

Metagenomic Characterization of Bacterial Communities Associated with Lung Cancer

Sweksha Ranjan

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus,
Lucknow, UP, India

 Ruchi Yadav

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus,
Lucknow, UP, India

Received: 11 August 2026 | Received Revised Version: 28 August 2026 | Accepted: 26 September 2026 | Published: 01 October 2026

Volume 08 Issue 10 2026 | Crossref DOI: 10.37547/tajir/Volume08Issue10-01

Abstract

*This research investigates the bacterial diversity associated with lung cancer through metagenomic analysis. The study utilized metagenomic data from the European Nucleotide Archive (ENA), specifically accession number PRJNA592147, examining ten FASTQ files representing two lung cancer samples. The 16S lung cancer metagenome was sequenced using Illumina MiSeq paired-end technology, with data processing conducted on the Galaxy server. Data quality was enhanced through FastQC quality control and Trimmomatic sequence trimming. Kraken 2 was employed for taxonomic classification and bacterial community abundance assessment, while the Krona tool facilitated data visualization. The analysis revealed three notable bacterial species: *Mycobacterium chubuense*, *Arthrobacter ramosus*, and *Streptosporangium longisporum*. *M. chubuense*, a member of the Mycobacteriaceae family, is characterized as a non-motile, spore-free bacterium that thrives in various media at 37°C. *A. ramosus*, belonging to the Micrococcaceae family, is an obligate aerobe that flourishes in multiple culture media at temperatures ranging from 28–30°C. *S. longisporum*, part of the Streptosporangiaceae family, forms sporangia and exhibits ideal progress at 30°C in specific media. The identification of these bacteria in lung cancer samples suggests their potential involvement in the disease's development and progression. This investigation offers insights into the microbial landscape associated with lung cancer, potentially contributing to a well understanding of the disease's etiology and the development of novel therapeutic approaches. Further research is required to validate these results and elucidate the mechanisms connecting these bacteria to lung cancer.*

Keywords: Metagenomics, Bacterial Diversity, Lung Cancer, Metagenomic Analysis, Korona pie chart.

© 2026 Sweksha Ranjan, & Ruchi Yadav. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). The authors retain copyright and allow others to share, adapt, or redistribute the work with proper attribution.

Cite This Article: Ranjan, S., & Yadav, R. (2026). Metagenomic Characterization of Bacterial Communities Associated with Lung Cancer. The American Journal of Interdisciplinary Innovations and Research, 8(10), 01–08. <https://doi.org/10.37547/tajir/Volume08Issue10-01>

1. Introduction

One of the primary drivers of disease related mortalities internationally is cellular breakdown in the lungs, which

is much of the time welcomed on by smoking however can likewise be welcomed on by openness to radon, asbestos, and contamination. It typically begins in the respiratory and other lining cells of the lung. Non-small

cell lung cancer (NSCLC) and small cell lung cancer (SCLC) are the two essential structures, with NSCLC being wilder. Cellular breakdowns in the lungs are ordinarily undetected, and when they are found, most of patients as of now have developed phases of the ailment [1]. Although the fact that tobacco smoking is by far the biggest risk factor, tobacco usage is mostly preventable, accounting for over 85% of deaths from the disease [2]. Over the past century, there have been two main attempts to screen for lung cancer: the first was based on chest radiographs and carried out in the 1970s and 1980s; the second, uses low-dose computed tomography (CT) [3].

In the past 20 years, immunotherapies and targeted treatments have enhanced NSCLC survival, especially in advanced stages. However, early-stage biomarkers for predicting recurrence are urgently needed [4]. There has been a renewed interest in recent years in cancer immunosurveillance, which has expanded into the field of cancer immunoediting. The process of cancer immunoediting consists of three stages: equilibrium, escape, and elimination (also referred to as cancer immunosurveillance) [5]. More recently, it has been demonstrated that immune treatment that targets molecules known as inhibitory checkpoints, such as programmed death 1 (PD-1), can alter T cell responses to neoantigens and increase lung cancer patient survival. However, between 40% and 60% of patients will either not respond to immunotherapy or may grow resistant to it [6].

The metagenomics is the field that can be used to study the genetic material and has uncovered connections between lung cancer and the microbiome of the lungs. Studies reveal that irregularities in the lung microbiome, like an excess of some bacteria, may influence immunological responses or cause inflammation, which in turn may lead to the development of cancer. Handelsman and Rodon first reported about metagenomics in 1998, and it quickly gained popularity as additional DNA sequencing method for examining the intricate gut microbiome [7].

By analyzing the DNA of microorganisms, metagenomics offers a potent investigation and study tool for prokaryotes and viruses that live in nearby. The identification of diseases brought on by microorganisms has also profited from the study of metagenomics. The detection of gut microbial species is aided by metagenomics, which also provides information on the evolution of probiotics. By observing the bacterial communities linked to homosapiens, it is possible to

determine how to modify them in order to improve human health [8]. The rapid development of immensely parallel sequencing technology is primarily responsible for the tremendous rise of metagenomics during the last 2 to 3 years [9].

Few studies have been done to monitor variations in the microbiota of the gut from the emergence of lung cancer. The gut microbiota regulates the host metabolism and immune response, which plays a vital role in tumor pathogenesis [10]. It has recently been discovered that early-stage lung tumor was linked to specific gut bacterial groups from lung cancer patients' gut microbiome, including *Ruminococcus* and *Bacteroides* [11].

The human gut microbiome has been widely researched in a medical context, revealing intricate interactions between the microbiome and the host immune system and metabolism [12]. Metagenomic next-generation sequencing (mNGS) is rapidly making the transition from laboratories to medical settings. Due to its fast turnaround time (TAT), high throughput, and culture independency, (mNGS) has emerged as a feasible approach for infectious disease screening on the Illumina platform [13]. Large-scale sequencing data can now be produced by small research groups owing to the arrival of new generation sequencing technologies [14].

The main functions of metagenome databases needed for taxonomic classification are annotation, mining, and preservation of individually sequenced bacterial genomes. The NCBI, ENA, and DDBJ, together make up the International Nucleotide Sequence Database Collaboration, which serves as the key universal source of genome sequences from all domains of life, as well as viruses [15]. EBI Metagenomics offers ENA archive services as well as unique accession numbers for datasets that are submitted. [16].

The accessibility of user-friendly interfaces is becoming an increasingly popular choice for the rising number of non-bioinformaticians working with NGS data. These interfaces usually have easily available pre-installed software packages and are connected to considerable computational resources. Genomics Virtual Lab and Galaxy are two instances of strong stages that provide non-bioinformaticians with access to the most recent NGS techniques [17].

Galaxy offers analytical tools that may be integrated into intricate workflows or utilized separately. The

intermediate data outputs of the tools can be used to initiate logic conditionals inside the process. Recent improvements enable researchers to conduct workflows on variable- quality data, with the workflow buffered so that experimental variability may be systematically investigated [18].

MATERIALS AND METHODS

1. Data retrieval (Metagenomics)

Data related to lung cancer (Metagenomics data) was obtained from the European Nucleotide Archive (ENA) database <https://www.ebi.ac.uk/ena/browser/home> with accession number PRJNA592147 and project id 592147 <https://www.ebi.ac.uk/ena/browser/view/PRJNA592147>. Ten fastq files were downloaded with two different lung tumor metagenomic samples. In this experiment, Illumina MiSeq paired-end sequencing was performed on the 16S lung cancer metagenome. These 10 sample files were chosen for the metagenomic analysis of breast

cancer in the current research.

2. Metagenomics tool-based analysis

The Galaxy server <https://usegalaxy.org/> and the metagenomics tools at Galaxy server <https://metagenomics.usegalaxy.eu/> were utilized for investigating the metagenomics and annotation of sample data. **FastQC tool** was employed to read data from each file, identify statistical parameters, and do quality control analysis. **Trimmomatic** was utilised to screen out the adapter sequence and improve the raw file quality for usage in subsequent analyses, it was employed to trim the fastqc file.

Kraken 2 tool was used in which all samples underwent taxonomic analysis, and the abundance of each bacterial community was projected. **Convert kraken tool** was utilized to transform the taxonomic file produced by the Kraken 2 tool into a file compatible with Krona, enabling it to be utilized in the **Krona tool** for the purpose of visualizing the variety of bacteria in each of the sample files examined.

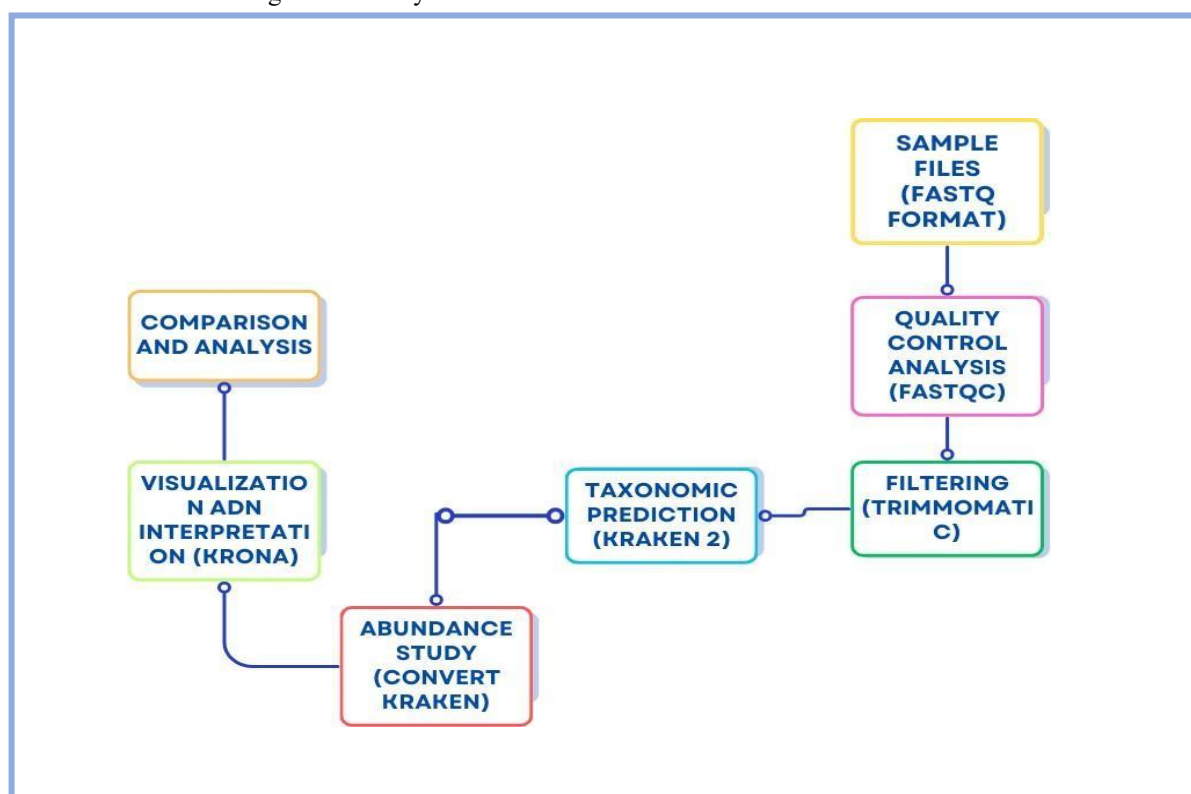


Figure 1: Flowchart for metagenomic analysis and tools

The procedures and tools utilized in metagenomics analysis are shown in Figure 1. The initial step is **quality control**, which is carried out for the data's timeliness,

accuracy, completeness, relevance, and validity.

Second step is the data trimming, which handles different trimming responsibilities for paired- end and single-end

data. Use the **Trimmomatic tool** found in the Galaxy tool panel to conclude this procedure. The Trimmomatic program is used to run the FastQC output file.

Next step is taxonomy prediction that is performed via **Kraken 2 tool**. Short DNA reads are given taxonomic classifications by the taxonomic sequence classifier Kraken 2 tool. It checks the k-mers found in a read to a taxonomy database after examining them. The Trimmomatic tool's output acts as the Kraken 2 tool's input. For visualization, a taxonomic abundance file was created using the **Convert Kraken tool**. With the use of this tool, the NCBI taxonomy is comprehensively expresses the Kraken metagenomic classifier findings. The Convert Kraken tool was used in which the output file from Kraken 2 was used as input.

Eventually, a zoomable pie chart was created using the **Krona pie chart tool** to display the metagenomic results. The Krona tool's input file was the result of the Convert Kraken tool.

RESULTS AND DISCUSSION

The Krona visualization tool extracts results of metagenomic outlining as a zoomable pie chart. It facilitates the exploration of hierarchical data, specifically taxonomic levels, through the utilization of zooming, multilayered pie charts. In our investigation, we observed that *Mycolicibacterium chubuense*, *Streptosporangium longisporum*, and *Arthrobacter ramosus* were crucial microorganisms involved in lung cancer. Table 1 shows the bacteria name, their percentage and run accession.

Table 1: Shows the bacterial community retrieved from metagenomic analysis

S. No.	Bacteria name	Run Accession	Percentage
1.	<i>Mycolicibacterium chubuense</i>	SRR10776868	73%
2.	<i>Arthrobacter ramosus</i>	SRR10776914	31%
3.	<i>Streptosporangium longisporum</i>	SRR10776914	20%

Mycolicibacterium chubuense [19] belongs to the *Mycobacteriaceae* family of Gram- positive bacteria with a high G+C content. It is a rod-shaped bacterium that is a member of the *Actinomycetota* phylum, *Actinomycetes* class, *Mycobacteriales* order, and *Mycobacteriaceae* family within the Bacteria domain. With a high degree of confidence in its motility and spore generation properties, it is non-motile and does not create spores. *M. chubuense* colonies usually take 10–14 days to incubate and have a light orange hue. It grows well at 37°C on a variety of culture media, including Middlebrook medium, BHI, and MB7H9. It also grows well on Löwenstein-Jensen medium. *M. chubuense* does not grow at 10°C or 45°C, however it may grow aerobically in the range of 22–37°C.

Arthrobacter ramosus [20] belongs to the family

Micrococcaceae and is a species of Gram- positive bacterium with a high G+C. The bacterium is rod-shaped and a member of the *Actinomycetota* phylum, *Actinomycetes* class, *Micrococcales* order, and *Micrococcaceae* family. *A. ramosus*, which was first identified by Jensen in 1960 and was added to the Approved Lists in 1980. It is an obligate aerobe, which means that it needs oxygen to develop and is a non-motile. The bacteria grow on a number of culture media, such as CIP Mediums 72 and 3, Trypticase Soy Yeast Extract Medium (DSMZ Medium 92), and Columbia agar (Medium 3). DSMZ and Media Dive provide recipes for these media. At 28– 30°C, *A. ramosus* begins to develop positively, indicating its adaptability to mild conditions. Its ability to adapt to a particular niche and its metabolic capacities are reflected in the particular growth circumstances and oxygen needs. These

characteristics are consistent with the obligate aerobe's physiological makeup, which emphasizes the need for oxygen for cellular functions and energy generation.

Streptosporangium longisporum [21] belongs to the *Streptosporangiaceae* family of Gram- positive bacteria with a high G+C content. It belongs to the class Actinomycetes, phylum Actinomycetota, domain Bacteria, order *Streptosporangiales*, and family *Streptosporangiaceae*. It appears in the 1980 Approved Lists and was originally described by Schäfer in 1969. After an incubation period of 10–14 days, this species develops orange-red colonies on ISP 2 and ISP 7 media,

whereas pink colonies are detected on ISP 3, ISP 4, ISP 5, and ISP 6 media. Aerial mycelium is a kind of multicellular complex that *S. longisporum* may develop; on ISP 2, 3, and 4, it appears sparse and white, while on ISP 7, it is thicker and white. ISP 5 and ISP 6 do not create any aerial mycelium. In a variety of culture mediums, including as Czapek Peptone Agar, Rolled Oats Mineral Medium, and GYM Streptomyces Medium, the bacteria grow best at a temperature of about 30°C. In addition, *S. longisporum* produces sporangia with a smooth surface, demonstrating its capacity to generate spores.

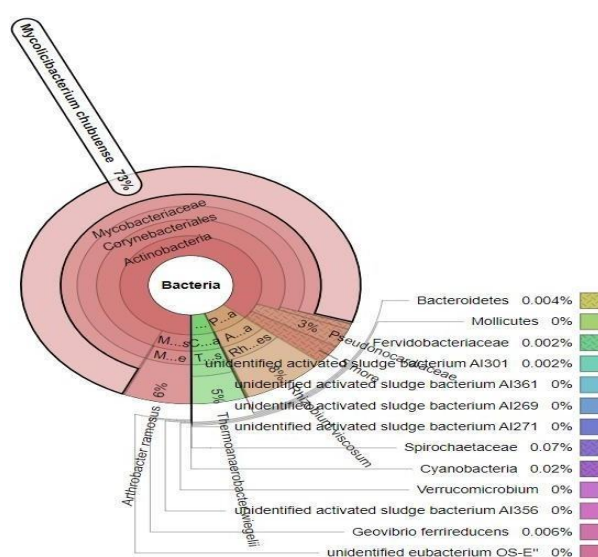


Figure 2: Bacterial diversity visualization through Krona pie chart of paired end data of SRR10776887_1 and SRR10776881_2

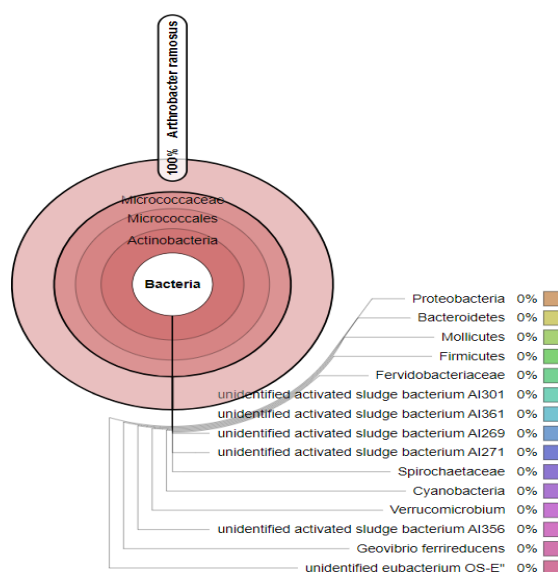


Figure 3: Bacterial diversity visualization through Krona pie chart of paired end data of SRR10776868_1

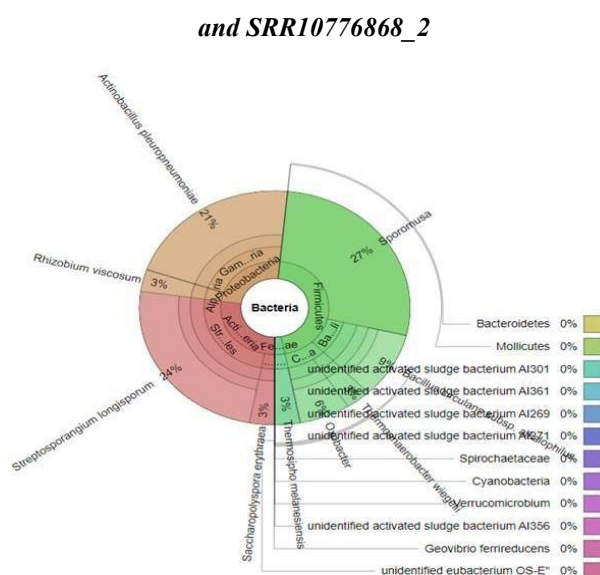
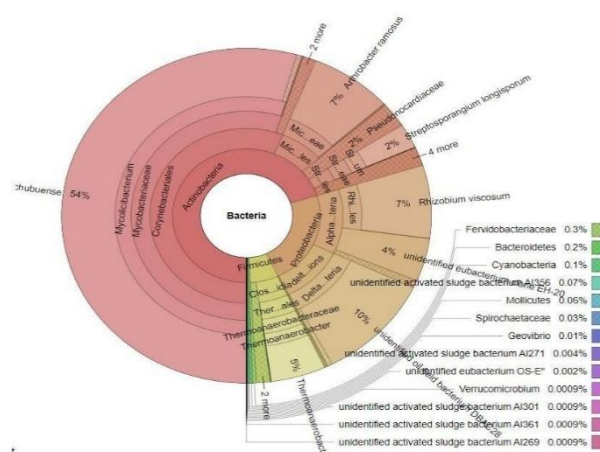


Figure 4: Bacterial diversity visualization through Krona pie chart of paired end data of SRR10776895_1 and SRR10776895_2



The metagenomic analysis performed using the Krona visualization tool enabled comprehensive exploration of microbial diversity across lung cancer samples through interactive hierarchical pie charts. The analysis revealed a distinct bacterial composition, with *Mycolicibacterium chubuense* (73%), *Arthrobacter ramosus* (31%), and *Streptosporangium longisporum* (20%) identified as the predominant taxa across the analyzed datasets (Table 1).

The dominance of *Mycolicibacterium chubuense* suggests a potential association between Actinomycetota members and lung tumor microenvironments. Members of the Mycobacteriaceae family are known for their resilience, metabolic versatility, and ability to survive under stress conditions, which may contribute to their

persistence in tumor-associated niches (Ref. 19). Similarly, *Arthrobacter ramosus*, belonging to the Micrococcaceae family, is recognized for its adaptability to diverse environmental conditions and its metabolic plasticity, which may play a role in host–microbe interactions within cancer-affected tissues.

The presence of *Streptosporangium longisporum*, a spore-forming actinomycete, further supports the hypothesis that soil- and environment-associated bacteria may colonize or influence lung tissues under pathological conditions. These organisms are known for producing bioactive compounds, which could potentially modulate inflammatory pathways or immune responses within the tumor microenvironment.

The Krona plots (Figures 2–5) demonstrated variation in bacterial abundance across different samples, indicating inter-sample heterogeneity in microbial composition. Such variability may reflect differences in disease stage, host immune response, environmental exposure, or sample-specific microbial colonization patterns.

Importantly, the predominance of Actinobacteria-related species observed in this study aligns with previous findings that suggest microbial dysbiosis plays a significant role in cancer progression through mechanisms such as chronic inflammation, immune modulation, and metabolic alterations. The identified bacterial taxa may contribute to tumorigenesis either directly, through production of carcinogenic metabolites, or indirectly, by influencing host immune responses.

However, it is important to note that this study is based on metagenomic data analysis and does not establish a causal relationship between these microorganisms and lung cancer. The observed associations require further validation through experimental studies, including functional metagenomics, transcriptomics, and clinical correlation analyses.

CONCLUSION

In the present study, a comprehensive metagenomic approach was employed to investigate bacterial diversity associated with lung cancer using publicly available sequencing datasets. The analysis successfully identified key bacterial taxa, including *Mycolicibacterium chubuense*, *Arthrobacter ramosus*, and *Streptosporangium longisporum*, suggesting their potential involvement in the lung cancer microenvironment.

The findings highlight the presence of distinct microbial communities in lung cancer samples and emphasize the possible role of microbial dysbiosis in tumor development and progression. The predominance of Actinomycetota-related bacteria indicates their potential contribution to inflammation, immune modulation, and metabolic interactions within tumor tissues.

Furthermore, this study demonstrates the effectiveness of computational metagenomic tools such as Galaxy, Kraken 2, and Krona in analyzing complex microbial datasets and generating meaningful biological insights. The use of such tools provides a cost-effective and scalable approach for microbiome-based cancer research. Despite these promising findings, several limitations exist. The study is primarily based on

secondary data analysis and lacks experimental validation. Additionally, the functional roles of the identified bacterial species in lung cancer progression remain unclear. Therefore, further investigations involving multi-omics approaches, in vitro and in vivo experiments, and clinical validation are required to establish causal relationships.

Future research should focus on identifying microbial biomarkers for early detection, prognosis, and therapeutic targeting of lung cancer. Understanding the interaction between host immunity and microbial communities may also facilitate the development of microbiome-based therapeutic strategies, including probiotics, microbial modulation, and personalized medicine approaches.

Overall, this study contributes to the growing body of evidence supporting the role of the microbiome in cancer biology and provides a foundation for future research aimed at exploiting microbial diversity for improved diagnosis and treatment of lung cancer.

References

1. M. Zhang et al., "Intestinal flora characteristics of advanced non-small cell lung cancer in China and their role in chemotherapy based on metagenomics: A prospective exploratory cohort study," *Thorac Cancer*, vol. 12, no. 24, pp. 3293–3303, Dec. 2021, doi: 10.1111/1759- 7714.14199.
2. R. S. Fontana et al., "Screening for lung cancer. A critique of the mayo lung project," *Cancer*, vol. 67, no. 4 S, pp. 1155–1164, 1991, doi: 10.1002/1097-0142(19910215)67:4+<1155::AID-CNCR2820671509>3.0.CO;2-0.
3. S. C. Spiro and G. A. Silvestri, "One hundred years of lung cancer," Sep. 01, 2005. doi: 10.1164/rccm.200504-531OE.
4. B. A. Peters et al., "The lung microbiome, peripheral gene expression, and recurrence-free survival after resection of stage II non-small cell lung cancer," *Genome Med*, vol. 14, no. 1, Dec. 2022, doi: 10.1186/s13073-022-01126-7.
5. G. P. Dunn, L. J. Old, and R. D. Schreiber, "Review The Immunobiology of Cancer Immunosurveillance and Immunoediting tective versus tumor-sculpting actions of the immune system in cancer (Shankaran et al we summarize recent work on the cancer immunosurveillance and immunoediting pro-cesses-underscoring a new optimism that an enhanced," 2004.

6. J. C. J. Tsay et al., "Lower airway dysbiosis affects lung cancer progression," *Cancer Discov*, vol. 11, no. 2, pp. 293–307, Feb. 2021, doi: 10.1158/2159-8290.CD-20-0263.
7. W. L. Wang, S. Y. Xu, Z. G. Ren, L. Tao, J. W. Jiang, and S. Sen Zheng, "Application of metagenomics in the human gut microbiome," *Jan*, vol. 21, 2015, WJG Press. doi: 10.3748/wjg.v21.i3.803.
8. M. Jain and R. Yadav, "Metagenomics Analysis of Breast Cancer to Study Bacterial Diversity," *Journal of Drug Delivery and Therapeutics*, vol. 12, no. 2-S, pp. 119–126, Apr. 2022, doi: 10.22270/jddt.v12i2-s.5433.
9. S. K. Pond et al., "Windshield splatter analysis with the Galaxy metagenomic pipeline," *Genome Res*, vol. 19, no. 11, pp. 2144–2153, Nov. 2009, doi: 10.1101/gr.094508.109.
10. C. Feng, N. Li, G. Gao, Q. He, L. Y. Kwok, and H. Zhang, "Dynamic Changes of the Gut Microbiota and Its Functional Metagenomic Potential during the Development of Non-Small Cell Lung Cancer," *Int J Mol Sci*, vol. 25, no. 7, Apr. 2024, doi: 10.3390/ijms25073768.
11. Y. Guo et al., "Association Studies on Gut and Lung Microbiomes in Patients with Lung Adenocarcinoma," *Microorganisms*, vol. 11, no. 3, Mar. 2023, doi: 10.3390/microorganisms11030546.
12. J. Jovel et al., "Characterization of the gut microbiome using 16S or shotgun metagenomics," *Front Microbiol*, vol. 7, no. APR, Apr. 2016, doi: 10.3389/fmicb.2016.00459.
13. Y. Guo et al., "Metagenomic next-generation sequencing to identify pathogens and cancer in lung biopsy tissue," *EBioMedicine*, vol. 73, Nov. 2021, doi: 10.1016/j.ebiom.2021.103639.
14. P. ten Hoopen et al., "The metagenomic data life-cycle: Standards and best practices," *Aug. 01*, 2017, Oxford University Press. doi: 10.1093/gigascience/gix047.
15. S. Yen and J. S. Johnson, "Metagenomics: a path to understanding the gut microbiome," *Aug. 01*, 2021, Springer. doi: 10.1007/s00335-021-09889-x.
16. A. Oulas et al., "Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies," *Bioinform Biol Insights*, vol. 9, pp. 75–88, May 2015, doi: 10.4137/BBI.S12462.
17. M. W. C. Thang, X.-Y. Chua, G. Price, D. Gorse, and M. A. Field, "MetaDEGalaxy: Galaxy workflow for differential abundance analysis of 16s metagenomic data," *F1000Res*, vol. 8, p. 726, May 2019, doi: 10.12688/f1000research.18866.1.
18. L. A. L. Abueg et al., "The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update," *Nucleic Acids Res*, vol. 52, no. W1, pp. W83–W94, Jul. 2024, doi: 10.1093/nar/gkae410.
19. B. Verslyppe, W. De Smet, B. De Baets, P. De Vos, and P. Dawyndt, "StrainInfo introduces electronic passports for microorganisms," *Syst Appl Microbiol*, vol. 37, no. 1, pp. 42–50, 2014, doi: https://doi.org/10.1016/j.syapm.2013.11.002.
20. A. C. Parte, J. S. Carbasse, J. P. Meier-Kolthoff, L. C. Reimer, and M. Göker, "List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ," *Int J Syst Evol Microbiol*, vol. 70, no. 11, pp. 5607–5612, 2020, doi: 10.1099/ijsem.0.004332.
21. J. M. Wink, "Compendium of Actinobacteria from Dr."