**UNIVERSIDADE FEDERAL DO PARÁ**
**INSTITUTO DE CIÊNCIAS BIOLÓGICAS**
**PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR**

**MARIANA SOUZA DE LIMA**

# ANÁLISE DE RNAs LONGOS NÃO CODIFICANTES NO CÂNCER DE PULMÃO NÃO PEQUENAS CÉLULAS

**BELÉM – PA**
**2024**

**MARIANA SOUZA DE LIMA**

# ANÁLISE DE RNAs LONGOS NÃO CODIFICANTES NO CÂNCER DE PULMÃO NÃO PEQUENAS CÉLULAS

Defesa do projeto apresentado ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Federal do Pará, como requisito para obtenção de título de doutora em Genética e Biologia Molecular.

**Orientador:** Prof. Dr. Paulo Assumpção

**BELÉM – PA**

**2024**

**MARIANA SOUZA DE LIMA**

**ANÁLISE DE RNAs LONGOS NÃO CODIFICANTES NO CÂNCER DE PULMÃO NÃO PEQUENAS CÉLULAS**

Presidente da Banca:

---

Prof. Dr. Paulo Pimentel de Assumpção
Núcleo de Pesquisas em Oncologia/UFPA- Presidente da Banca

Banca examinadora:

---

Prof. Dr. Leandro Lopes de Magalhães
Laboratório de Genética Humana e Médica/UFPA- Membro Interno

---

Prof. Dr. Fabiano Cordeiro Moreira
Núcleo de Pesquisa em Oncologia/UFPA- Membro Interno

---

Prof. Dr. Samir Mansour Moraes Casseb
Núcleo de Pesquisa em Oncologia/UFPA – Membro interno

---

Profª. Drª. Aline Maria Pereira Cruz Ramos
Faculdade de Enfermagem - ICS/UFPA- Membro Externo

## RESUMO

O Câncer de Pulmão (CP) foi o mais comum em todo o mundo, com 2,5 milhões de novos casos, representando 12,4% de todos os novos casos. O ressurgimento do CP como o tipo de maior frequência provavelmente está relacionado à persistência do uso de tabaco, frequentemente associado à idade (atualmente, risco maior entre 50 e 74 anos), em razão do tempo de exposição e carga tabágica. No Brasil, o CP é o terceiro tipo de câncer de maior incidência em homens (18.020 casos novos) e o quarto em mulheres (14.540 novos casos), é também o tipo de câncer com maior mortalidade para o sexo masculino. Alterações genéticas e epigenéticas influenciam na iniciação e no desenvolvimento do tumor pulmonar, incluindo mutações, instabilidade cromossômica, expressão anormal de RNAs não codificantes (ncRNAs) e evidências sugerem que disbioses na microbiota modulam a suscetibilidade a neoplasias malignas através da liberação de metabólitos bacterianos promotores de câncer e indutores de vias inflamatórias do hospedeiro regulando a expressão de mediadores e citocinas. Com o desenvolvimento das tecnologias multiômicas, hoje sabe-se que a microbiota pulmonar é única e transitória. Os mecanismos microbianos na etiopatogenia do câncer são múltiplos, dentre os quais foi identificada uma interação contínua e bidirecional entre o microbioma e a expressão de ncRNAs. Em estudos recentes baseados em ômicas, os ncRNAs, mostraram-se como moduladores de doenças, como o CP, em resposta à microbiota e sua disbiose. Nesse contexto, especificamente os RNAs longos não codificantes (lncRNAs) oferecem vantagens sobre as proteínas em algumas formas de regulação epigenética. Ainda há poucos estudos que investiguem as relações bidirecionais existentes entre a microbiota e os lncRNAs no CP, porém é possível que haja diferenças nas comunidades bacterianas entre indivíduos saudáveis e pacientes com câncer e a carga bacteriana pulmonar desempenha um papel chave na carcinogênese e na resposta imune contra células cancerosas, influenciando na sensibilidade à quimioterapia e imunoterapia. Para isso, este estudo visa analisar os transcritos de amostras de tecido pulmonar não tumoral adjacente e tumoral para caracterizar a interação entre os RNA não codificantes longos e a microbiota pulmonar através da metatranscriptômica.



**Palavras-chave:** Câncer de pulmão; lncRNA; Microbiota pulmonar; Biomarcadores; Metatranscriptomica.

## ABSTRACT

Lung cancer was the most common worldwide, with 2.5 million new cases, representing 12.4% of all new cases. The resurgence of CP as the most common type is probably related to the persistence of tobacco use, often associated with age (currently, higher risk between 50 and 74 years old), due to the length of exposure and smoking history. In Brazil, PC is the third highest incidence type of cancer in men (18,020 new cases) and the fourth in women (14,540 new cases), it is also the type of cancer with the highest mortality rate for males. Genetic and epigenetic changes influence lung tumor initiation and development, including mutations, chromosomal instability, abnormal expression of non-coding RNAs (ncRNAs) and evidence suggests that dysbiosis in the microbiota modulate susceptibility to malignant neoplasms through the release of bacterial metabolites that promote cancer and inducers of host inflammatory pathways by regulating the expression of mediators and cytokines. With the development of multi-omics technologies, it is now known that the lung microbiota is unique and transient. The microbial mechanisms in the etiopathogenesis of cancer are multiple, among which a continuous and bidirectional interaction between the microbiome and the expression of ncRNAs has been identified. In recent studies based on omics, ncRNAs were shown to modulate diseases, such as lung cancer, in response to the microbiota and its dysbiosis. In this context, specifically long non-coding RNAs (lncRNAs) offer advantages over proteins in some forms of epigenetic regulation. There are still few studies that investigate the bidirectional relationships between the microbiota and lncRNAs in PC, however it is possible that there are differences in the bacterial communities between healthy individuals and cancer patients and the lung bacterial load plays a key role in carcinogenesis and immune response against cancer cells, influencing sensitivity to chemotherapy and immunotherapy. To this end, this study aims to analyze transcripts from adjacent non-tumor and tumor lung tissue samples to characterize the interaction between long non-coding RNAs and lung microbiota through metatranscriptomics.



**Keywords:** Lung cancer; lncRNA; Pulmonary microbiota; Biomarkers; Metatranscriptomics.

## LISTA DE ILUSTRAÇÕES

**SUMÁRIO**

# 1 INTRODUÇÃO

## 1.1 CONSIDERAÇÕES GERAIS

O Câncer de Pulmão (CP) é considerado um dos tipos de câncer de maior incidência no mundo, onde o principal fator de risco prevenível para o seu desenvolvimento é o tabagismo e a estimativa de mortes anuais é de 1,8 milhões em ambos os sexos (representando 18.7% de todas as mortes por câncer), sendo no homem, a principal causa de morte (MORGAM, et al, 2023).

De acordo com a Organização Pan-Americana da Saúde (OPAS, 2022), o CP foi o mais comum em todo o mundo, com 2,5 milhões de novos casos, representando 12,4% de todos os novos casos. O ressurgimento do CP como o tipo de maior frequência provavelmente está relacionado à persistência do uso de tabaco, frequentemente associado à idade (atualmente, risco maior entre 50 e 74 anos), em razão do tempo de exposição, carga tabágica, bem como características sociodemográficas como sexo, raça e escolaridade (ARAUJO et al., 2018; CAMPOS et al., 2024).

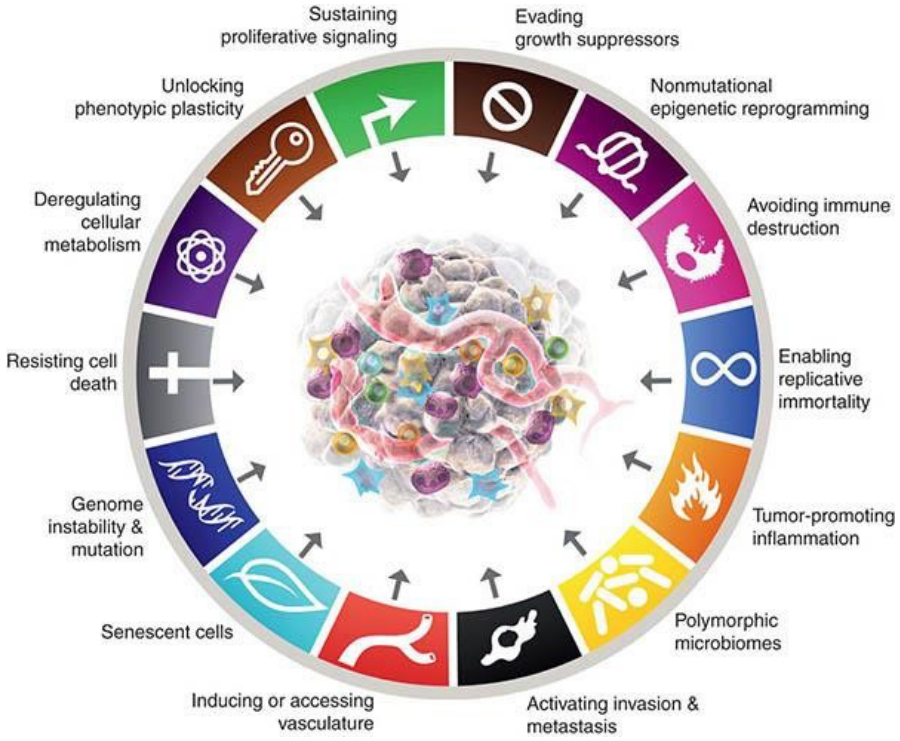
No Brasil, o CP é o terceiro tipo de câncer de maior incidência em homens (18.020 casos novos) e o quarto em mulheres (14.540 novos casos), é também o tipo de câncer com maior mortalidade para o sexo masculino (INCA, 2023). Isso significa dizer que há cerca de 13 casos novos de CP a cada 100mil homens e 9 casos novos a cada 100mil mulheres, com expressiva taxa de mortalidade à medida que a população envelhece sendo 12/100 mil em homens e 8.5/100mil em mulheres (FUNDAÇÃO DO CÂNCER, 2024).

O câncer pode ser causado por alterações genéticas, fatores relacionados a risco ocupacional, exposição ambiental e características sociodemográficas (IARC, 2004; HANAHAN & WEINBERG, 2011; INCA, 2023). No caso do CP, por exemplo, o tabagismo contribui para alterações epigenéticas (modificações químicas que afetam a expressão dos genes sem alterar a sequência do DNA em si), em genes supressores tumorais (*p16*), genes relacionados ao ciclo celular (*DAPK* e *TERT*), genes envolvidos com a apoptose (*Bcl2*) e outros genes, envolvidos com diferentes vias do metabolismo da célula (*RARâ, FHIT, CYP1A1, P450, PTGS2, EDNRB, SOCS1*, entre outros) (OLIVEIRA, 2011).

Dentro do CP, o câncer de pulmão de não pequenas células (NSCLC) é responsável por mais de 80% dos casos e é caracterizado como incidental e agressivo geralmente diagnosticado em estágio avançado e com baixa taxa de sobrevida; essa categoria ampla abrange vários subtipos, dos quais o mais prevalente é o adenocarcinoma (OLIVEIRA et al., 2019).

Os atributos das células neoplásicas podem ser descritos pelos *hallmarks* do câncer, conceito introduzido por Hanahan e Weinberg (2022), referem-se a características biológicas fundamentais que permitem às células tumorais proliferar de maneira descontrolada e sustentar o crescimento tumoral. Entre os hallmarks clássicos, destacam-se a evasão da morte celular programada, a ativação de vias de sinalização para a proliferação celular, a promoção da angiogênese e a capacidade de invasão e metástase (SAMENTINO et a., 2024) (Figura 1). No câncer de pulmão, essas características estão frequentemente associadas a mutações em genes como KRAS, EGFR, TP53 e ALK, que impulsionam o crescimento tumoral e a resistência ao tratamento.

**Figura 1** – *Hallmarks* do câncer.




**Fonte:** Hanahan, 2022.

Alterações genéticas e epigenéticas influenciam na iniciação e no desenvolvimento do tumor, incluindo mutações, instabilidade cromossômica, expressão anormal de RNAs não codifcantes (ncRNAs) e evidências sugerem que disbioses na microbiota modulam a suscetibilidade a neoplasias malignas através da liberação de metabólitos bacterianos promotores de câncer e indutores de vias inflamatórias do hospedeiro regulando a expressão de mediadores e citocinas (como por exemplo, IL-1, IL-23, TNF e IL-17) que agem nem padrões moleculares associados a patógenos (PAMPs), esses mediadores ou citocinas desencadeiam

vias de sinalização críticas (por exemplo, vias STAT3 e NF-kB e vias ERK e PI3K), o que promove a carcinogênese das células hospedeiras (GONZALES et al., 2008; WILLIAMS, T et al., 2017; WILLIAMS, M et al., 2017; RATTI ET et al., 2020; XU et al., 2020).

RNA não codificante (ncRNA) pode ser definido como sequências genéticas não relacionadas à codificação de proteínas que atuam como fatores transcricionais e pós-transcricionais, interferindo e regulando a expressão proteica (STATELLO et al., 2021). Desde a sua descoberta, os ncRNAs emergiram como importantes reguladores de múltiplas funções biológicas em uma variedade de tipos de células e tecidos, e sua desregulação tem sido implicada em doenças, especialmente no câncer (NEMETH et al., 2023). Os ncRNAs com funções reguladoras são divididos em duas categorias por tamanho: RNAs não codificantes curtos (sncRNA) e RNAs não codificantes longos (lncRNAs), este último são maiores que 200 nt (HE et al., 2023).

No campo de estudo do RNA, o Sequenciamento de Nova Geração (NGS) trouxe uma mudança de paradigma na pesquisa genômica, oferecendo capacidades incomparáveis para análise da molécula de maneira econômica e de alto rendimento. O NGS não permitiu apenas o sequenciamento abrangente do genoma, mas também facilitou estudos de transcriptômica, epigenômica, metagenômica e outros estudos ômicos e, com isso foi evidenciado que o genoma é maioritariamente transcrito em RNA mais longos (CHEN et al., 2024; MATTICK et al., 2023; SATAM et al., 2023).

A microbiota, por sua vez, corresponde a uma população de microrganismos que vivem biologicamente em todas as superfícies mucosas. Predominantemente, os estudos focam na microbiota intestinal e de outros órgãos, uma vez que o pulmão era considerado estéril devido às dificuldades de cultivo do tecido pulmonar. No entanto, com o desenvolvimento das tecnologias multiômicas, hoje sabe-se que a microbiota pulmonar é única e transitória, e está associada ao desenvolvimento não só de doenças infecciosas, mas de outras patologias incluindo doença pulmonar obstrutiva crônica, fibrose cística, asma e o câncer de pulmão (ALLEGRA et al., 2020; MADDI et al., 2019).

Em humanos, o pulmão é colonizado, predominantemente por bactérias dos gêneros *Prevotella, Rothia* e *Veillonella*, que são correlacionadas positivamente com os níveis de múltiplas citocinas TH17, como IL-1α, IL-1β, IL-6, e IL-17, e espécies de Proteobacteria que são associadas à ativação de vias associadas ao TH1711 (NATALINI, SINGH & SEGAL, 2023).

Os mecanismos microbianos na etiopatogenia do câncer são múltiplos, dentre eles foi identificado uma interação contínua e bidirecional entre o microbioma e a expressão de

ncRNAs (RATTI et al., 2020). Sabe-se que cerca de 98% dos transcritos do genoma humano correspondem a ncRNAs (PALAZZO et al., 2015). Em estudos recentes baseados em ômicas, os ncRNAs, mostraram-se como moduladores de doenças, como o CP, em resposta à microbiota e sua disbiose (MALMUTHUGE & GUAN, 2021). Nesse contexto, especificamente os lncRNAs oferecem vantagens sobre as proteínas em algumas formas de regulação epigenética, como a facilidade de se ligarem ao sítio de transcrição, promovendo especificidades regulatórias (LEE, 2012).

O RNA desempenha múltiplas funções cruciais nas células além da síntese de proteínas, entretanto, os RNAs codificantes de proteínas são apenas 3% do genoma humano, são eles: RNA mensageiro (mRNA), fundamental para a tradução de informações genéticas do DNA em proteínas nos ribossomos; RNA ribossômico (rRNA), que compõe a estrutura do ribossomo e o RNA transportador (tRNA) que identifica e transporta os aminoácidos até o ribossomo (SCHMITT & CHANG HY, 2016; MA, 2024).

O sequenciamento de RNA total (RNA-seq) tem sido uma ferramenta valiosa na investigação da relação microbiota-ncRNA, no qual através da caracterização dos transcritos encontrados em uma amostra, sejam eles de origem humana ou não, é possível identificar e quantificar a expressão gênica e assim estabelecer as correlações através da caracterização do perfil de expressão gênica global e perfis taxonômicos e funcionais da microbiota (DESHPANDE D et al., 2023; ZHANG et al., 2019).

Avanços recentes nos trabalhos com RNA-Seq permitiram aos pesquisadores elucidar ainda mais a complexidade funcional da transcrição, haja vista que, além dos transcritos de mRNA, o RNA-Seq pode ser aplicado para investigar diferentes populações de RNA, incluindo lncRNA (KUKURBA & MONTGOMERY, 2015).

A identificação da causa subjacente do início e progressão do CP é um passo imperativo no desenvolvimento de novas terapias para tratar este tipo de malignidade; além disso, pode facilitar a concepção de novos biomarcadores para a detecção precoce do câncer e os lncRNAs são transcrições promissoras para ambos os propósitos. Alterações na expressão de lncRNAs podem ser representativas em determinadas fases da progressão tumoral e podem ser usadas para prever o crescimento precoce do câncer ou a indução de vias de sinalização (TAHERI et al., 2023; XIA et al., 2022; SUN et al., 2022). Estudos baseados em ômicas revelaram que ncRNAs modulam fenótipos de doenças em resposta à microbiota intestinal/metabólitos microbianos durante patologias ligadas ao microbioma, como câncer (DESHPANDE D et al., 2023; MA, 2024).

Os dados sobre os papéis dos lncRNAs nas interações hospedeiro-microbiota foram

gerados principalmente por modelos murinos tratados com antibióticos e sem germes, bem controlados (usando cepas consanguíneas criadas em ambientes subcontrolados). Embora esses estudos tenham mostrado o impacto do microbioma em lncRNAs reguladores, sistemas imunológicos e no desenvolvimento de órgãos, seus resultados não refletem diretamente a complexidade (genética, ambiente, comorbidades) que existe em outras espécies de mamíferos (COSTA et al., 2010; DESHPANDE D et al., 2023; MA, 2024).

A microbiota de tumores metastáticos, evidenciou que camundongos com tumor pulmonar, após o transplante de microbiota livre de patógenos alterou os perfis de lncRNA na microbiota do tumor pulmonar, inibindo o seu crescimento (COSTA et al., 2010). Embora haja poucos estudos que investiguem as relações bidirecionais existentes entre a microbiota e os lncRNAs no CP, hoje sabe-se que há diferenças nas comunidades bacterianas entre indivíduos saudáveis e pacientes com câncer e a carga bacteriana pulmonar desempenha um papel chave na carcinogênese e na resposta imune contra células cancerosas, influenciando na sensibilidade à quimioterapia e imunoterapia (SANCHEZ-MARTÍNEZ ET AL., 2014).

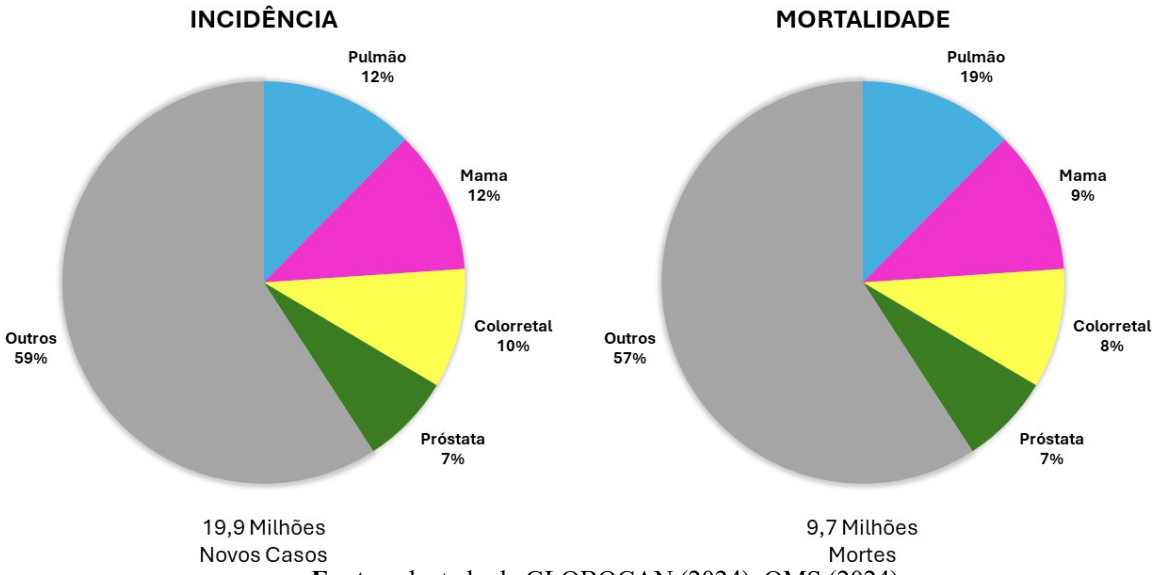
## 1.2 CÂNCER DE PULMÃO

O CP é a doença maligna mais comum em todo o mundo desde 1985, tanto em incidência quanto em mortalidade (OMS, 2020). Segundo a Organização Mundial da Saúde (2022), estima-se que em cerca de 2.2 milhões de casos tenham ocorrido em 2020 e a incidência cresça 2,0% a cada ano.

Nos últimos cinco anos, a sobrevida para essa doença variou entre 13,0% e 21,0% em países desenvolvidos e entre 7,0% e 10,0% nos países em desenvolvimento (INCA, 2023). Entretanto, estes números são largamente subestimados devido a uma elevada taxa de subdiagnóstico e subnotificação (MATHIAS et al., 2020).

A idade média do diagnóstico é de 70 anos. Os homens têm duas vezes mais probabilidades de serem diagnosticados com CP, o que reflete, em grande parte, ao consumo maior de tabaco, embora as mulheres possam ser mais susceptíveis devido a proporções mais elevadas de mutações nos receptores do fator de crescimento epidermal (EGF) e aos efeitos do estrogênio (THANDRA et al., 2021).

**Figura 2** – Incidência e mortalidade de câncer no mundo em 2024.




**Fonte**: adaptado de GLOBOCAN (2024); OMS (2024).

Fatores genéticos, ambientais e ocupacionais estão associados ao aumento na incidência e mortalidade do CP. Cerca de 8% dos casos é consequência de fatores genéticos, 80-90% é decorrente do tabagismo, seja exposição ativa ou passiva, além disso, a exposição ao radônio, riscos ocupacionais, fumaça de combustível doméstico e doenças infecciosas como tuberculose também desempenham um papel chave no desenvolvimento do câncer (CORRALES et al., 2020; SCHABATH; COTE, 2019; SHANKAR et al., 2019).

Mutações de CP podem ser identificadas em mutações no gene KRAS, no receptor do fator de crescimento epidérmico (EGFR), BRAF, oncogenes da via paralela da fosfatidilinositol 3-quinase (PI3K) e, mais recentemente, em MEK e HER2 (COOPER et al., 2013).

A maioria dos polimorfismos identificados no CP são de genes que codificam proteínas associadas à atividade de metabolizar carcinógenos da fumaça do tabaco e de suprimir mutações induzidas por esses carcinógenos (YOKOTA, SHIRAISHI & KOHNO, 2010).

Um estudo feito em 2017 por Joyce et al., referente a uma metanálise com 99 estudos de coorte constatou que o hábito de fumar, associado a idade do indivíduo, está atrelado a um risco maior de CP tanto em homens (7,48) quanto em mulheres (5,30); ainda o risco de CP aumenta de acordo com a carga tabágica.

Baseado na histologia, O CP pode ser dividido em três categorias: câncer de pulmão de pequenas células (SCLC) e o câncer de pulmão de não pequenas células (NSCLC). Sendo NSCLC o mais prevalente, cerca de 80% dos casos, dos quais aproximadamente 40% são adenocarcinoma (LUAD), 25 a 30% são carcinoma de células escamosas (LUSC), 10 a 15% são carcinomas de células grandes e o restante são outros subtipos histológicos menos comuns

(AHN et al.,2020).

Aproximadamente 75% de todos os NSCLC sofrem duplicação do genoma, um evento importante que permite que as células cancerosas tolerem, posteriormente, a instabilidade cromossômica. Os eventos causadores de câncer que ocorrem no início do desenvolvimento celular são clonais; eles são iniciadores de tumores, frequentemente associados a carcinógenos do tabagismo e são principalmente específicos para um determinado subtipo histológico como, por exemplo, no adenocarcinoma, onde estas mutações precoces ocorrem em BRAF, EGFR ou MET, enquanto no carcinoma escamoso existem mutações NOTCH1 ou FGFR1. As mutações TP53 são principalmente clonais e ocorrem em ambos os subtipos (NAGL et al., 2022; TAN et al., 2022).

Durante a maturação do mRNA, ocorre um processo em que os íntrons são removidos do pré-mRNA e os éxons são unidos para formar o mRNA maduro, isso acontece no núcleo da célula e é realizado por um complexo chamado spliceossoma, recebendo o nome de *splicing* do RNA (YOKOTA, SHIRAISHI & KOHNO, 2010). O *splicing* alternativo (AS), a seleção e uso de diferentes locais de *splicing*, ocorre frequentemente devido à regulação do mesmo e suas alterações em genes estão relacionados ao CP, especialmente no receptor do fator de crescimento epidérmico (EGFR), no receptor 2 do fator de crescimento de fibroblastos (FGFR2) e no cluster de diferenciação 44 (CD44), em ambos os tipos de NSCLC (YAN et al., 2023).

Embora os NSCLCs estejam fortemente associados ao uso de cigarro, os adenocarcinomas podem ser encontrados em pacientes que nunca fumaram; como este tipo de CP é menos sensível à quimioterapia e à radioterapia, ele apresenta maior índice de mortalidade em todo o mundo; os principais esforços genômicos elucidaram seus cenários de mutação e expressão gênica, e essas descobertas foram traduzidas em terapias direcionadas eficientes (AHN et al.,2020).

Com múltiplos subtipos, o CP apresenta alta taxa de heterogeneidade inter e intratumoral, resultante do acúmulo de alterações genéticas e epigenéticas que influenciam na iniciação e no desenvolvimento do tumor, incluindo alterações nos padrões de expressão gênica, mutações *drivers*, variação no número de cópias e assinaturas genômicas. A detecção dessas alterações permitiu traçar um perfil molecular do tumor, que por sua vez melhorou a estratificação terapêutica e auxiliou a definir uma abordagem considerando cada paciente como único (DESHPANDE D et al., 2023; WANG et al., 2019). A Tabela 1 descreve as principais alterações moleculares caracterizadas no NSCLC.

**Tabela 1** – Principais alterações moleculares relatadas no NSCLC.

| Gene | Tipo de alteração | Frequência em NSCLC |
|---|---|---|
| *EGFR* | Mutação | 10-35% |
| *KRAS* | Mutação | 15–25% |
| *FGFR1* | Amplificação | 20% |
| *PTEN* | Mutação | 4-8% |
| *DDR2* | Mutação | ~ 4% |
| *ALK* | Rearranjo | 3-7% |
| *HER2* | Mutação | 2–4% |
| *CONHECEU* | Amplificação | 2–4% |
| *BRAF* | Mutação | 1–3% |
| *PIK3CA* | Mutação | 1–3% |
| *AKT1* | Mutação | 1% |
| *MEK1* | Mutação | 1% |
| *NRAS* | Mutação | 1% |
| *RET* | Rearranjo | 1% |
| *ROS1* | Rearranjo | 1% |

**Fonte**: adaptado de SCHABATH e COTE (2019).

Estratégias terapêuticas para o tratamento do CP devem considerar alterações genéticas precoces para promover o seu reparo ou eliminar as células tumorais (DUARTE & PASCHOAL, 2006). Tradicionalmente, para estágios avançados, a quimioterapia era a principal opção, mas o índice de cura era baixo e os efeitos colaterais enormes. A identificação de alvos terapêuticos moleculares norteou o desenvolvimento de terapias direcionadas e imunoterapia, que reduziram as comorbidades associadas aos tratamentos e aumentaram a sobrevida global (SHAH & MASTERS, 2020).

O primeiro alvo molecular identificado foi o EGFR, onde cerca de 10-35% dos casos de NSCLC abrigam mutações e essas alterações estão associadas a sensibilidade a inibidores tirosina quinase (TKI), como gefitinibe ou erlotinib. EGFR-TKIs melhoraram a sobrevivência em comparação com a quimioterapia em pacientes avançados, no entanto, mesmo com a boa taxa de resposta, a resistência à terapia é comum, e de 9 a 13 meses a doença costuma progredir (YAN et al., 2023). A mutação pontual na posição 790 envolvendo a substituição de treonina por metionina (T790M) é um dos principais mecanismos de resistência, no qual Osimertinibe (medicamento contra o câncer que inibe a tirosina quinase) mostrou excelente eficácia (SHA; MASTERS, 2020; CASTELLANOS et al., 2017).

Atualmente, o arsenal de estratégias de tratamento abrange remoção cirúrgica, quimioterapia, terapia direcionada ou terapia alvo-específica e radioterapia. No entanto, o prognóstico permanece lamentavelmente ruim, com uma baixa taxa de sobrevivência de 5 anos (LI, YAN & HE, 2023). A eficácia clínica é frequentemente limitada pela resistência inata e

adquirida, permitindo a progressão e recorrência do tumor resultante da ausência de sintomas nos estágios iniciais da doença, diagnóstico tardio, resistência a medicamentos, e alta heterogeneidade inter e intratumoral (AHN et al., 2020; BENUSIGLIO et al., 2021).

É imperativo compreender melhor os fatores moleculares da tumorigênese, metástase e resistência à terapia no câncer de pulmão para desenvolver estratégias terapêuticas melhoradas. Os lncRNAs estão emergindo como moléculas importantes no câncer devido às suas funções oncogênicas ou supressoras tumorais (GENCEL-AUGUSTO, WU & BIVONA, 2023). Uma abordagem personalizada baseada na compreensão da tumorigênese sob aspectos genéticos e epigenéticos é necessária para mudar esse cenário. Destaca-se assim, o papel da medicina genômica e epigenômica nas investigações da biologia do câncer, determinação dos mecanismos de resistência, assim como no desenvolvimento de novas opções terapêuticas (ALLEGRA et al., 2020; SHAH; MASTERS, 2020).

### 1.2.1 Estadiamento NSCLC

O estadiamento é uma etapa crítica na avaliação de pacientes com CP onde ocorre a identificação precisa do estágio com base nas características do tumor primário (T), nódulos regionais (N) e doença metastática (M), o chamado critério anatômico do sistema TNM, tem como finalidade de determinar o cuidado apropriado (TANOUE, 2020).

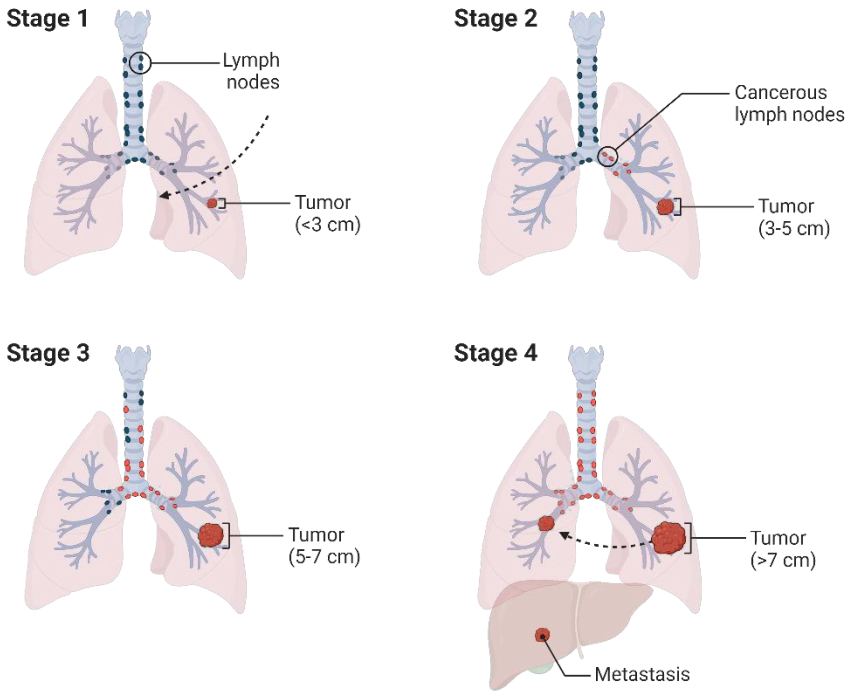
Esta etapa não inclui outros fatores prognósticos como condição cardiorrespiratória, idade, tipo histológico ou invasão extracapsular e reflete a possibilidade de remover cirurgicamente a neoplasia (COSTA et al., 2020).

Independente da fase em que o câncer é detectado, há necessidade de realizar o estadiamento, este pode ser clínico ou patológico: o clínico é baseado na avaliação da extensão do câncer feita antes da cirurgia ou outros tratamentos invasivos, utilizando informações obtidas através de exames físicos, exames de imagem, e outros testes diagnósticos e o patológico é realizado após a remoção cirúrgica do tumor ou de amostras de tecido (biopsias) e é baseado na análise histopatológica dessas amostras (INCA, 2023; SILVA et al., 2021).

O grau de extensão da doença é habitualmente classificado em estádio precoce (I e II), localmente avançado (III) e avançado/metastático (IV). Esses parâmetros recebem graduações, geralmente de T0 a T4; N0 a N3; e de M0 a M1. Via de regra, quanto menor o número do estádio, menos metástase houve (INCA, 2023; FENG & YANG, 2019a) (Figura 3).

Tal classificação é padronizada internacionalmente pela *Union for International Cancer Control* (UICC) e teve sua última atualização em 2017 (8ª edição).

**Figura 3** – Estádio/grau de extensão NSCLC de acordo com a UICC.




**Fonte:** adaptado de BIORENDER, 2024; MINISTÉRIO DA SAÚDE, 2022.

Esta última edição trouxe algumas modificações importantes como a classificação T com base no incremento de 1 cm, *downstage* do descritor T incluindo tumor endobrônquico desconsiderando sua distância da carina (T2), mesclando atelectasia/pneumonite total e parcial na mesma Categoria T (T2), invasão diafragmática ascendente para T4, novo conceito de classificação de adenocarcinoma in situ e adenocarcinoma minimamente invasivo para nódulos em vidro fosco puros e parcialmente sólidos e divisão adicional de metástase extratorácica em M1b e M1c com base no número e locais (FENG & YANG, 2019b; RAMI-PORTA et al., 2017).

A classificação N é categorizada com base na localização anatômica e não inclui a quantificação dos linfonodos envolvidos e é categorizado como N0 (sem linfonodos envolvidos), N1 (envolvimento ipsilateral de linfonodos peribrônquicos, hilares, interlobares ou intrapulmonares), N2 (envolvimento ipsilateral de linfonodos mediastinais ou subcarinais) ou N3 (mediastinal contralateral, hilar contralateral, escaleno ipsilateral ou contralateral ou envolvimento do nódulo supraclavicular) (HUANG et al., 2024).

Por fim, a classificação indica Metástases a distância, sendo: M1 Metástase a distância presente; M1a, nódulo (s) tumoral (ais) em lobo contralateral, tumor com nódulo pleural ou

pericárdico ou derrame pleural ou pericárdico maligno; M1b, metástase a distância única (fora do tórax) e M1c, múltiplas metástases extratorácicas em um ou mais órgãos (COSTA et al., 2020; SILVA et al., 2021).

No NSCLC, o estadiamento preciso é fundamental, pois é o preditor mais importante de sobrevivência e determina as opções de tratamento, incluindo a possibilidade de ressecção cirúrgica potencialmente curativa. Para isso, é necessário que haja uma combinação cuidadosa de métodos de imagem e métodos invasivos que, geralmente inicia com a tomografia computadorizada de tórax, para identificação do tumor primário. A Tomografia por Emissão de Pósitrons/Tomografia Computadorizada (PET-CT), fornece informações funcionais com maior sensibilidade e especificidade permitindo a identificação de locais metastáticos extratorácicos. Apesar disso, os resultados positivos no PET-CT devem ser confirmados por amostragem de tecido, bem como envolvimento dos linfonodos mediastinais (RUESCHHOFF, MOORE & JASAHUI, 2024).

### 1.2.1 Microbiota

O organismo humano é colonizado por uma complexa diversidade de microrganismos. O conjunto destes microrganismos, que inclui as bactérias, fungos, vírus e protozoários, é conhecido por microbioma humano em *Homo sapiens* (ALLEGRA et al., 2020; LIU et al., 2021; MADDI et al., 2019). Cada órgão possui uma comunidade microbiana específica que pode favorecer o desenvolvimento de diversas patologias, bem como auxiliar o corpo em resposta a elas (ALLEGRA et al., 2020; MADDI et al., 2019).

As comunidades microbianas nas doenças respiratórias são diferentes das dos indivíduos saudáveis, e o microbioma pulmonar pode não só influenciar a suscetibilidade ou as causas das doenças respiratórias, mas também ser afetado pelas atividades patológicas das doenças respiratórias, bem como nas respostas ao tratamento (YAGI et al., 2021). Nesse cenário, a microbiota pulmonar está correlacionada com o início e progressão do CP, por isso investigar os mecanismos carcinogênicos mediados por microorganismos implica na compreensão da correlação à causalidade, haja vista que eles intervêm nos processos fisiológicos, incluindo a modulação imunológica, metabolismo e inflamação (LIU et al., 2024).

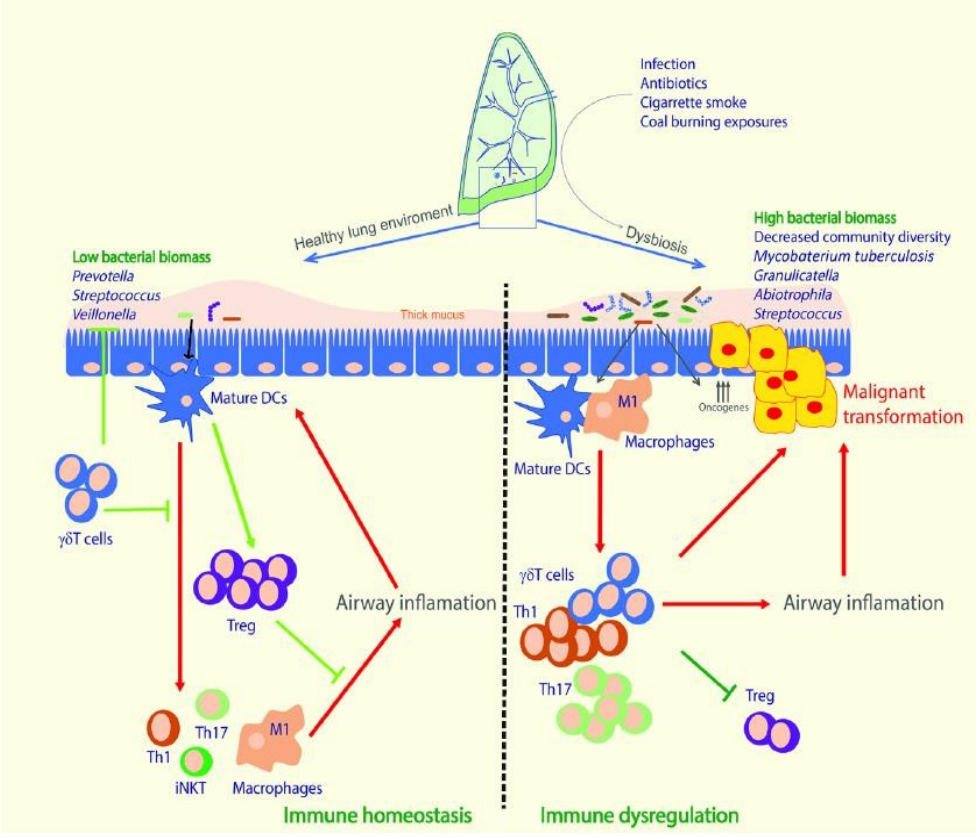
A relação entre microbiota pulmonar, sistema imunológico pulmonar e o microambiente do câncer de pulmão é complexa. A comunicação é mediada por metabólitos, padrões moleculares associados a microrganismos (MAMPs), receptores de reconhecimento de padrões (PRRs), mediadores inflamatórios, fatores de crescimento e disponibilidade de nutrientes

(HUYNH, CRANE & JAMIESON, 2023).

Em um pulmão saudável, o microbioma é composto por bactérias, vírus e fungos. *Streptococcus, Veillonella* e *Prevotella* são os gêneros mais comuns de bactérias, enquanto *Haemophilus* são exclusivos do pulmão como habitantes residentes e são raros em outros microbiomas. Os fungos são menos numerosos, com predominância de Candida, seguida por *Saccharomyces* e *Penicillium*. As famílias de vírus que dominam são *Paramyxoviridae, Picornaviridae* e *Orthomyxoviridae* (LI & ZHOU, 2024; ZERDAN et al., 2022).

Estudos em modelos animais demonstraram que a carga bacteriana nos pulmões aumenta durante as primeiras 2 semanas de vida. Acredita-se que tais alterações na microbiota estejam associadas ao acúmulo de uma população de células T reguladoras dependentes de PD-L1 que promove tolerância a alérgenos ambientais. Portanto, a microbiota pulmonar seria ser um evento-chave no início da vida, necessário para moldar o sistema imunológico pulmonar e proteger o órgão de respostas inflamatórias prejudiciais a antígenos inalados. Este processo inicial pode promover tolerância da microbiota comensal, evitando danos pulmonares relacionados à inflamação e, provavelmente, carcinogênese pulmonar (RAMÍREZ-LABRADA et al., 2020) (Figura 4).

**Figura 4** – Relação entre microbiota pulmonar, homeostase e câncer de pulmão.




**Fonte:** RAIREZ-LABRADA et al., 2020.

A composição diversa e equilibrada da microbiota promove homeostase no organismo por ter função de barreira e estimuladora da resposta imune. Portanto, em um quadro onde esta homeostase se apresenta comprometida pelo desiquilíbrio da microbiota, pode facilitar a entrada de produtos tóxicos e endotoxinas, tais como o lipopolissacarídeos (LPS), os quais podem gerar inflamação; existem outras associações em estudos que relacionam a ativação de células via receptores toll-like (TLR) por LPS circulantes no plasma sanguíneo, com algumas metástases na tireoide, no câncer colorretal e no câncer de mama (MADDI et al., 2019).

A inflamação e os processos infecciosos correspondem a 25% dos fatores causadores de câncer, em processos inflamatórios crônicos desregulações na microbiota desempenham um papel chave na manutenção do estado de inflamação, por meio da produção de espécies reativas de oxigênio/nitrogênio (ROS/RNS) e exposição a toxinas microbianas que podem causar danos diretos no DNA, resultando em mutações que geram em respostas imunológicas ineficientes e garantem vantagens proliferativas às células tumorais (CASTELLANOS et al., 2017; QIU et al., 2017; TIANSHENG et al., 2020).

No CP, o equilíbrio entre imigração e eliminação de microorganismos encontra-se perturbado. Assim, a microbiota pulmonar é alterada, a abundância de bactérias simbióticas diminui e predominam as bactérias patogênicas. Esta alteração leva a uma diminuição na diversidade da microbiota pulmonar e está associada à progressão tumoral (XU et al., 2020).

De acordo com as análises taxonômicas, o microbioma pulmonar em pacientes com CP possui a presença de *Proteobacteria, Firmicutes, Acinobacteria, Bacteroides* e *Verrucomicrobia*. Destes, *Firmicutes* é o mais abundante (YAGI et al., 2021). Huang et al. (2015), documentaram a correlação entre o microbioma pulmonar e a histologia, bem como o risco de progressão da doença. Streptococcus foi significativamente menor no adenocarcinoma metastático do que no adenocarcinoma não metastático, e *Veillonella* e *Rothia* foram maiores no carcinoma de células escamosas metastático em comparação com o não metastático.

A oncologia clínica avança no que diz respeito ao conhecimento das correlações entre microorganismos e CP. Segundo Costa et al. (2018), marcadores moleculares como EGFR, proteína de morte celular programada e quinase do linfoma anaplásico já são descritos na literatura e o novo alvo terapêutico é o microbioma pulmonar que pode se apresentar como possível marcador de doença.

Embora muitos oncogenes tenham sido identificados como condutores endógenos do CP, tal como KRAS no NSCLC, o papel da microbiota na manipulação de vias estabelecidas de oncogênese ainda não é tão elucidado. Embora uma grande variedade de microrganismos

bacterianos, virais e fúngicos estejam associados a tumores pulmonares, os microrganismos que causam oncogênese direta no câncer de pulmão não foram estabelecidos na literatura (HUYNH, CRANE & JAMIESON, 2023; ZAKI et al., 2021).

Portanto, há uma série de mecanismos plausíveis e possíveis pelos quais o microbioma pode influenciar o CP, um estudo extensivo do papel dessas bactérias pode fornecer alvos terapêuticos para a sua prevenção e manejo.

### 1.2.4 Métodos de investigação e análise do microbioma humano

O estudo do microbioma humano — e, mais recentemente, o do sistema respiratório — através de sofisticadas técnicas de biologia molecular, desvendou a imensa diversidade de colonização microbiana nos seres humanos, sejam saudáveis, sejam portadores de diferentes doenças (COSTA et al., 2017).

Muito do atraso no conhecimento da microbiota pulmonar se deve à dificuldade de se representar o habitat do pulmão humano através de técnicas de cultura convencional, haja vista que as amostras do trato respiratório requerem consideração especial durante a amostragem, processamento e análise, pois o trato respiratório inferior possuiu um sistema de baixa biomassa microbiana o que induz a resultados falso-positivos. Este erro é introduzido por sinais bacterianos da orofaringe durante a amostragem, por contaminação técnica e estocasticidade durante o sequenciamento, pode impactar a interpretação e as conclusões dos estudos do microbioma pulmonar (CARNEY et al., 2020; COSTA et al., 2017).

Os estudos que investigam o microbioma pulmonar utilizam diferentes tipos amostrais, dentre eles: amostras de escarro, líquido de lavagem broncoalveolar (BALF) e biópsia sólida do tecido pulmonar, além de diferentes métodos de investigação e técnicas de análise computacional (FENN et al., 2022; PORTEZAN & CAPEL, 2021).

Apesar dos avanços nas técnicas de microbiologia laboratorial, ainda existem limitações significativas na caracterização de microbiomas por métodos de cultivo, uma vez que os meios de cultura tradicionais não são capazes de suprir as necessidades específicas de todos os microrganismos presentes nas amostras (HA & DEVKOTA, 2020).

As biópsias pulmonares (Solid Biopsy) são descritas na literatura como um procedimento seguro, com sensibilidade e especificidade acima de 90% em casos de definição do tipo de tumor frente a lesões brônquicas (NUNES et al., 2023). Biópsias pulmonares para estadiamento de câncer de pulmão tem como principal benefício a exclusão de contaminações orais, por isso são a melhor fora de coleta de amostra para estudo de microbioma pulmonar.

Com os constantes avanços das tecnologias de NGS, novas abordagens emergiram para caracterizar, de forma taxonômica e genética, o microbioma humano e ambiental. Nesse contexto, a metagenômica foi a primeira técnica a integrar a genômica moderna ao estudo de comunidades microbianas diretamente em seus ambientes naturais, sem a necessidade de isolamento ou cultivo das espécies individualmente em laboratório.

Há diferentes estratégias de sequenciamento e as abordagens bioinformáticas usadas para estudar o microbioma humano, as principais incluem o sequenciamento de 16S rRNA, a metagenômica e a metatranscriptômica, cada uma com suas vantagens, limitações e aplicações específicas.

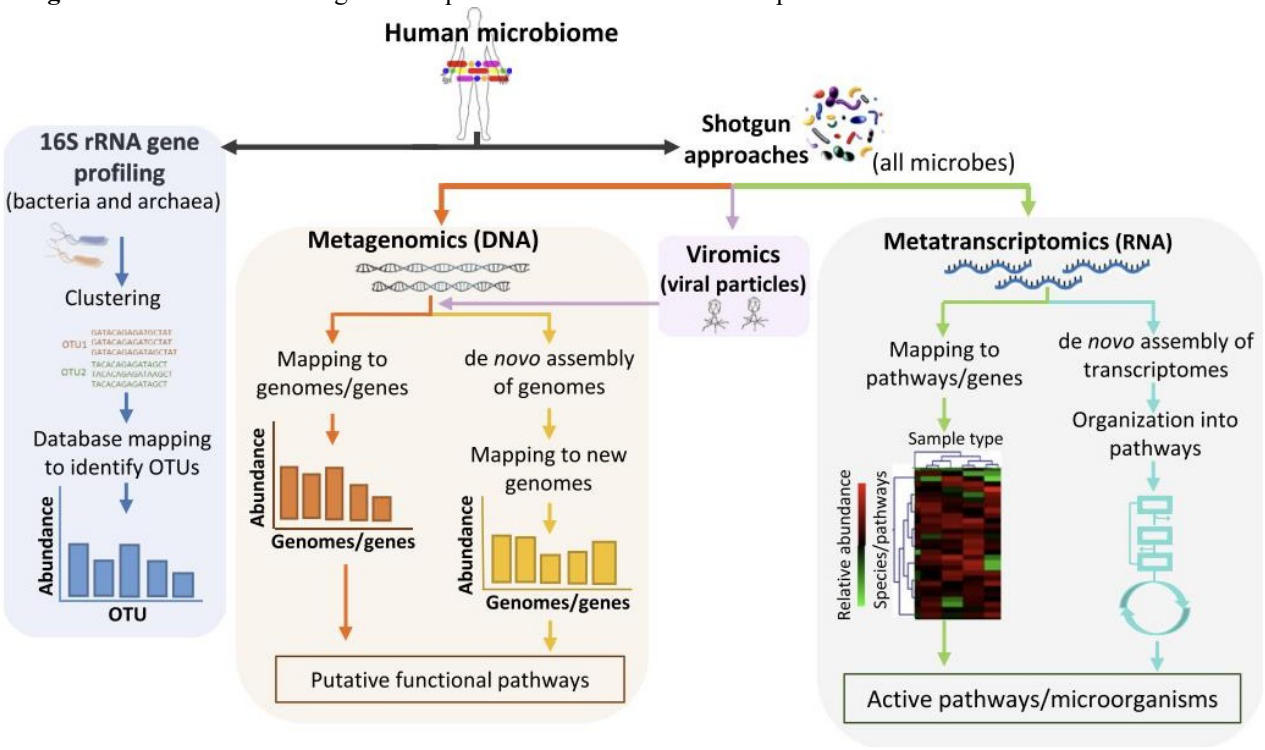
O sequenciamento de 16S rRNA é uma das abordagens mais amplamente utilizadas para caracterizar a composição taxonômica do microbioma. Ele se baseia na amplificação e sequenciamento de uma região conservada do gene que codifica o RNA ribossômico 16S, que está presente em todas as bactérias e arqueias, mas com variação suficiente para permitir a identificação de diferentes espécies (FROMENTIN et al., 2022; ROGERS, 2017).

A metagenômica envolve o sequenciamento direto do DNA total de uma amostra ambiental (como amostras de microbioma). Isso permite a identificação de todos os tipos de microrganismos presentes, incluindo bactérias, arqueias, vírus, fungos e outros eucariotos. Ao contrário do sequenciamento de 16S rRNA, a metagenômica fornece uma visão mais ampla, incluindo tanto a diversidade taxonômica quanto a funcionalidade do microbioma (CHEN et al., 2023; HU et al., 2020; RITCHIE & SINGANAYAGAM, 2020).

A metatranscriptômica vai além da análise do DNA do microbioma e se concentra na análise dos transcritos (mRNA) presentes nas amostras. Esta abordagem fornece uma visão sobre a expressão gênica ativa dos microrganismos, ou seja, sobre os genes que estão sendo transcritos em um determinado momento. Ao sequenciar os RNAs presentes em uma amostra, podemos estudar a atividade metabólica do microbioma e entender melhor como ele interage com o hospedeiro (CHANG et al., 2021; WANG et al., 2023).

A Figura 5 apresenta um panorama das principais técnicas computacionais utilizadas para análise do microbioma humano. Tendo como principal destaque, a metatranscriptômica.

**Figura 5** – Diferentes estratégias de sequenciamento e bioinformática para análise do microbioma humano.




**Fonte:** BIKEL et al., 2015.

## 1.3 RNAS LONGOS NÃO CODIFICANTES (LNCRNA)

### 1.3.1 Características e funções

A identificação de RNAs que não são traduzidos em proteínas foi um importante avanço haja vista que passou a definir a diversidade de moléculas envolvidas na regulação eucariótica da expressão gênica (FERNANDES et al., 2019). Nesse contexto, os lncRNAs oferecem benefícios e vantagens quando comparados as proteínas no que diz respeito a algumas formas de regulação epigenética, como a capacidade de se ligarem ao sítio de transcrição, permitindo uma determinada especificidade regulatória (LEE, 2012).

Definidos como moléculas com mais de 200nt, os lncRNAs, embora já tenham sido considerados subprodutos da transcrição da RNA polimerase II sem funções biológicas, foram considerados um grande repertório que servem para regular processos celulares, como modificação de cromossomos e genoma, ativação e interferência de transcrição e transporte nuclear (BHAT et al., 2016; CHEN et al., 2021).

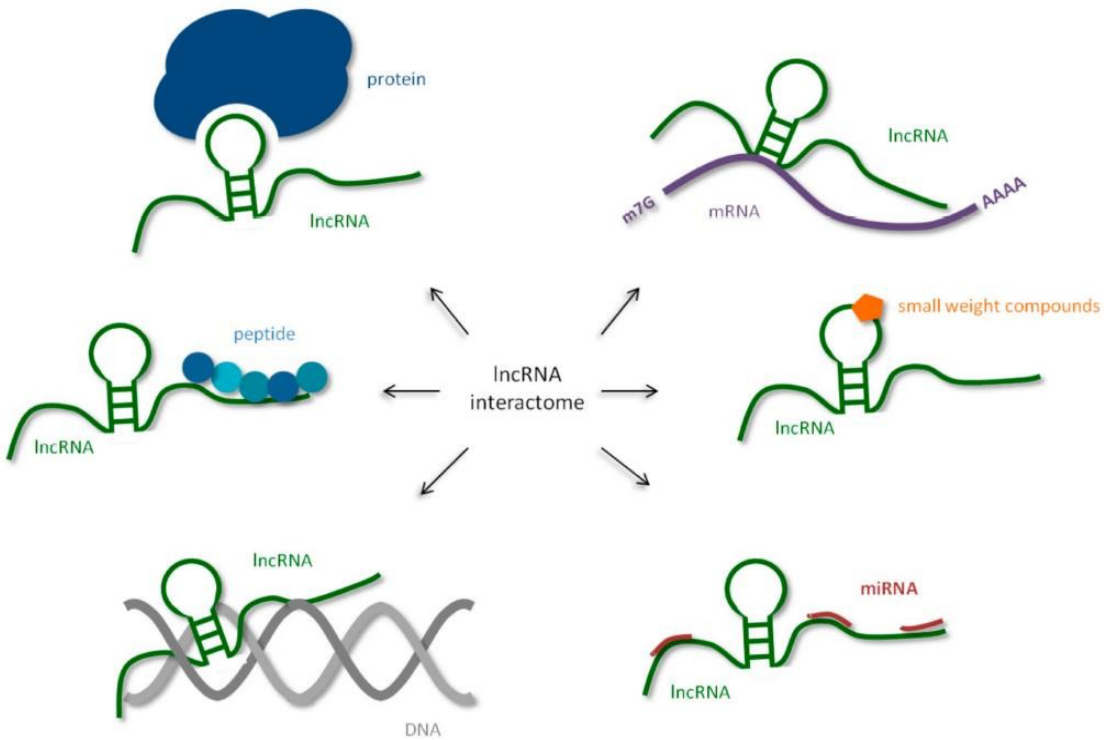
A natureza molecular dos lncRNAs varia e tem o potencial de influenciar a sua função e localização. Aproximadamente 50% dos lncRNAs possuem uma cauda poliA e 98% dos lncRNAs humanos são semelhantes aos mRNAs. Estas características semelhantes às do

mRNA contribuem para a capacidade de muitos lncRNAs saírem do núcleo e entrarem em vias nas quais os mRNAs participam (TSAGAKIS et al., 2020). Porém, ao contrário de outros mRNAs codificadores de proteínas, os lncRNAs exibem singularidade funcional ao participar e modular os vários processos celulares, como modificação de histonas, metilação do DNA e transcrição celular. Existem quatro principais tipos de interação (DAHARIYA et al., 2019).

LncRNAs interagem com mRNA (um tipo de RNA codificante) e ncRNA; no caso da interação lncRNA-mRNA, o lncRNA regula a expressão do mRNA direta ou indiretamente, além de ter como alvo direto o mRNA para regular seu *splicing*, edição, distribuição subcelular e estabilidade. Ainda, estes dois podem fazer *crosstalk* através de elementos responsivos ao micro RNA (miRNA) compartilhados, formando redes concorrentes de RNA endógeno (ceRNA) (CAO et al., 2017; ZHANG, JIA & KWOH, 2021).

No caso da interação lncRNA-ncRNA, o lncRNA atua como precursores de pequenos ncRNAs como miRNA, snoRNA e piRNA. Além disso, na associação entre lncRNA e miRNA, este último regula a expressão e estabilidade do lncRNA (STATELLO et a., 2021; ZHANG, JIA & KWOH, 2021; ZHENG, 2019) (Figura 6).

**Figura 6** – Interação lncRNA humano com biomoléculas celulares.




**Fonte:** KAZIMIERCZYK et al., 2020.

Estudos demonstram que os lncRNAs desempenham um papel significativo em diversos processos biológicos importantes como: transcrição, *splicing*, transdução, localização de

proteínas, integridade da estrutura celular, *imprinting*, ciclo celular e apoptose (CHEN et al, 2021; TSAGAKIS et al., 2020).

O lncRNA também desempenha funções a partir da interação com proteínas. Em geral, eles atuam como sinais, guias, a partir das respostas a estímulos demarcando a expressão de genes e identificando os danos no DNA, o que direciona o recrutamento da proteína com atividade translesão (TLS) para o promotor de um tipo de lncRNAs chamado CCND1 (gene de controle da proliferação celular) (SMITH, 2021; ZHENG, 2019; ZHANG, JIA & KWOH, 2021).

Na interação com o DNA, o lncRNA interatua principalmente através de três vias: (I) lncRNA associa ele mesmo com proteínas de ligação ao DNA; (II) lncRNA forma loops R com DNA e (III) lncRNA forma triplexes com DNA (STATELLO et a., 2021; TSAGAKIS et al., 2020).

No que diz respeito aos genes codificadores de proteínas, os lncRNAs podem ser 'intergênicos', antisense ou intrônicos. Eles também são derivados de 'pseudogenes' (sequência nucleotídica similar a um gene normal, mas que não dá como resultado um produto funcional, quer dizer, que não se expressa), e quase 15.000 foram identificados no genoma humano, alguns dos quais demonstraram ser funcionais. Os lncRNAs também incluem RNAs circulares gerados por *back-splicing* de transcritos codificantes e não codificantes e RNAs reguladores de ação trans derivados de sequências que convencionalmente atuam nas regiões 3' não traduzidas do mRNA (FERNANDES et al., 2019; MATTICK et al., 2023).

Os lncRNAs podem ser classificados de acordo com suas diferentes características, incluindo (1) localização e contexto do genoma, (2) efeito exercido nas sequências de DNA, (3) mecanismo de funcionamento e (4) mecanismo de direcionamento (MA, BAJIC & ZHANG, 2013). Entretanto, costumam ser classificados com base nas suas origens em (AGOSTINI et al., 2021; DARRIEN et al., 2012; REHMAN et al., 2023):

- RNAs antisenso – interceptam qualquer éxon de um lócus codificador de proteína oposta;
- RNAs longos intergênicos (lincRNAs) – provenientes de *loci* de RNAs não codificadores intergênicos, com um gene codificador dentro de um íntron na mesma fita;
- lncRNAs intrônicos – contidos em um íntron de um gene codificador de proteína;
- Transcritos processados – contém uma janela aberta de leitura.

O recente sequenciamento total do transcriptoma revelou um grande número de

supostos lncRNAs, entretanto a função de alguns foi estabelecida em um escopo limitado de processos biológicos, como regulação do ciclo celular, pluripotência de células-tronco embrionárias (ESC) e progressão do câncer, como, por exemplo, o CP (LUO et al., 2016).

### 1.3.2 lncRNA no Câncer de Pulmão

Estudos recentes demonstram que vários lncRNAs têm um papel disfuncional ou desregulador no câncer, podendo ter um potencial pró ou antitumoral. Na patologia do câncer, a indução da angiogênese é um ponto chave para o crescimento, invasão e metástase tumoral. A geração de novos vasos sanguíneos no sítio do tumor é mediada por vários fatores angiogênicos e vias de sinalização complexas, incluindo o fator de crescimento endotelial vascular (VEGF), Notch e vias de sinalização PI3K / AKT (GINN et al., 2020).

Os LncRNAs tornaram-se os principais fatores reguladores no desenvolvimento e progressão do CP, funcionando como oncogenes (por exemplo, MALAT1, HOTAIR, H19 e ANRIL) ou supressores de tumor (por exemplo, MEG3, GAS5 e TUG1). Esses lncRNAs são regulados em tecidos de CP e significativamente correlacionados com estágios patológicos, metástases e sobrevida do paciente. A expressão direcionada ou o silenciamento dos lncRNAs nas células de CP afetam a proliferação, migração e invasão celular, e induzem alterações no epitélio mesenquimal transitório (EMT) e resistência aos medicamentos. Alguns lncRNAs, como o H19, são detectados e elevados na circulação sanguínea de pacientes com câncer de pulmão (CAO et al., 2022; GENCEL-AUGUSTO, WU & BINOVA, 2023).

Duffy, Bouchier-Hayes e Harmey (2013) afirmam que o fator VEGF é o mediador primário da proliferação e migração de células endoteliais vasculares, ele é reconhecido como fator fundamental para o crescimento tumoral. Os lncRNAs estão envolvidos na sinalização de VEGF, sendo assim, contribuem para o crescimento tumoral. Um exemplo é o lnc00173.v1, o qual pode promover a migração de células endoteliais vasculares por meio da regulação positiva da expressão de VEGF em carcinoma de células escamosas do pulmão (WANG et al., 2019)

A hipóxia é fundamental para a angiogênese tumoral, pois as células cancerosas passam a se multiplicar de forma desordenada quando não está chegando suprimento de oxigênio (DUFFY; BOUCHIER-HAYES; HARVEY, 2013). Um estudo realizado por Jiang et al (2021), mostrou que a superexpressão de lncRNA-p21 em câncer de pulmão sob condições hipóxicas produziu um forte efeito pró-angiogênico. Diante disto, o estudo conclui que o lncRNA-p21 aumenta significativamente a formação de tubos em células endoteliais e estimula a adesão de células tumorais.

Nieminen et al. (2018) afirmam que o silenciamento de lncRNA-p21 reduziu a expressão de VEGFA e outros fatores pró-angiogênicos, como Matrix Metalopeptidase 2, Subunidade B do Fator de Crescimento Derivado de Plaquetas e Fator de Crescimento de Fibroblastos 2. De acordo com Anastasiadou, Jacob e Slack (2017), os lncRNAs estão amplamente envolvidos na regulação da proliferação de células tumorais e apoptose. No entanto, os mecanismos específicos pelos quais os lncRNAs afetam os processos biológicos subjacentes do câncer permanecem inconclusivos.

Segundo Palazzo e Lee (2015), vários lncRNAs estão implicados em diferentes tipos de câncer de pulmão e na maioria deles, os lncRNAs funcionam como promotores do câncer. Segundo Cao et al., (2022), todos os lncRNAs oncogênicos são regulados positivamente no câncer de pulmão, enquanto os inibidores de tumor são regulados negativamente.

Slack e Chinnaiyan (2019) afirmam que, como uma das marcas do câncer, a instabilidade genômica é caracterizada por altas cargas de mutação e participa de diferentes pontos de verificação do ciclo celular após o dano ao DNA. Corroborando com isto, Yao e Dai (2015) afirmam que isso pode ser observado em uma variedade de tumores malignos e lesões pré-cancerosas. A instabilidade genômica pode se manifestar como ganho ou perda de todo o cromossomo (aneuploidia), translocação cromossômica (gene de fusão), rearranjo ou alterações na estrutura do telômero.

Além disso, foi relatado no estudo de Jiang et al (2021) que os lncRNAs participam da maquinaria de metilação do DNA para mediar a regulação gênica baseada na cromatina. O fator de transcrição 21 (TCF21) é um supressor de tumor bem caracterizado. CGI1 e CGI3, dois exons que codificam para TCF21, mostraram ser não metilados em amostras de tecido pulmonar normal e hipermetilados em amostras de NSCLC.

Arab et al. (2014) realizou um estudo e foi demonstrado que o promotor indutor de RNA anti-sense lncRNA TCF21 (TARID) medeia a expressão, desmetilação e ativação de TCF21 ao interagir com os reguladores de desmetilação de DNA GADD45A, TDG e TET. O silenciamento da TARID em células cancerosas resultou na metilação aberrante de CGI1 e CGI3. O lncRNAs também mostraram estar envolvidos na inibição da proliferação de células tumorais no câncer de pulmão por meio de uma via associada ao P53. Por exemplo, o P53 pode suprimir a tumorigênese ao inibir a transcrição *Myc* dependente de PVT1b.

Segundo Lim e Kaldis (2013), as quinases dependentes de ciclina (CDKs) e os inibidores de CDK (CKIs) desempenham papéis reguladores vitais no ciclo celular, que é universalmente alterado no câncer. Os lncRNAs lnc00152 e lnc00511, que foram encontrados para ser regulados positivamente em tecidos de adenocarcinoma, podem desempenhar papéis

oncogênicos por inibir potentemente o supressor de tumor IL24 e o CKI p57(GINN et al, 2020). Outro lncRNA oncogênico bastante discutido na literatura é o transcrito 1 de adenocarcinoma de pulmão (LUADT1). LUADT1 mostrou ser altamente expresso em adenocarcinoma (LUAD) e também pode promover a proliferação de células cancerosas após a ligação com SUZ12 e mediar a trimetilação de H3K27 na região promotora do supressor de tumor p27, constatou-se que isto regula negativamente a expressão de P27 epigeneticamente (WANG et al., 2019). De acordo com Qiu et al (2015), o LUADT1 é um lncRNA oncogênico que regula a progressão de LUAD, sugerindo que lncRNAs desregulados podem servir como fatores regulatórios chave na progressão de LUAD.

LncRNAs também atuam como reguladores negativos da transcrição, como por exemplo, a regulação da dihidrofolato redutase (DHFR). O gene que codifica DHFR contém um promotor DHFR contém um promotor principal e um promotor secundário, sendo o último silenciado nas células quiescentes. Recentemente, foi demonstrado que o p53 ativa a expressão de vários lincRNAs (KAIKKONEM; LAM; GLASS, 2011).

Já no câncer de pulmão, o transcrito 1 de adenocarcinoma de pulmão associado à metástase (MALAT1), também conhecido como transcrito abundante enriquecido com núcleo 2 (NEAT2), é um lncRNA altamente é expresso em câncer. Verificou- se que MALAT1 manipula a proliferação, migração e apoptose em muitos cânceres humanos diferentes, como câncer de pâncreas, câncer de pulmão e câncer de ovário (TIANSHENG et al., 2020). Um estudo recente relatou MALAT1foi um supressor de tumor que prejudicou a capacidade metastática das células do câncer de mama pela ligação e inativação do oncogene TEAD, exigindo a retificação dos efeitos deste lncRNA prometastático clássico (KIM et al., 2018).

# 2 OBJETIVOS

## 2.1 GERAL

Analisar os transcritos de amostras de tecido pulmonar não tumoral adjacente e tumoral para traçar os lncRNAs como possíveis biomarcadores.

## 2.2 ESPECÍFICOS

Determinar o perfil de expressão de lncRNAs no CP;
Descrever a microbiota presente no tecido pulmonar tumoral e não tumoral adjacente;
Correlacionar os resultados obtidos, com os dados clínicos e de estadiamento;
Identificar biomarcadores para potenciais uso clínico.

# 3 CAPÍTULO 1 - A NARRATIVE REVIEW OF THE LUNG CANCER MICROBIOME: HURDLES IN THE SEARCH FOR CONSISTENT BIOMARKERS

# A Narrative Review of the Lung Cancer Microbiome: Hurdles in the Search for Consistent Biomarkers

Lorena Duarte Fernandes[1,2*]
Sérgio Augusto Antunes Ramos[1,3*]
Camila Tavares Carvalho Uchôa[1,2]
Mariana Souza Lima[1,3]
Jéssica Manoelli Silva[1,3]
Eliel Barbosa Teixeira[1,4]
Geraldo Roger Normando Jr.[5]
Samir Mansour Casseb[1]
Fabiano Cordeiro Moreira[1]
Paulo Pimentel Assumpção[1+]

[1] Núcleo de Pesquisa em Oncologia, Universidade Federal do Pará, Belém, PA, Brasil
[2] Programa de Pós-graduação em Oncologia e Ciências Médicas, Belém, PA, Brasil
[3] Programa de Pós-graduação em Genética e Biologia Molecular, Belém, PA, Brasil
[4] Programa de Pós-graduação em Biotecnologia, Belém, PA, Brasil
[5] Hospital Universitário João de Barros Barreto, Universidade Federal do Pará, Belém, PA, Brasil

[+] Correspondence author
assumpcaopp@gmail.com

[*] First shared authorship

## Abstract

**Background and objective:** Lung cancer (LC) is recognized as the most prevalent and lethal malignant disease globally, influenced by both environmental and genetic factors. Recent research has identified notable associations between the composition of the pulmonary microbiota and the pathogenesis of lung cancer. These findings suggest potential pathways for the development of targeted therapies. This review aims to examine the array of investigative methodologies employed to explore the intricate relationship between lung microbiota and lung cancer, its impact on inflammation and immune responses, and the critical challenges confronting the scientific community in this field.

**Methods:** we searched NCBI (via PubMed) using our keywords "lung cancer", "microbiome profiling", "microbial biomarkers" and "metagenomics analysis" from the inception of the database to May 2024. Our search was restricted to publications in English language. We retrieved other eligible studies by manual searching of the reference lists of included studies. We then selected studies that provided sequencing data in NCBI under an accession code (Bioproject).

**Key content and findings:** This review discusses various studies employing different sample types, including bronchoalveolar lavage fluid (BALF), sputum, and solid lung biopsies, along with diverse research methodologies and computational analysis techniques. Given the lack of consensus in the literature on the pulmonary microbiome in the context of lung cancer, this study addresses the inconsistencies reported by different authors and highlights potential biases that should be avoided to prevent misinterpretation of the data. Notably, sputum samples revealed significant differences in the abundances of predominant bacterial genera between lung cancer patients and healthy controls, with *Neisseria, Prevotella, Fusobacterium, Streptococcus,* and *Haemophilus* being more abundant in the cancer group.

**Conclusions:** These findings highlight the potential of sputum microbiome profiling as a non-invasive approach for lung cancer screening, with the identified genera serving as candidate biomarkers for further investigation, if validated in larger and independent cohorts.

**Key-words:** Lung cancer; microbiome profiling; metagenomics analysis; microbial biomarkers.

## Introduction

At the end of the 19th century, lung cancer (LC) was rare, accounting for just 1% of all tumors found in autopsies in 1878. However, its prevalence surged to 10-15% by the late 1920s, with a sharp increase following World War I [1]. Since 1985, LC has become the most common and deadly malignant disease globally, with 2.48 million new cases and over 1.81 million deaths reported in 2022 [2].

The genetic and epigenetic changes play crucial roles in the tumor's pathophysiology. These include driver mutations in genes related to cell growth, tumor suppression, and DNA repair, making LC one of the tumors with the highest number of driver mutations [3, 4]. Environmental factors such as exposure to contaminants, toxins, pollution, respiratory infections, and especially smoking, significantly influence LC development [5].

Traditionally, the lung was considered a sterile organ due to negative results from conventional microbiological culture techniques. However, over the past decade, culture-independent molecular techniques have disproven this belief, revealing that the lung hosts a unique and dynamic microbiota, comprising bacteria, fungi, and viruses. This microbial community exists in both healthy individuals and patients with respiratory diseases, challenging established views in respiratory medicine [6].

Recent studies have highlighted compelling associations between the composition of the lung microbiota and the pathogenesis of lung cancer. Elucidating these interactions presents a promising opportunity for the development of targeted therapeutic strategies. This review aims to examine the array of investigative methodologies employed to explore the intricate relationship between lung microbiota and lung cancer, its impact on inflammation and immune responses, and the critical challenges confronting the scientific community in this field. We present the study in accordance with the Narrative Review reporting checklist (Available at ).

## Methods

To obtain relevant literature, we searched NCBI (via PubMed) using keywords “lung cancer”, “microbiome profiling”, “microbial biomarkers” and “metagenomics analysis” from the inception of the database to May 2024. Our search was restricted to publications in English language. We retrieved other eligible studies by manual searching of the reference lists of

included studies and then selected studies that provided sequencing data in NCBI under an accession code (Bioproject) (Table 1).

**Table 1.** The search strategy summary.

## Exploring the Pulmonary Microbiome: From Health to Dysbiosis

The human body is colonized by a diverse range of microorganisms, including bacteria, fungi, viruses, and protozoa [7]. Both environmental and genetic factors shape the composition, diversity, and stability of the microbiome [8]. In the lungs, microbial harmony results from a dynamic balance between immigration (via inhalation, microaspiration, and mucosal dispersion) and elimination (through coughing, mucociliary clearance, and immune system activity). Factors influencing the growth and survival of commensal bacteria include nutrient availability, temperature, pH, oxygen tension, and the activity of host immune cells [9].

The upper respiratory tract (URT) is the primary source of the pulmonary microbiome, complemented by the oral cavity microbiome [10]. However, only a portion of this community becomes established in the lungs due to clearance mechanisms. Ibironke et al. [11] found that 95% of microorganisms in bronchoalveolar lavage, throat swab, mouthwash, and nasal swab samples were transported from outside the lung, with less than 5% of bacterial species restricted to the bronchoalveolar lavage. This highlights the compositional differences between the upper and lower respiratory tracts [12], with lung bacterial communities showing 2 to 4 logs lower biomass compared to the URT [13].

Biomass is higher in more proximal pulmonary regions, with modest regional differences suggesting differential clearance and possibly limited local replication. Therefore, it remains uncertain whether long-term, self-sustaining bacterial residents exist in the lungs. The nature and dynamics of the pulmonary microbiome differ from those in the gut, skin, vaginal, and oral cavities, which have robust, self-sustaining microbial communities [14].

Each microbiome is specific and can vary with changes in influencing factors. For instance, the nasopharyngeal microbiome in children is seasonal, as is the vaginal microbiome throughout the menstrual cycle [15]. Studies have shown differences in microbial composition and diversity between the lungs of healthy and diseased individuals [16] . In healthy individuals, the pulmonary microbiota is characterized by low abundance and high diversity, primarily from the phyla Bacteroidetes, Firmicutes, and Proteobacteria. At the genus level, microorganisms

such as Pseudomonas, Streptococcus, Prevotella, Fusobacteria, and Veillonella, which have lower pathogenic potential, dominate. In contrast, diseased individuals experience a loss of bacterial diversity, potentially contributing to disease progression [17, 18].

Fungal and viral communities in the lung, known as the mycobiome and virome, are less studied due to challenges such as low biomass, difficulty in extracting fungal DNA, and the lack of conserved viral regions analogous to the bacterial 16S and fungal 18S genes. Inaccurate genomic annotations further complicate identification [19]. Nonetheless, studies have identified Ascomycota and Basidiomycota as the most common fungal taxa, with genera like Candida, Saccharomyces, Penicillium, Cladosporium, and Fusarium frequently found [20, 21]. Young et al. [22] reported that Anelloviridae is the predominant eukaryotic virus, with herpesviruses, papillomaviruses, retroviruses, and other respiratory viruses detected sporadically. Ruiz et al. [23] linked active HCMV infection to early recurrence in postoperative lung cancer patients.

Humans are exposed to microorganisms immediately after birth, establishing microbial profiles that vary with the type of delivery, impacting the development and maturation of the newborn's immune system [24]. The microbiota plays a critical role in the induction, training, and function of the host's immune system. In turn, the immune system evolved to maintain a symbiotic relationship with these diverse and evolving microbes. Under homeostatic conditions, the immune system-microbiota interaction allows for protective responses to pathogens while maintaining tolerance to harmless antigens [25].

Dysbiosis, or disruption of the normal microbiome, and the overgrowth of pathogenic species can compromise physiological homeostasis, shifting mutualistic relationships to parasitism. Commensal microorganisms help maintain host health by producing molecules that affect the survival and virulence of pathogenic microorganisms [26]. Pulmonary dysbiosis is linked to various pathological conditions, including cardiovascular and respiratory diseases, diabetes, and is routinely associated with lung cancer development [27].

Recent studies connect dysbiosis with disease severity in asthma, COPD exacerbation, cystic fibrosis, and bronchiectasis pathophysiology. Chronic bacterial colonization and infection are common in bronchiectasis patients, with organisms like Pseudomonas aeruginosa associated with prognosis and symptom exacerbation frequency [9]. There is growing evidence

that individual variability in microbiomes can profoundly impact cancer phenotypes, interacting with genome instability, mutation, and tumor-promoting inflammation [28].

Studies have demonstrated that the microbiota can influence tumor progression by sustaining chronic inflammation [29]. Additionally, microbial dysbiosis is linked to resistance to primary PD-L1 checkpoint inhibition therapy, suggesting that antibiotics in advanced cancer patients affect the clinical benefits of checkpoint inhibitors—drugs that block signals suppressing immune system activity, particularly T cells, thereby enhancing their effectiveness against tumor cells [30, 31].

**Role of Lung Microbiota in Inflammation and Immunity**

The lung, as a critical gas exchange organ, possesses one of the largest interfaces between the human host and the external environment, making it a key mucosal site for host-microbiota interactions [32]. In healthy lungs, the environment is typically inhospitable to bacterial community development, leading to a relatively low bacterial replication rate [33]. However, similar to the role of commensal bacteria in intestinal immunity, pulmonary commensal bacteria are crucial in preventing chronic inflammation under normal conditions. They contribute to creating an immunotolerant environment by facilitating the development, activation, and recruitment of immunoregulatory immune cells, such as regulatory T cells (Treg), M2 polarized alveolar macrophages, and tolerogenic dendritic cells (DCs) [34].

The majority of immune system interactions with microorganisms, and the resulting immune responses, are a product of the host's symbiotic relationship with its microbiota. Contrary to early beliefs that these microbes were largely ignored by the immune system, it is now understood that microbes on all barrier surfaces actively induce immune responses, including those targeting the microbiota itself. What distinguishes these responses from those against pathogenic microbes is that immune responses to the microbiota and the accumulation of commensal-specific lymphocytes in tissues occur without inflammation, a process known as homeostatic immunity [25].

In pathological conditions, the stability of the lung's microbial community is compromised, and its species composition often shifts due to immune system activity in response to innate and adaptive immune responses [35]. This inflammatory state can lead to epithelial fibroproliferation, increased immune cell migration to the lungs due to cytokine and chemokine production, and an elevated Th2 immune response, among other potential

pathological mechanisms. Even dead bacteria can provoke responses through mediators such as microbial metabolites like short-chain fatty acids (SCFAs) and pattern recognition receptor (PRR) ligands [14].

Hosang et al. [36] reported that dysregulation in the pulmonary microbiome could significantly increase susceptibility to central nervous system autoimmune disease in mice. Additionally, a longitudinal multi-cohort analysis by Wang et al. [37] involving 1,706 sputum samples from 510 individuals with chronic obstructive pulmonary disease (COPD) found that neutrophilic inflammation was associated with two main bacterial community types during stability and exacerbations. One type resembled the bacterial community of the upper respiratory tract, while the other was dominated by pathogens such as Haemophilus, Moraxella, and Streptococcus. In contrast, eosinophilic inflammation in COPD was linked to fewer clinical bacterial infections, lower bacterial load, greater microbial diversity, and a lower prevalence of Haemophilus and Moraxella. These findings suggest distinct mechanisms by which the pulmonary microbiome may influence inflammation in different COPD endotypes.

Chronic inflammation is a key factor in cancer pathogenesis, promoting primary tumor development and establishing a metastatic microenvironment. However, the underlying mechanisms and their pro-inflammatory effects remain incompletely understood [38]. In a study by Tsay et al. [39], the lower airway transcriptome in patients with cancer exhibited significant differences from that of control subjects, including upregulation of ERK (extracellular signal-regulated kinase) and PI3K (phosphoinositide 3-kinase) signaling pathways..

Microbial metabolic products can either promote or suppress tumor growth by directly altering the local tissue microenvironment or indirectly through chronic stimulation of the immune system, leading to immune exhaustion and reduced immune surveillance. Consequently, the relationship between specific microorganisms and tumor behavior may vary depending on the tumor's molecular pathological signatures [40–42].

**Lung Microbiome in Lung Cancer**

Bacterial species are present not only within tumor tissues but also in adjacent areas and potential sites of tumor development, such as the intestine, genitourinary tract, and airways, suggesting a similarity in microbial profiles across these locations. These microbial populations, distributed across the tumor, surrounding areas, and more distant sites, interact and

play distinct roles in cancer progression, contributing to what is known as the "field of cancerization" [43, 44].

A study on lung cancer patients demonstrated a correlation between pulmonary colonization by pathogenic microorganisms and poorer prognosis, including reduced overall survival [45]. Tuberculosis, a chronic infectious disease caused by *Mycobacterium tuberculosis*, has been shown to promote inflammatory and immunological changes, significantly increasing the risk of lung cancer [46, 47].

In another study that examined 165 samples of tumor tissue and adjacent non-tumor tissue, the taxonomic and functional profiles of the pulmonary microbiota revealed distinct microbiomes between anatomical sites, with greater diversity in adjacent tissue compared to tumor tissue. In metastatic cases, the genera *Thermus* and *Legionella* were found to be more abundant in advanced stages (IIIB, IV), suggesting their potential importance in LC progression. Additionally, *Thermus* was more prevalent in lung adenocarcinoma tissue, while *Ralstonia* was less abundant compared to squamous cell carcinoma [48]. *Acidovorax* was identified as another significant genus, particularly abundant in LUSC cases with a TP53 mutation, and absent in LUAD cases [49].

Yan et al. [50] found that *Capnocytophaga*, *Selenomonas*, *Veillonella*, and *Neisseria* were significantly increased in saliva samples from patients with squamous cell carcinoma (SCC) and adenocarcinoma compared to controls. The study also highlighted *Capnocytophaga* and *Veillonella* as bacterial biomarkers, with sensitivity and specificity values capable of distinguishing patients with SCC and adenocarcinoma from controls.

In a prospective study analyzing airway brush samples from individuals undergoing diagnostic bronchoscopy for lung nodules, it was observed that airway samples from cancer patients frequently exhibited increased levels of *Streptococcus* and *Veillonella*. In contrast, samples from controls with non-cancerous conditions showed a predominance of *Streptophyta*, *Moraxellaceae*, and *Stenotrophomonas*, while healthy controls were enriched with *Acholeplasma* and *Acidocella*. The findings from uninvolved cancer airway samples were similar to those from involved cancer airway samples [39].

However, a study by Su et al. [51], which examined paired samples of tumor tissue, adjacent tissue, distal lung tissue, and bronchial tissue from 16 patients, found no significant differences in microbiome diversity between these tissues. Contrastingly, Dohlman et al. [52]

identified a higher prevalence of *Blastomyces* in lung tumor tissue compared to adjacent tissue, while Mao et al. [53] reported a greater abundance of *Modestobacter* in tumor tissue versus adjacent tissue. Another study highlighted intratumoral fungal diversity in smokers, revealing greater fungal richness of *Aspergillus* and *Agaricomycetes* in smokers compared to non-smokers [54].

Studies investigating the pulmonary microbiome employ various types of samples, including sputum, bronchoalveolar lavage fluid (BALF), and solid lung tissue biopsies, alongside diverse research methods and computational analysis techniques. These differences in approach provide multiple perspectives for assessing pulmonary microbial composition, as outlined in the methods described below.

**Methods of Investigation and Analysis of the Human Microbiome**

Despite significant advances in laboratory microbiology techniques, characterizing microbiomes through cultivation methods remains limited. Traditional media often fail to support the growth requirements of all microorganisms present in a sample [55]. The continuous improvements in Next-Generation Sequencing (NGS) technologies have led to innovative approaches for taxonomic and genetic characterization of human and environmental microbiomes. Among these, metagenomics was the first technique to bridge modern genomics with the study of microbial communities directly in their natural environments, bypassing the need for isolating and cultivating individual species in the laboratory [56].

Metagenomics enables the exploration of the structure, function, and dynamics of a community of organisms through two primary approaches. The first, shotgun-metagenome (also known as metaprofiling or amplicon sequencing), involves sequencing phylogenetic marker genes, such as the 16S ribosomal RNA (rRNA) gene in bacteria and the 18S rRNA gene in fungi. These genes, despite being genetically conserved, contain variable regions that allow for taxonomic identification up to the genus level [57, 58].

The second approach, whole-metagenome shotgun sequencing, covers the entire genomic content of a sample. This method offers broader coverage than marker-based techniques and provides additional insights, such as detecting novel organisms, understanding gene functionality for functional and metabolic profiling, and achieving more detailed taxonomic resolution down to the species or strain level [59].

However, both methodologies involve significant analytical complexity [60], particularly during genome assembly, and are prone to generating chimeras and including DNA from non-target organisms (contaminants) [61]. A study analyzing 7,000 bacterial genomes isolated from metagenomic sequencing data revealed numerous assembly errors, chimeras, and contaminants, underscoring the limitations of the technique [62].

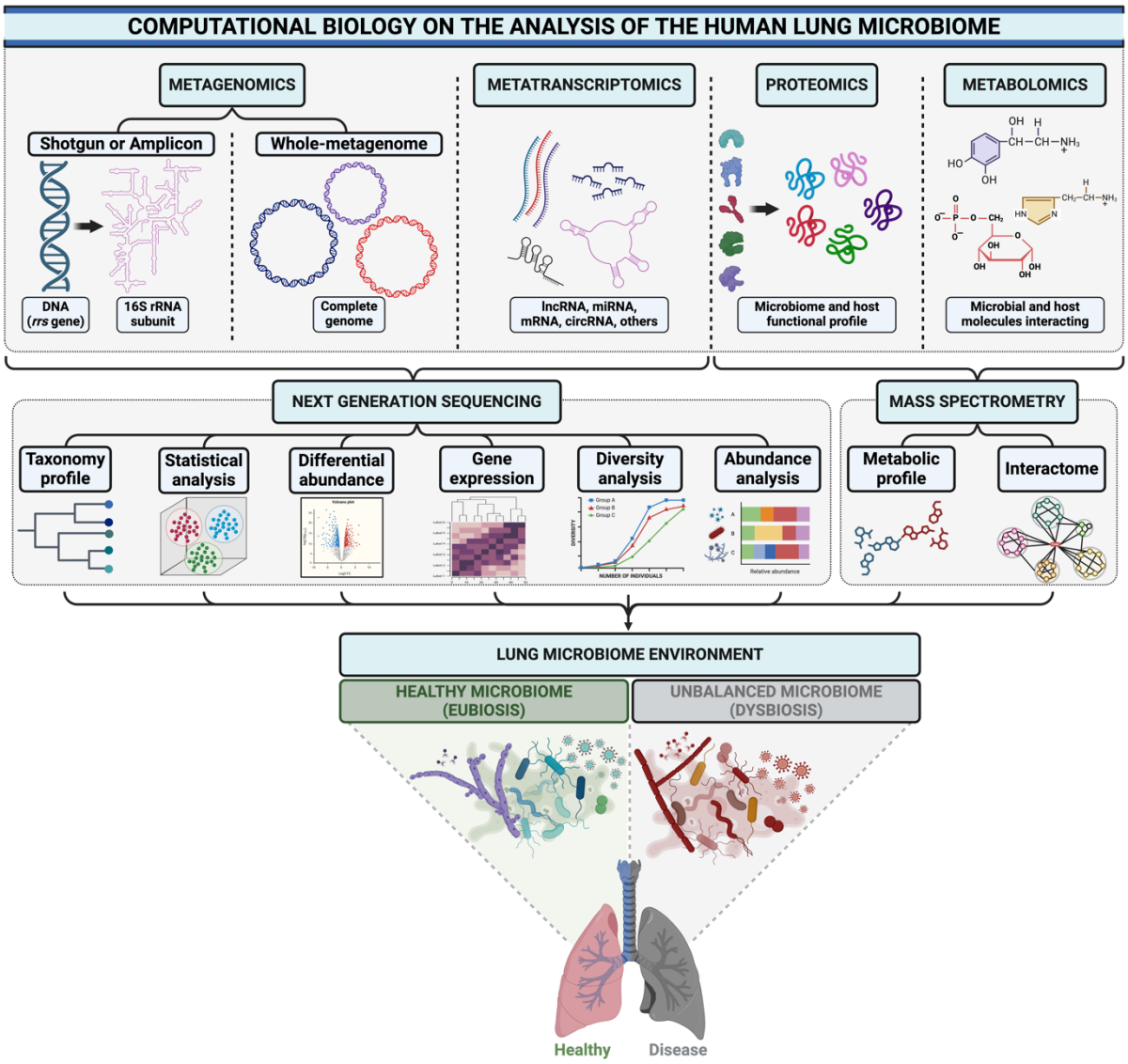
The interest in obtaining more precise and complementary information about microbial communities expanded with the advent of RNA sequencing techniques (RNA-seq), revolutionizing the study of microbial transcriptomes. Among the various RNA-seq approaches, metatranscriptomics has significantly advanced the prediction of transcriptional profiles, providing a more accurate representation of the functional activities of microbial populations, especially in gastrointestinal and respiratory tracts [63–67].

Metatranscriptomics allows researchers to assess transcriptionally active microbial populations and elucidate the functional roles these microorganisms play in their environments. This technique, in combination with genomic and transcriptomic studies using NGS approaches, enables the generation of taxonomic profiles, statistical analysis of microbial diversity, and differential abundance and gene expression analyses [68, 69].

However, implementing metatranscriptomics presents several challenges: 1. Difficulty in obtaining genetic material at the ideal concentrations and integrity levels for sequencing, due to RNA's high instability, particularly in certain sample types; 2. The use of outdated bioinformatics tools, which can hinder proper data interpretation; 3. The presence of translatable transcripts does not necessarily confirm the presence of their corresponding proteins, necessitating complementary techniques like metaproteomics and metabolomics to provide a more comprehensive understanding of the microbial community's interaction with its host or microhabitat [70].

Figure 1 provides an overview of the main computational techniques used for analyzing the human microbiome, with a focus on studying healthy and altered pulmonary microbiological profiles, particularly in the presence of pathological processes. The figure also illustrates methodologies based on mass spectrometry, including metabolomics and proteomics approaches, which are designed to assess the biochemical [71, 72] and functional [73, 74] profiles of the study target, respectively.

**Figure 1**. Computational biology in the evaluation of microbial data. Computational biology dedicated to the analysis of human microbiomes involves different analysis methods, and can use data from next-generation sequencing as well as mass spectrometry. In this way, it is possible to evaluate the nuances of the health-disease process, especially when it comes to the pulmonary microenvironment.




In the context of studying the pulmonary microbiome by NGS, species description encompasses three main sample types, such as BALF [75, 76], sputum collection [77], and solid biopsy (SB), obtained from lung epithelial resection [78]. Although differences may exist between the microbiota of different histological and molecular types, the literature mostly presents results in a unified manner, which limits the analysis of individual contexts.

This study aimed to assess the consistency and reproducibility of published articles' results, as well as reanalyze data deposited in NCBI through a unified in-house pipeline.

**Microbiome Data Inconsistencies: Advocating for Methodological Harmonization**

The primary analysis included ten studies (see Table S1 in the Supplementary Materials), whose inclusion criteria involved the investigation of the lung cancer microbiome and provided sequencing data in NCBI under an accession code (Bioproject).

The samples were categorized into three types: BALF, sputum, and solid biopsy. The results of the most abundant operational taxonomic units (OTUs) of each author, considering their respective analyses, were processed using RStudio [79].

Quantitative evaluation revealed that the most abundant OTUs ranged from 03 to 28 taxa. While a core microbiome was identified in the sputum group, no such core was found in the BALF and solid biopsy groups. However, four taxa—*Neisseria*, *Staphylococcus*, *Prevotella*, and *Haemophilus*—were consistently present across all sample types. For more details, see Supplementary Figures S1-S3.

Among the authors who assessed the lung microbiome in lung cancer patients from bronchoalveolar lavage fluid (BALF), Seixas et al. [80] characterized 106 samples, revealing that *Streptococcus, Prevotella, Salmonella,* and *Escherichia/Shigella* were the most prevalent taxa in the lung cancer (LC) group, with Streptococcus and Prevotella serving as strong predictors for lung cancer.

Qian et al. [81] examined microbiota distribution in patients with synchronous multiple primary lung cancer (sMPLC), finding significant differences between oral microbial communities and those from tumor regions. *Clostridium, Actinobacteria, Fusobacterium,* and *Rothia* were the most abundant taxa in the bronchial mucosa.

Ekanayake et al. [82] compared bacterial microbiota in the upper and lower respiratory tracts of lung cancer and bronchiectasis patients to healthy individuals. Only lung cancer patients exhibited disease-associated bacterial strains in BALF, including *Corynebacterium tuberculostearicum* and *Keratinibaculum paraultunense*.

Xia et al. [83] compared lung microbiota across respiratory diseases, noting higher alpha diversity and bacterial richness in LC patients than in those with primary pulmonary tuberculosis (PTB) and community-acquired pneumonia (CAP). *Sphingobium* and *Marseilla* showed significant increases in the LC group compared to PTB, while *Neisseria, Megamonas,* and *Fusobacterium* increased in LC compared to CAP.

Cameron et al. [84] analyzed sputum samples from ten patients (four with LC). Alpha diversity measures showed no significant differences between LC-positive and LC-negative

groups, but 16 bacterial species were found in all LC-positive samples, with higher abundances of *Granulicatella adiacens, Streptococcus intermedius,* and *Mycobacterium tuberculosis*.

Huang et al. [85] evaluated sputum samples from 85 patients with non-small cell lung cancer (NSCLC), identifying dominant genera: *Streptococcus, Prevotella, Rothia, Neisseria, Actinomyces, Leptotrichia, Porphyromonas, Veillonella, Granulicatella, Haemophilus, Atopobium, Peptostreptococcus, Capnocytophaga,* and *Fusobacterium.*

Four articles analyzed solid biopsies to identify microorganisms in lung neoplasia, among them, Apopa et al. [86], who compared the microbiome of normal lung tissue, lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD), identifying 18 genera from five phyla: *Proteobacteria, Bacteroidetes, Actinobacteria, Firmicutes,* and *Cyanobacteria*. Notable genera included *Achromobacter, Acinetobacter, Actinomyces, Elizabethkingia, Rothia, Sphingobacterium, Kocuria, Pseudomonas, Staphylococcus, Streptococcus, Finegoldia,* and *Phyllobacterium*, with high relative abundance of Cyanobacteria in LUAD.

Peters et al. [87] found that *Clostridia* and *Bacteroidia* are present in normal lung tissue and can serve as biomarkers for early-stage NSCLC patients post-tumor resection. Additionally, TAP1, TAPBP, CSF2RB, and IFITM2 expression in peripheral blood may correlate with improved survival when combined with lung microbiome and peripheral gene expression biomarkers.

Kovaleva et al. [88] investigated the lung tumor microbiome and its correlation with the microenvironment affecting prognosis. They found no significant taxonomic diversity difference between adjacent normal lung tissue and tumor tissue, with *Actinobacteria, Proteobacteria, Firmicutes,* and *Bacteroidetes* predominating in both. In adenocarcinoma and SCC patients, 37 bacterial genera showed contrasting correlations with these lung cancer subtypes.

Chang et al. [89] conducted a metatranscriptomic analysis of lung cancer metagenomes, finding that *Rhodococcus, Pseudomonas, Acinetobacter, Bacillus, Mycobacteroides, Brevundimonas,* and *Staphylococcus* were prevalent in lung tissues of cancer patients and associated with reduced survival.

**Pooled Analysis of Published Data**

Due to the limited common findings among authors, this analysis indicates that methodological choices in characterizing microorganisms in lung cancer (LC) using various in silico approaches may result in non-reproducibility of observations regarding the microbiome-

lung cancer relationship, as shown in intra-group comparative results (Figures S1-3). Variations in sampling techniques, bioinformatics pipelines, and patient populations underscore the complexity of interpreting microbiome studies and highlight the need for standardization in methodologies, particularly in computational biology, to facilitate better comparisons across studies.

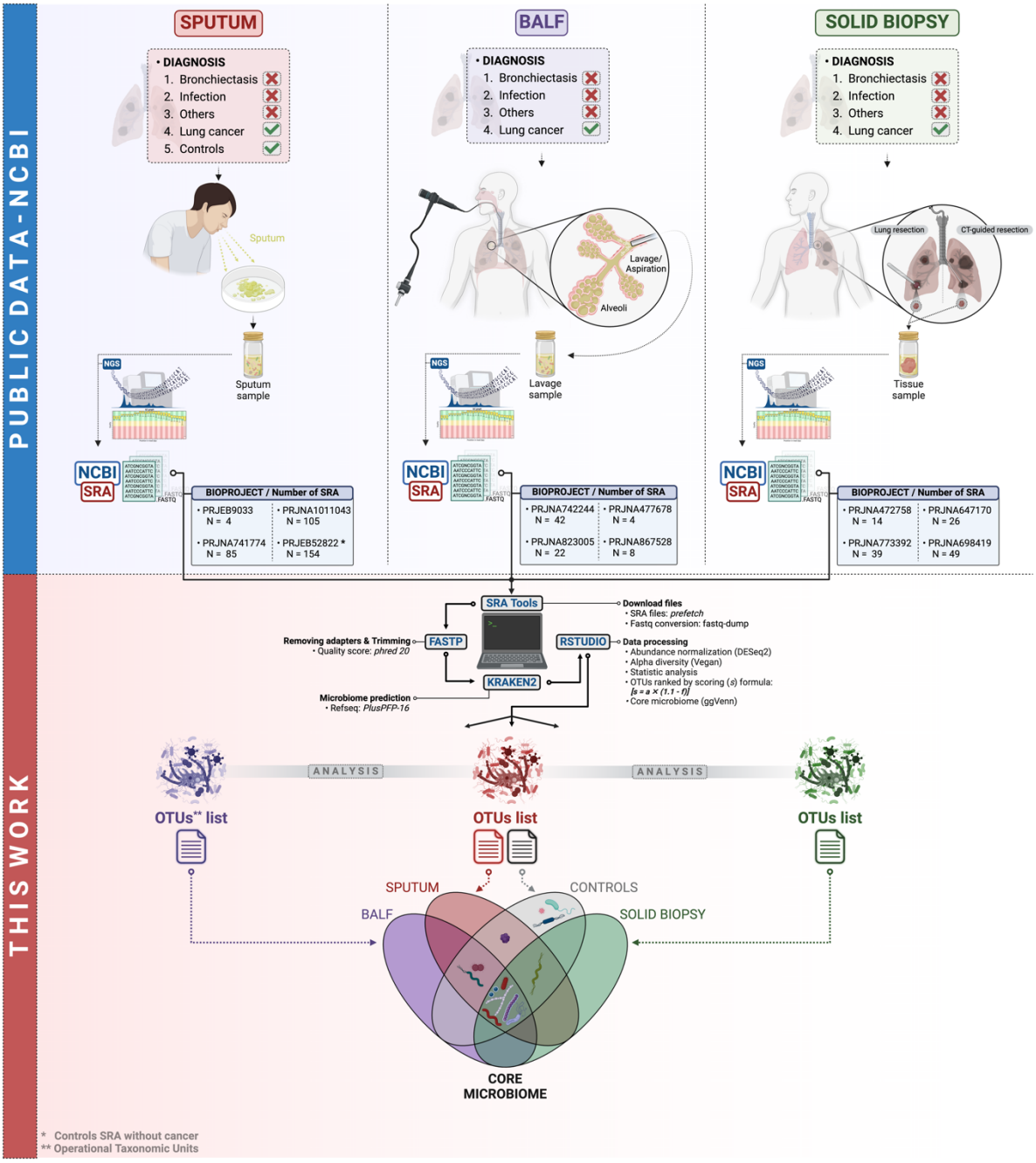
Figure 2 illustrates the flowchart of our reanalysis method, which involved selecting scientific articles focused on the lung cancer microbiome, utilizing samples from bronchoalveolar lavage fluid (BALF), sputum, and solid biopsy. Samples from normal/adjacent tissue trauma and bronchiectasis patients were excluded, except for control data (non-cancer) sourced from sputum of healthy donors. We subsequently retrieved all Sequence Read Archive (SRA) codes from the authors of the selected studies.

Our pipeline, developed in Shell scripting language, included commands for downloading SRA data from NCBI and converting it to fastq format using SRA Tools. Adapter removal and data trimming were performed with Fastp [90], followed by sequence alignment against a microbiome database using Kraken2 software [91].

Taxonomic reports generated by Kraken2 for each sample underwent filtering to eliminate OTUs with fewer than 10 reads in over 50% of samples from each author/bioproject. This filtering ensured that remaining OTUs were more representative of their respective bioprojects, reducing potential binning errors or false-positive OTUs. This filtering and figure generation were conducted using R language [79], limited to the genus taxonomic level.

Finally, OTUs from samples of the same collection origin were grouped by tissue type, allowing for the creation of a Venn diagram summarizing the number of shared genera among the four sample groups, termed the core microbiome. For relative abundance analysis, alpha diversity (Shannon index, H') and multivariate dispersion homogeneity were assessed by pooling samples from all study groups. Read counts were normalized using the DESeq2 package [92] in RStudio.

**Figure 2.** Methodological analysis model for predicting the microbial core in lung cancer. Sequenced data from different sample types (SPUTUM, BALF and Solid Biopsy) were tested for the diagnosis of LC. Part of the SPUTUM data also served as Control (non-LC). The SRA of the Bioprojects were downloaded, treated and detailed in several programs and, finally, the microbiome was predicted and plotted in a Venn graph, as well as in other graphical models. BALF, bronchoalveolar lavage fluid. BALF, bronchoalveolar lavage fluid.



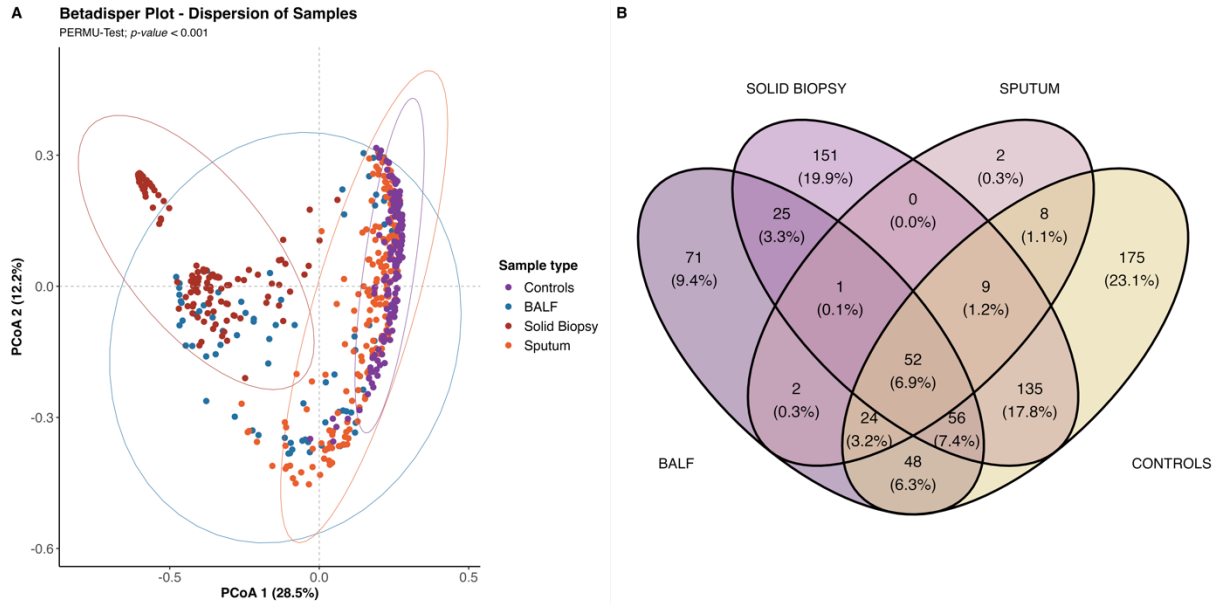

## Critical Evaluation of the Data Analyzed

The Bioprojects utilized in this study, including the corresponding number of SRAs obtained from each author, the sample groups (LC and Controls), the methodological approach, and the NGS platform employed, are detailed in Table S2 (Supplementary Materials).

The overview of the analyses conducted by our pipeline is illustrated in Figure 3. The analysis of dispersion homogeneity revealed that the Sputum and Control groups cluster together, reflecting their shared sample type and similar microbial profiles. In contrast, the Solid Biopsy group is more distantly clustered from the Sputum groups, while the BALF group encompasses all samples, indicating substantial heterogeneity in microbial composition associated with this collection method. Notably, there is a significant difference in group variances (PERMU-Test; p-value < 5%) (Figure 3A).

When assessing the microbial core of sample groups originating from LC and Controls, a significant overlap of 52 bacterial genera between these conditions is observed. However, there was not an equivalent count when considering only the cancer subset, as only one genus (*Mycoplasmoides*) occurred among the three LC-related groups (Figure 3B).

**Figure 3.** Summary of the analyses proposed by the pipeline of the present study. Homogeneity analysis of sample dispersions (A), microbial core between LC and healthy controls (B). BALF, bronchoalveolar lavage fluid.



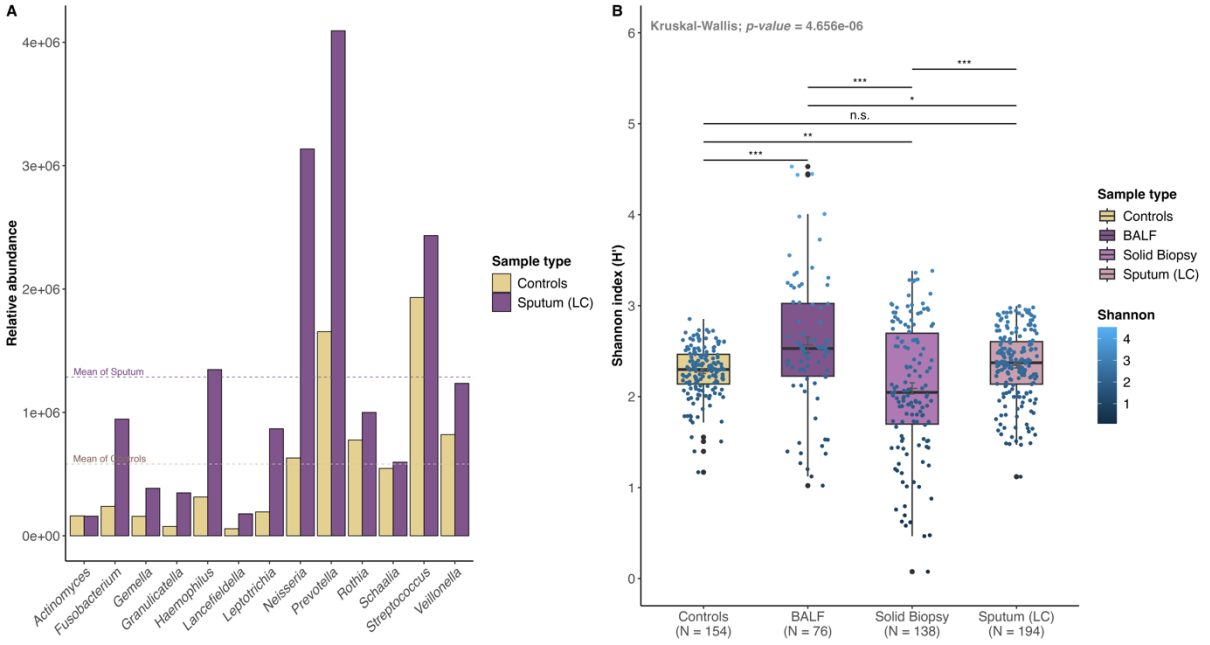

The analysis of sputum samples revealed significant differences in the abundances of predominant bacterial genera between lung cancer (LC) patients and healthy controls. The genera Neisseria, Prevotella, Fusobacterium, Streptococcus, and Haemophilus were found to be more abundant in the sputum of LC patients compared to controls (Figure 4A).

When assessing microbial diversity, statistically significant differences were observed among the different sample groups (Kruskal-Wallis/Dunnett's test; p-value < 5%), except for the comparison between Controls and Sputum (LC) samples, which did not show significant differences in diversity profiles (Figure 4B). This suggests that the microbial composition in

sputum samples from LC patients is distinct from healthy controls, with specific genera being enriched in the cancer group.

These findings highlight the potential of using sputum microbiome profiling as a non-invasive approach for lung cancer screening. The identified genera, such as Neisseria, Prevotella, Fusobacterium, Streptococcus, and Haemophilus, may serve as candidate biomarkers for further investigation. However, it is important to note that these results are based on a single analysis, and validation in larger, independent cohorts is necessary to confirm the robustness and clinical utility of these microbial biomarkers for lung cancer detection.

**Figure 4.** Normalized abundance of bacteria with the best ranking scores (Sputum and Controls) (A). Alpha diversity of genera by the Shannon index for each sample group (B). BALF, bronchoalveolar lavage fluid; LC, lung cancer; Dunnett's test; n.s., non-significant; **p-value adjusted* < 0.05; ***p-adj* < 0.01; ****p-adj* < 0.001.




Following data normalization, OTUs from each sample group were ranked according to the scoring model proposed by Barra et al. [93]. This model classifies taxa from lowest to highest rank, providing the abundance position of each OTU within its respective group without affecting the rankings of taxa from other groups.

Table 2 indicates that in most sample groups, the genera *Streptococcus*, *Prevotella*, and *Veillonella* consistently occupy the top three positions in this analysis. This ranking suggests that these genera are predominant in the microbial profiles of the samples, potentially serving as key indicators for further investigation into their roles in lung cancer and its microbiome.

**Table 2.** Most abundant bacteria by sample group, according to the proposed pipeline, including both normalized abundance and ranking position scores (n = 759). A score [$s = a \times (1.1 - f)$] was created, where *s* is the score value; *a* is the mean of the taxon ranking position in each sample in a given group; and *f* is the taxon frequency in each group. BALF, bronchoalveolar lavage fluid.

| Order | BALF | SOLID BIOPSY | SPUTUM | HEALTHY CONTROL |
|---|---|---|---|---|
| 1 | *Streptococcus* | *Staphylococcus* | *Streptococcus* | *Streptococcus* |
| 2 | *Prevotella* | *Bacillus* | *Prevotella* | *Prevotella* |
| 3 | *Veillonella* | *Pseudomonas* | *Neisseria* | *Veillonella* |
| 4 | *Escherichia* | *Acinetobacter* | *Rothia* | *Rothia* |
| 5 | *Sodalis* | *Corynebacterium* | *Porphyromonas* | *Schaalia* |
| 6 | *Gemella* | *Streptococcus* | *Leptotrichia* | *Neisseria* |
| 7 | *Enterococcus* | *Klebsiella* | *Veillonella* | *Haemophilus* |
| 8 | *Clostridium* | *Escherichia* | *Granulicatella* | *Fusobacterium* |
| 9 | *Bacillus* | *Enterococcus* | *Schaalia* | *Actinomyces* |
| 10 | *Lactobacillus* | *Streptomyces* | *Haemophilus* | *Gemella* |
| 11 | *Staphylococcus* | *Burkholderia* | *Gemella* | *Leptotrichia* |
| 12 | *Corynebacterium* | *Sphingomonas* | *Fusobacterium* | *Granulicatella* |
| 13 | *Haemophilus* | *Salmonella* | *Actinomyces* | *Lancefieldella* |
| 14 | *Pseudomonas* | *Cupriavidus* | *Campylobacter* | *Mogibacterium* |
| 15 | *Acinetobacter* | *Brevundimonas* | *Lancefieldella* | *Capnocytophaga* |

A score [s = a × (1.1 − f)] was created, where s is the score value; a is the mean of the taxon ranking position in each sample in a given condition; and f is the taxon frequency in each condition.

## Conclusions

As the importance of microorganisms in health and disease becomes clearer, their role in protecting against pathogenic microorganisms that can disrupt the resident microbiota and cause harm to host organisms is increasingly evident. Research into the carcinogenic mechanisms of intratumoral microorganisms provides valuable insights for the development of more effective therapeutic interventions, alternatives to non-invasive and low-cost approaches, such as sputum samples for lung cancer screening.

In addition, it is important to emphasize that microbiome analysis from the perspective of bioinformatics techniques also presents challenges such as the lack of reproducibility and standards of care and data visualization. On the other hand, evaluating the molecular aspects of lung cancer from the perspective of transcriptional profiling could show some relationship with the microbial profile present there, making this investigation even more intriguing.

## Funding

This study was Financially supported by Universidade Federal do Pará, Pró-reitoria de Pesquisa e Pós-graduação and Conselho Nacional de Desenvolvimento Científico e Tecnologia (CNPq; PP, 313303/2021-5)

## Ethical Statement

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Acknowledgments

We would like to thank the High Performance Computing Center of the Federal University of Pará (CCAD) for providing the Apolo cluster necessary to perform all analyses; Dr. Edivaldo Costa Sousa Júnior for contributing to the improvement of the data and results generated; and MsC Daniel de Souza Avelar for technical assistance in the process of downloading the NCBI files.

## Footnotes

*Reporting Checklist:* The authors have completed the Narrative Review reporting checklist.

*Conflicts of Interest:* The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. All of them have completed the ICMJE uniform disclosure form.

## References

1 Witschi H. A short history of lung cancer. *Toxicol Sci Off J Soc Toxicol* 2001; 64: 4–6.

2 Sung H, Ferlay J, Siegel RL, *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021; 71: 209–249.

3 Ostroverkhova D, Przytycka TM, Panchenko AR. Cancer driver mutations: predictions and reality. *Trends Mol Med* 2023; 29: 554–566.

4 Calabrese C, Davidson NR, Demircioğlu D, *et al.* Genomic basis for RNA alterations in cancer. *Nature* Nature Publishing Group; 2020; 578: 129–136.

5 INCA IN de. Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro, RJ: Instituto Nacional De Câncer; 2023.

6 Faner R, Sibila O, Agustí A, *et al.* The microbiome in respiratory medicine: current challenges and future perspectives. *Eur Respir J* 2017; 49: 1602086.

7 Altveş S, Yildiz HK, Vural HC. Interaction of the microbiota with the human body in health and diseases. *Biosci Microbiota Food Health* 2020; 39: 23–32.
8 Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012; 486: 207–214.
9 Yagi K, Huffnagle GB, Lukacs NW, *et al.* The Lung Microbiome during Health and Disease. *Int J Mol Sci* 2021; 22: 10872.
10 Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* Nature Publishing Group; 2017; 15: 259–270.
11 Ibironke O, McGuinness LR, Lu S-E, *et al.* Species-level evaluation of the human respiratory microbiome. *GigaScience* 2020; 9: giaa038.
12 Natalini JG, Singh S, Segal LN. The dynamic lung microbiome in health and disease. *Nat Rev Microbiol* 2023; 21: 222–235.
13 Charlson ES, Bittinger K, Haas AR, *et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011; 184: 957–963.
14 Whiteside SA, McGinniss JE, Collman RG. The lung microbiome: progress and promise. *J Clin Invest* [Internet] American Society for Clinical Investigation; 2021 [cited 2023 Oct 16]; 131. Available from: https://www.jci.org/articles/view/150473.
15 Cho I, Blaser MJ. The Human Microbiome: at the interface of health and disease. *Nat Rev Genet* 2012; 13: 260–270.
16 Dickson RP, Erb-Downward JR, Huffnagle GB. The role of the bacterial microbiome in lung disease. *Expert Rev Respir Med* 2013; 7: 245–257.
17 Costa AN, da Costa FM, Campos SV, *et al.* The pulmonary microbiome: challenges of a new paradigm. *J Bras Pneumol* 2018; 44: 424–432.
18 Dickson RP, Erb-Downward JR, Martinez FJ, *et al.* The Microbiome and the Respiratory Tract. *Annu Rev Physiol* 2016; 78: 481–504.
19 Chorlton SD. Ten common issues with reference sequence databases and how to mitigate them. *Front Bioinforma* [Internet] Frontiers; 2024 [cited 2024 Aug 22]; 4. Available from: https://www.frontiersin.org/journals/bioinformatics/articles/10.3389/fbinf.2024.1278228/full.
20 Sharma A, Laxman B, Naureckas ET, *et al.* Associations between fungal and bacterial microbiota of airways and asthma endotypes. *J Allergy Clin Immunol* Elsevier; 2019; 144: 1214-1227.e7.
21 Soret P, Vandenborght L-E, Francis F, *et al.* Respiratory mycobiome and suggestion of inter-kingdom network during acute pulmonary exacerbation in cystic fibrosis. *Sci Rep* 2020; 10: 3589.
22 Young JC, Chehoud C, Bittinger K, *et al.* Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am J Transplant Off J Am Soc Transplant Am Soc Transpl Surg* 2015; 15: 200–209.
23 Ruiz JC, Lücke L, Minarowski Ł, *et al.* Human cytomegalovirus infection effects on lung cancer prognosis after surgical resection. *Eur Respir J* [Internet] European Respiratory Society; 2019

[cited 2024 Jul 15]; 54. Available from: https://erj.ersjournals.com/content/54/suppl_63/PA3053.
24 Pattaroni C, Watzenboeck ML, Schneidegger S, *et al.* Early-Life Formation of the Microbial and Immunological Environment of the Human Airways. *Cell Host Microbe* 2018; 24: 857-865.e4.
25 Ansaldo E, Farley TK, Belkaid Y. Control of Immunity by the Microbiota. *Annu Rev Immunol* Annual Reviews; 2021; 39: 449–479.
26 Belkaid Y, Hand TW. Role of the Microbiota in Immunity and Inflammation. *Cell* 2014; 157: 121–141.
27 Hou K, Wu Z-X, Chen X-Y, *et al.* Microbiota in health and diseases. *Signal Transduct Target Ther* 2022; 7: 135.
28 Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* 2022; 12: 31–46.
29 Xu N, Wang L, Li C, *et al.* Microbiota dysbiosis in lung cancer: evidence of association and potential mechanisms. *Transl Lung Cancer Res* 2020; 9: 1554–1568.
30 Routy B, Le Chatelier E, Derosa L, *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* 2018; 359: 91–97.
31 Shiravand Y, Khodadadi F, Kashani SMA, *et al.* Immune Checkpoint Inhibitors in Cancer Therapy. *Curr Oncol* 2022; 29: 3044–3060.
32 Dong Q, Chen ES, Zhao C, *et al.* Host-Microbiome Interaction in Lung Cancer. *Front Immunol* 2021; 12: 679829.
33 Sommariva M, Le Noci V, Bianchi F, *et al.* The lung microbiota: role in maintaining pulmonary immune homeostasis and its implications in cancer development and therapy. *Cell Mol Life Sci* 2020; 77: 2739–2749.
34 Wang J, Li F, Sun R, *et al.* Bacterial colonization dampens influenza-mediated acute lung injury via induction of M2 alveolar macrophages. *Nat Commun* 2013; 4: 2106.
35 Kovaleva OV, Romashin D, Zborovskaya IB, *et al.* Human Lung Microbiome on the Way to Cancer. *J Immunol Res* 2019; 2019: 1394191.
36 Hosang L, Canals RC, van der Flier FJ, *et al.* The lung microbiome regulates brain autoimmunity. *Nature* Nature Publishing Group; 2022; 603: 138–144.
37 Wang Z, Locantore N, Haldar K, *et al.* Inflammatory Endotype–associated Airway Microbiome in Chronic Obstructive Pulmonary Disease Clinical Stability and Exacerbations: A Multicohort Longitudinal Analysis. *Am J Respir Crit Care Med* 2021; 203: 1488–1502.
38 Francescone R, Hou V, Grivennikov SI. Microbiome, inflammation, and cancer. *Cancer J Sudbury Mass* 2014; 20: 181–189.
39 Tsay J-CJ, Wu BG, Badri MH, *et al.* Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer. *Am J Respir Crit Care Med* 2018; 198: 1188–1198.
40 Liu H-X, Tao L-L, Zhang J, *et al.* Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. *Infect Causes Cancer* 2018; .

41 Pflughoeft KJ, Versalovic J. Human Microbiome in Health and Disease. *Annu Rev Pathol Mech Dis* Annual Reviews; 2012; 7: 99–122.
42 Zitvogel L, Daillère R, Roberti MP, *et al.* Anticancer effects of the microbiome and its products. *Nat Rev Microbiol* 2017; 15: 465–478.
43 Picardo SL, Coburn B, Hansen AR. The microbiome and cancer for clinicians. *Crit Rev Oncol Hematol* 2019; 141: 1–12.
44 Pirlog R, Cismaru A, Nutu A, *et al.* Field Cancerization in NSCLC: A New Perspective on MicroRNAs in Macrophage Polarization. *Int J Mol Sci* 2021; 22: 746.
45 Chang Y-S, Hsu M-H, Tu S-J, *et al.* Metatranscriptomic Analysis of Human Lung Metagenomes from Patients with Lung Cancer. 2021; .
46 Fogel N. Tuberculosis: a disease without boundaries. *Tuberc Edinb Scotl* 2015; 95: 527–531.
47 Mur LA, Huws SA, Cameron SJ, *et al.* Lung cancer: a new frontier for microbiome research and clinical translation. *Ecancermedicalscience* 2018; 12: 866.
48 Yu G, Gail MH, Consonni D, *et al.* Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome Biol* 2016; 17: 163.
49 Greathouse KL, White JR, Vargas AJ, *et al.* Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol* 2018; 19: 123.
50 Yan X, Yang M, Liu J, *et al.* Discovery and validation of potential bacterial biomarkers for lung cancer. *Am J Cancer Res* 2015; 5: 3111–3122.
51 Su K, Gao Y, He J. A comparison of the microbiome composition in lower respiratory tract at different sites in early lung cancer patients. *Transl Lung Cancer Res* 2023; 12: 1264–1275.
52 Dohlman AB, Klug J, Mesko M, *et al.* A pan-cancer mycobiome analysis reveals fungal involvement in gastrointestinal and lung tumors. *Cell* 2022; 185: 3807-3822.e12.
53 Mao Q, Ma W, Wang Z, *et al.* Differential flora in the microenvironment of lung tumor and paired adjacent normal tissues. *Carcinogenesis* 2020; 41: 1094–1103.
54 Narunsky-Haziza L, Sepich-Poore GD, Livyatan I, *et al.* Pan-cancer analyses reveal cancer-type-specific fungal ecologies and bacteriome interactions. *Cell* 2022; 185: 3789-3806.e17.
55 Ha CWY, Devkota S. The new microbiology: cultivating the future of microbiome-directed medicine. *Am J Physiol Gastrointest Liver Physiol* 2020; 319: G639–G645.
56 Schloss PD, Handelsman J. Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 2003; 14: 303–310.
57 Joseph TA, Pe'er I. An Introduction to Whole-Metagenome Shotgun Sequencing Studies. *Methods Mol Biol Clifton NJ* 2021; 2243: 107–122.
58 Quince C, Walker AW, Simpson JT, *et al.* Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 2017; 35: 833–844.
59 Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Front Genet* 2015; 6: 348.

584 60 Pérez-Cobas AE, Gomez-Valero L, Buchrieser C. Metagenomic approaches in microbial ecology:
585 an update on whole-genome and marker gene sequencing analyses. *Microb Genomics* 2020; 6:
586 mgen000409.
587 61 Orakov A, Fullam A, Coelho LP, *et al.* GUNC: detection of chimerism and contamination in
588 prokaryotic genomes. *Genome Biol* 2021; 22: 178.
589 62 Chen L-X, Anantharaman K, Shaiber A, *et al.* Accurate and complete genomes from metagenomes.
590 *Genome Res* 2020; 30: 315–333.
591 63 Benítez-Páez A, Belda-Ferre P, Simón-Soro A, *et al.* Microbiota diversity and gene expression
592 dynamics in human oral biofilms. *BMC Genomics* 2014; 15: 311.
593 64 Franzosa EA, Morgan XC, Segata N, *et al.* Relating the metatranscriptome and metagenome of the
594 human gut. *Proc Natl Acad Sci U S A* 2014; 111: E2329-2338.
595 65 Gosalbes MJ, Durbán A, Pignatelli M, *et al.* Metatranscriptomic approach to analyze the functional
596 human gut microbiota. *PloS One* 2011; 6: e17447.
597 66 Rajagopala SV, Bakhoum NG, Pakala SB, *et al.* Metatranscriptomics to characterize respiratory
598 virome, microbiome, and host response directly from clinical samples. *Cell Rep Methods* 2021; 1:
599 100091.
600 67 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*
601 2009; 10: 57–63.
602 68 Aguiar-Pulido V, Huang W, Suarez-Ulloa V, *et al.* Metagenomics, Metatranscriptomics, and
603 Metabolomics Approaches for Microbiome Analysis. *Evol Bioinforma Online* 2016; 12: 5–16.
604 69 Shakya M, Lo C-C, Chain PSG. Advances and Challenges in Metatranscriptomic Analysis. *Front*
605 *Genet* 2019; 10: 904.
606 70 Bashiardes S, Zilberman-Schapira G, Elinav E. Use of Metatranscriptomics in Microbiome
607 Research. *Bioinforma Biol Insights* 2016; 10: 19–25.
608 71 Raja G, Gupta H, Gebru YA, *et al.* Recent Advances of Microbiome-Associated Metabolomics
609 Profiling in Liver Disease: Principles, Mechanisms, and Applications. *Int J Mol Sci* 2021; 22: 1160.
610 72 Zarei I, Koistinen VM, Kokla M, *et al.* Tissue-wide metabolomics reveals wide impact of gut
611 microbiota on mice metabolite composition. *Sci Rep* 2022; 12: 15018.
612 73 Pan S, Chen R. Metaproteomic analysis of human gut microbiome in digestive and metabolic
613 diseases. *Adv Clin Chem* 2020; 97: 1–12.
614 74 Qian X, Zhang H-Y, Li Q-L, *et al.* Integrated microbiome, metabolome, and proteome analysis
615 identifies a novel interplay among commensal bacteria, metabolites and candidate targets in non-
616 small cell lung cancer. *Clin Transl Med* 2022; 12: e947.
617 75 Jochum M, Lee MD, Curry K, *et al.* Analysis of bronchoalveolar lavage fluid metatranscriptomes
618 among patients with COVID-19 disease. *Sci Rep* 2022; 12: 21125.
619 76 Schneeberger PHH. Microbiota analysis optimization for human bronchoalveolar lavage fluid.
620 2019; .

77 Lemieux ME, Reveles XT, Rebeles J, *et al.* Detection of early-stage lung cancer in sputum using automated flow cytometry and machine learning. *Respir Res* 2023; 24: 23.
78 Carney SM, Clemente JC, Cox MJ, *et al.* Methods in Lung Microbiome Research. *Am J Respir Cell Mol Biol* 2020; 62: 283–299.
79 R Core Team. A Language and Environment for Statistical Computing. 2024.
80 Seixas S, Kolbe AR, Gomes S, *et al.* Comparative analysis of the bronchoalveolar microbiome in Portuguese patients with different chronic lung disorders. *Sci Rep* 2021; 11: 15042.
81 Qian K, Deng Y, Krimsky WS, *et al.* Airway Microbiota in Patients With Synchronous Multiple Primary Lung Cancer: The Bacterial Topography of the Respiratory Tract. *Front Oncol* 2022; 12: 811279.
82 Ekanayake A, Madegedara D, Chandrasekharan V, *et al.* Respiratory Bacterial Microbiota and Individual Bacterial Variability in Lung Cancer and Bronchiectasis Patients. *Indian J Microbiol* 2020; 60: 196–205.
83 Xia X, Chen J, Cheng Y, *et al.* Comparative analysis of the lung microbiota in patients with respiratory infections, tuberculosis, and lung cancer: A preliminary study. *Front Cell Infect Microbiol* [Internet] Frontiers; 2022 [cited 2024 Apr 25]; 12. Available from: https://www.frontiersin.org/articles/10.3389/fcimb.2022.1024867.
84 Cameron SJS, Lewis KE, Huws SA, *et al.* A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. *PLoS ONE* 2017; 12: e0177062.
85 Huang DH, He J, Su XF, *et al.* The airway microbiota of non-small cell lung cancer patients and its relationship to tumor stage and gene mutation. *Thorac Cancer* 2022; 13: 858–869.
86 Apopa PL, Alley L, Penney RB, *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.
87 Peters BA, Pass HI, Burk RD, *et al.* The lung microbiome, peripheral gene expression, and recurrence-free survival after resection of stage II non-small cell lung cancer. *Genome Med* 2022; 14: 121.
88 Kovaleva O, Podlesnaya P, Rashidova M, *et al.* Lung Microbiome Differentially Impacts Survival of Patients with Non-Small Cell Lung Cancer Depending on Tumor Stroma Phenotype. *Biomedicines* Multidisciplinary Digital Publishing Institute; 2020; 8: 349.
89 Chang Y-S, Hsu M-H, Tu S-J, *et al.* Metatranscriptomic Analysis of Human Lung Metagenomes from Patients with Lung Cancer. *Genes* Multidisciplinary Digital Publishing Institute; 2021; 12: 1458.
90 Chen S, Zhou Y, Chen Y, *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018; 34: i884–i890.
91 Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019; 20: 257.

92 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15: 550.

93 Barra WF, Sarquis DP, Khayat AS, *et al.* Gastric Cancer Microbiome. *Pathobiol J Immunopathol Mol Cell Biol* 2021; 88: 156–169.

# 4 CAPÍTULO 2 – TRANSCRIPTIONAL LANDSCAPE OF LONG NON-CODING RNAS AND THEIR REGULATORY TARGETS IN NON-SMALL CELL LUNG CANCER

# TRANSCRIPTIONAL LANDSCAPE OF LONG NON-CODING RNAS AND THEIR REGULATORY TARGETS IN NON-SMALL CELL LUNG CANCER

Mariana S. De Lima[1, 2], Sérgio A. Ramos[1, 2], Lorena D. Fernandes[1, 3, 4], Camila T. Uchôa[1, 3], Marcelli G. Da Silva[1, 2], Jéssica M. Silva[1, 2], Ronald M. Mourão[1, 2], Eliel B. Teixeira[1, 5, 6], Diego Pereira[1, 2], Valéria C. da Silva[1, 2], Daniel D. Avelar[1, 5], Alexandre F. Virla[4], Maria Morard[4], Samir M. Casseb[1], Fabiano C. Moreira[1], Paulo P. Assumpção[1*]

[1]Núcleo de Pesquisas em Oncologia, Universidade Federal do Pará, Brazil, [2]Programa de Pós-graduação em Genética e Biologia Molecular, Brazil, [3]Programa de Pós-graduação em Oncologia e Ciências Médicas, Brazil, [4]Universidade Federal do Estado do Rio de Janeiro, Brazil, [5]Programa de Pós-Graduação em Biotecnologia, Brazil, [6]Universidade Federal do Amapá, Brazil



*Submitted to Journal:*
Frontiers in Oncology

*Specialty Section:*
Thoracic Oncology

*Article type:*
Original Research Article

*Manuscript ID:*
1539927

*Received on:*
05 Dec 2024

*Journal website link:*
www.frontiersin.org

## Scope Statement

Non-small cell lung cancer (NSCLC) is one of the most prevalent and lethal malignancies, making it essential to study the underlying molecular mechanisms. Long non-coding transcripts (lncRNAs) have emerged as important regulators, but their role in NSCLC is still poorly understood, especially regarding interactions with coding transcripts and their impact on biological pathways. Our article addresses these gaps through differential expression analyses of lncRNAs and biological pathway enrichment in tumor and adjacent samples. The results reveal critical processes, such as increased activation of the KRAS pathway, angiogenesis, DNA replication, cell growth and proliferation, and immune, stress, and inflammatory responses. We also highlight how lncRNAs interact with their coding targets, contributing to the understanding of NSCLC mechanisms. Our approach provides relevant insights for the identification of biomarkers and therapeutic targets in NSCLC, being aligned with the scope of the journal and of interest to researchers and clinicians seeking advances in oncology and cancer biology.

## Conflict of interest statement

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

## Credit Author Statement

Paulo Pimentel de Assumpção: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing. Alexandre Finoquio Virla: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Daniel De Souza Avelar: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Diego Pereira: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Eliel Barbosa Teixeira: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Fabiano Cordeiro Moreira: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Jéssica Manoelli Costa Silva: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Lorena Duarte Fernandes: Conceptualization, Data curation, Formal Analysis, Investigation, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Maria Morard: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Mariana Sousa De Lima: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Marcelli Geisse De Oliveira Prata Da Silva: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Ronald Matheus Da Silva Mourão: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Samir Mansour Casseb: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing. Sérgio Augusto Antunes Ramos: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Camila Tavares Carvalho Uchôa: Conceptualization, Data curation, Formal Analysis, Investigation, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing. Valéria Cristiane Santos da Silva: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – review & editing.

## Keywords

lung cancer, Transcriptomics, lncRNA, lncRNA-mRNA interaction, biomarkers

## Abstract

Word count: 251

Lung cancer (LC), largely caused by smoking, is a leading cause of cancer deaths globally. Non-small cell lung cancer (NSCLC), the most common type, is often diagnosed at advanced stages with low survival rates. Genetic alterations, including those in non-coding RNAs, play a key role in tumor progression. Long non-coding RNAs (lncRNAs) are emerging as potential biomarkers and therapeutic

targets in LC. This study investigates the differential expression of genes and lncRNAs in NSCLC subtypes, LUAD and LUSC to identify potential biomarkers. Methods: Twenty-seven samples were used, of which 18 were tumors and 7 were adjacent tissue from patients diagnosed with lung cancer, who signed an informed consent form. Total RNA was extracted, quantified and sequenced on the NextSeq 500 platform (Illumina). After processing the readings, the data were processed in RStudio software, where differential expression, statistics, ROC curve and data visualization analyses were performed. The expression data of long non-coding RNAs were correlated with the expression data of coding RNAs, whose interaction was previously confirmed in the ENCORI webserver. Results: A total of 162 differentially expressed genes were identified, with distinct expression patterns between the subtypes. Pathway enrichment analysis highlighted the involvement of inflammation-related pathways such as TNFA signaling, IL6-JAK-STAT3, and apoptosis, which are crucial in cancer progression. Correlation analysis between lncRNAs and coding genes revealed significant interactions linked to tumor progression, metabolism, and immune response.The findings suggest that lncRNAs like MIR924HG and MIR22HG, along with specific coding genes, could serve as potential biomarkers and therapeutic targets for lung cancer.

## *Funding information*

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado do Pará (FAPESPA).

## *Funding statement*

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article.

## *Ethics statements*

### *Studies involving animal subjects*
Generated Statement: No animal studies are presented in this manuscript.

### *Studies involving human subjects*
Generated Statement: The studies involving humans were approved by National Health Council Research Ethics Committee 41667021430010017. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

### *Inclusion of identifiable human data*
Generated Statement: No potentially identifiable images or data are presented in this study.

## *Data availability statement*

Generated Statement: The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept a manuscript that does not adhere to our open data policies.

## *Generative AI disclosure*

No Generative AI was used in the preparation of this manuscript.

# TRANSCRIPTIONAL LANDSCAPE OF LONG NON-CODING RNAS AND THEIR REGULATORY TARGETS IN NON-SMALL CELL LUNG CANCER

**Mariana Souza De Lima[1,2+]; Sérgio Augusto Antunes Ramos[1,2+]; Lorena Duarte Fernandes[1,3,6]; Camila Tavares Carvalho Uchôa[1,3]; Marcelli Geisse De Oliveira Prata Da Silva[1,2]; Jéssica Manoelli Costa Silva[1,2]; Ronald Matheus da Silva Mourão[1,2]; Eliel Barbosa Teixeira[1,4,7]; Diego Pereira[1,2]; Valéria Cristiane Santos da Silva[1,2]; Daniel de Souza Avelar[1,4]; Alexandre Virla[6]; Maria Morard[6]; Samir Mansour Moraes Casseb[1]; Fabiano Cordeiro Moreira[1]; Paulo Pimentel de Assumpção[1*]**

[1] Núcleo de Pesquisas em Oncologia, Universidade Federal do Pará, Belém, PA, Brazil
[2] Programa de Pós-graduação em Genética e Biologia Molecular, Belém, PA, Brazil
[3] Programa de Pós-graduação em Oncologia e Ciências Médicas, Belém, PA, Brazil
[4] Programa de Pós-Graduação em Biotecnologia, Belém, PA, Brazil
[5] Programa de Pós-graduação em Neurologia, Rio de Janeiro, RJ, Brazil
[6] Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[7] Universidade Federal do Amapá, Macapá, AP, Brazil.

[+] **These authors share the first authorship**

[*]**Correspondence author:**

assumpcaopp@gmail.com



**Key-words:** lung cancer; transcriptomics; lncRNA; lncRNA-mRNA interaction; biomarkers.

## Abstract

**Introduction:** Lung cancer (LC), largely caused by smoking, is a leading cause of cancer deaths globally. Non-small cell lung cancer (NSCLC), the most common type, is often diagnosed at advanced stages with low survival rates. Genetic alterations, including those in non-coding RNAs, play a key role in tumor progression. Long non-coding RNAs (lncRNAs) are emerging as potential biomarkers and therapeutic targets in LC. This study investigates the differential expression of genes and lncRNAs in NSCLC subtypes, LUAD and LUSC to identify potential biomarkers. **Methods:** Twenty-seven samples were used, of which 18 were tumors and 7 were adjacent tissue from patients diagnosed with lung cancer, who signed an informed consent form. Total RNA was extracted, quantified and sequenced on the NextSeq 500 platform (Illumina). After processing the readings, the data were processed in RStudio software, where differential expression, statistics, ROC curve and data visualization analyses were performed. The expression data of long non-coding RNAs were correlated with the expression data of coding RNAs, whose interaction was previously confirmed in the ENCORI webserver. **Results:** A total of 162 differentially expressed genes were identified, with distinct expression patterns between the subtypes. Pathway enrichment analysis highlighted the involvement of inflammation-related pathways such as TNFA signaling, *IL6-JAK-STAT3*, and apoptosis, which are crucial in cancer progression. Correlation analysis between lncRNAs and coding genes revealed

significant interactions linked to tumor progression, metabolism, and immune response. **Conclusion:** The findings suggest that lncRNAs like MIR924HG and MIR22HG, along with specific coding genes, could serve as potential biomarkers and therapeutic targets for lung cancer.

## 1 Introduction

Lung cancer (LC) is one of the most prevalent cancer types worldwide, with smoking being the primary preventable risk factor. It is estimated to cause 1.8 million deaths annually across both sexes, accounting for 18.7% of all cancer-related deaths, and remains the leading cause of cancer mortality in men[1].

The high incidence of LC is likely linked to the persistent use of tobacco, often associated with age (currently a higher risk between 50 and 74 years), duration of exposure, smoking intensity, and sociodemographic factors such as sex, race, and education [2,3]. Smoking, as the primary environmental contributor to LC, induces epigenetic alterations in tumor suppressor genes (e.g., *TP16*), cell cycle regulators (e.g., *DAPK* and *TERT*), apoptosis-related genes (e.g., *BCL2*), and other genes involved in cellular metabolism, such as *FHIT, CYP1A1, P450, PTGS2, EDNRB*, and *SOCS1*[4].

Non-small cell lung cancer (NSCLC) accounts for over 80% of LC cases. It is characterized as incidental and aggressive, typically diagnosed at advanced stages with low survival rates. This broad category includes various subtypes, with adenocarcinoma being the most prevalent[5].

Genetic and epigenetic alterations significantly influence tumor initiation and progression, including mutations, chromosomal instability, abnormal expression of non-coding RNAs (ncRNAs), and microbiota dysbiosis, all of which can drive carcinogenesis[6-10].

Non-coding RNAs (ncRNAs) are genetic sequences that do not encode proteins but act as transcriptional and post-transcriptional regulators of protein expression[11]. Since their discovery, ncRNAs have emerged as critical regulators of numerous biological functions in various cell types and tissues. Their dysregulation is implicated in diseases, particularly cancer[12]. Regulatory ncRNAs are classified by size into short non-coding RNAs (sncRNAs) and long non-coding RNAs (lncRNAs), the latter being larger than 200 nucleotides[13].

In RNA research, Next-Generation Sequencing (NGS) has revolutionized genomic studies, enabling economical, high-throughput analysis. NGS has facilitated comprehensive genome sequencing, transcriptomics, epigenomics, metagenomics, and other omics studies, revealing that most of the genome is transcribed into long RNAs[14-16].

Understanding the underlying causes of LC initiation and progression is crucial for developing novel therapies and biomarkers for early detection. LncRNAs are promising candidates for both purposes, as their expression alterations can reflect specific stages of tumor progression or trigger signaling pathways[17-19].

Large-scale studies have analyzed lncRNA expression in LC, identifying differentially expressed lncRNAs. Most upregulated lncRNAs function as oncogenes, while downregulated lncRNAs often act as tumor suppressors. However, some lncRNAs remain controversial in their roles as tumor promoters or suppressors, underscoring the need for further research[20,21].

This study aims to characterize the profile of long non-coding RNAs and their targets in lung cancer samples, correlating their expression with coding transcripts to identify potential biomarkers.

## 2 Materials and Methods

### 2.1 Ethical aspects and patient selection

In Brazil, the ethical aspects involved in research activities involving human beings are regulated by the guidelines of Resolution 466/12 of the National Health Council (CNS), established on December, 2012[22].

First, the study was submitted to Plataforma Brasil for consideration by the Research Ethics Committee (CEP), under No. 41667021.4.3001.0017. Patients who voluntarily agreed to participate in the research were informed by the health professional about the objectives, benefits and possible risks of the research, as well as about the guarantee of their privacy and confidentiality regarding the information processed, so that they could freely and clearly choose to participate in the research. Individuals who agreed to participate in the research signed the Free and Informed Consent Form (FICF).

### 2.2 Obtaining samples and clinical data

Eighteen unpaired samples of tumor tissue and nine samples of adjacent non-tumor lung tissue were obtained from patients undergoing surgery at João de Barros Barreto University Hospital (HUJBB). Clinical data from participants were collected through medical record reviews, providing access to their medical history and enabling patients to complete a questionnaire.

Paired samples could not be utilized because tumor samples were obtained via CT-guided biopsy, which precluded the collection of adjacent tissue. Additionally, in cases where paired samples were collected, the quality of one of the samples was compromised during extraction or sequencing.

### 2.3 Total RNA extraction

Approximately 50–100 ng of tumor tissue from each sample was macerated for RNA extraction using the TRIzol Reagent method (Thermo Fisher Scientific) following the manufacturer's instructions. Total RNA was then assessed for integrity and concentration (ng/μL) using the Qubit 4 Fluorometer (Thermo Fisher Scientific). The established criteria for RNA Integrity Number (RIN) were $\geq 5$.

### 2.4 Library construction and RNA sequencing

Library construction was performed using the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold (Illumina, USA) according to the manufacturer's instructions, using approximately 1 μg of total RNA per sample in a total volume of 11 μL. After completion of the libraries, samples were evaluated for the integrity of the generated cDNA fragments using the 2200 TapeStation system (Agilent Technologies AG, Basel, Switzerland).

Sequencing was performed in paired-end mode on the NextSeq® 500 platform (Illumina®, USA), using the NextSeq® 500 MID Output V2 - 150 cycles kit (Illumina®) according to the manufacturer's instructions.

### 2.5 Reads quality control

After sequencing, the obtained RNA-Seq reads were subjected to quality control. Initially, the quality of the reads was assessed using FastQC. Subsequently, adapters and low-quality reads were removed using Trimmomatic v0.39, with a quality value (QV) threshold set above 15.

### 2.6 Alignment with the human genome

Human transcript reads were characterized through alignment and quantification using Salmon v1.5.2, with hg coding transcripts v38 (www.ensembl.org) used as a reference index and GENCODE v.42 (www.gencodegenes.org) used for annotation. Reads were imported from Salmon into Rstudio using the Tximport v3.14.0 package.



### 2.7 Differential and statistical analysis

Differential gene expression (DEG) analysis and data normalization were performed on the RStudio platform using the DESeq2 package [23, 24]. The selection of the best non-coding biomarkers was performed by ROC curve analysis, with AUC parameter >= 0.9 and p-value <= 5%, in which the remaining lncRNAs were plotted in heatmap by the ComplexHeatmap package, indicating the "spearman" correlation method. Each gene cluster was named according to the biological functions that the genes perform as documented in the literature.

The sample grouping assessment occurred by ordered principal component analysis (PCoA), using PERMANOVA as a statistical test to assess the significance of the groups.

### 2.8 Pathway enrichment analysis and correlation with coding transcripts

Functional enrichment analysis was performed using Gene Set Enrichment Analysis (GSEA) with the ClusterProfiler package in R, based on differential expression data [25]. Hallmark gene sets (category H) were obtained from the MSigDB database using the msigdbr package and converted to a list suitable for analysis. GSEA was then performed using the sorted gene array, with genes mapped to the MSigDB-defined sets. The top enriched sets were visualized in a dotplot, highlighting the 10 most significant ones, selected based on the lowest adjusted p-value (<0.05).

To verify the correlation of lncRNA expression and human coding genes, the same methodology of differential analysis and normalization for coding transcripts was performed as previously. After that, the ENCORI webserver was used to select which lncRNAs interact with coding transcripts found in the transcriptome data [26]. From a table with the lncRNAs and their coding targets, a correlation test

was performed using the Spearman method and p-value <= 5% and the result was represented in a corplot in RStudio.

## 3 Results and discussion

### 3.1 Sample characterization

Data from 24 lung cancer patients, including 14 females and 10 males, were evaluated using medical records and completed data forms. The age ranges of this population (**Figure 1A**) were organized into five-year intervals (50–55 to 76–80), with males represented in blue and females in pink. The mean age was 62.3 years for males (SD: 40.49) and 64.7 years for females (SD: 42.12), with a higher concentration of patients in the 61–65 age range, particularly among females.

Most LC cases occur between 50 and 70 years of age. The incidence intensifies after 40, often linked to comorbidities and/or lifestyle factors. In 2022, 85.6% of cases involved individuals aged over 40, with a predominance of females between 40–59 years and males aged 60 and above[27]. LC's impact on women is growing[28]. Domagala-Kulawik and Trojnar (2020) attribute this trend to differing hormonal states, suggesting that biological differences may explain the varied clinical manifestation in men and women[29]. Factors such as 17-β-estradiol levels, estrogen receptor activity, aromatase expression, and pituitary sex hormone receptors, along with genetic aberrations related to estrogen, are under investigation.

In **Figure 1B**, vertical bars depict the distribution of LC histopathological types (adenocarcinoma, squamous cell carcinoma, and undifferentiated) stratified by survival status (alive vs. deceased). Among deceased patients, adenocarcinoma was the most prevalent type, followed by squamous cell carcinoma and undifferentiated carcinoma. In surviving patients, adenocarcinoma was significantly more frequent, with a lower incidence of squamous cell carcinoma. The **Figure 1C** presents clinical staging and LC histopathological types. The vertical bar chart illustrates the relationship between LC stage (I to IV, with I being the less advanced and II - IV the most advanced) and histopathological type. Advanced-stage disease was predominant across both histological types.

Late-stage diagnosis significantly impacts life expectancy. A cohort study conducted in Brazil identified economic and structural barriers as contributors to advanced-stage diagnosis[30]. Most LC cases present late, with diagnosis typically taking 4–5 months. Delays are influenced by patient-related factors, such as low educational levels, and physician-related factors, such as unnecessary anti-tuberculosis treatments. Addressing these gaps is critical to reduce delays and ensure timely care and treatment[31, 32].

Following an LC diagnosis, smoking cessation is associated with improved survival, reduced postoperative complications, enhanced therapy response, and better quality of life. Survival in NSCLC correlates strongly with cancer stage at diagnosis. For instance, five-year survival rates are 68.4% for stage I and 5.8% for stage IV patients. Unfortunately, most LC cases are detected at stage IV, leading to lower survival rates and increased symptom burden[33,34].

In Brazil, mortality rates increase steadily with age, becoming more pronounced after 60 years. In 2019, the mortality rate was 0.11 for individuals under 49 years, rising to 6.25 for those aged 60–69, and reaching 16.65 among individuals aged 80 or older[3].

Finally, **Figure 1D** examines the smoking habits of men and women, correlating histopathological types with smoking status. The categories analyzed were smokers and non-smokers. The data in panel D indicate that the percentage of smokers is higher regardless of histopathological type.

Smoking is the primary risk factor for lung cancer, and former smokers have an increased risk compared to individuals who have never smoked. Over the past two decades, adult smoking prevalence has declined globally across all socioeconomic classes due to effective anti-smoking campaigns and restrictions. However, it is estimated that one in five adults worldwide continues to smoke, and nearly one-third of smokers have quit smoking in the last 20 years[35].

Lung cancer screening methods can reduce mortality risk by 20%, and when combined with successful smoking cessation, this benefit increases to an estimated 38%. Smoking cessation decreases LC mortality risk (6–11) and incidence (12–15) in both smokers and former smokers. Nonetheless, the risk of LC remains elevated for years compared to those who have never smoked[36].

Global data from the International Agency for Research on Cancer (IARC) indicate that, if current smoking behavior trends persist, LC incidence will increase by over 65%, and mortality will rise by 74% by 2040 compared to 2022[37].

A 2024 study by Malta et al. on tobacco use among adolescents in Brazil highlighted stable cigarette use between 2015 (6.6%) and 2019 (6.8%), but noted an increase in the use of any tobacco products (from 10.6% in 2015 to 14.8% in 2019), with waterpipe smoking being the most common (7.8%). This trend contributes to increased mortality from diseases such as LC and remains one of its principal risk factors[38].

**Figure 1** – Sample characterization of the present study. Graph of age groups by sex with respective means (A), number of living and deceased patients for each histopathological subtype of LC (B), staging of LC in each histopathological subtype, being "less advanced" (stage 1) and "more advanced" (stages 2-4) (C), smoking habit by histopathological type of LC (D).

### 3.2 Transcriptomic analysis of LC

The RNA-seq analysis yielded an average of approximately 3,400,000 reads per sample in the quantified read counts. The volcano plot illustrates the differential gene expression (DGE) between tumor and adjacent tissues, identifying 1,147 DEGs, with 447 underexpressed and 700 overexpressed genes (**Figure 2A**). Principal Coordinate Analysis (PCoA) demonstrated a clear and statistically significant separation between sample types (PERMANOVA; p-value < 5%), with principal component 1 accounting for 40.5% of intergroup variance (**Figure 2B**). Furthermore, the ROC curve analysis classified 58 DEGs (AUC ≥ 0.9), represented in the heatmap (**Figure 2C**), which highlights distinct expression profiles between tumor and adjacent tissues, forming distinct gene clusters (cluster 1 vs. clusters 2 and 3).

The heatmap also reveals critical cellular processes associated with each gene cluster, including cell proliferation, signaling, immune response, inflammation, angiogenesis, and cell motility. Notably, adjacent tissues already exhibit inflammatory characteristics, commonly linked to tumor microenvironments, suggesting a potential influence of the tumor microenvironment on neighboring tissues. Specific genes from the heatmap, including *MT1JP, SFTA1P, CASC19, CTTN-DT, IL12A-AS1, LCAL1, MIR3677HG,* and *PCAT6*, were selected for further analysis. Among these, *MT1JP* and

*SFTA1P* were downregulated in tumor tissues compared to adjacent tissues, whereas *CASC19, CTTN-DT, IL12A-AS1, LCAL1, MIR3677HG*, and *PCAT6* were upregulated in tumor tissues (**Figure 2D**).

**Figure 2 –** The volcano plot reveals the differentially expressed genes comparing tumor and adjacent tissue samples (A). The multidimensional PCoA plot shows two groups of tumor tissue (B). The heatmap shows the differentially expressed genes involved in biological processes by tissue type (C). The boxplot shows a plot of hypo- and hyperexpressed genes present in the heatmap (D).

### 3.3 Molecular distinctions and heterogeneity among NSCLC subtypes

The most common histological subtypes of NSCLC are LUAD and LUSC, each displaying distinct biological signatures. Differential gene expression analysis of these histological subtypes identified 162 DEGs, with 134 downregulated and 28 upregulated (**Figure 3A**). The PCoA plot comparing the histopathological subtypes showed a clear separation between LUAD and LUSC clusters (PERMANOVA; *p-value* = 0.001), suggesting notable molecular differences between the groups (**Figure 3B**).

The heatmap also highlights important cellular processes associated with each gene cluster, including homeostasis, response to oxidative stress, cell cycle control, regulation of the cell cycle, epithelial-mesenchymal transition, regulation of ion channels, and interaction with the tumor microbiota. *SFTA3, MIR205HG, UBQLN1-AS1*, and *NMRAL2P* are genes that are overexpressed in LUAD, while *MIR205HG, K6C, CALML3-AS1*, *CBLL1-AS1*, and *MSC-AS1* are overexpressed in LUSC (**Figure 3C**). Notably, distinct gene expression patterns between LUAD and LUSC suggest fundamental biological differences between the subtypes (**Figure 3D**).

It is evident that targets such as *MIR205HG* may be useful for targeted therapies due to their relevance in both subtypes, and in epithelial alterations, genes like *K6C* and *CALML3-AS1* could serve as therapeutic targets in LUSC, while *SFTA3* could be investigated in LUAD.

**Figure 3** – Volcano plot with differentially expressed genes between LUAD and LUSC histological types (A), PCoA graph showing separation of sample groups according to histopathological type (B), heatmap for DEGs of each histopathological subtype of LC (C), expression profile in boxplots of the DEGs illustrated in the heatmap (D).

Regarding the enrichment pathway analysis, **Figure 4** highlights *TNFA* Signaling via *NFκB* and inflammatory response, which present the highest proportions of genes, indicating that these pathways are strongly impacted in the context of lung cancer. These signatures may be related to inflammation and pro-tumor signaling, processes known to contribute to cancer progression.

*IL6-JAK-STAT3* signaling and *IL2-STAT5* signaling represent other significant pathways and may be involved in the regulation of inflammatory and immune responses, possibly reflecting the role of the tumor microenvironment and cytokines in cancer. Interferon gamma response, in turn, suggests a relationship with antitumor immunity; its activation may be associated with resistance mechanisms or immune activation. Apoptosis refers to changes in the apoptosis pathway, which are central to cancer,

as they affect the ability of cells to undergo programmed cell death[39]. This pathway, in turn, had the smallest proportion of genes linked to it.

Lung cancer is frequently characterized by an inflammatory microenvironment, as reflected in the enrichment of pathways such as *TNFA* Signaling via *NFκB* and Inflammatory Response. Additionally, the *IL6-JAK-STAT3* signaling pathway indicates the important involvement of complexes that modulate the expression of genes involved in cell proliferation, inflammation, and immune response, often associated with various inflammatory diseases and cancers due to its hyperactivation[40]. The epithelial-mesenchymal transition and *Kras* signaling pathways reflect processes involved in tumor invasion, metastasis, and activation of oncogenes linked to cell cycle activation.

**Figure 4** – Enrichment analysis of pathways associated with carcinogenesis processes.

These results reinforce the complexity and interconnectedness of the biological pathways involved in lung cancer progression, highlighting the relevance of inflammatory, immune, and tissue remodeling processes in tumor development. This underscores the central role of the tumor microenvironment and cellular signaling alterations in promoting tumor cell proliferation, invasion, and resistance. Thus, these signatures not only enhance the understanding of lung cancer pathophysiology but also point to potential therapeutic targets that could be explored for more effective treatment and intervention strategies.

The correlation plot shows differentially expressed lncRNAs and coding genes, highlighting those with previously confirmed biological interactions. The data reveal significant relationships between these elements, with correlations represented by colors and circle sizes (**Figure 5**).

Among the highlighted lncRNAs, MIR924HG shows a strong positive correlation with genes such as *FERMT2, IL6ST*, and *SLC39A8*, which are involved in crucial biological functions such as modulating pathways associated with tumor progression, the microenvironment, and the metabolic adaptation of cancer cells[41]. Notably, *SLC39A8*, a zinc transporter implicated in cellular metabolism, suggests that regulation by lncRNAs may be closely linked to tumor metabolic adaptation[42].

Another lncRNA of interest is *MIR22HG*, which correlates positively with *SGMS2*, a gene involved in lipid metabolism, modulating cell proliferation and apoptosis[43]. This suggests that *MIR22HG* plays a role in regulating the balance between cell growth and death in tumors. On the other hand, negative correlations, such as the one observed between *NTNG1* and lncRNAs like *LINC00641* and *GPRC5D-AS1*, indicate a possible repression of *NTNG1* function, which is associated with cell adhesion. The negative regulation of this gene may impact critical processes such as migration and cell differentiation in the tumor context[44,45].

These correlations highlight the relevance of lncRNAs in regulating essential tumor pathways, such as inflammation, metabolism, cell adhesion, and migration. Genes like IL6ST and *SLC39A8*, associated with lncRNAs, emerge as potential therapeutic targets for interventions in the tumor microenvironment. Furthermore, *MIR22HG* and *MIR924HG* stand out as promising biomarkers, given their extensive interaction with functionally critical genes for cancer. The identification of negative correlations, such as between NTNG1 and *LINC00641*, reveals additional regulatory mechanisms that may be explored, such as gene silencing in targeted therapies. These findings

reinforce the central role of lncRNAs in tumor biology and open pathways for the development of personalized therapies in cancer treatment.

**Figure 5** – Correlation graph between coding genes and differentially expressed lncRNAs that show confirmed biological interaction.

**Figure 6** presents an overview of differentially expressed lncRNAs in lung cancer, dividing them into two main categories: downregulated and upregulated lncRNAs, both of which biologically interact with coding transcripts. It is important to note that the interaction between these two classes of RNAs can have significant repercussions in carcinogenesis, as they play fundamental roles in biological processes relevant to tumor development, such as tissue remodeling, stress response, metabolism, inflammation, and regeneration.

The downregulation of lncRNAs, such as *RNF32-AS1, GPRC5D-AS1,* and *LINC00641*, appears to impact functions related to responses to hypoxia, tissue development, extracellular matrix, as well as cell adhesion and cytoskeleton processes. The interacting genes listed, such as *EPAS1, MAMDC2, ADAMTSL3,* and *NTNG1*, point to mechanisms regulating processes like the hypoxic response and angiogenesis (EPAS1), tissue function and extracellular matrix remodeling (*ADAMTSL3* and *MAMDC2*), and cell adhesion (*TMOD1* and *PDE3A*)[46,47].

Among the upregulated lncRNAs, such as *MIR924HG*, *CASC19*, and *ZFPM2-AS1*, functions related to tissue remodeling, inflammation, regeneration, and metabolic homeostasis stand out. These lncRNAs interact with genes like *FERMT2, IL6ST, SLC39A8,* and *SGMS2*, which are associated with cell migration and adhesion (*FERMT2*), inflammatory response and tissue regeneration (*IL6ST*), and metabolic homeostasis and oxidative stress (*SLC39A8*).

**Figura 6 –** Expression overview of differentially expressed lncRNAs in lung cancer and their targets.

## 4 Conclusions

Integrated information on lung cancer (LUAD and LUSC), including demographic analyses, gene expression, and lncRNA regulation patterns. The demographic and pathological analysis indicates that adenocarcinoma and squamous cell carcinoma (LUSC) similarly affect both sexes. However, there is a predominance of mortality associated with adenocarcinoma, especially in more advanced stages. Additionally, smoking appears as a factor associated with both LUAD and LUSC, highlighting its relevance in the development of these conditions.

lncRNA expression profiles show significant differential regulation, with some markers being upregulated and others downregulated in LUAD and LUSC subtypes. This differentiation is closely related to important biological processes such as cell migration, adhesion, and inflammatory response, suggesting a significant impact on the tumor microenvironment. In terms of correlation between lncRNAs and functional genes, these interactions play fundamental roles in critical pathways such as angiogenesis, metabolism, and tissue remodeling. Upregulated lncRNAs, such as MIR924HG, appear

to have essential functions in tumor progression, while downregulated lncRNAs, such as EPB41L4A-AS1, are associated with suppressed regulatory functions.

Finally, expression mapping and tumor subtypes reveal specific clusters of lncRNAs associated with LUAD and LUSC. lncRNAs such as MIR205HG and SUNO1 stand out in LUAD, suggesting their connection to epithelial transitions and metabolic alterations. These findings reinforce the critical role of lncRNAs in the progression of LUAD and LUSC, highlighting their relevance as potential diagnostic biomarkers and therapeutic targets specific to each tumor subtype.

## Conflit of interest

The authors declare that the research was conducted without commercial or financial relationships that could create a conflict of interest.

## Author Contributions

All authors contributed to all steps of the article's development, from research design to data analysis, manuscript writing, and final revisions.



## Funding

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado do Pará (FAPESPA).

## Acknowledgments

We thank the High-Performance Computing Center of the Federal University of Pará (CCAD) for providing the Apollo 2000 cluster necessary to perform all analyses.

## References

1. Morgan E, et al., Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. **Gut** 2023;72:338-344.
2. Araujo, L. H., Baldotto, C., Castro, G., Jr, Katz, A., Ferreira, C. G., Mathias, C., Mascarenhas, E., Lopes, G. L., Carvalho, H., Tabacof, J., Martínez-Mesa, J., Viana, L. S., Cruz, M. S., Zukin, M., Marchi, P., Terra, R. M., Ribeiro, R. A., Lima, V. C. C., Werutsky, G., Barrios, C. H., … Grupo Brasileiro de Oncologia Torácica (2018). Lung cancer in Brazil. *Jornal brasileiro de pneumologia : publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia*, *44*(1), 55–64.

3. Campos, M. R. et al. Smoking, mortality, access to diagnosis, and treatment of lung cancer in Brazil. **Revista De Saúde Pública**. 58(18). 2024.

4. Oliveira, N. F. P. Alterações no epigenoma e o hábito de fumar. **Revista Brasileira de Ciências da Saúde**, João Pessoa, v. 14, n. 4 ,p. 101-106, out./dez. 2011.

5. Oliveira, A. C. et al. Molecular profile of non-small cell lung cancer in northeastern Brazil. **Jornal Brasileiro de Pneumologia,** v. 45, n.3, p. 320180181, 2019.

6. Gonzalez, S, Pisano, David G; Serrano, Manuel. Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs. **Cell Cycle**, [*s. l.*], v. 7, n. 16, 15 ago. 2008.

7. Williams, T, Ngo L H, Wickramasinghe V O. Nuclear export of RNA: Different sizes, shapes and functions. **Semin Cell Dev Biol**. 2018 Mar;75:70-77. doi: 10.1016/j.semcdb.2017.08.054. Epub 2017 Sep 1. PMID: 28866329.

8. Williams M, CHENG YY, BLENKIRON C, REID G. Exploring Mechanisms of MicroRNA Downregulation in Cancer. **Microrna**. 6(1):2-16. 2017.

9. Ratti, Margherita *et al*. MicroRNAs (miRNAs) and Long Non-Coding RNAs (lncRNAs) as New Tools for Cancer Therapy: First Steps from Bench to Bedside. **Targeted Oncology**, [*s. l.*], v. 15, n. 3, 2020.

10. Xu, N., et al. Microbiota dysbiosis in lung cancer: evidence of association and potential mechanisms. **Translational lung cancer research**, *9*(4), 1554–1568. 2020.

11. Statello, Luisa *et al*. Gene regulation by long non-coding RNAs and its biological functions. **Nature Rev. Mol. Cell biol.**, [*s. l.*], 22 dez. 2020.

12. Nemeth, K., et al. Non-coding RNAs in disease: from mechanisms to therapeutics. **Nature reviews**. *Genetics*, *25*(3), 211–232. 2024.

13. He, Xiangyang et al. Epigenetic modifications in radiation-induced non-targeted effects and their clinical significance. **Biochimica et Biophysica Acta (BBA) - General Subjects**, Volume 1867, Issue 8, 2023.

14. Chen, Z. et al. feito Prottor do RNA não codificante longo OXCT1-AS1 na Apoptose de Células Miocárdicas Humanas Induzida pela Doxorrubicina pelo Padrão Competitivo de RNA endógeno. **Arquivos Brasileiros de Cardiologia**, v. 121, n. 6, p. e20230675, 2024.

15. Mattick, J.S., et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. **Nat Rev Mol Cell Biol** 24, 430–447. 2023

16. Satam, H., et al., Next-Generation Sequencing Technology: Current Trends and Advancements. **Biology**, *12*(7), 997. 2023.

17. Taheri M, Badrlou E, Hussen BM, Kashi AH, Ghafouri-Fard S and Baniahmad A (2023) Importance of long non-coding RNAs in the pathogenesis, diagnosis, and treatment of prostate cancer. *Front. Oncol.* 13:1123101.

18. Xia K G, et al. LncRNA NEAT1-associated aerobic glycolysis blunts tumor immunosurveillance by T cells in prostate cancer. **Neoplasma**. 69:594–602. 2022.

19. Sun X, et al. Long non−coding RNA CASC11 interacts with YBX1 to promote prostate cancer progression by suppressing the p53 pathway. **Int J Oncol**. 61. 2022.

20. Gencel-Augusto, J., Wu, W., & Bivona, T. G. Long Non-Coding RNAs as Emerging Targets in Lung Cancer. **Cancers**, *15*(12), 3135. 2023.

21. Ginn, Lucy *et al*. LncRNAs in Non-Small-Cell Lung Cancer. **Noncoding RNA**, [*s. l.*], v. 6, n. 3, 2020. DOI https://dx.doi.org/10.3390%2Fncrna6030025. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7549371/. Acesso em: 4 jan. 2022.

22. (Brasil). Resolução n º 466, de 12 de dezembro de 2012. Brasília, 2012 [citado 2023 Mar 11].

23. RStudio Team (2020). RStudio: Integrated Development for R. **RStudio**, PBC, Boston, MA URL http://www.rstudio.com/.

24. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome biology**, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8.

25. Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. **Omics: a journal of integrative biology**, 16(5), 284–287. https://doi.org/10.1089/omi.2011.0118.

26. Li, J. H., Liu, S., Zhou, H., Qu, L. H., & Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research*, 42(Database issue), D92–D97. https://doi.org/10.1093/nar/gkt1248.

27. INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro: INCA, 2023.

28. Nogueira, F.J et al. Perfil Epidemiológico do câncer de pulmão no Brasil entre os anos de 2013 e 2020. Research, Society and Development, v. 10, n. 16, e203101623566, 2021.

29. Domagala-Kulawik J, Trojnar A. Lung cancer in women in 21th century. **J Thorac Dis**. 12(8):4398-4410. Aug. 2020.

30. Valerio, T. S., Emmerick, I. C. M., & Sobreira-Da-Silva, M. J. Factors associated with late-stage diagnosis and overall survival for lung cancer: An analysis of patients treated in a Brazilian hospital and a US-hospital from 2009 to 2019. *Cancer epidemiology*, *86*, 102443. 2023.

31. El Alam, R., Hammer, M. M., & Byrne, S. C. Factors Associated With Delay in Lung Cancer Diagnosis and Surgery in a Lung Cancer Screening Program. **Journal of thoracic imaging**, *39*(5), 293–297. 2024.

32. Garg, A., et al. Evaluation of delays during diagnosis and management of lung cancer in India: A prospective observational study. **European journal of cancer care**, *31*(5). 2022.

33. Garinet S, et al. Updated Prognostic Factors in Localized NSCLC. **Cancers (Basel).** 9;14(6):1400. Mar. 2022

34. Mithoowani H, Febbraro M. Non-Small-Cell Lung Cancer in 2022: A Review for General Practitioners in Oncology**.** *Curr Oncol.* 9;29(3):1828-1839. Mar. 2022.

35. Tas, F., & Erturk, K. Online public interest in smoking and lung cancer: A comparative study in Google Trends. *Journal of cancer research and therapeutics*, 19(Suppl 2), S614–S617. 2023.

36. Tindle Ha, et al. Lifetime Smoking History and Risk of Lung Cancer: Results From the Framingham Heart Study**.** *J Natl Cancer Inst.* 1;110(11):1201-1207. Nov. 2018.

37. OMS, Organização Mundial da Saúde. International Agency for Research on Cancer. The Global Cancer Observatory Report. Lyon: International Agency for Research on Cancer; 2024.

38. Malta, Deborah Carvalho et al. Mudanças no uso do tabaco entre adolescentes brasileiros e fatores associados: Pesquisa Nacional de Saúde do Escolar. *Ciência & Saúde Coletiva* [online]. v.

29, n. 9. 2024.

39. Eskiler, G. G., et al. IL-6 mediated JAK/STAT3 signaling pathway in cancer patients with cachexia. **Bratislavske lekarske listy,** *66*(11), 819–826. 2019.

40. Huang B, Lang X, Li X. The role of IL-6/JAK2/STAT3 signaling pathway in cancers. *Front Oncol.* 16;12:1023177. Dec. 2022.

41. Ying, L., Du, L., Zou, R., Shi, L., Zhang, N., Jin, J., Xu, C., Zhang, F., Zhu, C., Wu, J., Chen, K., Huang, M., Wu, Y., Zhang, Y., Zheng, W., Pan, X., Chen, B., Lin, A., Tam, J. K. C., van Dam, R. M., … Su, D. (2020). Development of a serum miRNA panel for detection of early stage non-small cell lung cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(40), 25036–25042.

42. Zhou H, Zhu Y, Qi H, Liang L, Wu H, Yuan J, Hu Q. Evaluation of the prognostic values of solute carrier (SLC) family 39 genes for patients with lung adenocarcinoma. Aging (Albany NY). 2021 Feb 1;13(4):5312-5331.

43. Wang, S., Wang, Y., Li, S., Nian, S., Xu, W., & Liang, F. (2023). Long non-coding RNA MIR22HG inhibits the proliferation and migration, and promotes apoptosis by targeting microRNA-9-3p/ SOCS1 axis in small cell lung cancer cells. *Molecular biology reports*, *50*(9), 7445–7456.

44. Su W, Guo C, Wang L, Wang Z, Yang X, Niu F, Tzou D, Yang X, Huang X, Wu J, Chen X, Zou L, Yang Z, Chen G. LncRNA *MIR22HG* abrogation inhibits proliferation and induces apoptosis in esophageal adenocarcinoma cells via activation of the STAT3/c-Myc/FAK signaling. *Aging* (Albany NY). 2019 Jul 10;11(13):4587-4596.

45. Zhang, L., Li, C., & Su, X. (2020). Emerging impact of the long noncoding RNA MIR22HG on proliferation and apoptosis in multiple human cancers. *Journal of experimental & clinical cancer research : CR*, *39*(1), 271

46. Zhen, Q., Zhang, Y., Gao, L., Wang, R., Chu, W., Zhao, X., Li, Z., Li, H., Zhang, B., Lv, B., & Liu, J. (2021). EPAS1 promotes peritoneal carcinomatosis of non-small-cell lung cancer by enhancing mesothelial-mesenchymal transition. *Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft ... [et al]*, *197*(2), 141–149.

47. Yafang LI, et al. Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population, *Carcinogenesis***,** Volume 39, Issue 3, Pages 336–346. March 2018.

Figure 1.TIFF

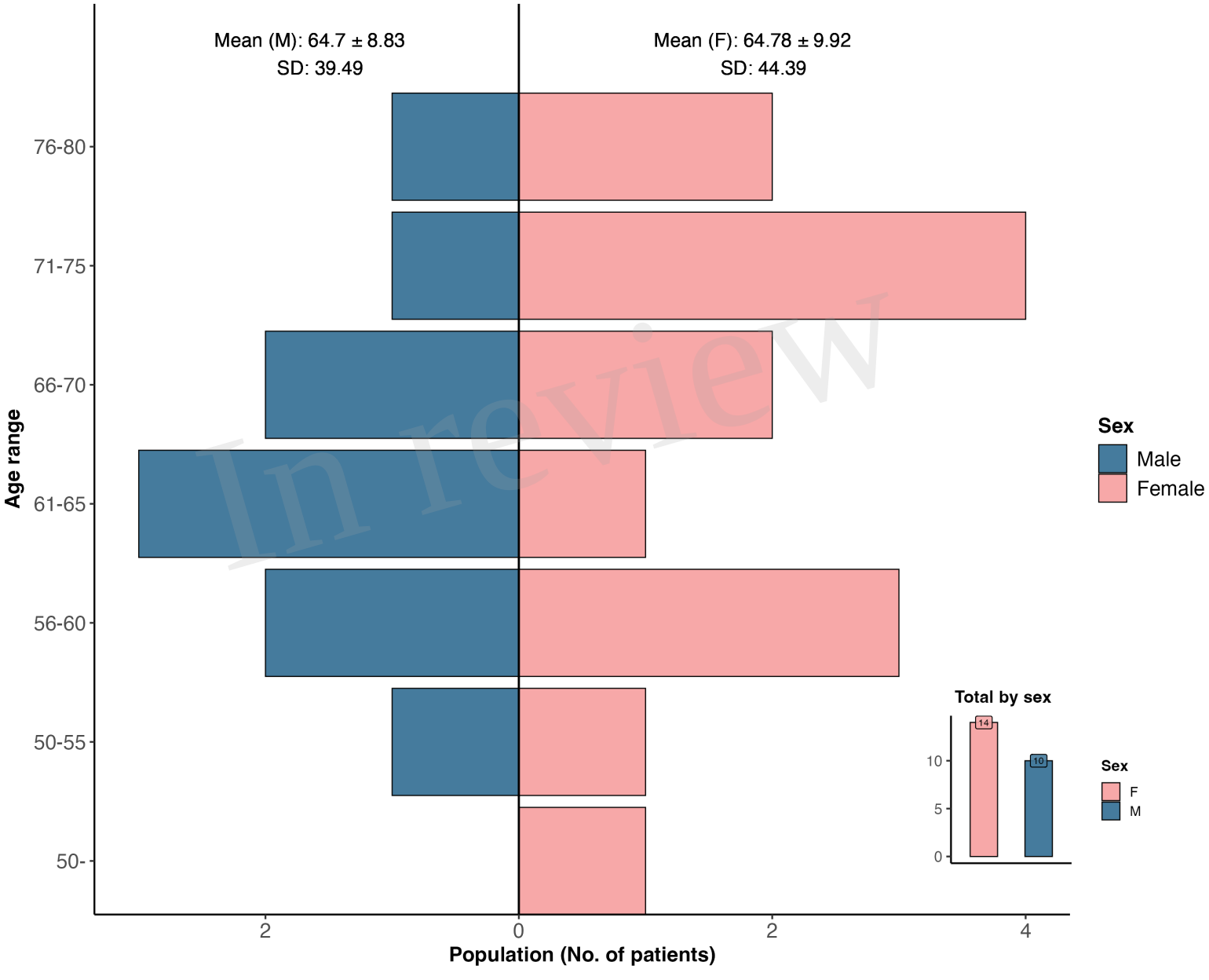


Figure 2.TIFF

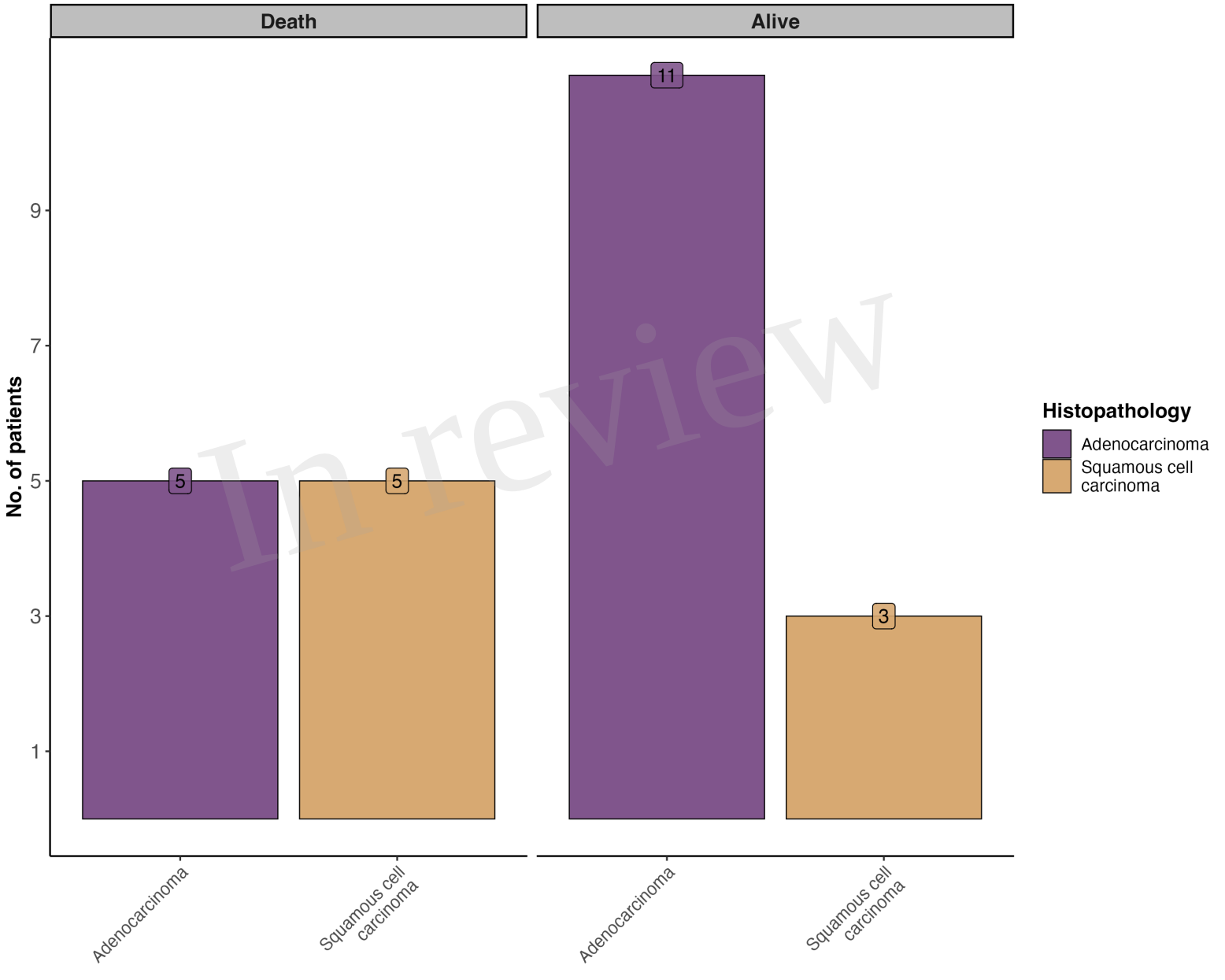


Figure 3.TIFF

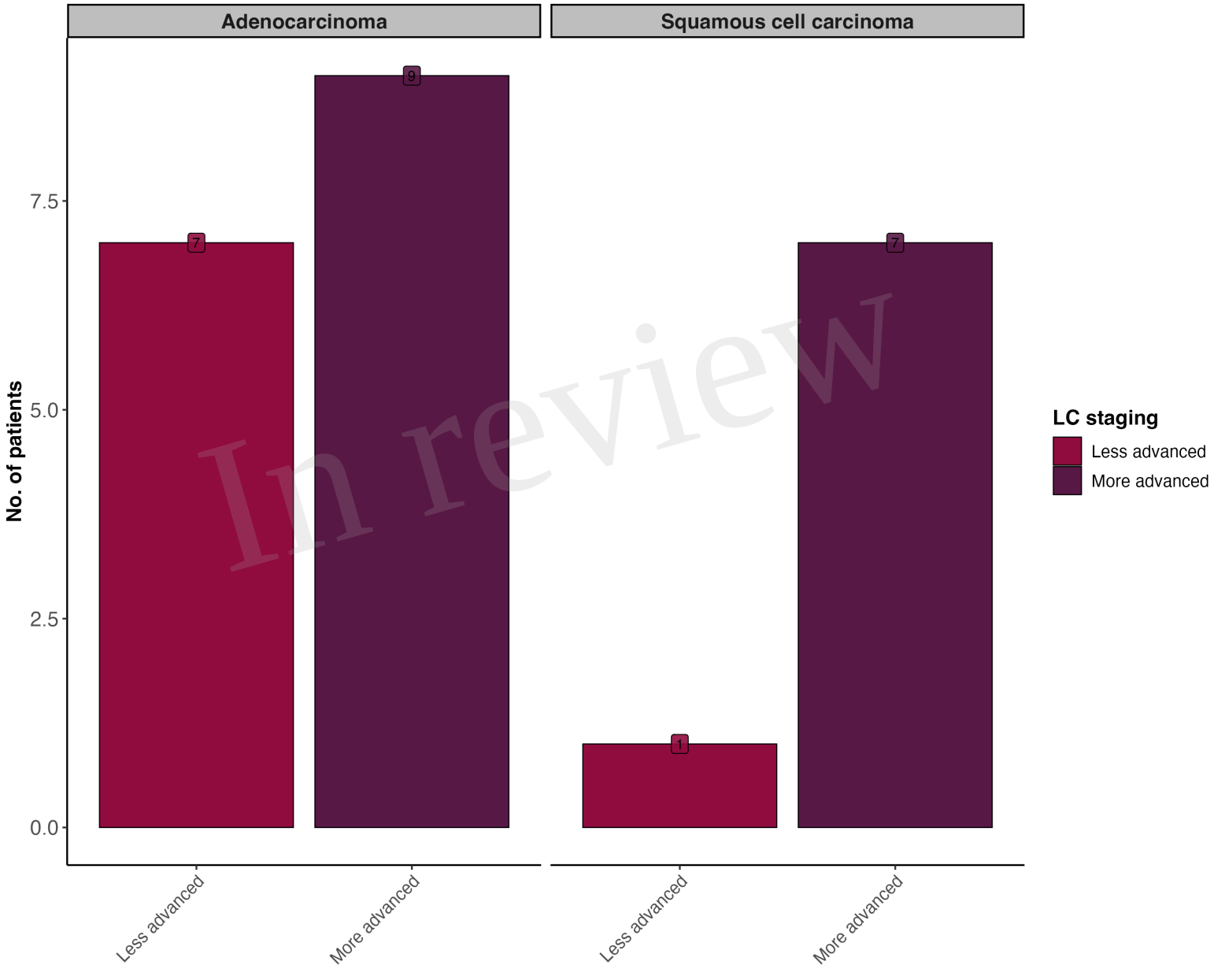


Figure 4.TIFF

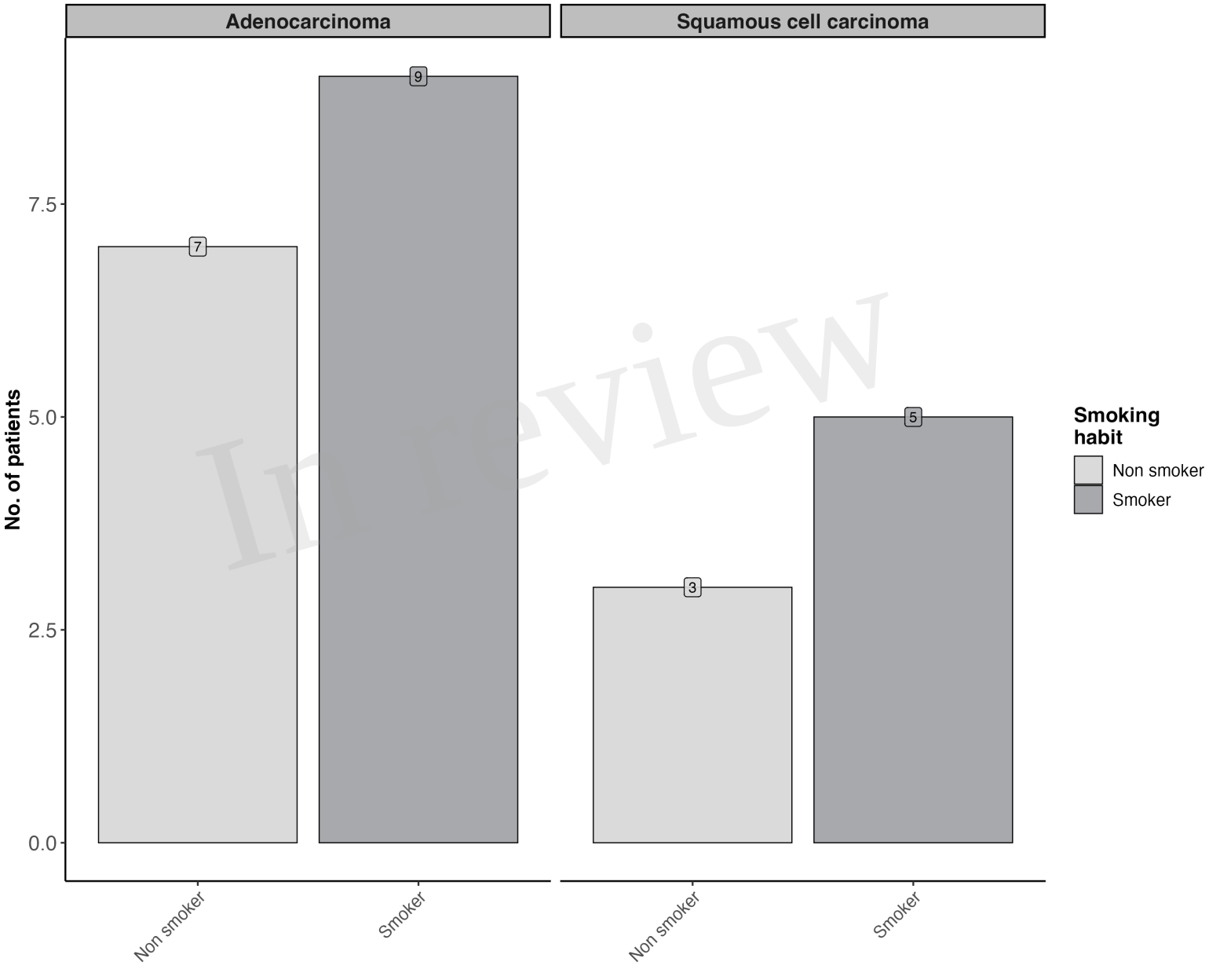


Figure 5.TIFF

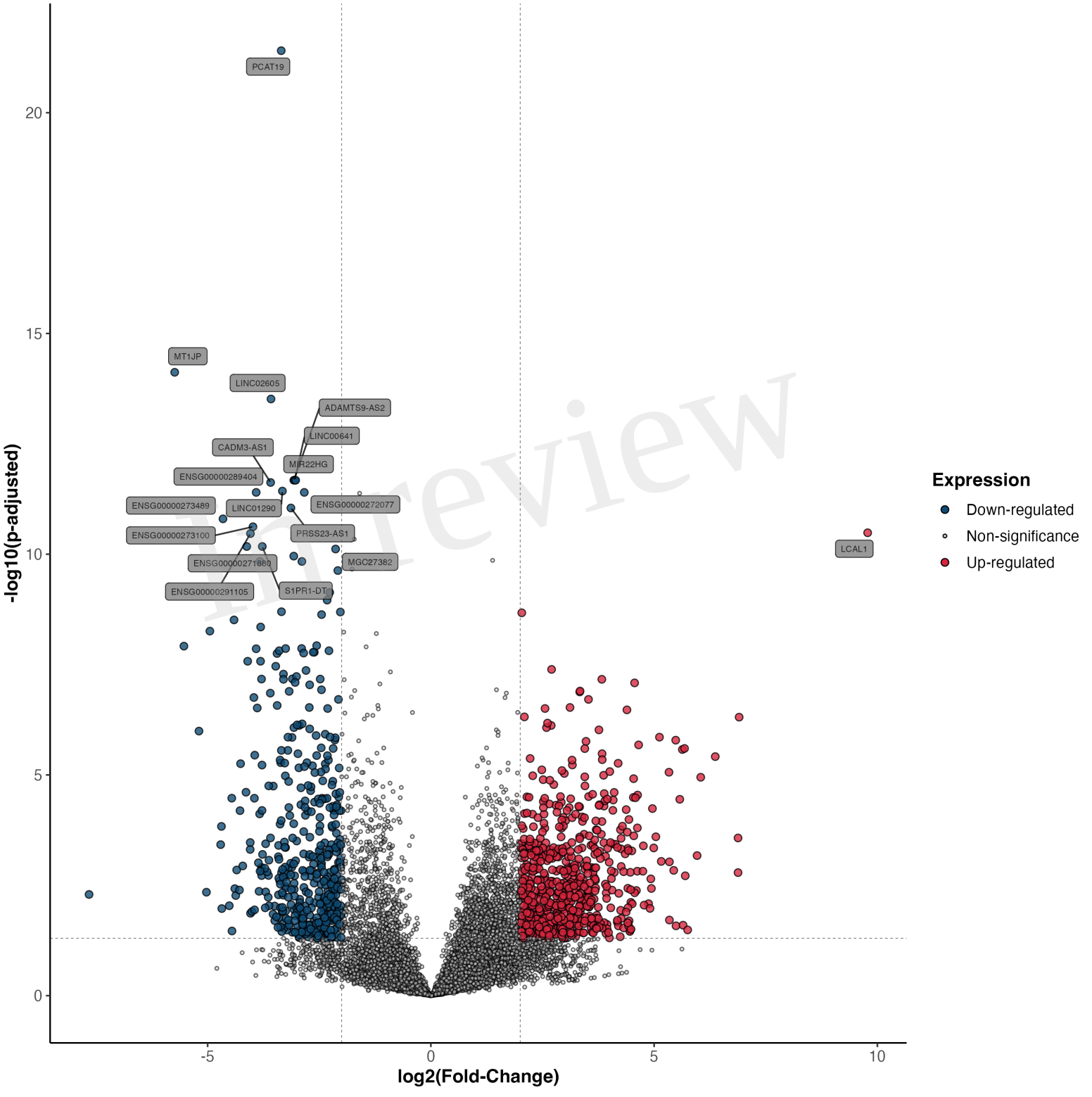


Figure 6.TIFF

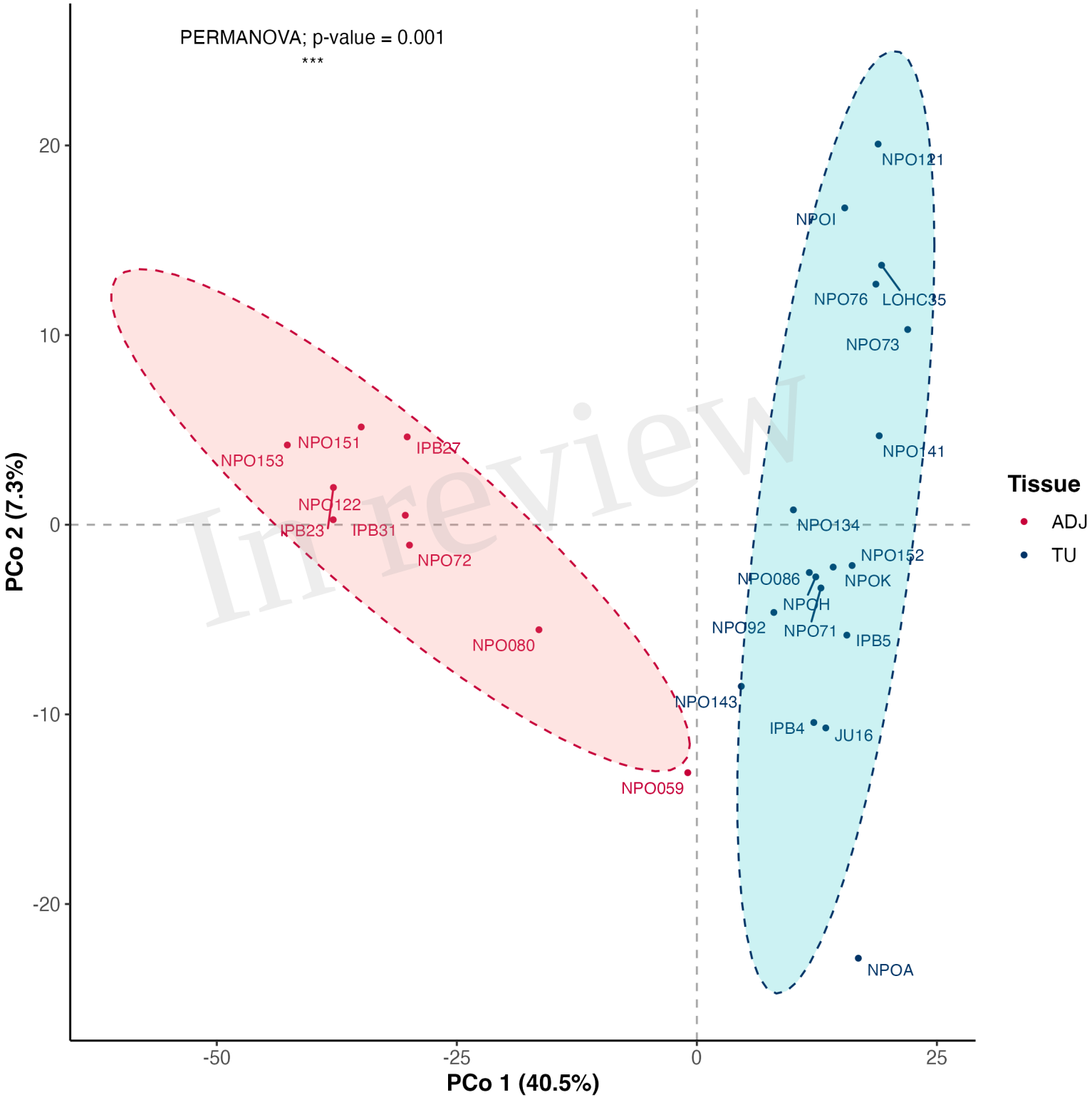


Figure 7.TIFF

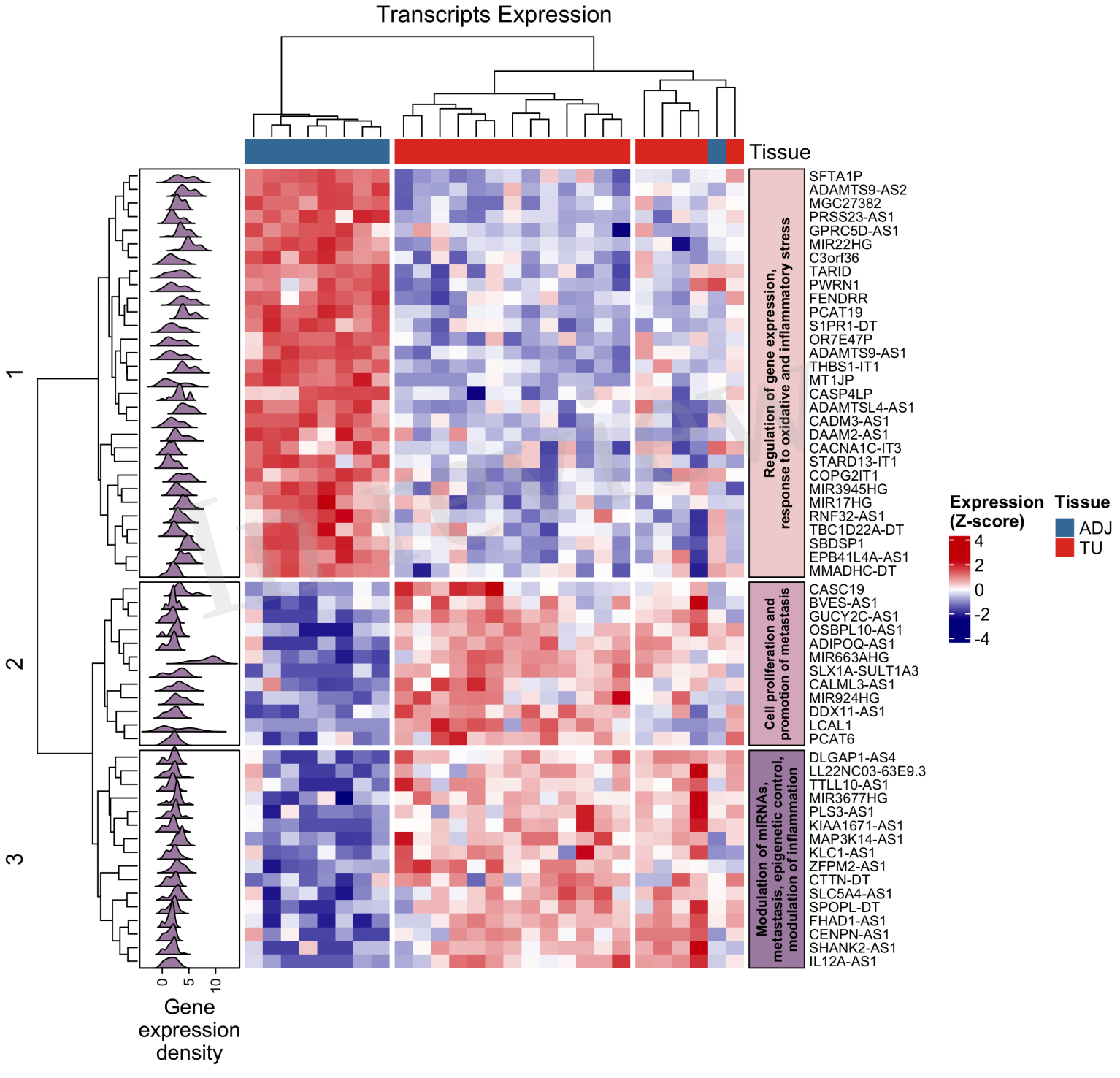


Figure 8.TIFF

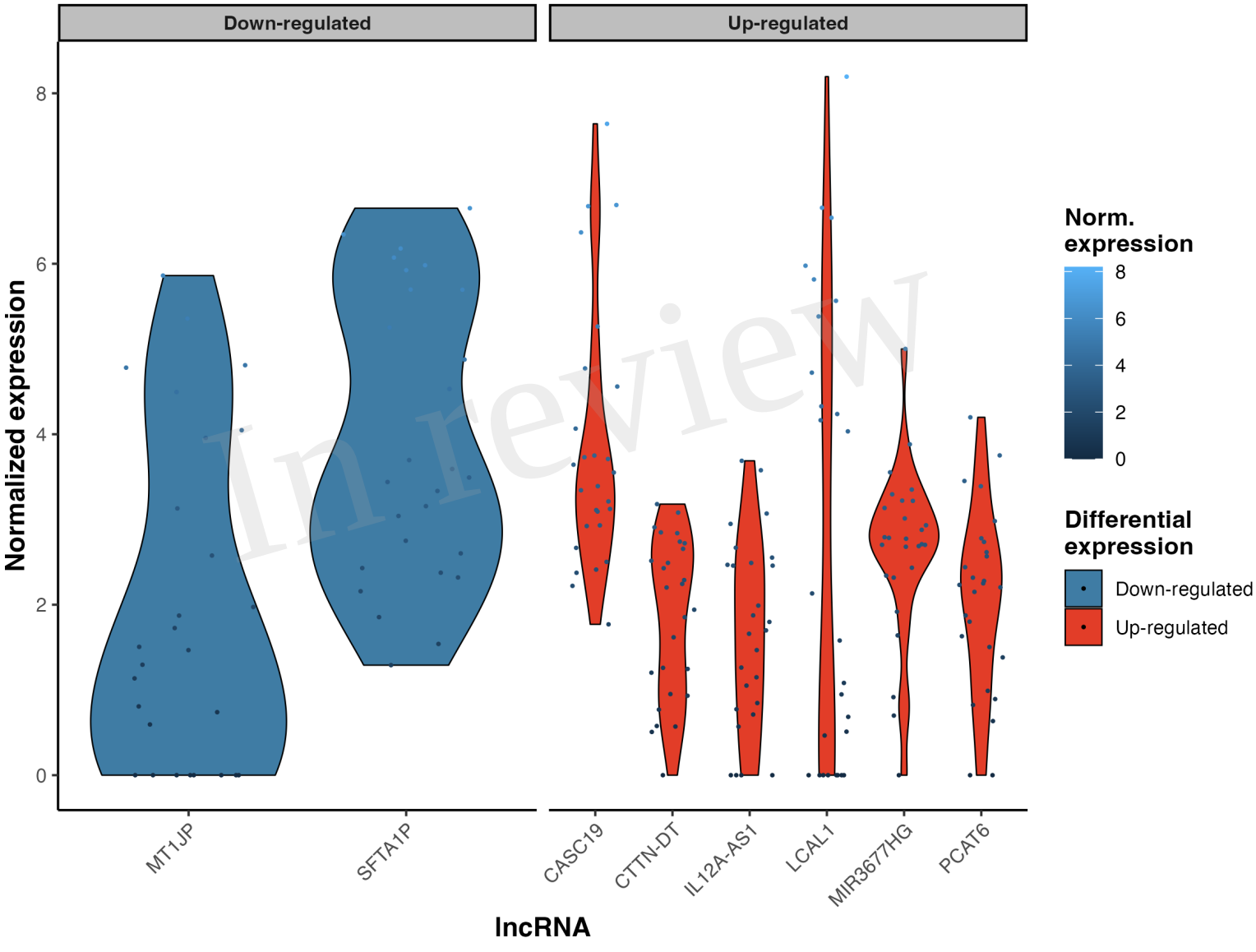


Figure 9.TIFF

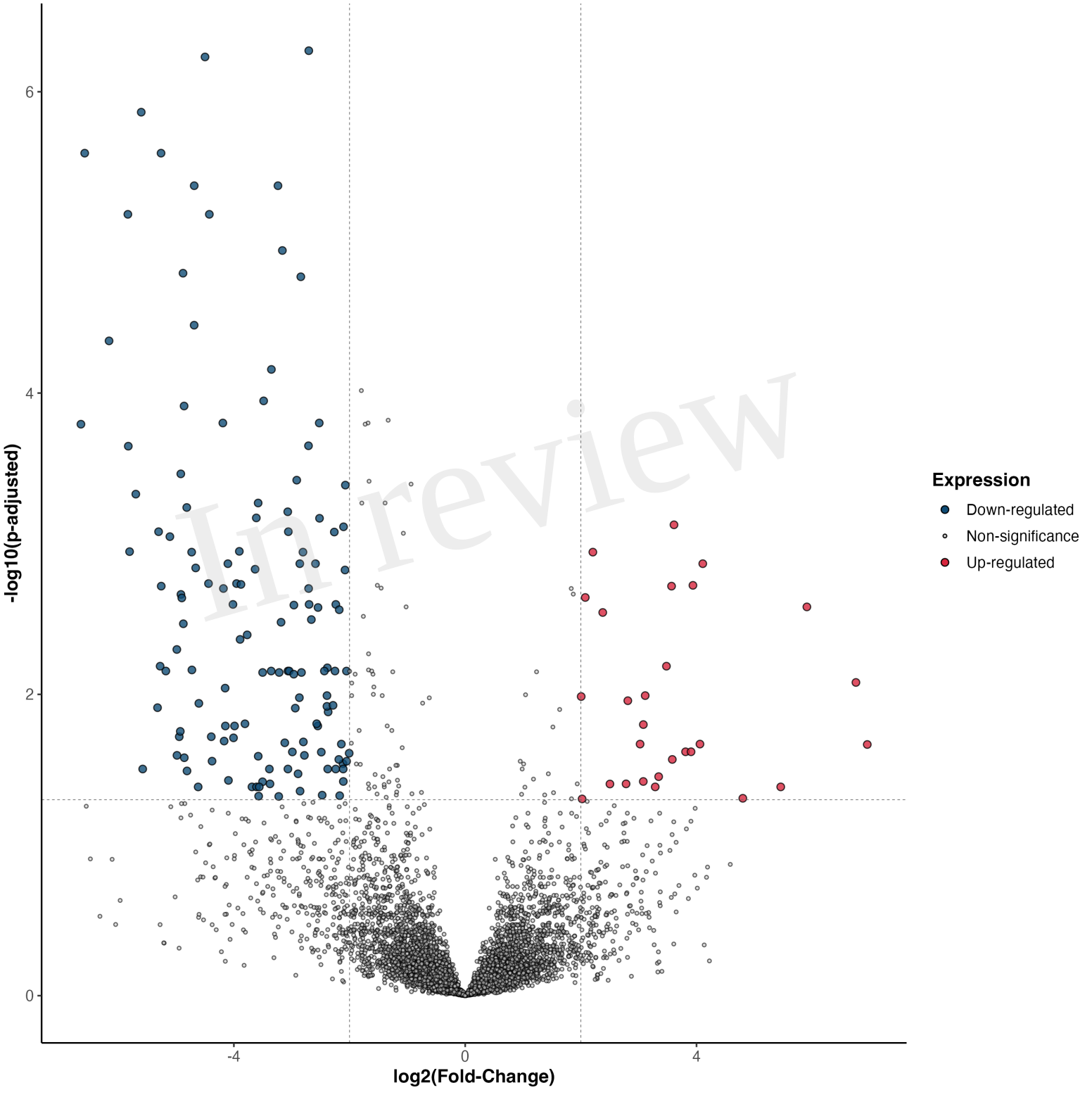


Figure 10.TIFF

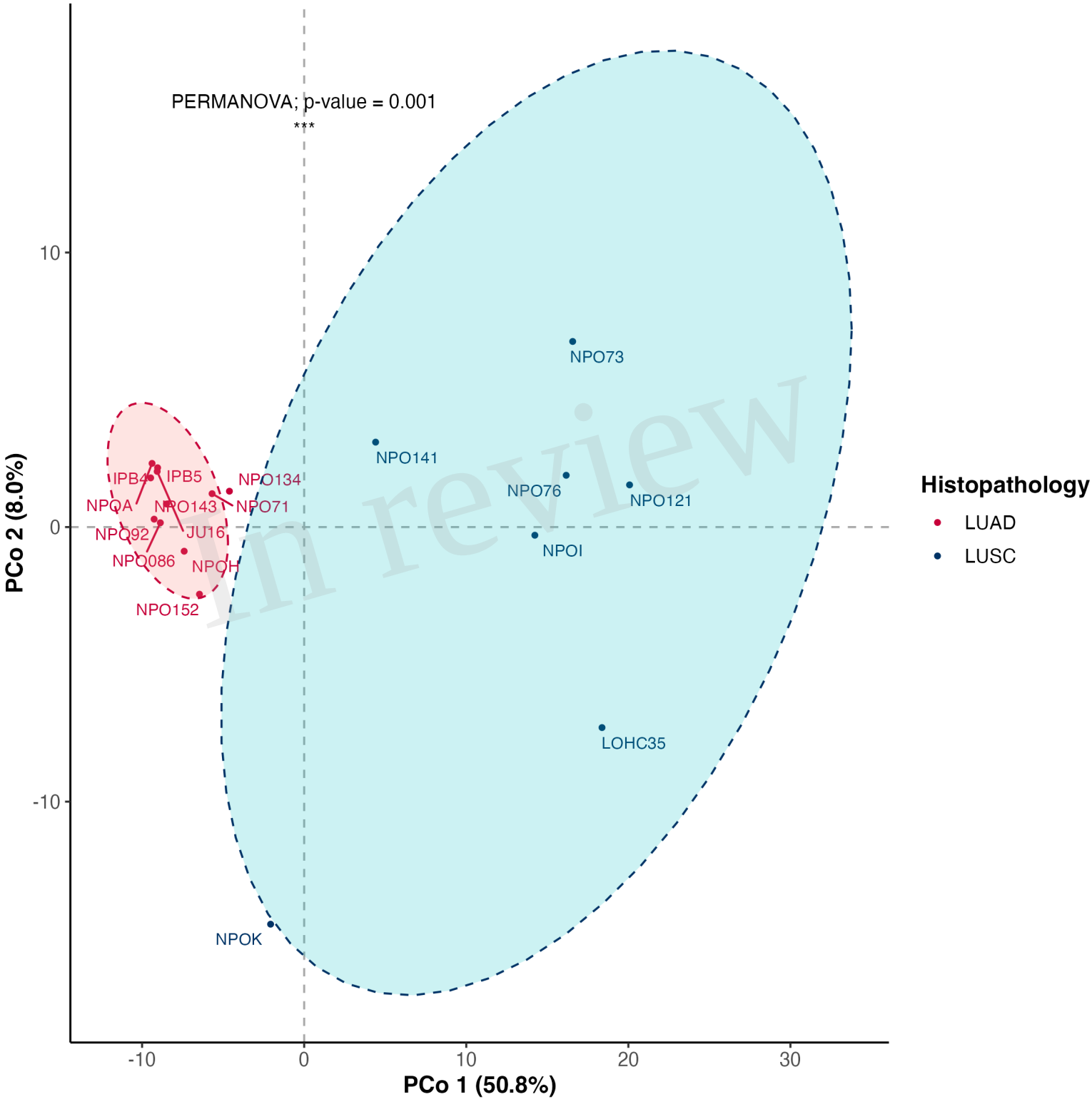


Figure 11.TIFF

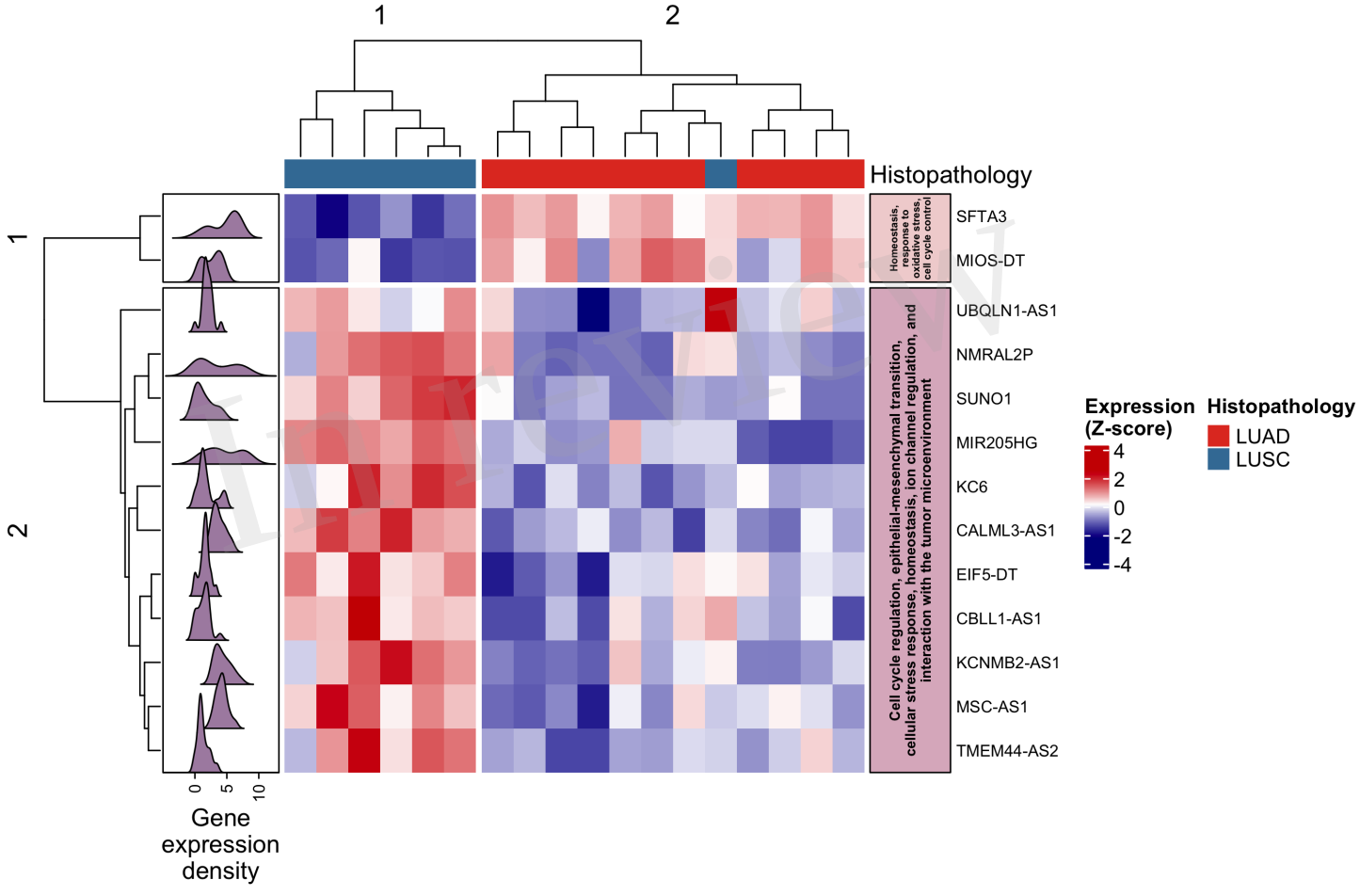


Figure 12.TIFF

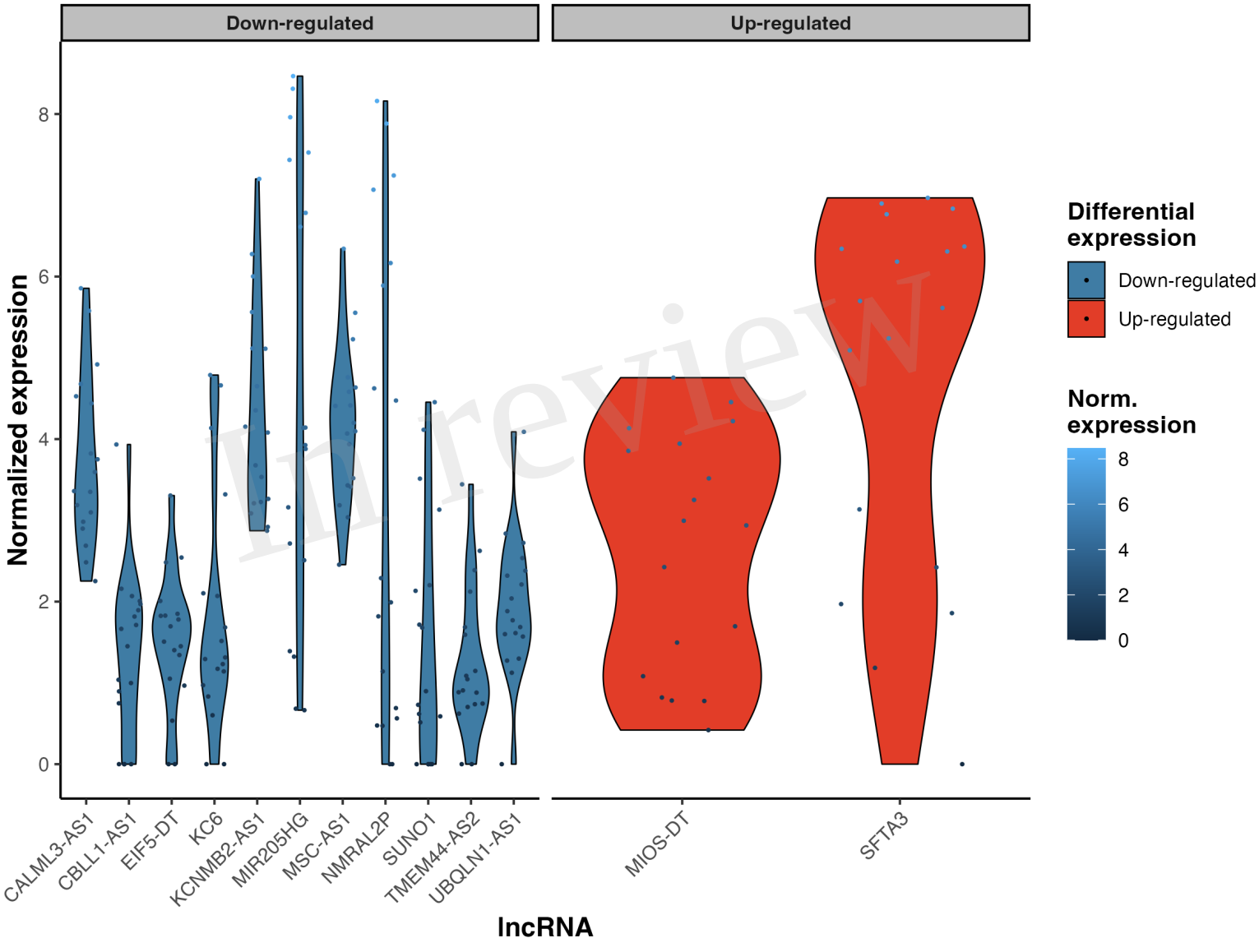


Figure 13.TIFF

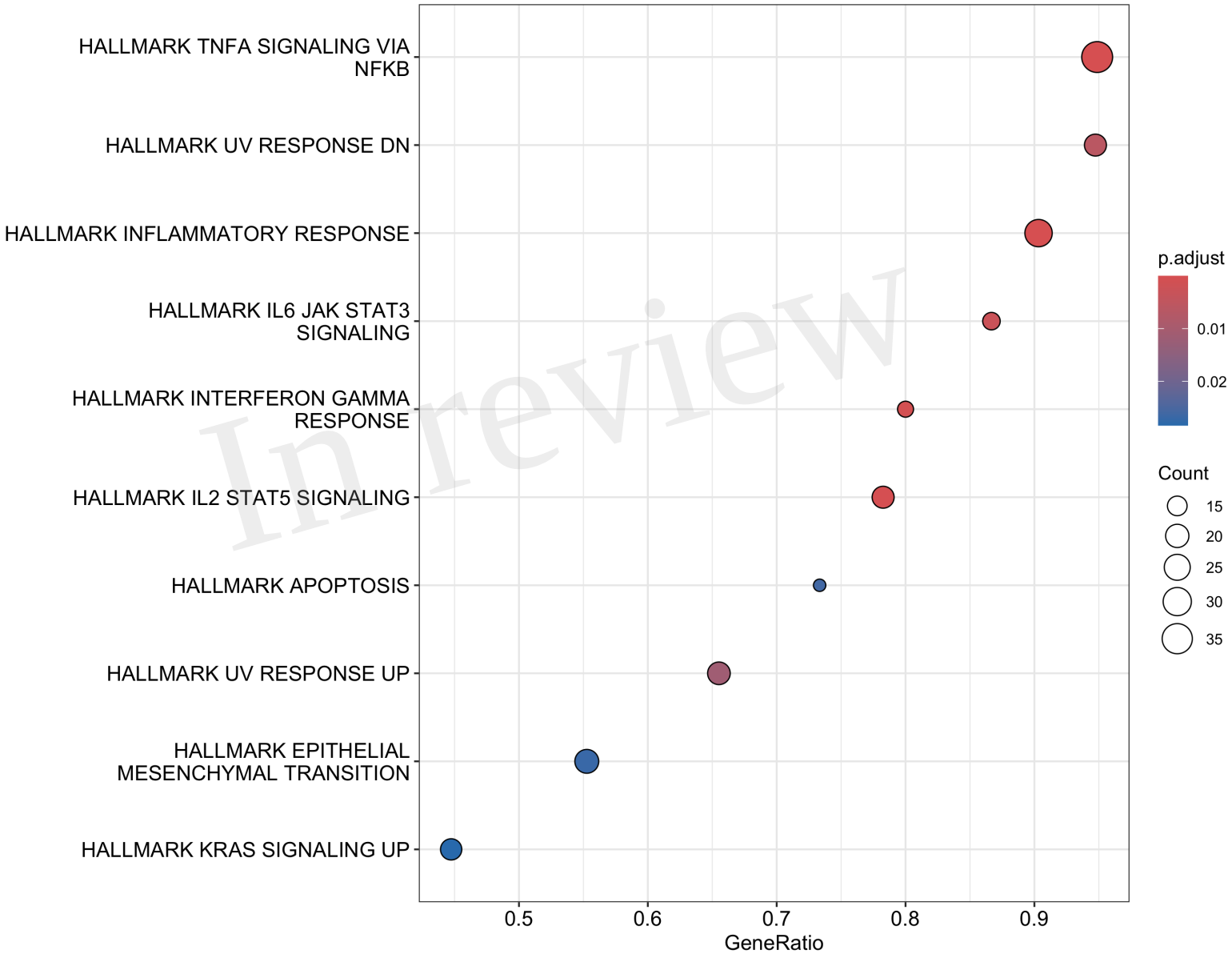


Figure 14.TIFF

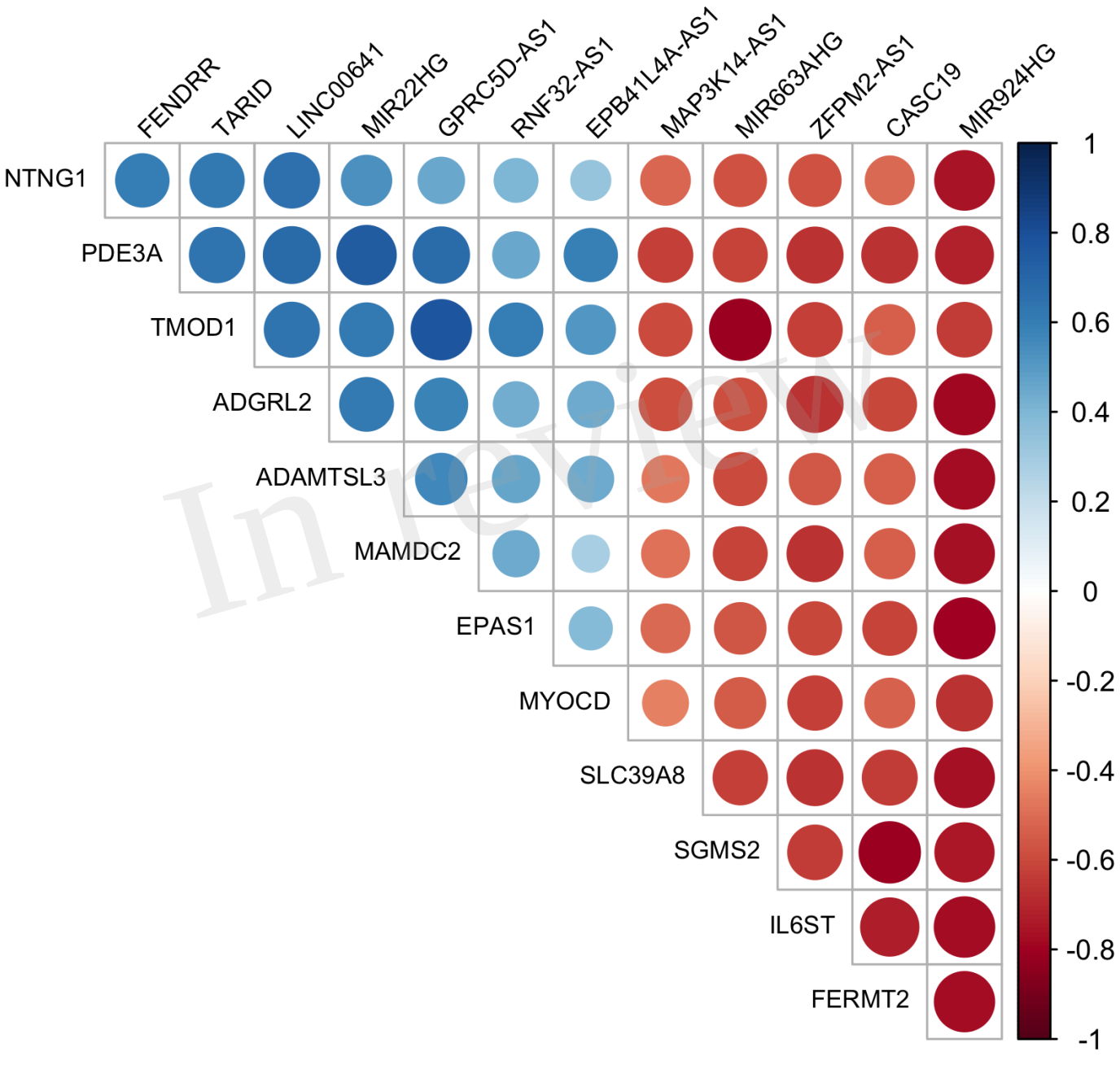
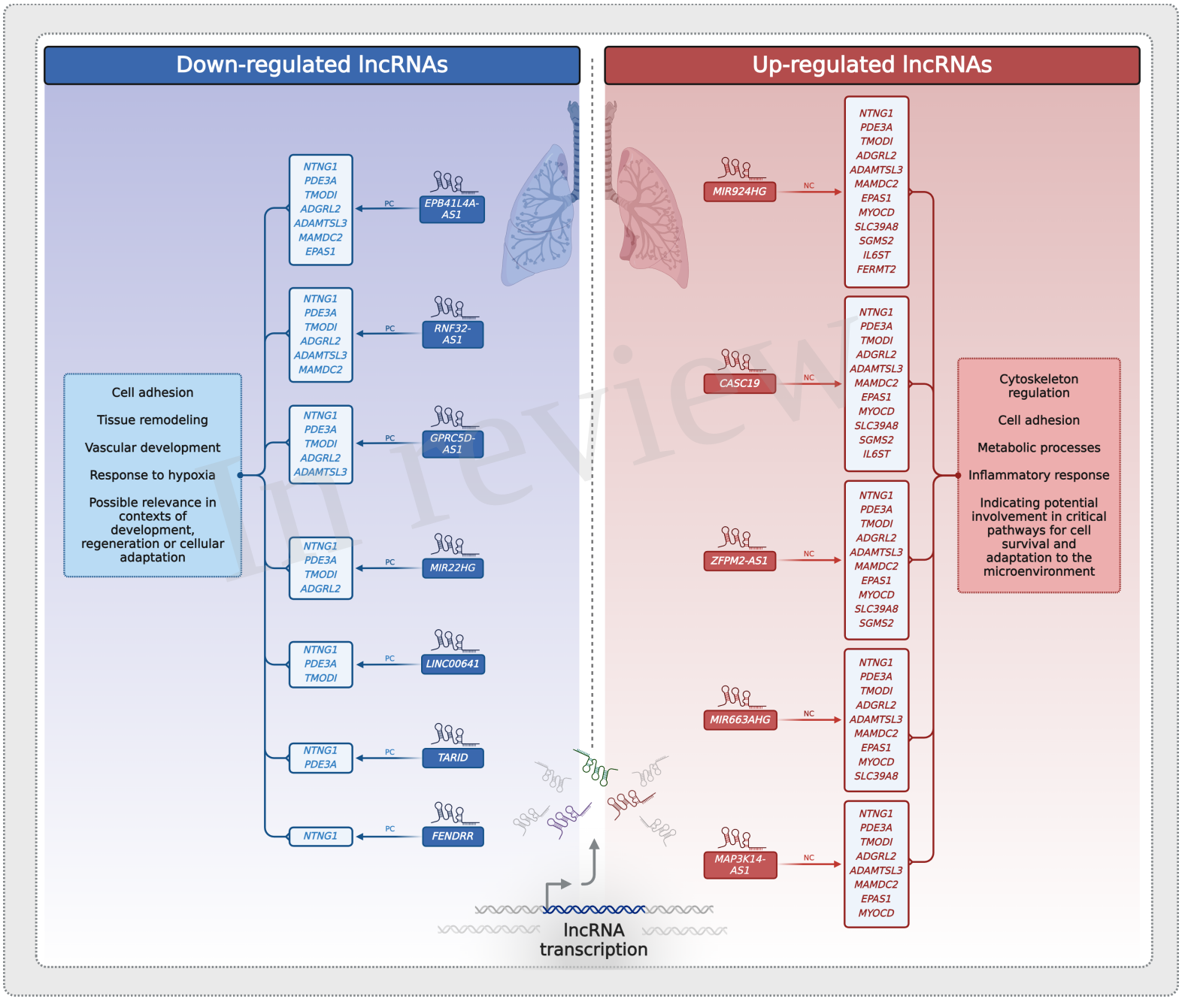


Figure 15.TIFF



# 5 DISCUSSÃO

Os dados corroboram com o fato de que CP promove a ativação de vias inflamatórias e imunológicas, além de desregular a apoptose, o que favorece a progressão tumoral e a evasão imunológica. Esses achados podem ajudar a identificar possíveis alvos terapêuticos, como inibidores de *NFκB* ou moduladores de sinalização *JAK/STAT* haja vista que a via *NFκB* regula a expressão de genes antiapoptóticos, como *BCL-2*, que favorecem a sobrevivência das células tumorais. Além disso, promove a liberação de citocinas pró-inflamatórias, como IL-6 e TNF-α, que criam um microambiente favorável ao crescimento tumoral, e regula moléculas de adesão e fatores de remodelação da matriz extracelular, facilitando a invasão e a metástase (ABDEL-FATTAH et al., 2023; HU et al., 2021). Por outro lado, a sinalização *JAK/STAT* é essencial para transduzir sinais de citocinas e fatores de crescimento, regulando processos como proliferação, diferenciação e imunidade. No câncer de pulmão, sua hiperativação está associada ao crescimento tumoral e à evasão imunológica (CHOUDHARY et al., 2023).

Os resultados confirmam achados recentes que apontam diferenças moleculares significativas entre LUAD e LUSC, reforçando a necessidade de abordagens mais específicas para cada subtipo. Por exemplo, enquanto genes relacionados à queratinização são proeminentes em LUSC, LUAD parece depender mais de vias ligadas à proliferação e diferenciação celular, como as reguladas por *DLK1*. Além disso, o microambiente tumoral, influenciado por genes como CLCA1, tem emergido como um fator crítico para a progressão e resposta terapêutica nos dois subtipos de NSCLC.

Um estudo realizado por ZENGIN & ONAL-SUZEK (2021), sobre a compreensão do perfil genômico e transcriptômico e a diferença entre grupos de risco de pacientes com LUAD e LUSC evidenciou que ambos os grupos de CP têm vias imunológicas reguladas negativamente em maior quantidade e vias metabólicas reguladas positivamente em maior número, entretanto é necessário avaliar também o grau de risco entre pacientes bem como o perfil clínico associado.

O entendimento das interações entre esses lncRNAs e seus genes-alvo é fundamental para identificar potenciais biomarcadores e alvos terapêuticos no manejo do câncer de pulmão, especialmente em subtipos como LUAD e LUSC. A expressão diferencial de genes pode permitir a classificação molecular de subtipos de CP, auxiliando na escolha de tratamentos específicos e mais eficazes haja vista que genes diferencialmente expressos podem estar envolvidos em mecanismos de resistência à quimioterapia ou imunoterapia, ajudando no desenvolvimento de novas abordagens terapêuticas (CHEN & DHAHBI, 2021).

No que diz respeito ao microambiente tumoral, os achados fornecem uma visão abrangente sobre a expressão de lncRNAs destacando seu impacto no microambiente tumoral. Os lncRNAs regulados negativamente e positivamente, indicando interações específicas com genes codificadores de proteínas. Os lncRNAs regulados negativamente estão associados a funções como resposta à hipóxia e regulação metabólica, processos fundamentais no microambiente tumoral, especialmente em condições de hipóxia, característica comum em tumores sólidos. Por outro lado, os lncRNAs regulados positivamente estão relacionados a mecanismos de remodelação tecidual, migração celular e inflamação, sugerindo um papel relevante na progressão tumoral, na evasão imune e na regeneração tecidual (STELLA et al., 2022; YANG et al.,2022).

O microambiente tumoral, caracterizado por inflamação crônica, alterações metabólicas e imunomodulação, é amplamente influenciado por esses lncRNAs, o que repercute na microbiota residente. Por exemplo, lncRNAs relacionados a genes como *IL6ST* e *FERMT2*, envolvidos em processos inflamatórios e remodelação tecidual, promovem um ambiente inflamatório persistente que pode favorecer a disbiose, reduzindo a diversidade microbiana e estimulando o crescimento de microrganismos oportunistas (RAHAL et a., 2024).

Além disso, lncRNAs que regulam genes associados à resposta à hipóxia, como EPAS1, criam um ambiente com baixos níveis de oxigênio, característico de tumores sólidos. Essa condição favorece a proliferação de bactérias anaeróbicas, alterando significativamente a composição da microbiota pulmonar. Os lncRNAs também influenciam o metabolismo energético local, por meio de genes como *PDE3A* e *TMOD1*, modificando o perfil metabólico do tecido pulmonar e impactando os substratos disponíveis para os microrganismos residentes. Essas alterações podem contribuir para a sobrevivência de cepas bacterianas específicas e a exclusão de outras, desestabilizando o equilíbrio microbiano (CAO et al., 2024; FORDER et al., 2023).

Outro aspecto importante é a remodelação da matriz extracelular, regulada por genes como ADGRL2 e TMOD1, que modifica o microambiente estrutural do tecido pulmonar. Essa remodelação pode impactar a adesão microbiana, influenciando a ecologia local e facilitando a colonização por microrganismos patogênicos. Adicionalmente, alterações na integridade da barreira epitelial, frequentemente associadas à EMT promovida por lncRNAs, enfraquecem as junções intercelulares, permitindo a invasão de patógenos e exacerbando a inflamação local (LIU et al., 2023; XU et al., 2021).

A produção de mediadores inflamatórios, como citocinas e quimiocinas, regulada por

genes relacionados a lncRNAs, também exerce um impacto direto na microbiota pulmonar. Citocinas pró-inflamatórias, como aquelas associadas a *IL6ST*, criam um ambiente que favorece o crescimento de determinadas populações bacterianas enquanto inibe outras, contribuindo para o desequilíbrio microbiano (LI et al., 2024). Além disso, as alterações provocadas pelo microambiente tumoral podem ser potencializadas por terapias direcionadas ao câncer, como quimioterápicos ou imunomoduladores, que frequentemente levam à disbiose secundária.

# 6 CONCLUSÃO

Este projeto apresentou um conjunto de estudos que contribuem para a compreensão dos lncRNA no CP, abordando de forma integrada a expressão de transcritos codificantes aos quais interagem na busca de biomarcadores. A partir das análises realizadas, foi possível confirmar e evidenciar que os lncRNAs desempenham papéis críticos no microambiente tumoral do câncer de pulmão, influenciando processos como inflamação, angiogênese, remodelação da matriz extracelular e evasão imunológica. As interações específicas entre lncRNAs e genes codificadores de proteínas apontadas fornecem informações valiosas que podem auxiliar no desenvolvimento de novos alvos terapêuticos e biomarcadores para os diferentes subtipos de câncer de pulmão.

Além disso, os dados apresentados elucidaram a diferença molecular entre os subtipos de CP com base na expressão gênica, destacando a importância de investigar sobre os diferentes mecanismos biológicos envolvidos no desenvolvimento e na invasividade do câncer. Esses mecanismos são refletidos nas características clínicas dos tumores de NSCLC, que resultam da desregulação de vias de sinalização evolutivas e de seus efeitos a jusante.

Esses achados trazem implicações significativas ao identificar padrões de expressão gênica, o que ajuda a direcionar, futuramente, estudos de novas drogas e a entender a biologia complexa do câncer, promovendo avanços na oncologia de precisão.

**REFERÊNCIAS**


ABDEL-FATTAH, M. M., et al. Role of NF-κB/ICAM-1, JAK/STAT-3, and apoptosis signaling in the anticancer effect of tangeretin against urethane-induced lung cancer in BALB/c mice. **Life sciences**, *325*, 121749. 2023.

AGOSTINI, F., et al. Intergenic RNA mainly derives from nascent transcripts of known genes. **Genome biology**, *22*(1), 136. 2021.

AHN, Young-Ho; KO, Yoon Ho. Diagnostic and Therapeutic Implications of microRNAs in Non-Small Cell Lung Cancer. **Int. J. Mol. Sci.**, [*s. l.*], 2020. DOI https://dx.doi.org/10.3390%2Fijms21228782. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7699705/. Acesso em: 5 jan. 2022.

ALLEGRA A, et al., Interactions between the MicroRNAs and Microbiota in Cancer Development: Roles and Therapeutic Opportunities. **Cancers (Basel)**. 2020 Mar 27;12(4):805. doi: 10.3390/cancers12040805. PMID: 32230762; PMCID: PMC7225936.

AMARAL, Bruna Aguiar do *et al*. MicroRNAs – biogênese, funções e seu papel potencial na carcinogênese oral. **Odontologia Clínico-Científica**, Recife, v. 10, n. 2, 2010. Disponível em: http://revodonto.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-38882010000200003. Acesso em: 5 jan. 2022.

ANASTASIADOU, Eleni; JACOB, Leni S.; SLACK, Frank J. Non-coding RNA networks in cancer. **Nature Reviews Cancer**, [*s. l.*], 24 nov. 2017.

ARAB, Khelifa *et al*. Long Noncoding RNA TARID Directs Demethylation and Activation of the Tumor Suppressor TCF21 via GADD45A. **Molecular Cell**, [*s. l.*], 21 ago. 2014. DOI https://doi.org/10.1016/j.molcel.2014.06.031. Disponível em: https://www.sciencedirect.com/science/article/pii/S1097276514005656. Acesso em: 5 jan. 2022.

BECK, Cristian et al. Daily Expression Pattern of Protein-Encoding Genes and Small Noncoding RNAs in *Synechocystis* sp. Strain PCC 6803. **ASM Journals**., [*s. l.*], 2014.

BENETTI, Roberta *et al*. A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. **Nat. Struct. Mol. Biol.**, [*s. l.*], 2008.

BENUSIGLIO PR, et al. Lung cancer is also a hereditary disease. **Eur Respir Rev**. Oct 20;30(162):210045. 2021.

BHAT, S. A., et al. Long non-coding RNAs: Mechanism of action and functional utility. **Non-coding RNA research**, *1*(1), 43–50. 2016

BROCKDORFF, Neil *et al*. Non-coding RNAs as regulators of gene expression and epigenetics. **RNA**, [*s. l.*], v. 19, n. 4, p. 429–442, 19 maio 2014. DOI https://dx.doi.org/10.1261%2Frna.037598.112. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3677253/. Acesso em: 7 jan. 2022.

CAMPOS, M. R. et al. Smoking, mortality, access to diagnosis, and treatment of lung cancer in Brazil. **Revista De Saúde Pública**. 58(18). 2024.

CAO, M. X., et al. The crosstalk between lncRNA and microRNA in cancer metastasis: orchestrating the epithelial-mesenchymal plasticity. **Oncotarget**, *8*(7), 12472–12483. 2017.

CAO, Y., *et al.* Intratumoural microbiota: a new frontier in cancer development and

therapy. **Sig Transduct Target Ther.** 9, 15. 2024.

CARNEY SM, et al. Methods in Lung Microbiome Research. **Am J Respir Cell Mol Biol**. 62(3):283-299. Mar. 2020.

CASTELLANOS E, FELD E, HORN L. Driven by Mutations: The Predictive Value of Mutation Subtype in EGFR-Mutated Non-Small Cell Lung Cancer . **J Thorac Oncol.** Apr;12(4):612-623. 2017

CENTEIO, R., et al. CLCA1 Regulates Airway Mucus Production and Ion Secretion Through TMEM16A**. International journal of molecular sciences**, *22*(10), 5133. 2021.

CHANG YS, et al. Metatranscriptomic Analysis of Human Lung Metagenomes from Patients with Lung Cancer. **Genes (basel).** 21;12(9):1458. Sep. 2021.

CHEN, J.W. & DHAHBI, J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. **Sci Rep.** 11, 13323. 2021.

CHEN, Q., et al. Screening of potential microbial markers for lung cancer using metagenomic sequencing. **Cancer Medicine**, *12*(6), 7127–7139. 2023.

CHEN, Y., et al. Long non-coding RNAs: From disease code to drug role. **Acta pharmaceutica Sinica**. *B*, *11*(2), 340–354. 2021

CHEN, Z. et al. feito Prottor do RNA não codificante longo OXCT1-AS1 na Apoptose de Células Miocárdicas Humanas Induzida pela Doxorrubicina pelo Padrão Competitivo de RNA endógeno. **Arquivos Brasileiros de Cardiologia**, v. 121, n. 6, p. e20230675, 2024.

CHOUDHARY, N., et al. Targeting Cell Signaling Pathways in Lung Cancer by Bioactive Phytocompounds. ***Cancers,*** *15*(15), 3980. 2023.

COOPER, W. A., et al. Molecular biology of lung cancer. **Journal of thoracic disease**, 5 Suppl 5(Suppl 5), S479–S490. 2013.

CORRALES L, Lung cancer in never smokers: The role of different risk factors other than tobacco smoking. **Crit Rev Oncol Hematol**. Apr;148:102895. 2020.

COSTA, FF. Non-coding RNAs: Meet thy masters. **Bioessays**. 2010 Jul;32(7):599-608. 2010.

COSTA, G. J. et al., Tumor-node-metastasis staging and treatment patterns of 73.167 patients with lung cancer in Brazil. **Jornal Brasileiro de Pneumologia,** v. 46, n. 1, p. e20180251, 2020.

DAHARIYA, S., et al. Long non-coding RNA: Classification, biogenesis and functions in blood cells. **Molecular immunology**, *112*, 82–92. 2019.

DESHPANDE, D, et al., RNA-seq data science: From raw data to effective interpretation. **Front Genet**. 2023 Mar 13;14:997383. doi: 10.3389/fgene.2023.997383. PMID: 36999049; PMCID: PMC10043755.

DOMAGALA-KULAWIK J, TROJNAR A. Lung cancer in women in 21th century. **J Thorac Dis**. 12(8):4398-4410. Aug. 2020.

DU, D., et al. Expression of long non-coding RNA SFTA1P and its function in non- small cell lung cancer. **Pathology, research and practice**, *216*(9), 153049. 2020.

DUARTE RL, PASCHOAL ME. Molecular markers in lung cancer: prognostic role and relationship to smoking. **J Bras Pneumol**. Jan-Feb;32(1):56-65. English, Portuguese. 2006.

DUFFY, Angela M.; BOUCHIER-HAYES, David J.; HARMEY, Judith H. Vascular

Endothelial Growth Factor (VEGF) and Its Role in Non-Endothelial Cells: Autocrine Signalling by VEGF. **Madame Curie Bioscience Database**, [*s. l.*], 2013. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK6482/. Acesso em: 4 jan. 2022.

EL ALAM, R., HAMMER, M. M., & BYRNE, S. C. Factors Associated With Delay in Lung Cancer Diagnosis and Surgery in a Lung Cancer Screening Program. **Journal of thoracic imaging**, *39*(5), 293–297. 2024.

ESKILER, G. G., et al. IL-6 mediated JAK/STAT3 signaling pathway in cancer patients with cachexia. **Bratislavske lekarske listy,** *66*(11), 819–826. 2019.

FABBRI, Muller *et al*. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. **Proc. Natl. Acad. Sci. USA.**, [*s. l.*], 2 out. 2007. DOI 10.1073/pnas.0707628104. Disponível em: https://pubmed.ncbi.nlm.nih.gov/17890317/. Acesso em: 7 jan. 2022.

FENG, S. H., & YANG, S. T. The new 8th TNM staging system of lung cancer and its potential imaging interpretation pitfalls and limitations with CT image demonstrations. **Diagnostic and interventional radiology** (Ankara, Turkey), *25*(4), 270–279. 2019.

FENN D, et al. Composition and diversity analysis of the lung microbiome in patients with suspected ventilator-associated pneumonia. **Crit Care**. V. 6 n. 26 p. 33-45. Jul 2022.

FERNANDES, Juliane C. R. *et al*. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. **Noncoding RNA**, [*s. l.*], 2019. DOI https://dx.doi.org/10.3390%2Fncrna5010017. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6468922/. Acesso em: 5 jan. 2022.

FORDER A, et al. Mechanisms Contributing to the Comorbidity of COPD and Lung Cancer. **Int J Mol Sci**. 2;24(3):2859. Feb. 2023.

FROMENTIN, M., et al. The 16S rRNA lung microbiome in mechanically ventilated patients: a methodological study. **Experimental lung research,** *48*(1), 23–34. 2022.

GARG, A., et al. Evaluation of delays during diagnosis and management of lung cancer in India: A prospective observational study. **European journal of cancer care**, *31*(5). 2022.

GARINET S, et al. Updated Prognostic Factors in Localized NSCLC. **Cancers (Basel).** 9;14(6):1400. Mar. 2022

GENCEL-AUGUSTO, J., WU, W., & BIVONA, T. G. Long Non-Coding RNAs as Emerging Targets in Lung Cancer. **Cancers**, *15*(12), 3135. 2023.

GINN, Lucy *et al*. LncRNAs in Non-Small-Cell Lung Cancer. **Noncoding RNA**, [*s. l.*], v. 6, n. 3, 2020. DOI https://dx.doi.org/10.3390%2Fncrna6030025. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7549371/. Acesso em: 4 jan. 2022.

GONZALEZ, Susana; PISANO, David G; SERRANO, Manuel. Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs. **Cell Cycle**, [*s. l.*], v. 7, n. 16, 15 ago. 2008. DOI https://doi.org/10.4161/cc.7.16.6541. Disponível em: https://pubmed.ncbi.nlm.nih.gov/18719372/. Acesso em: 7 jan. 2022.

GREATHOUSE KL, et al., Interaction between the microbiome and TP53 in human lung cancer. **Genome Biol.** Aug 24;19(1):123. doi: 10.1186/s13059-018-1501-6. Erratum in: Genome Biol. 2020 Feb 20;21(1):41. 2018.

GRILLONE, Katia *et al*. Non-coding RNAs in cancer: platforms and strategies for investigating the genomic “dark matter”. **Journal of Experimental & Clinical Cancer Research**, [*s. l.*], 20 jul. 2020.

HAN, L., et al LncRNA PCAT6 is a predictor of poor prognosis of colorectal cancer. **Journal of gastrointestinal oncology,** *15*(1), 190–202. 2024.

HANAHAN D. Hallmarks of Cancer: New Dimensions. **Cancer discovery**, *12*(1), 31–46. 2022.

HANG, Dong *et al*. A novel plasma circular RNA circFARSA is a potential biomarker for non-small cell lung cancer. **Cancer Medicine**, [*s. l.*], 2018. DOI https://pubmed.ncbi.nlm.nih.gov/29722168/. Disponível em: https://pubmed.ncbi.nlm.nih.gov/29722168/. Acesso em: 5 jan. 2022.

HE, Xiangyang et al. Epigenetic modifications in radiation-induced non-targeted effects and their clinical significance. **Biochimica et Biophysica Acta (BBA) - General Subjects**,

HEWITT, R. J., et al. Lung extracellular matrix modulates KRT5$^{+}$ basal cell activity in pulmonary fibrosis. **Nature communications**, *14*(1), 6039. 2023.

HU, X., *et al.* The JAK/STAT signaling pathway: from bench to clinic. **Sig Transduct Target Ther,** 6, 402. 2021.

HU, Y., et al. Metagenomic analysis of the lung microbiome in pulmonary tuberculosis - a pilot study. **Emerging microbes & infections**, *9*(1), 1444–1452. 2020.

HUANG B, LANG X, LI X. The role of IL-6/JAK2/STAT3 signaling pathway in cancers. **Front Oncol**. 16;12:1023177. Dec. 2022.

HUANG, J., et al., The International Association for the Study of Lung Cancer Staging Project for Lung Cancer: Proposals for the Revision of the N Descriptors in the Forthcoming Ninth Edition of the TNM Classification for Lung Cancer. **Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer**, *19*(5), 766–785. 2024.

HUANG, Y.J., et al. The airway microbiome in patients with severe asthma: Associations with disease features and severity. **J. Allergy Clin. Immunol.** 136:874–884. 2015.

HUYNH, M., CRANE, M. J., & JAMIESON, A. M. The lung, the niche, and the microbe: Exploring the lung microbiome in cancer and immunity. **Frontiers in immunology**, *13*, 1094110. 2023.

IARC. Globocan 2022: Incidence and mortality worldwide. Lyon: International Agency for Research on Cancer, 2022. Disponível em: https://gco.iarc.fr/. Acesso em: 2 maio 2024.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro: INCA, 2023. Disponível em: https://www.inca.gov.br/numeros-de-cancer/estimativa. Acesso em: 2 maio 2024.

JIANG , Jun *et al*. Os papéis emergentes de RNAs não codificantes longos como marcas do câncer de pulmão. **Frontiers in Oncology**, [*s. l.*], 7 out. 2021.

JIANG, Lin; LI, Zheng; WANG, Ranran. Long non-coding RNAs in lung cancer: Regulation patterns, biologic function and diagnosis implications (Review). **Int J Oncol.**, [*s. l.*], 2019.

JOYCE EA, et al., Imaging features of extrapulmonary small cell carcinoma. **Clin Radiol**. 2013 Sep;68(9):953-61. doi: 10.1016/j.crad.2013.04.006. Epub 2013 Jun 19. PMID: 23790688.

KAIKKONEN, Minna U.; LAM, Michael T.Y.; GLASS, Christopher K. Non-coding RNAs as regulators of gene expression and epigenetics. **Cardiovasc. Res.**, [*s. l.*], 1 jun. 2011.

KANWAL, Rajnee; GUPTA, Karishma; GUPTA, Sanjay. Cancer epigenetics: an

introduction. **Methods Mol. Biol.**, [*s. l.*], 2015.

KIM, Jongchan *et al*. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. **Nat. Genet**, [*s. l.*], v. 50, n. 12, 2018.

KUKURBA, K. R., & MONTGOMERY, S. B. RNA Sequencing and Analysis. **Cold Spring Harbor protocols**, (11), 951–969. 2015.

LEE, J.T. Epigenetic regulation by long noncoding RNAs. **Science.** V.338, n. 6113, p. 1435-1439. 2012.

LEI, Zhengwen *et al*. MiR-185 inhibits non-small cell lung cancer cell proliferation and invasion through targeting of SOX9 and regulation of Wnt signaling. **Mol. Med. Rep.**, [*s. l.*], 2018.

LI J, et al. Interaction between long noncoding RNA and microRNA in lung inflammatory diseases. **Immun Inflamm Dis**. 12(1):e1129. Jan. 2024.

LI, J. Y., & LUO, Z. Q. LCAL1 enhances lung cancer survival via inhibiting AMPK- related antitumor functions. **Molecular and cellular biochemistry**, *457*(1-2), 11–20. 2019.

LI, R., LI, J. & ZHOU, X. Lung microbiome: new insights into the pathogenesis of respiratory diseases. **Sig Transduct Target Ther** 9, 19. 2024.

LI, Y., YAN, B., & HE, S. Advances and challenges in the treatment of lung cancer. **Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie**, 169, 115891. 2023.

LIANG, Zhuo-Zheng *et al*. CircRNA-miRNA-mRNA regulatory network in human lung cancer: an update. **Cancer Cell International**, [*s. l.*], 19 maio 2020.

LIM, Shuhui; KALDIS, Philipp. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. **Development**, [*s. l.*], 2013.

LING, Hui; FABBRI, Muller; CALIN, George A. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. **Nat. Rev. Drug Discov.**, [*s. l.*], 2013.

LIU X., et al., The role of gut microbiota in lung cancer: From carcinogenesis to immunotherapy. **Front. Oncol**. 11:3118. 2021.

LIU Y, et al. Non-coding RNAs in lung cancer: molecular mechanisms and clinical applications. **Front Oncol**. 8; 13:1256537. Sep. 2023.

LIU, W., et al., Microbiome and lung cancer: carcinogenic mechanisms, early cancer diagnosis, and promising microbial therapies. **Critical reviews in oncology/hematology**, *196*, 104322. 2024.

LUO, M., et al. Long non-coding RNAs control hematopoietic stem cell function. **Cell stem cell,** *16*(4), 426–438. 2016.

MA T, et al. Research progress of human key DNA and RNA methylation-related enzymes assay. **Talanta**. 2024 Jun 1;273:125872. doi: 10.1016/j.talanta.2024.125872. Epub 2024 Mar 5. PMID: 38471421.

MA, J., et al. Long-Chain Non-Coding RNA (lncRNA) MT1JP Suppresses Biological Activities of Lung Cancer by Regulating miRNA-423-3p/Bim Axis. **Medical science monitor : international medical journal of experimental and clinical research**, *25*, 5114–5126. 2019.

MA, L., BAJIC, V. B., & ZHANG, Z. On the classification of long non-coding RNAs. **RNA**

**biology**, *10*(6), 925–933. 2013.

MADDI A, et al.,. The microbiome and lung cancer. **J Thorac Dis**. 2019 Jan;11(1):280-291. doi: 10.21037/jtd.2018.12.88. PMID: 30863606; PMCID: PMC6384374.

MALTA, Deborah Carvalho et al. Mudanças no uso do tabaco entre adolescentes brasileiros e fatores associados: Pesquisa Nacional de Saúde do Escolar. **Ciência & Saúde Coletiva [online]**. v. 29, n. 9. 2024.

MARSLAND, B.; GOLLWITZER, E. Host-microorganism interactions in lung diseases. Nature Publishing Group, [*s. l.*], 2014 DOI 10.1038/nri3769. Disponível em: https://pubmed.ncbi.nlm.nih.gov/25421702/

MATHIAS, C., et al. Lung Cancer in Brazil. **Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer**, *15*(2), 170–175.

MATTICK, J.S., et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. **Nat Rev Mol Cell Biol** 24, 430–447. 2023.

MITHOOWANI H, FEBBRARO M. Non-Small-Cell Lung Cancer in 2022: A Review for General Practitioners in Oncology**. Curr Oncol.** 9;29(3):1828-1839. Mar. 2022.

MORGAN E, et al., Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. **Gut** 2023;72:338-344.

NAGL, L., *et al.* Molecular profiling in lung cancer. **memo 15**, 201–205 (2022).

NATALINI, J. G., SINGH, S., & SEGAL, L. N. The dynamic lung microbiome in health and disease. **Nature reviews. Microbiology**, *21*(4), 222–235. 2023.

NEMETH, K., et al. Non-coding RNAs in disease: from mechanisms to therapeutics. **Nature reviews**. *Genetics*, *25*(3), 211–232. 2024.

NIEMINEN, Tiina *et al*. Long Non-Coding RNA Modulation of VEGF-A during Hypoxia. **Frontier in Genetics**, [*s. l.*], v. 4, n. 4, 20 nov. 2018. DOI https://doi.org/10.3390/ncrna4040034. Disponível em: https://pubmed.ncbi.nlm.nih.gov/30463374/. Acesso em: 5 jan. 2022.

OLIVEIRA, A. C. et al. Molecular profile of non-small cell lung cancer in northeastern Brazil. **Jornal Brasileiro de Pneumologia,** v. 45, n.3, p. 320180181, 2019.

OLIVEIRA, N. F. P. Alterações no epigenoma e o hábito de fumar. **Revista Brasileira de Ciências da Saúde**, João Pessoa, v. 14, n. 4 ,p. 101-106, out./dez. 2011.

OMS, Organização Mundial da Saúde. International Agency for Research on Cancer. The Global Cancer Observatory Report. Lyon: International Agency for Research on Cancer; 2020

PALAZZO, Alexander F.; LEE, Eliza S. Non-coding RNA: what is functional and what is junk? **HYPOTHESIS AND THEORY article**, [*s. l.*], 26 jan. 2015. DOI https://doi.org/10.3389/fgene.2015.00002. Disponível em: https://www.frontiersin.org/articles/10.3389/fgene.2015.00002/full. Acesso em: 4 jan. 2022.

PENG, Yong; CROCE, Carlo M. The role of MicroRNAs in human cancer. **Signal Transduction and Targeted Therapy**, [*s. l.*], 28 jan. 2016. DOI https://doi.org/10.1038/sigtrans.2015.4. Disponível em: https://www.nature.com/articles/sigtrans20154. Acesso em: 5 jan. 2022.

PRIETO TG, et al. Pulmonary Neuroendocrine Neoplasms Overexpressing Epithelial-Mesenchymal Transition Mechanical Barriers Genes Lack Immune-Suppressive Response

and Present an Increased Risk of Metastasis. **Front Oncol**. 45(12). 2021.

QIU, M *et al*. A novel lncRNA, LUADT1, promotes lung adenocarcinoma proliferation via the epigenetic suppression of p27. **Cell Death Dis.**, [*s. l.*], 20 ago. 2015. DOI https://doi.org/10.1038/cddis.2015.203. Disponível em: https://pubmed.ncbi.nlm.nih.gov/26291312/. Acesso em: 5 jan. 2022.

RAHAL, Z., et al. Inflammation Mediated by Gut Microbiome Alterations Promotes Lung Cancer Development and an Immunosuppressed Tumor Microenvironment**. Cancer immunology research**, *12*(12), 1736–1752. 2024.

RAMI-PORTA, R., et al., Lung cancer - major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. **CA: a cancer journal for clinicians**, *67*(2), 138–155. 2017.

RATTI, Margherita *et al*. MicroRNAs (miRNAs) and Long Non-Coding RNAs (lncRNAs) as New Tools for Cancer Therapy: First Steps from Bench to Bedside. **Targeted Oncology**, [*s. l.*], v. 15, n. 3, 2020.

REHMAN, S. U., et al. Recent insights into the functions and mechanisms of antisense RNA: emerging applications in cancer therapy and precision medicine. **Frontiers in chemistry**, *11*, 1335330. 2024.

RITCHIE AI & SINGANAYAGAM A. Metagenomic Characterization of the Respiratory Microbiome. A Pièce de Résistance. **Am J Respir Crit Care Med** 1;202(3):321-322. Aug. 2020.

ROGERS, G. B. The lung microbiome. **Emerging topics in life sciences**, *1*(4), 313–324. 2017.

RUESCHHOFF, ALI B., et al. Lung Cancer Staging—A Clinical Practice Review. ***Journal of Respiration,*** *v.* 4, no. 1: 50-61. 2024

SANCHEZ-MARTÍNEZ D, et al. All-trans retinoic acid (ATRA) induces miR-23a expression, decreases CTSC expression and granzyme B activity leading to impaired NK cell cytotoxicity. **Int J Biochem Cell Biol**. Apr;49:42-52. 2014.

SATAM, H., et al., Next-Generation Sequencing Technology: Current Trends and Advancements. **Biology**, *12*(7), 997. 2023.

SCHABATH MB, COTE ML. Cancer Progress and Priorities: Lung Cancer. Cancer Epidemiol **Biomarkers Prev**. 2019 Oct;28(10):1563-1579. doi: 10.1158/1055-9965.EPI-19-0221. PMID: 31575553; PMCID: PMC6777859.

SCHMITT AM, CHANG HY. Long Noncoding RNAs in Cancer Pathways. **Cancer Cell**. 2016 Apr 11;29(4):452-463. doi: 10.1016/j.ccell.2016.03.010. PMID: 27070700; PMCID: PMC4831138.

SEMENTINO, E., et al. AKT and the Hallmarks of Cancer. **Cancer research**, 10.1158/0008-5472.CAN-24-1846. Advance online publication. 2024.

SHAH DR, MASTERS GA. Precision Medicine in Lung Cancer Treatment. **Surg Oncol Clin N Am.** 2020 Jan;29(1):15-21. doi: 10.1016/j.soc.2019.08.002. Epub 2019 Oct 16. PMID: 31757310.

SHANKAR A, et al. Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward. **Transl Lung Cancer Res.** 2019 May;8(Suppl 1):S106-S121. doi: 10.21037/tlcr.2019.03.03. PMID: 31211111; PMCID: PMC6546626.

SILVA, Pamela Alcantara Borges Santana et al. O novo 8° sistema TNM de câncer de pulmão: um panorama geral das mudanças. **ULAKES J Med**. v.1 n.2 p. 81-98, 2021.

SLACK, Frank J.; CHINNAIYAN, Arul M. The Role of Non-coding RNAs in Oncology. **Leading Edge Rev.**, [*s. l.*], 14 nov. 2019. DOI https://doi.org/10.1016/j.cell.2019.10.017. Disponível em: https://www.cell.com/cell/pdf/S0092-8674(19)31167-5.pdf. Acesso em: 5 jan. 2022.

STATELLO, Luisa *et al*. Gene regulation by long non-coding RNAs and its biological functions. **Nature Rev. Mol. Cell biol.**, [*s. l.*], 22 dez. 2020. DOI https://doi.org/110.1038/s41580-021-00330-4. Disponível em: https://www.nature.com/articles/s41580-020-00315-9. Acesso em: 5 jan. 2022.

STELLA GM, et al. Pragmatic Expectancy on Microbiota and Non-Small Cell Lung Cancer: A Narrative Review. **Cancers (Basel).** 26;14(13):3131. Jun, 2022.

SUN X, et al. Long non−coding RNA CASC11 interacts with YBX1 to promote prostate cancer progression by suppressing the p53 pathway. **Int J Oncol**. 61. 2022.

TAKAGI, H., et al. Delta-like 1 homolog (DLK1) as a possible therapeutic target and its application to radioimmunotherapy using $^{125}$I-labelled anti-DLK1 antibody in lung cancer models (HOT1801 and FIGHT004). **Lung cancer**. Amsterdam, Netherlands, *153*, 134–142. 2021.

TAN, Aaron C. et al. Targeted Therapies for Lung Cancer Patients With Oncogenic Driver Molecular Alterations. **JCO** 40, 611-625. (2022).

TANOUE L. T. Lung Cancer Staging. **Clinics in chest medicine**, *41*(2), 161–174. 2020.

TAS, F., & ERTURK, K. Online public interest in smoking and lung cancer: A comparative study in Google Trends. **Journal of cancer research and therapeutics**, 19(Suppl 2), S614–S617. 2023.

THAI, Alesha A *et al*. Lung Cancer. **Lancet**, [*s. l.*], p. P535-554, 14 jul. 2021.

THANDRA, K. C. et al., Epidemiology of lung cancer. **Contemporary oncology** *(Poznan, Poland)*, *25*(1), 45–52. 2021.

TIANSHENG, Guo *et al*. LncRNA Metastasis-Associated Lung Adenocarcinoma Transcript 1 Promotes Proliferation and Invasion of Non-Small Cell Lung Cancer Cells via Down-Regulating miR-202 Expression. **Cell J.**, [*s. l.*], v. 3, p. 375-385, 2020.

TINDLE HA, et al. Lifetime Smoking History and Risk of Lung Cancer: Results From the Framingham Heart Study**. J Natl Cancer Inst**. 1;110(11):1201-1207. Nov. 2018.

TSAGAKIS, I., DOUKA, K., BIRDS, I., & ASPDEN, J. L. Long non-coding RNAs in development and disease: conservation to mechanisms. **The Journal of pathology**, *250*(5), 480–495. 2020

TUTAR, Esen; TUTAR, Yusuf. Non-coding RNAs in lung cancer. **J. Thoracic Dis.**, [*s. l.*], v. 11, n. 3, 2019.

VALERIO, T. S., EMMERICK, I. C. M., & SOBREIRA-DA-SILVA, M. J. Factors associated with late-stage diagnosis and overall survival for lung cancer: An analysis of patients treated in a Brazilian hospital and a US-hospital from 2009 to 2019. **Cancer epidemiology**, *86*, 102443. 2023.

Volume 1867, Issue 8, 2023.

WANG L, et al. Metatranscriptome of human lung microbial communities in a cohort of

mechanically ventilated COVID-19 Omicron patients. **Signal Transduct Target Ther**. 10;8(1):432. Nov. 2023.

WANG, Chengdi *et al*. CircRNAs in lung cancer - Biogenesis, function and clinical implication. **Cancer Letters**, [*s. l.*], 2020.

WANG, Dingxue *et al*. LncRNA LUADT1 sponges miR-15a-3p to upregulate Twist1 in small cell lung cancer. **BMC Pulm. Med.**, [*s. l.*], 2019.

WANG, J., *et al.* lncRNA ZNRD1-AS1 promotes malignant lung cell proliferation, migration, and angiogenesis via the miR-942/TNS1 axis and is positively regulated by the m$^6$A reader YTHDC2. **Mol Cancer 21**, 229. 2022.

WANG, L., et al. Long non-coding RNA CASC19 facilitates non-small cell lung cancer cell proliferation and metastasis by targeting the miR-301b-3p/LDLR axis. **The journal of gene medicine**, *22*(12), e3254. 2020.

WANG, Maolong *et al*. Long non-coding RNAs in non-small cell lung cancer: functions and distinctions from other malignancies. **Translational Cancer Research** , [*s. l.*], v. 8, n. 7, 2019.

WANG, S., et al. The Role of lncRNA PCAT6 in Cancers. **Frontiers in oncology**, *11*, 701495. 2021.

WEI, Jian-Wei *et al*. Non-coding RNAs as regulators in epigenetics (Review). **Oncology Reports**, [*s. l.*], 8 nov. 2016.

WILLIAMS M, CHENG YY, BLENKIRON C, REID G. Exploring Mechanisms of MicroRNA Downregulation in Cancer. **Microrna**. 6(1):2-16. 2017.

WILLIAMS T, NGO LH, WICKRAMASINGHE VO. Nuclear export of RNA: Different sizes, shapes and functions. **Semin Cell Dev Biol**. 2018 Mar;75:70-77. doi: 10.1016/j.semcdb.2017.08.054. Epub 2017 Sep 1. PMID: 28866329.

XIA KG, et al. LncRNA NEAT1-associated aerobic glycolysis blunts tumor immunosurveillance by T cells in prostate cancer. **Neoplasma**. 69:594–602. 2022.

XU T, et al. Crosstalk between Environmental Inflammatory Stimuli and Non-Coding RNA in Cancer Occurrence and Development. **Cancers (Basel).** 2;13(17):4436. Sep. 2021

XU, N., et al. Microbiota dysbiosis in lung cancer: evidence of association and potential mechanisms. **Translational lung cancer research**, *9*(4), 1554–1568. 2020.

YAFANG LI, et al. Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population, **Carcinogenesis,** Volume 39, Issue 3, Pages 336–346. March 2018.

YAGI, K., et al. The Lung Microbiome during Health and Disease. **International journal of molecular sciences**, *22*(19), 10872. 2021.

YAN, Yueren et al., RNA splicing alterations in lung cancer pathogenesis and therapy. **Cancer Pathogenesis and Therapy**. Volume 1, Issue 4, Pages 272-283, ISSN 2949-7132. 2023.

YANG F, et al. The gut microbiota mediates protective immunity against tuberculosis *via* modulation of lncRNA. **Gut Microbes**. 14(1):2029997. Jan-Dec. 2022

YAO, Yixin; DAI, Wei. Genomic Instability and Cancer. **J. Carcinog Mutagen.**, [*s. l.*], 25 fev. 2015.

YE X, et al. Identification of significant genes in non-small cell lung cancer by bioinformatics analyses**. Transl Cancer Res**. 9(7):4330-4340. Jul, 2020.

YOKOTA, J., SHIRAISHI, K., & KOHNO, T. Genetic basis for susceptibility to lung cancer: Recent progress and future directions. **Advances in cancer research,** *109*, 51–72. 2010.

ZAKI, A., et al. Long non-coding RNA (lncRNA): A potential therapeutic target in acute lung injury. **Genes & diseases,** *9*(5), 1258–1268. 2021.

ZENG, Yujian *et al*. Altered expression profiles of circular RNA in colorectal cancer tissues from patients with lung metastasis. **Int. J. Mol. Med.**, [*s. l.*], v. 40, n. 6, 2017.

ZENGIN T & ÖNAL-SÜZEK T. Comprehensive Profiling of Genomic and Transcriptomic Differences between Risk Groups of Lung Adenocarcinoma and Lung Squamous Cell Carcinoma. **J Pers Med**. 23;11(2):154. Fev. 2021.

ZHANG, Baohong *et al*. MicroRNAs as oncogenes and tumor suppressors. **Dev. Biol.**, [*s. l.*], 2007.

ZHANG, Peijing *et al*. Non-Coding RNAs and their Integrated Networks. **J. Integr. Bioinform.**, [*s. l.*], v. 16, n. 3, 2019.

ZHANG, Y., JIA, C., & KWOH, C. K. Predicting the interaction biomolecule types for lncRNA: an ensemble deep learning approach. **Briefings in bioinformatics**, *22*(4), bbaa228. 2021.

ZHAO, Jiangmin *et al*. CircRNA Expression Profile in Early-Stage Lung Adenocarcinoma Patients. **Cell Physiol. Biochem**, [*s. l.*], 2017.

ZHAO, Shikang *et al*. Circular RNA Signature in Lung Adenocarcinoma: A MiOncoCirc Database-Based Study and Literature Review. **Frontiers in Oncology**, [*s. l.*], 9 out. 2020.

ZHU, B., *et al.* LncRNA SFTA1P mediates positive feedback regulation of the Hippo-YAP/TAZ signaling pathway in non-small cell lung cancer. **Cell Death Discov**. **7**, 369. 2021.

**ANEXOS**

ANEXO 1

**Termo de Consentimento Livre e Esclarecido**

Você está sendo convidado(a) para participar, como voluntário, em uma pesquisa. Após ser esclarecido(a) sobre as informações a seguir, no caso de aceitar fazer parte do estudo assine ao final deste documento, que está em duas vias e o mesmo será assinado e rubricado em todas as suas páginas. Uma delas é sua e a outra é do pesquisador responsável. Em caso de recusa você não será penalizado (a) de forma alguma. Se você tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) do Nucleo de Pesquisas em Oncologia, localizado no endereço: rua dos mundurucus, n. 4487, prédio da UNACON do Hospital Barros Barreto, telefone 3201-6778, email cep.npo@gmail.com. Título de Projeto: "Transcriptoma Total do câncer de pulmão".

O estudo pretende caracterizar a microbiota pulmonar no tecido com e sem tumor, e realizar a análise da expressão gênica destes biomas, comparando-os e relacionando-os com os dados clínico-patológicos dos pacientes estudados. Adicionalmente objetiva-se analisar a expressão gênica dos pacientes, de modo a encontrar correlações entre a expressão gênica bacteriana e humana no câncer de pulmão e em indivíduos sem câncer com o intuito de elucidar os mecanismos que favoreçam o desenvolvimento da neoplasia pulmonar.

Serão coletados pequenos fragmentos de tecido neoplásico removido por cirurgia e biópsias (análise de pequenas amostras de tecido para o diagnóstico do tipo do tumor). Não havendo despesas pessoais para cada participante em qualquer fase do estudo. A obtenção das amostras para pesquisa não implicará riscos adicionais no tratamento ou na cirurgia. Serão convidados a participar da pesquisa, pacientes com câncer de pulmão sem histórico de outro tipo de câncer.

Não participarão do estudo, pacientes com história de outro tipo de câncer e doenças infecciosas. Estas coletas serão realizadas mediante sua livre aceitação e assinatura do deste Termo de Consentimento Livre e Esclarecido.

Os possíveis riscos encontrados neste projeto estão relacionados a confidencialidade de dados do paciente para os estudos. Para diminuição deste risco, serão utilizadas estratégias

para manter o sigilo de informações, como associar dados dos pacientes à códigos de identificação. Quanto ao procedimento de coleta, a cirurgia e a biópsia pulmonar são procedimentos indicados e realizados para estabelecimento de diagnóstico e/ou tratamento do câncer de pulmão, assim como, de outras doenças pulmonares, sendo para tanto, necessária a remoção de material biológico relacionado à esta enfermidade. Parte do material retirado será encaminhada para as análises moleculares. Não há benefício direto para o participante. Trata-se de estudo experimental investigando o papel da microbiota no início e desenvolvimento do câncer de pulmão. Os resultados dessa pesquisa buscam gerar informações que visam ampliar o conhecimento sobre a carcinogênese e o microbioma no câncer pulmonar e posteriormente promover campos de pesquisa para busca não apenas de novas drogas no combate ao câncer de pulmão, mas principalmente métodos menos invasivos diagnósticos para detecção precoce da doença, assim como métodos prognósticos e preditivos.

Os dados coletados serão mantidos sob sigilo, avaliados e utilizados apenas para esta pesquisa e você não será recompensado com qualquer tipo de pagamento por ter participado. Em qualquer etapa do estudo você poderá ter acesso ao pesquisador responsável para esclarecimento de quaisquer dúvidas. OS principais investigadores são Prof. Dr. Paulo Pimentel Assumpção e Ma. Mariana Souza de Lima, que podem ser encontrados no Núcleo de Pesquisas em Oncologia, 1º Piso da UNACON. Av. Mundurucus, 4487, Guamá, Belém-PA, CEP: 66073-000, telefone (91) 3201-6776, e-mail: assumpcaopp@gmail.com e enf_mariana_souza@outlook.com

É garantida a liberdade de retirar o consentimento a qualquer momento e deixar de participar do estudo, sem qualquer prejuízo à continuidade do seu tratamento na instituição. As informações obtidas serão analisadas em conjunto com outros pacientes, não sendo divulgada a identificação de nenhum paciente. Esses materiais serão utilizados somente para este estudo. A eventual inclusão dos resultados em uma publicação científica será feita de modo a garantir o anonimato do participante. Na eventualidade de qualquer dano pessoal causado direta ou indiretamente após procedimentos propostos neste estudo (anexo causal comprovado), o participante terá direito a tratamento médico na instituição, bem como indenizações estabelecidas.

Acredito ter sido suficientemente esclarecido a respeito das informações que li ou que foram lidas para mim, descrevendo um estudo “Transcriptoma Total do câncer de pulmão”. Eu discuti com _____sobre a minha decisão de participar do estudo. Ficaram claros para mim quais são os propósitos do estudo, os procedimentos a serem realizados, seus desconfortos e riscos, as garantias de confidencialidade e de esclarecimentos permanentes.

Concordo voluntariamente em participar deste estudo e poderei retirar meu consentimento a qualquer momento, antes ou durante o processo, sem penalidades, prejuízo ou perda de qualquer benefício que possa ter adquirido. Ressaltando que tal documento será emitido em duas vias, permanecendo uma com o participante e uma via com o pesquisador.

Data: ____/____/_____

Nome do participante: _____ Assinatura do participante ou responsável legal:_____

Declaro que obtive de forma apropriada e voluntária o consentimento livre e esclarecido deste paciente ou representante legal para participação neste estudo.

Assinatura do responsável pelo estudo: _____

ANEXO II

**Universidade Federal do Pará**

***PRONTUÁRIO***

**Nome do Paciente:** _____

**Telefone para Contato:** _____

**Data de nascimento:** _____/_____/_____

***INFORMAÇÕES***

**Gênero:** ☐ Masculino ☐ Feminino

**Peso:** _____ **Altura:** _____

**Idade:** _____

***PACIENTE***

**Estado Civil:** ☐ Solteiro ☐ Casado ☐ Separado/Divorciado ☐ Viúvo

**Ocupação:** _____

**Naturalidade:** Cidade: _____ Estado: _____

**Endereço atual:** Cidade: _____ Estado: _____

**Etnia:** ☐ Amarelo ☐ Branco ☐ Pardo ☐ Negro

**Consumo de bebida alcoólica?** ☐ Sim ☐ Parou ☐ Nunca

Frequência: _____

Há quanto tempo parou de beber? _____

**Tabagismo:** ☐ Sim ☐ Parou ☐ Nunca fumou

Frequência (cigarros/dia/anos): _____

Há quanto tempo parou de fumar? _____

**Doenças Pulmonares:**
- ☐ DPOC
- ☐ Fibrose Cística
- ☐ Tuberculose
- ☐ Infecção Respiratória Recente (menos de 1 mês)
- ☐ Necessitou uso de antibiótico. Qual? _____
- ☐ Outras _____

**Medicamentos (uso regular):** _____

Há quantos anos: _____

**Presença de outros tipos de câncer:** ☐ Positivo ☐ Negativo

Qual? _____

Quando? _____

Realizou algum tratamento? Qual? _____

**Histórico de câncer na família:** ☐ Positivo ☐ Negativo

Qual o câncer? _____

Grau de parentesco: _____

Responsável: _____ Data: _____/_____/_____