ISSN: 0067-2904

Exploring the Potency of Lower Respiratory Tract Microbiome as Biomarkers for Lung Cancer

Sama Kamal Rasheed*, Ahmed Salim Kadhim Al-Khafaji
Department Biology, College of Science, University of Baghdad, Baghdad, Iraq

Received: 10/5/2023 Accepted: 19/9/2023 Published: 30/11/2024

Abstract
Lung cancer is the first killer among all the known malignancies. Late detection of this malignancy is a major contributor to advanced-stage diagnosis and poor outcomes. Pulmonary microbiota has been recently reported as one of lung cancer risk hallmarks that is still to be fully understood. This study aimed to explore the potential role of microbiome in predicting lung carcinogenesis. Microbiome Library Construction of the *16SrRNA* variable region (V3–V4) was conducted by gene amplicon sequencing using the Illumina sequencing platform. Microbiome data was analysed using Version QIIME2-202006 software for species annotations. The sequences were denoised by the DADA2 plugin implemented in QIIMETM2. The bacterial amplicon sequence variants (ASVs) were then identified. The results of *16SrRNA* sequencing and gene library bioinformatic analysis indicated that *Haemophilus, Prevotella, and Streptococcus* were on the top of abundant genera. The species *Haemophilus influenza* has also been identified among the top 12 bacterial species in malignant and non-malignant lung samples. Microbiome-based identification of the human airway microbiota may provide effective predictive biomarkers for lung carcinogenesis.

Keywords: *16SrRNA*; Microbiome; Microbiota; Lung cancer; Respiratory tract

استكشاف فاعلية مايكروبيوم الجهاز التنفسي السفلي كمؤشرات حيوية لسرطان الرئة

سما كمال رشيد* , احمد سالم كاظم الخفاجي
قسم علوم الحياة , كلية العلوم , جامعة بغداد , بغداد , العراق

الخلاصة
سرطان الرئة هو القاتل الأول بين جميع الأورام الخبيثة المعروفة. يعد الاكتشاف المتأخر لهذا الورم الخبيث عاملاً رئيسياً في التشخيص في المرحلة المتقدمة والنتائج السيئة له. اشارت التقارير مؤخرًا بأن الميكروبات الرئوية قد تمثل واحدة من عوامل خطورة الإصابة بسرطان الرئة، ولكن لا يزال يتعين فهمها بالكامل. تهدف هذه الدراسة إلى استكشاف الدور المحتمل للميكروبيوم في التنبؤ بتسرطن الرئة. تم إنشاء مكتبة Microbiome للمنطقة المتغايرة (V3–V4) للجين *16SrRNA* عن طريق تسلسل amplicon الجين باستخدام منصة تسلسل Illumina. تم تحليل بيانات الـ Microbiome باستخدام برنامج QIIME2– 202006 للتعليقات التوضيحية للأنواع. تم إزالة التشويش على التسلسلات الجينية بواسطة برنامج DADA2 المنفذ في QIIMETM ثم تم تحديد متغيرات تسلسل Amplicon البكتيريا (ASVs). أشارت نتائج تسلسل

*Email: sama.kamal9977@gmail.com

الجين 16SrRNA وتحليل المعلومات الحيوية لمكتبة الجينات إلى أن الاجناس البكتيرية *Haemophilus* و *Prevotella*و *Streptococcus* كانت على رأس الأجناس ذات الوفرة. كما تم تحديد النوع *Haemophilus influenza* من بين ألاكثر 12 نوعً بكتيري في عينات الرئة الخبيثة وغير الخبيثة. قد يوفر التحديد المستند إلى الميكروبيوم للميكروبات في مجرى الهواء البشري مؤشرات حيوية تنبؤية فعالة لتسرطن الرئة.

Introduction

Lung cancer ranks as the third most incident malignancy after breast cancer in females and prostate cancer in males. However, it represents the first cause of cancer related deaths among all the known cancers up to now. Late detection of lung cancer is the main reason for increasing mortalities among lung cancer patients [1]. Having been one of the most prevalent cancers in the world, lung cancer poses a serious threat to human health due to its high rates of morbidity (1.82 million) and mortality (1.59 million) per year [2]. Increasing evidence has revealed a connection between lung cancer and microbial-related pulmonary disorders such as chronic obstructive pulmonary disease (COPD), pneumonia and cystic fibrosis, and lung cancer. Therefore, the pulmonary dysbiosis of certain microorganisms is believed to be linked to lung tumorigenesis [3].

Generally speaking, the microbial community of archaea, bacteria, and eukarya that inhabit particular sites inside the body, mainly the aerodigestive tract, alongside outside the body (i.e., skin) is termed as 'microbiota' [4]. Nevertheless, lungs were believed to be free of microbiota until the last decade. The human microbiota that lives on and in the human body are thought to outnumber human cells 10-fold.

Bacterial microbiota has a significant impact on the etiology of health and disease because of the large bacterial communities found inside the human body and the range of activities they participate in [5]. Both smokers and non-smokers have had bacterial populations found in their broncho-alveolar lavage samples [6], pleural fluid [7], and lung tissues, where lung-based microorganisms contributed to the etiology of non-malignant respiratory disorders [8,9]. Although recent studies have identified the lung cancer microbiome in samples collected from respiratory tract fluids and tissues, the profile variation of the pulmonary microbiota between normal and malignant lung tissues alongside their role in the prediction of lung carcinogenesis is not fully uncovered [10].

Discovering the pathophysiological mechanisms of pulmonary illness's progression, particularly in patients with chronic obstructive pulmonary disease (COPD), requires data collection about the lung microbiota and their alterations during disease courses [11]. Managing the lung microbiome information may help develop new preventative therapeutic strategies for various malignant pulmonary diseases and for a better understanding of the lung microbiome association between malignant and non-malignant pulmonary diseases [12,13].
Traditional culture techniques no longer represent the standard for microbial queries since it has been predicted that most of the bacterial species of human microbiota cannot be cultivated using the traditional isolation methods [14,15]. The breadth and depth of the microbiota present in both the healthy and diseased lung have been shown by advanced microbiome techniques that identify bacterial DNA sequences, including the 16S ribosomal RNA gene (16S rRNA) [16,17,18].

Several phyla have been previously identified in the microbiome of lung cancer such as *Bacteroidetes*, *Proteobacteria*, and *Firmicutes*, in addition to several bacterial genera

Prevotella, Neisseria, and *Streptococcus* [19]. Moreover, it has been recently reported that particular genera are dominated in malignant lung tissues with advanced metastases stage, which may suggest their role in cancer progression [20]. On the other hand, the domination of the families *Lachnospiraceae, Ruminococcaceae,* and *Bacteroidaceae* may be associate with normal lung tissues, linking the normal microbiota with good prognosis of lung cancer patients [21].

Predicting molecular biomarkers has recently been explored in respiratory tract cancers [22,23]and other malignant diseases [24]. Despite recent increases in extensive research focusing on pulmonary microbiome, the key role behind involving lung microbiota as one of the lung cancer risk hallmarks remains to be fully known. We, therefore, hypothesized that the respiratory microbiota can possess tumorigenic activity through dysregulation or even modulating oncogenic genes and thus they are important to unravel the potential role of the microbiome in lung cancer [25]. Dysbiosis of microbiota could be a determinator of early pathohistological events in lung tissue and thus microbiome could act as a potential biomarker for predicting lung carcinogenesis [26].

We, therefore, aimed to explore the comparison between microbiomes in samples of lung cancer patients and healthy controls, and how the identified microbiota could serve as predictive biomarkers for lung carcinoma.

Materials and Methods

2.1. Patients and Sample Collection

A total of 151 clinical samples were collected from patients admitted to Ghazi al-Hariri Surgical Specialities Hospital and individuals attending Tuberculosis Institute in Baghdad during the period from September 2021 to February 2022. The samples were gathered based on various inclusion criteria; such as age, sample type, smoking status, and health status of recruited individuals. The current study was ethically approved by the College of Science Research Ethics Committee at the University of Baghdad under the reference number "CSEC/0122/0015". All samples were anonymously gathered from people who provided their informed consent.

2.2 Extraction of Genomic DNA

The QIAamp DNA Microbiome Kit (cat. no. 51704 - QIAGEN) was utilized to extract the genomic DNA of microbiota from the collected samples The Qubit® dsDNA Assay Kit in Qubit® 2.0 Flurometer (cat. no. Q32851 -Life Technologies, CA, USA) and NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc., USA) were used to measure the DNA integrity, concentration and purity. Quality control of DNA extracts was further analysed on 1% agarose gel electrophoresis at 100 voltages run for 40 minutes. The extracted DNA was diluted to 1 ng/uL with deionized distilled water to achieve the concentration needed to amplify the variable region (V3–V4) of *the 16S rRNA* gene.

2.3. 16S rRNA Gene Amplicon Sequencing

Polymerase chain reaction (PCR) amplification was conducted using particular primers tagged with sample-specified barcodes and targeting the V3-V4 region of the bacterial *16S rRNA* gene with modified Linker Primer Sequences (forward: 5'-CCTAYGGGRBGCASCAG -3'; reverse: 5'-GGACTACNNGGGTATCTAAT -3') [27].

2.4. Library Construction, Quality Control and Sequencing

The PCR products with the proper sizes between 400 and 450 bp were selected on 2% agarose

gel electrophoresis at 80 voltages run for 40 minutes and purified with a Qiagen Gel Extraction Kit (cat. no. 28706X4 -Qiagen).

The same amount of PCR products from each sample were pooled, end-repaired, A-tailed, and further ligated with Illumina adapters. Microbiome libraries were sequenced and created on a paired-end Illumina platform to generate 250bp paired-end raw reads.

2.5. Microbiome Data Analysis

Microbiome sequences obtained from the examined samples were analyzed using the QIIME™ (Quantitative Insights into Microbial Ecology) software (Version QIIME2-202006) for species annotation. The sequences were denoised using the Divisive Amplicon Denoising Algorithm 2 (DADA2) plugin implemented in QIIME^{TM}2. The bacterial amplicon sequence variants (ASVs) were then identified.

Alpha diversity was assessed by Chao1, the Shannon index, and phylogenetic diversity [28]. Beta diversity was evaluated by weighted and unweighted UniFrac distances [29]. The differences in microbiota content between normal and cancerous samples were investigated by principal coordinates analysis (PCoA) and represented as box-and-whiskers plots [30]. Linear discriminant analysis (LDA) was employed to identify the differential taxonomy by effect size (LEfSe) [29]. *16S rRNA* abundance-based metagenomic function prediction was carried out using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2) software [31].

2.6. Statistical Analysis

All data was calculated as mean values alongside their standard deviation or 95% confidence intervals. Data categorized according to the sample type (malignant or normal) were compared using a one-way analysis of variance or the χ^2 test (for categorical variables) [32]. Similarity analysis was carried out to identify the differences in the microbiota communities. The Wilcoxon rank-sum test or *t*-test was employed to indicate any significant differences in alpha or beta diversity and the UniFrac dissimilarities [33]. The Kruskal-Wallis rank-sum test was used to distinguish the significant differential abundance among different groups in LEfSe analysis. All significance tests were two-sided, and *p*-values < 0.05 indicated statistical significance [34]. All statistical analysis tests were conducted using SPSS software (Version 27.0; SPSS, USA).

Results and Discussion

3.1. Clinical Distribution of the Collected Samples

All collected clinical samples of sputum and pleural fluid were distributed among 3 groups of recruited individuals. 1st group represented lung cancer patients (n=21; 14%). 2nd group demonstrated cases of lower respiratory tract infections (n=86; 57%). The rest of the samples (n=44; 29%) were collected from apparently healthy individuals.

3.2. 16SrRNA Amplification

The genomic DNA of the microbiota was subjected to purification from the collected samples. Twenty-two out of 30 samples, selected according to specified inclusion criteria, passed the quality control (QC) analysis. For the other 8 failed samples, no band was obtained from the PCR after running the product on the gel (Figure 1). The PCR amplification results of 16SrRNA variable regions V3-V4 (16SV34) showed that one-third of the examined samples (n=10; 33.3%) met the sample quality requirements for library preparation (construction) and sequencing amplification since the amplified microbial DNA required nucleic acid fragments in the target region only (Figure 2).

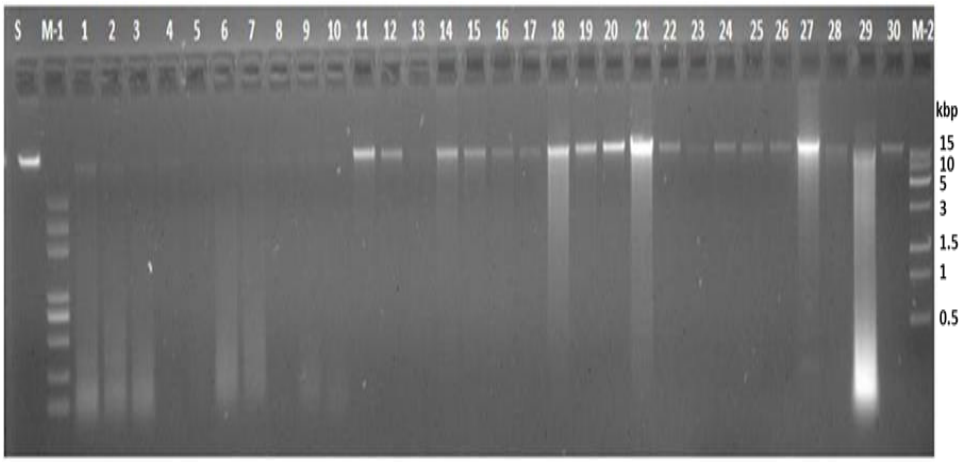


Figure 1: Representative image showing the bands of extracted DNA on 1% agarose gel electrophoresis, where (S) lane represents a standard sample (50ng), (M1 and M2) lanes represent the Trans 15k plus DNA molecular weight marker, which is 2µL loaded premixed containing eight different sizes of linear double-stranded DNA fragments, while (1-30) lanes demonstrate samples ranged in the order (All loaded with 2µL of DNA samples).

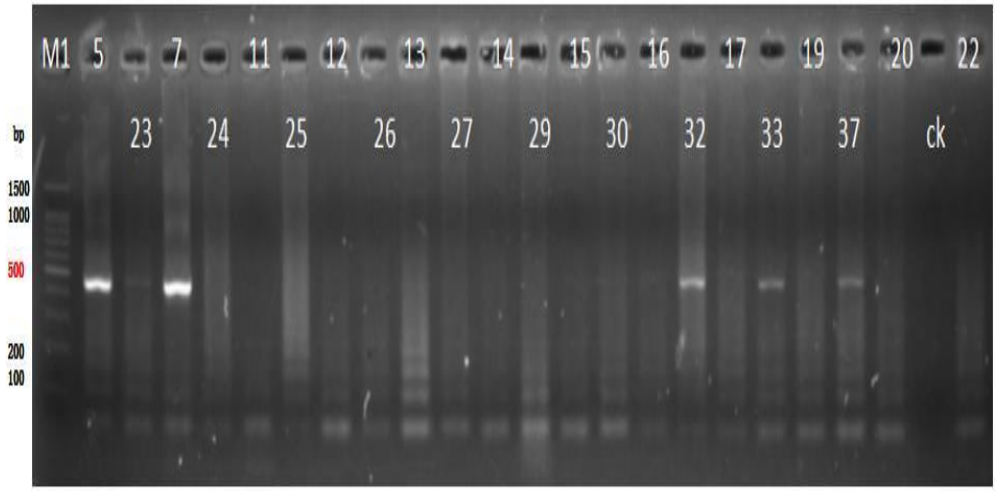


Figure 2: The image depicts PCR products on 2% agarose gel electrophoresis, where (M1) illustrates Trans 100bp DNA ladder loaded as 1µL per well; samples ranged in the order all loaded as 3µL per well.

Based on these results, the passed samples were chosen for DNA library construction of 16SV34 and sequencing analysis afterward. The DNA QC and subsequent 16SV34 amplification are fundamental stages to ensure the accuracy and reliability of sequencing data achieved from the bacterial DNA source. Therefore, the quality data of the expected output could be obtained as high as possible [35].

3.3. OUTs Clustering and ASVs Annotations

To explore the potential changes in the lung microbiome, ten representative samples represented 2 groups: lung cancer patients (n=3) and individuals with and without benign pulmonary diseases (n=7). Both groups were subjected to sequencing analysis according to the demographic data of the recruited individuals. The raw data obtained for the elected samples by sequencing had a total of 512,975 paired-end reads (PE). Around 69.78% of reads out of the PEs were combined to 357,955 resulting in an average of 404 bp merged reads per sample (Table 1).

Table 1: Raw reads and merged tags for each sample after pair-end reads merging by sequencing analysis.

Sample	Raw PE Reads	Combined[1]	Uncombined[2]	Combined (%)	Combined (bp)	Avg._len.(bp)
LP1	45,862	41,394	4,468	90.26	17,489,577	423
LP2	48,709	23,333	25,376	47.9	8,132,394	349
LP3	51,282	31,508	19,774	61.44	10,944,468	347
P1	47,304	34,225	13,079	72.35	14,080,811	411
P2	49,641	37,704	11,937	75.95	15,885,747	421
P3	42,270	29,891	12,379	70.71	12,113,801	405
S1	100,494	58,542	41,952	58.25	23,402,899	400
S2	42,296	34,362	7,934	81.24	14,569,207	424
S3	44,349	31,901	12,448	71.93	13,159,768	413
S4	40,768	35,095	5,673	86.08	14,884,207	424
Total:	**512,975**	**357,955**	**155,020**	**69.78**	**144,662,879**	**404**

[1] (The combined tags sequences achieved by trimming), [2] the leftover uncombined sequences.

Usually, the raw data gained by sequencing has a particular rate of dirty data. To produce precise and trustworthy outcomes for bioinformatic analysis, splicing and filtering of raw data were firstly conducted to achieve clean data using DADA2 implemented in QIIME^{TM}2 to lessen noise [36]. Initially, the reads that excluded bacterial primer sequences or that included low-quality primer sequences were removed [37]. ASVs feature sequence table corresponding to the representative sequence of Operational Taxonomic Units (OTUs) for predicting functional profiles was conducted using DADA2 to denoise the sequences (Table 2), where the ASVs were identified, and any sequences with abundance < 5 were filtered out to achieve the final ASVs [38].

Out of the total combined (n=357,955), qualified PE reads (n=353,791) were achieved after filtration of the low-quality reads (n=4,164), followed by removing the barcodes, adaptors, primers, and chimeras. The latter represented artifact sequences produced by incorrectly joined sequences that happened during the PCR process utilizing a mixture of templates. Effective tags, after excluding the chimera sequences known as 'Nochime', were obtained (n=317,915) for the next steps of bioinformatic analysis at an average length of (400 nt) with 52.5% GC ratio and 97.93% quality; which was statistically valid for OUT clustering (Table 2).

Table 2 : Statistical outcomes of processed data achieved after splicing and filtering the low-quality reads bioinformatically.

Sample	Qualified[1]	Nochime[2]	Base (nt)[3]	Avg. Len.(nt)[4]	Qual%[5]	GC%[6]
LP1	40,912	31,870	13,458,594	422	98.11	52.23
LP2	22,913	21,509	7,391,483	344	97.77	52.97
LP3	31,119	30,096	10,362,668	344	98.09	50.43
P1	33,873	32,606	13,400,095	411	98.2	52.78
P2	37,296	35,873	15,115,887	421	97.57	52.97
P3	29,686	26,058	10,537,997	404	97.84	54.82
S1	57,795	45,017	17,788,804	395	98.22	53.43
S2	33,918	32,601	13,819,717	424	97.46	51.48
S3	31,560	29,381	12,100,721	412	98.05	52.41
S4	34,719	32,904	13,954,697	424	97.99	51.9
Total	**353,791**	**317,915**	**127,930,663**	**400.1**	**97.93**	**52.542**

[1] the qualified sequences after Raw Tag sequences carried out filtering short length sequences with low quality; [2] the Tags filtrating the chimera sequences (i.e. Effective Tags ultimately utilized for subsequent bioinformatics analysis; [3] the number of DNA base pairs in the ultimate Effective Tags; [4] the length average of the Effective Tags; [5]the nucleotide sequencing proportion that has quality values in Effective Tags more than 97%; [6]GC (%) the ratio of GC content in Effective Tags.

3.4. Distance Variation of Microbial Community

Based on the OUTs clustering and the findings of ASVs annotations characteristics of each examined sample, the relative bacterial abundance was obtained at 7 taxonomic levels representing the kingdom, phyla, class, order, family, genus, and species. Principal components analysis (PCA) was conducted to estimate the distance variation in the structure of the microbial community. PCA is a statistical procedure to extract structures in data by orthogonal transformation and reducing dimensionalities of data [39]. It extracts the first two axes reflecting the variety of samples to the most extent and thus can reflect high-dimensional data's variation in a two-dimensional graph, which reveals the simple principle embedding in complex data. The more similar composition of the community among the samples is the closest distance of their corresponding data points on the PCA graph. As shown in Figure 3, the proportion of the 1st principal component (PC1) is 36.45% ; which spanned the most variation in distance differences in the structure of the studied microbiota community. The 2nd principal component plotted spanned the second most variation accounting for 19.14%. The PCA results exhibited that OTUs clusters were distributed among all the plotted quadrants. Surprisingly, the OTUs clustering distance of the pleural samples was the closest, followed by OTUs of lung cancer samples. All examined samples were overlapped in two of the four quadrants. However, sputum samples showed higher divergence in comparison to other malignant and non-malignant lung samples. These results suggested that the subsets of malignant and non-malignant lung samples shared OTUs components at a closer distance than that of normal sputum samples.

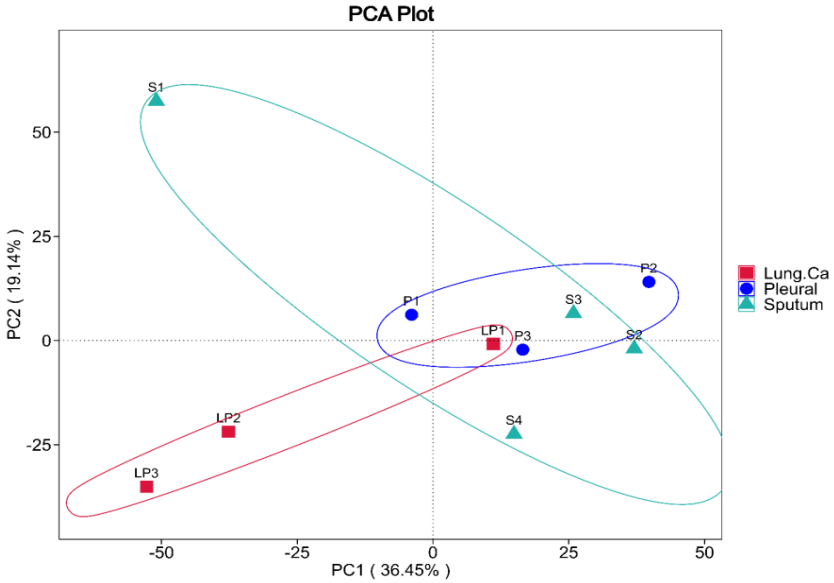


Figure 3: PCA diagram representing the principal components (PCs) samples data of each related group. The first principal component (PC1= 36.45%) plotted on the X-axis spans the most variation, while the second principal component (PC2=19.14%) plotted on the Y-axis spans the second most variation. PCA plot elucidates that differences in the microbial community of the lung cancer samples and non-malignant pleural ones are less variable when comparing with sputum ones.

3.5. Microbiome Taxonomy

A histogram of investigated phyla was depicted based on their abundance in each examined sample. Bacterial phyla abound in all the tested samples Actinobacteriota, Bacteroidota, Cyanobacteria, Firmicutes, Fusobacteriota, and Proteobacteria (Figure 4). Recent studies have reported Proteobacteria and Firmicutes abundance in lung cancer [40]. It is worth mentioning that an abundance of Cyanobacteria has been reported in the blood microbiome of breast cancer patients [41]. However, detecting chloroplast in the tested samples represented a challenge as its microbiome data and even mitochondrial ones tended to be filtered in a recent relative study [42]. Therefore, discussing the related microbiome data was overlooked in the current study. The bacterial genera were explored using the aligned representative sequences. The findings further demonstrated that the explored phyla had included 31 genera as follows; Actinobacteriota (*Atopobium, Actinomyces, Rothia*), Bacteroidota (*Muribaculaceae, Bacteroides,Rikenellaceae, Capnocytophaga, Porphyromonas, Prevotella, Alloprevotella*), Proteobacteria (*Acinetobacter, Pantoea, Neisseria, Pseudomonas, Haemophilus, Achromobacter, Stenotrophomonas, Escherichia/Shigella*), Firmicutes (*Lactobacillus, Blautia, Faecalibacterium, Monoglobus, Allobaculum, Oribacterium, Solobacterium, Streptococcus, Veillonella, Gemella*) and Fusobacteriota (*Fusobacterium, Leptotrichia*).

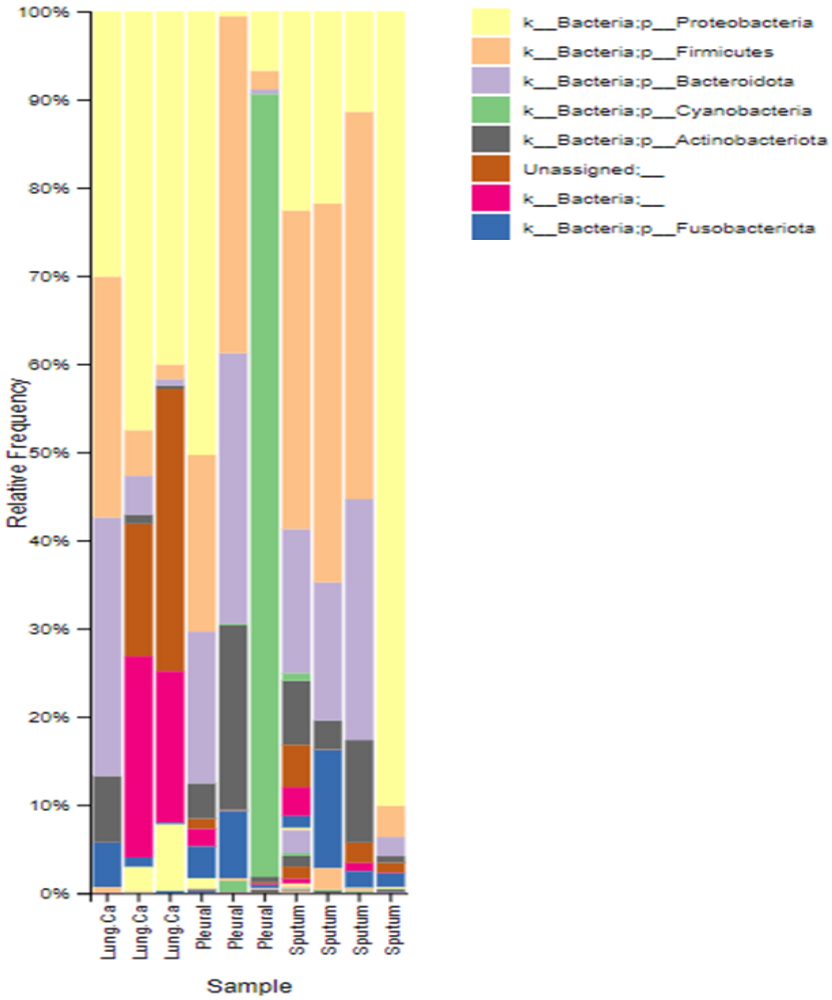


Figure 4: Histogram of Relative Abundance of Microbiota at the phyla level based on representative sequences (ASVs) after denoise and frequency distribution of ASVs in each sample as well as visualization of annotation results of ASVs at the genus level of the bacterial microbiota.

3.6. Microbiota Abundance and Phylogenetic Relationship

To further study the microbiome's relative abundance and their phylogenetic relationship, heatmaps alongside evolutionary trees of explored genera were drawn based on their annotated OTUs in each examined sample (Figure 5).

Interestingly, the abundant genus belonging to Proteobacteria in lung cancer samples was *Hemophilus* (LP1=11.866%, LP2=6.815%, and LP3=1.362%), while in non-cancerous control counterparts (P1=0.085%, P2=0.148% and P3=1.030%). Therefore, the existence of *Haemophilus spp*. in non-malignant pulmonary samples was less abundant than that in lung cancer samples. However, normal sputum controls varied between the low (S1=0.198%) to highest abundance (S4= 82.885%) of this bacterial genus. Indeed, the *Haemophilus spp* abundance in the sputum microbiome may reflect the microbial enrichment of the oral environment. This may, therefore, provide an informative clue about the potential relation between *Hemophilus* spp.-related airway diseases such as COPD and lung carcinogenesis [43].

Interestingly, two well-known oncogenic bacterial genera (*Fusobacterium* and *Porphyromonas*) have also been identified in the microbiome sequencing analysis. Even though these onco-bacteria usually colonize the upper airways and are associated with oral

malignancies, they have been reported to implicate lung cancer development due to their long-term colonization [44,45].

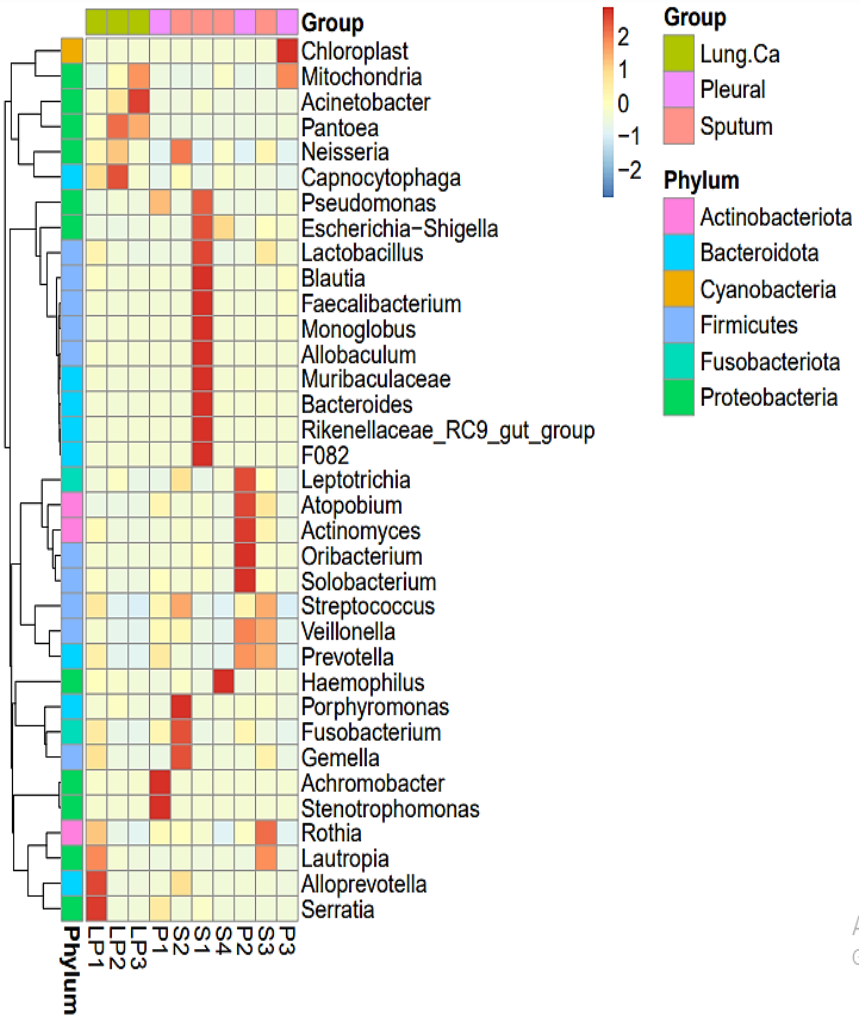
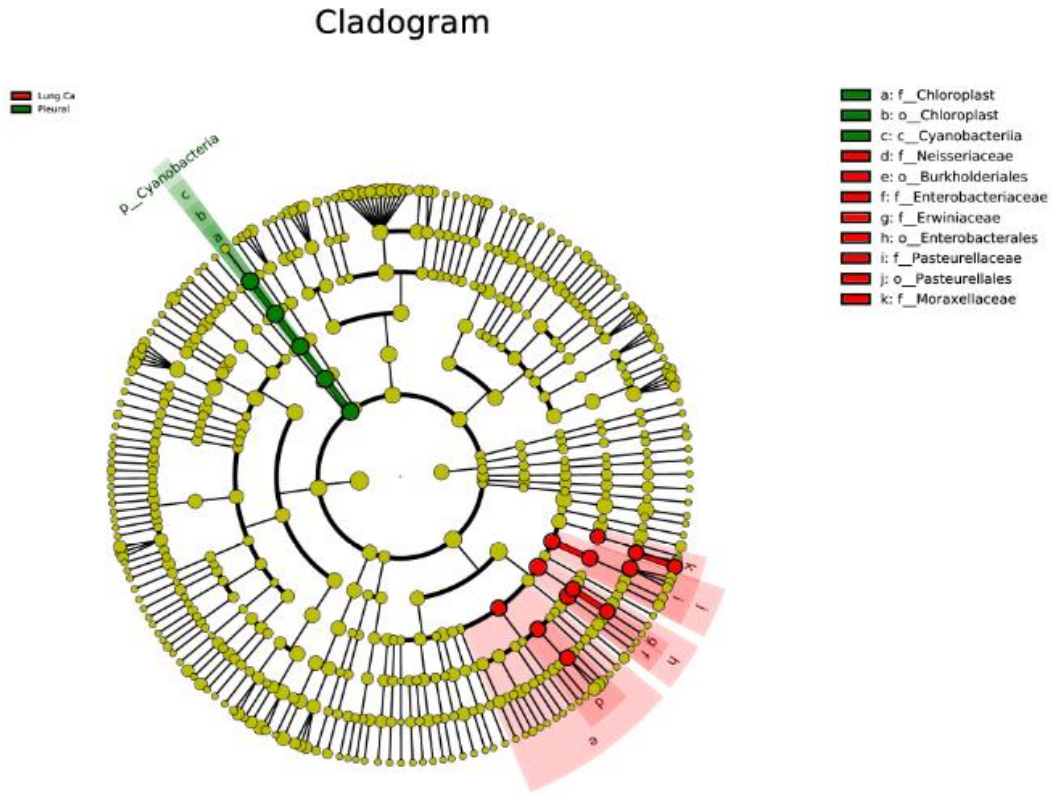


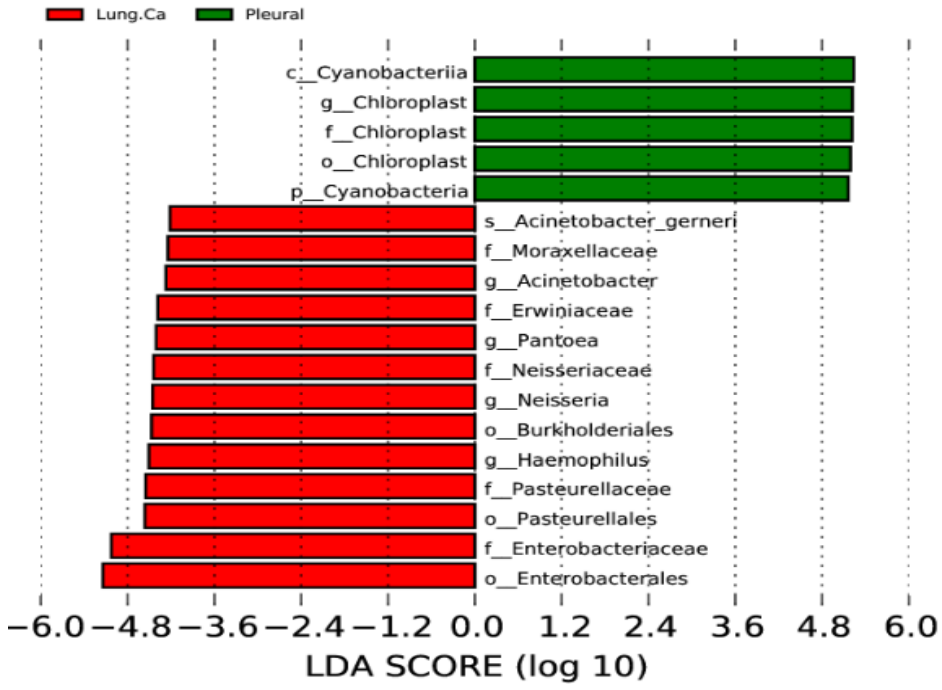
Figure 5: Heatmap showing genus OTUs Abundance. Sample symbols are shown on the x-axis, while their OTU annotations are presented. Genera clustering tree is presented to the left of the figure. The heatmap scale is measured by the Z score, which represents the taxonomic relative abundance of a bacterial genera calculated as a ratio between the mean of the relative abundance of all samples in the taxa and their standard deviation.

3.7. Airways Microbiome is a Potential Predictive Lung Cancer Biomarker

The cladistic analysis gives an accurate definition of organisms' taxonomy in which the organisms are grouped into 'clades' according to the latest common ancestor and are best illustrated by cladogram design referring to the relationship between the diverse levels of clades in different groups of samples [46]. Investigation of distinctive microbiota at the phylum level may, therefore, provide a better idea about potential effects on predicting lung carcinogenesis (Figure 6).



(A)



(B)

Figure 6: Cladogram and enrichment analysis among lung cancer and control samples. (A) Cladogram for phylogenetic relation of control and lung cancer genera. Cladogram was structured by the Linear Discrimination Analysis (LDA) Effect Size (LefSe) method to investigate the phylogenetic allocation of bacterial microbiota that was significantly indicated in lung cancer compared to control samples. (B) LDA scores illustrated significant bacterial differences within cancer and control samples at phyla to genera levels.

The algorithm of Linear Discrimination Analysis (LDA) Effect Size (LefSe) permits the recognition of biomarkers of high dimensional data among various examined categories [47]. LefSe was employed to explore the identifying taxa within lung cancer samples and their controls. The taxa levels revealed that *Haemophilus* alongside other 3 genera including *Acinetobacter, Pantoea, and Neisseria,* were significantly enriched in the lung cancer samples in comparison to the pleural controls.

A ternary plot, one of alpha diversity analyses, has been employed to show the bacterial genera distributed as circles among the groups on the triangle vertices. The circle sizes indicate the OTUs relative abundance of the dominant taxa, and the size of the circles is proportional to the relative abundance, while the location of the circles represents the proportional abundance of each group compared to other ones [48]. The ternary plot demonstrated that *Hemophilus, Prevotella, and Streptococcus* were the most enriched genera among the studied sample groups with almost similar proportions of around (20%) in lung cancer samples. The plot below depicts that the abundance of *Haemophilus spp.* accounted for almost (80%) of sputum samples, while it was around 4% in the pleural samples. Interestingly, *Veillonella* was identified with approximately 10% abundance in malignant lung samples, whereas in benign pleural and sputum ones, the abundance accounted for 52% and 38% respectively (Figure 7).

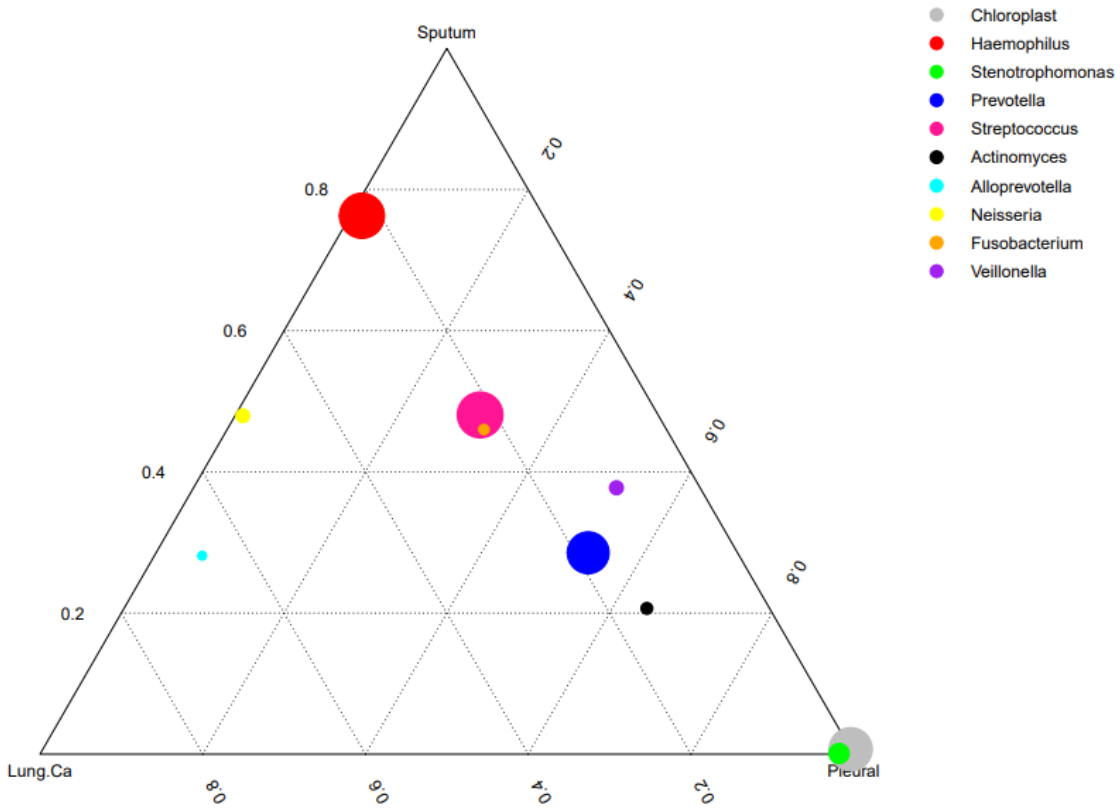


Figure 7: Ternary plot elucidating OTUs abundance landscape of the bacterial genera of malignant, non-malignant and healthy samples within the examined groups distributed on its three vertexes. The bigger and closer circle to a group vertex the higher genus abundance in that group. Ternary plot depicts that the most enriched genera are *Hemophilus, Prevotella, Streptococcus*.

Enrichment of the respiratory tract with *Hemophilus*, *Prevotella*, *and Streptococcus* microbiota has been recently determined to associate with ERK/PI3K pathway alterations in respiratory epithelial cells, which are relevant to lung cancer [49]. A very recent study has reported that *Prevotella* and *Veillonella* were strongly associated with lung cancer progression [50]. Taken together, these outcomes support our perception of how much microbiota inhabit the lower respiratory tract and may motivate early events in the neoplastic transformation of the epithelium layer.

Since *Haemophilus* was on the top of abundant genera of Proteobacteria, the species *Haemophilus influenza* was identified among the top 12 bacterial species in malignant and non-malignant lung samples. *H. influenza* represents the main bacterial cause of COPD-related smoking, which is reported as a potential risk factor for lung cancer since approximately 50–80% of lung cancer cases with smoking status have previously diagnosed with COPD [51, 52, 53]. Overall, regardless of smoking history, age, or sex, patients with COPD have a 4- to 6-fold higher chance of developing lung cancer [54]. Having been detected in lung cancer patients alongside the non-cancerous ones, *Haemophilus spp*. may potentiate their role as microbiome biomarkers for developing lung cancer.

Conclusion

Non-cultural 16SrRNA sequencing techniques have explored a complex microbiome harbored by the respiratory tract which is not detectable by conventional cultivation methods. Our findings of the lung microbiota genera Hemophilus spp. alongside the genera '*Prevotella, Streptococcus*' could potentiate their prediction role in earlier lung tumorigenesis and thus may offer a platform for further exploration of novel prognostic microbiome-based biomarkers for lung carcinomas. Foreseeable future investigation on a larger scale relative to studies is highly recommended.

Acknowledgement

Our thanks to Dr. Ahmed Al-Dhamin, Head of the Dept of Biology, College of Science, Baghdad of University, and Dr. Mais Emad, A staff member at the Dept. , for facilitating access to the related dept laboratories. We also sincerely thank Mrs Asmaa Alkabtan, a staff member of Tuberculosis Institute, The Medical City , for assisting to obtain the respiratory clinical samples.

Disclosure and Conflict of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Shankar, D. Saini, A. Dubey, S. Roy, S. J. Bharati, N. Singh, M. Khanna, C. P. Prasad, M. Singh, S. Kumar, B. Sirohi, T. Seth, M. Rinki, A. Mohan, R. Guleria, and G. K. Rath, "Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward," Transl Lung Cancer Res, vol. 8, no. Suppl 1, pp. S106-s121, May 2019, doi: 10.21037/tlcr.2019.03.03.

[2] B.A. Peters, R.B. Hayes, C. Goparaju, C. Reid, H.I. Pass, and J. Ahn, "The Microbiome in Lung Cancer Tissue and Recurrence-Free Survival," Cancer Epidemiol. Biomarkers Prev., vol. 28, no. 4, pp. 731-740, Apr. 2019, doi: 10.1158/1055-9965.EPI-18-0966.

[3] A. Maddi, A. Sabharwal, T. Violante, S. Manuballa, R. Genco, S. Patnaik, and S. Yendamuri, "The microbiome and lung cancer," J. Thorac. Dis., vol. 11, no. 1, pp. 280-291, Jan. 2019, doi: 10.21037/jtd.2018.12.88.

[4] N.S. Wolff, F. Hugenholtz, and W.J. Wiersinga, "The emerging role of the microbiota in the ICU," Crit. Care, vol. 22, no. 1, p. 78, Mar. 2018, doi: 10.1186/s13054-018-1999-8.

[5] H.D. Hosgood 3rd, A.R. Sapkota, N. Rothman, T. Rohan, W. Hu, J. Xu, R. Vermeulen, X. He, J.R. White, G. Wu, F. Wei, E.F. Mongodin, and Q. Lan, "The potential role of lung microbiota in lung cancer attributed to household coal burning exposures," Environ. Mol. Mutagen., vol. 55, no. 8, pp. 643-651, Oct. 2014, doi: 10.1002/em.21878.

[6] P.H.H. Schneeberger, J. Prescod, L. Levy, D. Hwang, T. Martinu, and B. Coburn, "Microbiota analysis optimization for human bronchoalveolar lavage fluid," Microbiome, vol. 7, no. 1, p. 141, Oct. 2019, doi: 10.1186/s40168-019-0755-x.

[7] S. Shojaee, A. Sharma, N. Gottel, T. Sanchez, J.A. Gilbert, and N.M. Rahman, "Microbiome profile associated with malignant pleural effusion," PLoS One, vol. 15, no. 5, p. e0232181, May 2020, doi: 10.1371/journal.pone.0232181.

[8] Y.J. Huang, E.S. Charlson, R.G. Collman, S. Colombini-Hatch, F.D. Martinez, and R.M. Senior, "The role of the lung microbiome in health and disease. A National Heart, Lung, and Blood Institute workshop report," Am. J. Respir. Crit. Care Med., vol. 187, no. 12, pp. 1382-1387, Jun. 2013, doi: 10.1164/rccm.201303-0488WS.

[9] A.A. Pragman, T. Lyu, J.A. Baller, T.J. Gould, R.F. Kelly, C.S. Reilly, R.E. Isaacson, and C.H. Wendt, "The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease," Microbiome, vol. 6, no. 1, p. 7, Jan. 2018, doi: 10.1186/s40168-017-0381-4. K.

[10] K. Wang, Y. Huang, Z. Zhang, J. Liao, Y. Ding, X. Fang, L. Liu, J. Luo, and J. Kong, "A Preliminary Study of Microbiota Diversity in Saliva and Bronchoalveolar Lavage Fluid from Patients with Primary Bronchogenic Carcinoma," Med Sci Monit, vol. 25, pp. 2819-2834, Apr 17 2019, doi: 10.12659/msm.915332.

[11] A.A. Pragman, T. Lyu, J.A. Baller, T.J. Gould, R.F. Kelly, C.S. Reilly, R.E. Isaacson, and C.H. Wendt, "The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease," Microbiome, vol. 6, no. 1, p. 7, Jan. 2018, doi: 10.1186/s40168-017-0381-4. PMID: 29316977; PMCID: PMC5759273.

[12] K. Yatera, S. Noguchi, and H. Mukae, "The microbiome in the lower respiratory tract," Respir. Investig., vol. 56, no. 6, pp. 432-439, Nov. 2018, doi: 10.1016/j.resinv.2018.08.003. PMID: 30392534.

[13] R. Ruan, X. Deng, X. Dong, Q. Wang, X. Lv, and C. Si, "Microbiota Emergencies in the Diagnosis of Lung Diseases: A Meta-Analysis," Front. Cell. Infect. Microbiol., vol. 11, p. 709634, Sep. 2021, doi: 10.3389/fcimb.2021.709634. PMID: 34621687; PMCID: PMC8490768.

[14] A. Diakite, G. Dubourg, N. Dione, P. Afouda, S. Bellali, I.I. Ngom, C. Valles, M.L. Tall, J.C. Lagier, and D. Raoult, "Optimization and standardization of the culturomics technique for human microbiome exploration," Sci. Rep., vol. 10, no. 1, p. 9674, Jun. 2020, doi: 10.1038/s41598-020-66738-8. PMID: 32541790; PMCID: PMC7295790.

[15] B. Ditz, S. Christenson, J. Rossen, C. Brightling, H.A.M. Kerstjens, M. van den Berge, and A. Faiz, "Sputum microbiome profiling in COPD: beyond singular pathogen detection," Thorax, vol. 75, no. 4, pp. 338-344, Apr. 2020, doi: 10.1136/thoraxjnl-2019-214168. PMID: 31996401; PMCID: PMC7231454.

[16] A. Morris, J.M. Beck, P.D. Schloss, T.B. Campbell, K. Crothers, J.L. Curtis, S.C. Flores, A.P. Fontenot, E. Ghedin, L. Huang, K. Jablonski, E. Kleerup, S.V. Lynch, E. Sodergren, H. Twigg, V.B. Young, C.M. Bassis, A. Venkataraman, T.M. Schmidt, and G.M. Weinstock, "Comparison of the respiratory microbiome in healthy nonsmokers and smokers," Am. J. Respir. Crit. Care Med., vol. 187, no. 10, pp. 1067-1075, May 2013, doi: 10.1164/rccm.201210-1913OC. PMID: 23491408; PMCID: PMC3734620.

[17] F. Liu, J. Li, Y. Guan, Y. Lou, H. Chen, M. Xu, D. Deng, J. Chen, B. Ni, L. Zhao, H. Li, H. Sang, and X. Cai, "Dysbiosis of the Gut Microbiome is associated with Tumor Biomarkers in Lung Cancer," Int. J. Biol. Sci., vol. 15, no. 11, pp. 2381-2392, Sep. 2019, doi: 10.7150/ijbs.35980. PMID: 31595156; PMCID: PMC6775324.

[18] J. Huang and J. Huang, "Microbial Biomarkers for Lung Cancer: Current Understandings and Limitations," J. Clin. Med., vol. 11, no. 24, p. 7298, Dec. 2022, doi: 10.3390/jcm11247298. PMID: 36555915; PMCID: PMC9782454.

[19] C. Jin, G.K. Lagoudas, C. Zhao, S. Bullman, A. Bhutkar, B. Hu, S. Ameh, D. Sandel, X.S. Liang, S. Mazzilli, M.T. Whary, M. Meyerson, R. Germain, P.C. Blainey, J.G. Fox, and T. Jacks, "Commensal Microbiota Promote Lung Cancer Development via γδ T Cells," Cell, vol. 176, no. 5, pp. 998-1013.e16, Feb. 2019, doi: 10.1016/j.cell.2018.12.040. Epub 2019 Jan 31. PMID: 30712876; PMCID: PMC6691977.

[20] K.L. Greathouse, J.R. White, A.J. Vargas, V.V. Bliskovsky, J.A. Beck, N. von Muhlinen, E.C. Polley, E.D. Bowman, M.A. Khan, A.I. Robles, T. Cooks, B.M. Ryan, N. Padgett, A.H. Dzutsev, G. Trinchieri, M.A. Pineda, S. Bilke, P.S. Meltzer, A.N. Hokenstad, T.M. Stickrod, M.R. Walther-Antonio, J.C. Mell, J.E. Krol, S.V. Balashov, A.S. Bhat, G.D. Ehrlich, A. Valm, C. Deming, S. Conlan, J. Oh, J.A. Segre, and C.C. Harris, "Interaction between the microbiome and TP53 in human lung cancer," Genome Biol., vol. 19, no. 1, p. 123, Aug. 2018, doi: 10.1186/s13059-018-1501-6. Erratum in: Genome Biol. 2020 Feb 20;21(1):41. PMID: 30143034; PMCID: PMC6109311.

[21] D. Huang, X. Su, M. Yuan, S. Zhang, J. He, Q. Deng, W. Qiu, H. Dong, and S. Cai, "The characterization of lung microbiome in lung cancer patients with different clinicopathology," Am. J. Cancer Res., vol. 9, no. 9, pp. 2047-2063, Sep. 2019, PMID: 31598405; PMCID: PMC6780665.

[22] S. Al-Khafaji, P. Pantazi, A. Acha-Sagredo, A. Schache, J. M. Risk, R. J. Shaw, and T. Liloglou, "Overexpression of HURP mRNA in head and neck carcinoma and association with in vitro response to vinorelbine," Oncology Letters, vol. 19, pp. 2502-2507, 2020.

[23] S. Al-Khafaji, I. M. Hade, M. A. Al-Naqqash, and G. O. Alnefaie, "Potential effects of miR-146 expression in relation to malondialdehyde as a biomarker for oxidative damage in patients with breast cancer," World Academy of Sciences Journal, vol. 5, pp. 1-9, 2023

[24] C. A. Ali, F. M. Lafta, M. M. Al Sayyid, and A.-A. N. G. Al-Rekabi, "BRCA1 Gene Expression is Down Regulated in Both Familial and Sporadic Breast Cancer Cases in Baghdad-Iraq," Iraqi Journal of Science, pp. 34-41, 2020.

[25] Y. Zhao, Y. Liu, S. Li, Z. Peng, X. Liu, J. Chen, and X. Zheng, "Role of lung and gut microbiota on lung cancer pathogenesis," J. Cancer Res. Clin. Oncol., vol. 147, no. 8, pp. 2177-2186, Aug. 2021, doi: 10.1007/s00432-021-03644-0.

[26] F. Liu, J. Li, Y. Guan, Y. Lou, H. Chen, M. Xu, D. Deng, J. Chen, B. Ni, L. Zhao, H. Li, H. Sang, and X. Cai, "Dysbiosis of the Gut Microbiome is associated with Tumor Biomarkers in Lung Cancer," Int. J. Biol. Sci., vol. 15, no. 11, pp. 2381-2392, Sep. 2019, doi: 10.7150/ijbs.35980.

[27] C. Sundberg, W. A. Al-Soud, M. Larsson, E. Alm, S. S. Yekta, B. H. Svensson, S. J. Sørensen, and A. Karlsson, "454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters," FEMS Microbiol. Ecol., vol. 85, no. 3, pp. 612-626, Sep. 2013, doi: 10.1111/1574-6941.12148.

[28] N. A. Bokulich, B. D. Kaehler, J. R. Rideout, M. Dillon, E. Bolyen, R. Knight, G. A. Huttley, and J. Gregory Caporaso, "Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin," Microbiome, vol. 6, no. 1, p. 90, May 2018, doi: 10.1186/s40168-018-0470-z.

[29] C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight, "UniFrac: an effective distance metric for microbial community comparison," ISME J., vol. 5, no. 2, pp. 169-172, Feb. 2011, doi: 10.1038/ismej.2010.133.

[30] A. Amir, D. McDonald, J. A. Navas-Molina, E. Kopylova, J. T. Morton, Z. Zech Xu, E. P. Kightley, L. R. Thompson, E. R. Hyde, A. Gonzalez, and R. Knight, "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns," mSystems, vol. 2, no. 2, pp. e00191-16, Mar. 2017, doi: 10.1128/mSystems.00191-16.

[31] D. Huang, X. Su, M. Yuan, S. Zhang, J. He, Q. Deng, W. Qiu, H. Dong, and S. Cai, "The characterization of lung microbiome in lung cancer patients with different clinicopathology," Am. J. Cancer Res., vol. 9, no. 9, pp. 2047-2063, Sep. 2019, PMID: 31598405; PMCID: PMC6780665.

[32] F. Xia, X. Zhou, Y. Liu, Y. Li, X. Bai, and X. Zhou, "Composition and predictive functional analysis of bacterial communities inhabiting Chinese Cordyceps insight into conserved core microbiome," BMC Microbiol., vol. 19, no. 1, p. 105, May 2019, doi: 10.1186/s12866-019-1472-0.

[33] M. Li, D. Shao, J. Zhou, J. Gu, J. Qin, W. Chen, and W. Wei, "Signatures within esophageal microbiota with progression of esophageal squamous cell carcinoma," Chin. J. Cancer Res., vol. 32, no. 6, pp. 755-767, Dec. 2020, doi: 10.21147/j.issn.1000-9604.2020.06.09.

[34] C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight, "UniFrac: an effective distance metric for microbial community comparison," ISME J., vol. 5, no. 2, pp. 169-172, Feb. 2011, doi: 10.1038/ismej.2010.133.

[35] B. J. Callahan, P. J. McMurdie, and S. P. Holmes, "Exact sequence variants should replace operational taxonomic units in marker-gene data analysis," ISME J., vol. 11, no. 12, pp. 2639-2643, Dec. 2017, doi: 10.1038/ismej.2017.119.

[36] R. Villette, G. Autaa, S. Hind, J. B. Holm, A. Moreno-Sabater, and M. Larsen, "Refinement of 16S rRNA gene analysis for low biomass biospecimens," Sci Rep., vol. 11, no. 1, p. 10741, May 2021, doi: 10.1038/s41598-021-90226-2.

[37] E. Bolyen, J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar et al., "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2," Nat Biotechnol, vol. 37, no. 8, pp. 852-857, Aug. 2019, doi: 10.1038/s41587-019-0209-9.

[38] B. J. Callahan, J. Wong, C. Heiner, S. Oh, C. M. Theriot, A. S. Gulati, S. K. McGill, and M. K. Dougherty, "High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution," Nucleic Acids Res., vol. 47, no. 18, p. e103, Oct. 2019, doi: 10.1093/nar/gkz569.

[39] B. He, R. Zhu, H. Yang, Q. Lu, W. Wang, L. Song, X. Sun, G. Zhang, S. Li, J. Yang et al., "Assessing the Impact of Data Preprocessing on Analyzing Next Generation Sequencing Data," Front Bioeng Biotechnol, vol. 8, p. 817, Jul. 2020, doi: 10.3389/fbioe.2020.00817.

[40] J. Lever, M. Krzywinski, and N. Altman, "Points of significance: Principal component analysis," Nature methods, vol. 14, no. 7, pp. 641-643, Jul. 2017, doi: 10.1038/nmeth.4346. Q. Dong, E. S. Chen, C. Zhao, and C. Jin, "Host-Microbiome Interaction in Lung Cancer," Front Immunol, vol. 12, p. 679829, May 2021, doi: 10.3389/fimmu.2021.679829.

[41] Q. Dong, E. S. Chen, C. Zhao, and C. Jin, "Host-Microbiome Interaction in Lung Cancer," Front Immunol, vol. 12, p. 679829, May 2021, doi: 10.3389/fimmu.2021.679829, PMID: 34108973; PMCID: PMC8183378.

[42] J. An, B. J. Kil, H. Kwon, and Y. J. Kim, "Analysis of the Impact of the Presence of Phylum Cyanobacteria in the Microbiome of Patients with Breast Cancer on Their Prognosis," J Clin Med, vol. 11, no. 24, p. 7272, Dec. 2022, doi: 10.3390/jcm11247272.

[43] E. Riquelme, Y. Zhang, L. Zhang, M. Montiel, M. Zoltan, W. Dong, P. Quesada, I. Sahin, V. Chandra, A. San Lucas et al., "Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes," Cell, vol. 178, no. 4, pp. 795-806.e12, Aug. 2019, doi: 10.1016/j.cell.2019.07.008.

[44] B. Ditz, S. Christenson, J. Rossen, C. Brightling, H. A. M. Kerstjens, M. van den Berge, and A. Faiz, "Sputum microbiome profiling in COPD: beyond singular pathogen detection," Thorax, vol. 75, no. 4, pp. 338-344, Apr. 2020, doi: 10.1136/thoraxjnl-2019-214168.

[45] Y. Liu, X. Yuan, K. Chen, F. Zhou, H. Yang, H. Yang, Y. Qi, J. Kong, W. Sun, and S. Gao, "Clinical significance and prognostic value of Porphyromonas gingivalis infection in lung cancer," Transl Oncol, vol. 14, no. 1, p. 100972, Jan. 2021, doi: 10.1016/j.tranon.2020.100972.

[46] X. Jiang, Y. Zhang, H. Wang, Z. Wang, S. Hu, C. Cao, and H. Xiao, "In-Depth Metaproteomics Analysis of Oral Microbiome for Lung Cancer," Research (Wash D C), vol. 2022, p. 9781578, Oct. 2022, doi: 10.34133/2022/9781578.

[47] R. González-José, I. Escapa, W. A. Neves, R. Cúneo, and H. M. Pucciarelli, "Cladistic analysis of continuous modularized traits provides phylogenetic signals in Homo evolution," Nature, vol. 453, no. 7196, pp. 775-778, Jun. 2008, doi: 10.1038/nature06891.

[48] S. Wang, F. Song, H. Gu, Z. Shu, X. Wei, K. Zhang, Y. Zhou, L. Jiang, Z. Wang, J. Li, H. Luo, and W. Liang, "Assess the diversity of gut microbiota among healthy adults for forensic application," Microb Cell Fact, vol. 21, no. 1, p. 46, Mar. 2022, doi: 10.1186/s12934-022-01769-6, PMID: 35331204; PMCID: PMC8944125.

[49] C. Y. Chiu, Y. L. Chan, M. H. Tsai, C. J. Wang, M. H. Chiang, C. C. Chiu, and S. C. Su, "Cross-talk between airway and gut microbiome links to IgE responses to house dust mites in childhood

airway allergies," Sci Rep, vol. 10, no. 1, p. 13449, Aug. 2020, doi: 10.1038/s41598-020-70528-7, PMID: 32778700; PMCID: PMC7417544.

[50] J. J. Tsay, B. G. Wu, M. H. Badri, J. C. Clemente, N. Shen, P. Meyn, Y. Li, T. A. Yie, T. Lhakhang, E. Olsen, V. Murthy, G. Michaud, I. Sulaiman, A. Tsirigos, A. Heguy, H. Pass, M. D. Weiden, W. N. Rom, D. H. Sterman, R. Bonneau, M. J. Blaser, and L. N. Segal, "Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer," Am J Respir Crit Care Med, vol. 198, no. 9, pp. 1188-1198, Nov. 2018, doi: 10.1164/rccm.201710-2118OC, PMID: 29864375; PMCID: PMC6221574.

[51] W. Zeng, C. Zhao, M. Yu, H. Chen, Y. Pan, Y. Wang, H. Bao, and S. Ma, "Alterations of lung microbiota in patients with non-small cell lung cancer," Bioengineered, vol. 13, no. 3, pp. 6665-6677, Mar. 2022, doi: 10.1080/21655979.2022.2045843, PMID: 35254206; PMCID: PMC8973753.

[52] R. P. Young, R. J. Hopkins, T. Christmas, P. N. Black, P. Metcalf, and G. D. Gamble, "COPD prevalence is increased in lung cancer, independent of age, sex and smoking history," Eur Respir J, vol. 34, no. 2, pp. 380-386, Aug. 2009, doi: 10.1183/09031936.00144208, Epub 2009 Feb 5, PMID: 19196816.

[53] J. P. de Torres, G. Bastarrika, J. P. Wisnivesky, A. B. Alcaide, A. Campo, L. M. Seijo, J. C. Pueyo, A. Villanueva, M. D. Lozano, U. Montes, L. Montuenga, and J. J. Zulueta, "Assessing the relationship between lung cancer risk and emphysema detected on low-dose CT of the chest," Chest, vol. 132, no. 6, pp. 1932-1938, Dec. 2007, doi: 10.1378/chest.07-1490, PMID: 18079226

[54] E. Calabrò, G. Randi, C. La Vecchia, N. Sverzellati, A. Marchianò, M. Villani, M. Zompatori, R. Cassandro, S. Harari, and U. Pastorino, "Lung function predicts lung cancer risk in smokers: a tool for targeting screening programmes," Eur Respir J, vol. 35, no. 1, pp. 146-151, Jan. 2010, doi: 10.1183/09031936.00049909, Epub 2009 Aug 13, PMID: 19679603.

[55] J. Gonzalez, M. Marin, P. Sánchez-Salcedo, and J. J. Zulueta, "Lung cancer screening in patients with chronic obstructive pulmonary disease," Ann Transl Med, vol. 4, no. 8, p. 160, Apr. 2016, doi: 10.21037/atm.2016.03.57, PMID: 27195278; PMCID: PMC4860483.