015**, *2*, 41–47.
8. Kostic, A.D.; Chun, E.; Robertson, L.; Glickman, J.N.; Gallini, C.A.; Michaud, M.; Clancy, T.E.; Chung, D.C.; Lochhead, P.; Hold, G.L.; et al. *Fusobacterium nucleatum* Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host Microbe* **2013**, *14*, 207–215, doi:10.1016/j.chom.2013.07.007.
9. Feng, Y.; Ramnarine, V.R.; Bell, R.; Volik, S.; Davicioni, E.; Hayes, V.M.; Ren, S.; Collins, C.C. Metagenomic and meta-transcriptomic analysis of human prostate microbiota from patients with prostate cancer. *BMC Genom.* **2019**, *20*, 1–8, doi:10.1186/s12864-019-5457-z.

10. Nejman, D.; Livyatan, I.; Fuks, G.; Gavert, N.; Zwang, Y.; Geller, L.T.; Rotter-Maskowitz, A.; Weiser, R.; Mallel, G.; Gigi, E.; et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* **2020**, *368*, 973–980, doi:10.1126/science.aay9189.
11. Poore, G.D.; Kopylova, E.; Zhu, Q.; Carpenter, C.; Fraraccio, S.; Wandro, S.; Kosciolk, T.; Janssen, S.; Metcalf, J.; Song, S.J.; et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* **2020**, *579*, 567–574, doi:10.1038/s41586-020-2095-1.
12. Yu, G.; Gail, M.H.; Consonni, D.; Carugno, M.; Humphrys, M.; Pesatori, A.C.; Caporaso, N.E.; Goedert, J.J.; Ravel, J.; Landi, M.T. Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome Biol.* **2016**, *17*, 1–12, doi:10.1186/s13059-016-1021-1.
13. Apopa, P.L.; Alley, L.; Penney, R.B.; Arnaoutakis, K.; Steliga, M.A.; Jeffus, S.; Bircan, E.; Gopalan, B.; Jin, J.; Patumcharoenpol, P.; et al. PARP1 Is Up-Regulated in Non-small Cell Lung Cancer Tissues in the Presence of the Cyanobacterial Toxin Microcystin. *Front. Microbiol.* **2018**, *9*, 1757, doi:10.3389/fmicb.2018.01757.
14. Liu, Y.; O'Brien, J.L.; Ajami, N.J.; Scheurer, M.E.; Amirian, E.S.; Armstrong, G.; Tsavachidis, S.; Thrift, A.P.; Jiao, L.; Wong, M.C.; et al. Lung tissue microbial profile in lung cancer is distinct from emphysema. *Am. J. Cancer Res.* **2018**, *8*, 1775–1787.
15. Peters, B.A.; Hayes, R.B.; Goparaju, C.; Reid, C.; Pass, H.I.; Ahn, J. The microbiome in lung cancer tissue and recurrence-free survival. *Cancer Epidemiol. Biomark. Prev.* **2019**, *28*, 731–740, doi:10.1158/1055-9965.epi-18-0966.
16. Bingula, R.; Filaire, E.; Molnar, I.; Delmas, E.; Berthon, J.-Y.; Vasson, M.-P.; Bernalier-Donadille, A.; Filaire, M. Characterisation of microbiota in saliva, bronchoalveolar lavage fluid, non-malignant, peritumoural and tumour tissue in non-small cell lung cancer patients: A cross-sectional clinical trial. *Respir. Res.* **2020**, *21*, 1–20, doi:10.1186/s12931-020-01392-2.
17. Shakya, M.; Lo, C.C.; Chain, P.S. Advances and Challenges in Metatranscriptomic Analysis. *Front. Genet.* **2019**, *10*, 904, doi:10.3389/fgene.2019.00904.
18. Pérez-Losada, M.; Castro-Nallar, E.; Bendall, M.; Freishtat, R.; Crandall, K.A. Dual Transcriptomic Profiling of Host and Microbiota during Health and Disease in Pediatric Asthma. *PLoS ONE* **2015**, *10*, e0131819, doi:10.1371/journal.pone.0131819.
19. Nowicki, E.M.; Shroff, R.; Singleton, J.A.; Renaud, D.E.; Wallace, D.; Drury, J.; Zirnheld, J.; Colleti, B.; Ellington, A.D.; Lamont, R.J.; et al. Microbiota and Metatranscriptome Changes Accompanying the Onset of Gingivitis. *mBio* **2018**, *9*, e00575-18, doi:10.1128/mbio.00575-18.
20. Schirmer, M.; Franzosa, E.A.; Lloyd-Price, J.; McIver, L.J.; Schwager, R.; Poon, T.W.; Ananthakrishnan, A.N.; Andrews, E.; Barron, G.; Lake, K.; et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.* **2018**, *3*, 337–346, doi:10.1038/s41564-017-0089-z.
21. Chang, Y.-S.; Tu, S.-J.; Chiang, H.-S.; Yen, J.-C.; Lee, Y.-T.; Fang, H.-Y.; Chang, J.-G. Genome-Wide Analysis of Prognostic Alternative Splicing Signature and Splicing Factors in Lung Adenocarcinoma. *Genes* **2020**, *11*, 1300, doi:10.3390/genes11111300.
22. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120, doi:10.1093/bioinformatics/btu170.
23. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915.
24. Kovaka, S.; Zimin, A.V.; Pertea, G.M.; Razaghi, R.; Salzberg, S.L.; Pertea, M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **2019**, *20*, 1–13, doi:10.1186/s13059-019-1910-1.
25. Aran, D.; Hu, Z.; Butte, A.J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **2017**, *18*, 220, doi:10.1186/s13059-017-1349-1.
26. Song, L.; Cohen, D.; Ouyang, Z.; Cao, Y.; Hu, X.; Liu, X.S. TRUST4: Immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **2021**, *18*, 627–630, doi:10.1038/s41592-021-01142-2.
27. Lefranc, M.P. IMGT, the International ImmunoGeneTics Information System. *Cold Spring Harb. Protoc.* **2011**, *2011*, 595–603.
28. Shugay, M.; Bagaev, D.V.; Turchaninova, M.; Bolotin, D.; Britanova, O.V.; Putintseva, E.; Pogorelyy, M.; Nazarov, V.I.; Zvyagin, I.V.; Kirgizova, V.I.; et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **2015**, *11*, e1004503, doi:10.1371/journal.pcbi.1004503.
29. Westreich, S.T.; Korf, I.; Mills, D.A.; Lemay, D.G. SAMSA: A comprehensive metatranscriptome analysis pipeline. *BMC Bioinform.* **2016**, *17*, 399, doi:10.1186/s12859-016-1270-8.
30. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 1–13, doi:10.1186/s13059-019-1891-0.
31. Zhang, J.; Kobert, K.; Flouri, T.; Stamatakis, A. PEAR: A fast and accurate Illumina Paired-End read mergeR. *Bioinformatics* **2014**, *30*, 614–620, doi:10.1093/bioinformatics/btt593.
32. Kopylova, E.; Noé, L.; Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012**, *28*, 3211–3217, doi:10.1093/bioinformatics/bts611.
33. Coordinators, N.R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44*, D7–D19, doi:10.1093/nar/gkv1290.
34. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745, doi:10.1093/nar/gkv1189.

35. Weng, S.-L.; Chiu, C.-M.; Lin, F.-M.; Huang, W.-C.; Liang, C.; Yang, T.; Yang, T.-L.; Liu, C.-Y.; Wu, W.-Y.; Chang, Y.-A.; et al. Bacterial Communities in Semen from Men of Infertile Couples: Metagenomic Sequencing Reveals Relationships of Seminal Microbiota to Semen Quality. *PLoS ONE* **2014**, *9*, e110152, doi:10.1371/journal.pone.0110152.
36. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072, doi:10.1128/aem.03006-05.
37. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272, doi:10.1038/s41592-019-0686-2.
38. Davidson-Pilon, C.; Kalderstad, J.; Zivich, P.; Kuhn, B.; Williamson, M.; Fiore-Gartland, A.; Moneda, L.; Willson, D.; Parij, A.; Stark, K.; et al. *CamDavidsonPilon/Lifelines: v0.22.3 (Late)*; Zenodo: Genève, Switzerland, 2019.
39. Ferdous, U.T.; Shishir, A.; Khan, S.N.; Hoq, M. *Bacillus* spp.: Attractive Sources of Anti-cancer and Anti-proliferative Biomolecules. *Microb. Bioact.* **2018**, *1*, E033–E045, doi:10.25163/microbbioacts.11005b0408130818.
40. Crisafulli, E.; Aredano, I.; Valzano, I.; Burgazzi, B.; Andrani, F.; Chetta, A. Pleuritis with pleural effusion due to a *Bacillus megaterium* infection. *Respirol. Case Rep.* **2019**, *7*, e00381.
41. Bocchi, M.B.; Cianni, L.; Perna, A.; Raffaele, V.; Greco, T.; Maccauro, G.; Perisano, C. A rare case of *Bacillus megaterium* soft tissues infection. *Acta Biomed.* **2020**, *91*, e2020013.
42. Simmon, K.E. *Mycobacterium chelonae*-abscessus Complex Associated with Sinopulmonary Disease, Northeastern USA. *Emerg. Infect. Dis.* **2011**, *17*, 1692–1700, doi:10.3201/eid1709.101667.