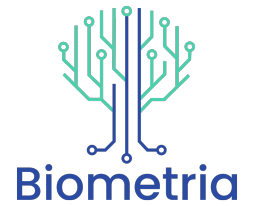


# RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese será disponibilizado somente a partir de 21/02/2026.



Universidade Estadual Paulista “Júlio de Mesquita Filho”  
Instituto de Biociências – Câmpus de Botucatu  
Programa de Pós-graduação em Biometria



# **Identificação de microRNAs e Genes Alvo como Biomarcadores de Câncer de Cabeça e Pescoço sob uma Abordagem de Bioinformática**

Saditt Rocio Robles Colonia

Botucatu  
2024

Saditt Rocio Robles Colonia

# **Identificação de microRNAs e Genes Alvo como Biomarcadores de Câncer de Cabeça e Pescoço sob uma Abordagem de Bioinformática**

Tese de Doutorado apresentada ao Programa de Pós-graduação em Biometria da Universidade Estadual Paulista “Júlio de Mesquita Filho” como parte dos requisitos necessários para a obtenção do título de Doutora em Biometria.

**Orientador:** Prof. Dr. Rogério Antonio de Oliveira

**Coorientadora:** Profa. Dra. Patrícia Pintor dos Reis

Botucatu  
2024

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: MARIA CAROLINA A. CRUZ E SANTOS-CRB 8/10188

Robles Colonia, Saditt Rocio.

Identificação de microRNAs e genes alvo como biomarcadores de câncer de cabeça e pescoço sob uma abordagem de bioinformática / Saditt Rocio Robles Colonia. - Botucatu, 2024

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Rogério Antonio de Oliveira

Coorientador: Patrícia Pintor dos Reis

Capes: 90194000

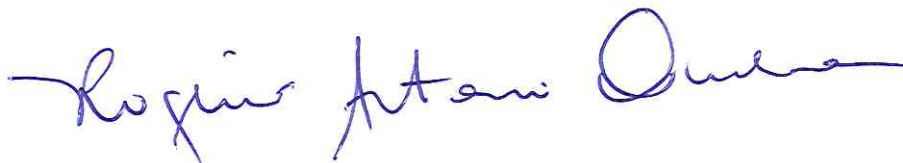
1. Árvores de decisão. 2. Análise de sobrevivência (Biometria). 3. Boca. 4. Laringe. 5. Marcadores bioquímicos.

Palavras-chave: Árvore de decisão; Análise de sobrevivência; Cavidade oral; Laringe; Redes de interação.

**ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE SADITT ROCIO ROBLES COLONIA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.**

Aos 21 dias do mês de fevereiro do ano de 2024, às 08:30 horas, no(a) Sala 7 da Central de Aulas 3, realizou-se a defesa de TESE DE DOUTORADO de SADITT ROCIO ROBLES COLONIA, intitulada **Identificação de microRNAs e genes alvo como biomarcadores de câncer de cabeça e pescoço sob uma abordagem de bioinformática**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. ROGERIO ANTONIO DE OLIVEIRA (Orientador(a) - Participação Presencial) do(a) Departamento de Biodiversidade e Bioestatística / Instituto de Biociências de Botucatu UNESP, Prof. Dr. ROBSON FRANCISCO CARVALHO (Participação Presencial) do(a) Departamento de Biologia Estrutural e Funcional / Instituto de Biociências de Botucatu - UNESP, Profa. Dra. SILVIA HELENA MODENESE GORLA DA SILVA (Participação Presencial) do(a) Departamento de Agronomia e Recursos Naturais / Faculdade de Ciências Agrárias do Vale do Ribeira - Câmpus de Registro - UNESP, Prof. Dr. ANGEL MAURICIO CASTRO GAMERO (Participação Virtual) do(a) Instituto de Ciências da Natureza (ICN) - Universidade Federal de Alfenas/MG, Profa. Dra. HILDETE PRISCO PINHEIRO (Participação Presencial) do(a) Departamento de Estatística / Instituto de Matemática, Estatística e Computação Científica - Universidade Estadual de Campinas. Após a exposição pela doutoranda e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, a discente recebeu o conceito final: APROVADA. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Prof. Dr. ROGERIO ANTONIO DE OLIVEIRA



*Dedico este trabalho à minha família, com um carinho especial à minha mãe, cujo maior legado para mim foi a valiosa herança da educação.*

# Agradecimentos

Em primeiro lugar, desejo expressar minha gratidão a Deus, que, nos momentos mais desafiadores da minha vida, sustentou-me e atendeu aos anseios do meu coração.

Agradeço aos meus pais, Manuel e Esperanza, à minha irmã Elianne e aos meus tios, pela constante inspiração, confiança e apoio ao longo desta jornada acadêmica, oferecendo seu apoio e amor incondicional. A meu amigo eterno Alberto Frank Lázaro Aguirre, por sua amizade constante, apoio inabalável e parceria que mantivemos desde a graduação. Desde o momento em que nos conhecemos, compartilhamos o sonho de realizar a Pós-Graduação no Brasil, e é gratificante finalmente alcançá-lo juntos.

Ao meu orientador, Prof. Dr. Rogério Antonio de Oliveira, pela confiança depositada em mim, pela constante motivação proporcionada, pela notável paciência demonstrada e pelo valioso auxílio oferecido ao longo do meu percurso acadêmico. À minha coorientadora, Prof. Dra. Patrícia Pintor dos Reis, pelo direcionamento deste trabalho.

Aos membros da banca examinadora, Prof. Dr. Robson Francisco Carvalho, Profa. Dra. Silvia Helena Modenese Gorla da Silva, Profa. Dra. Hildete Prisco Pinheiro e ao Prof. Dr. Angel Mauricio Castro Gamero pelas valiosas contribuições neste trabalho.

Ao corpo docente do Programa de Pós-Graduação em Biometria, pela oportunidade de ampliar meus conhecimentos e vivenciar experiências enriquecedoras. Aos funcionários, em especial a Luiz, expresso minha sincera gratidão pela atenção e apoio oferecidos.

Quero estender meus agradecimentos especiais a Juliana Gualberto e seus pais, que não apenas me acolheram, mas me consideraram parte de sua família. A Bruno Novaes e sua esposa Renata, pela amizade sincera e pelo constante apoio que sempre me proporcionaram.

Por fim, agradeço a meus amigos Alberto Lázaro, Julián Lázaro, Luis Baselly, Juliana Gualberto, Mario Lucas, Lara Cardeal e Abrantes Mussafo, com os quais compartilhei momentos inesquecíveis. A presença e apoio de todos vocês tornaram esta jornada acadêmica ainda mais significativa e memorável.

Este trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*Quem te disse que você não pode? Você não entende que Deus está contigo! Deus está ao nosso lado e somos mais que vencedores.*

*M.A.A. Fuego de Dios*

*Não basta fazer análises complicadas, os resultados devem ser úteis na prática e traduzidos para o pesquisador da outra o que os números que encontramos querem dizer.*

*Luzia Trinca*

# Resumo

O câncer de cabeça e pescoço (CCP) apresenta uma baixa taxa de sobrevivência, principalmente devido a diagnósticos tardios, falta de compreensão biológica e carência de biomarcadores confiáveis. Neste contexto, os microRNAs (miRNAs ou miRs) e seus alvos surgiram como biomarcadores relevantes para a formação, progressão e prognóstico de diversos tipos de cânceres, incluindo o CCP, sendo objeto de investigação constante. Este estudo teve como objetivo identificar miRNAs e seus genes alvo como potenciais biomarcadores para diagnóstico, prognóstico e tratamento do CCP, utilizando diversas ferramentas de bioinformática. A análise das expressões de miRNA-Seq e mRNA-Seq de pacientes com CCP foi conduzida com base nos dados do projeto Atlas do Genoma do Câncer (TCGA), juntamente com as informações clinicopatológicas correspondentes. As análises abrangeram a expressão diferencial, componentes principais, agrupamento, construção de redes de interação miRNA-mRNA, redes de co-expressão de proteínas, análises de enriquecimento funcional e classificação baseada em árvores, bem como análise de sobrevivência. No contexto da cavidade oral, identificaram-se 55 miRNAs expressos diferencialmente (DEmiRNAs) associados a 103 mRNAs expressos diferencialmente (DEmRNAs), e na laringe, 40 DEmiRNAs associados a 63 DEmRNAs. Alterações transcricionais foram observadas, com destaque para o gene COL1A1, regulado principalmente pelos miR-133a-3p, miR-29c-3p e miR-133b, indicando sua importância na digestão e absorção de proteínas. As árvores de decisão baseadas na classificação destacaram a relevância dos miR-101-3p, miR-455-3p, miR-99a-5p e miR-30a-5p como potenciais biomarcadores no diagnóstico do câncer de cavidade oral, distinguindo o tumor primário do tecido normal adjacente. No câncer de laringe, miR-338-5p, miR-1910-5p, miR-99a-5p, miR-196b-5p e miR-139-3p emergiram como candidatos a biomarcadores independentemente do estadiamento da doença. A análise de riscos proporcionais de Cox univariada identificou cinco miRNAs e uma assinatura combinada de 24 miRNAs associadas à sobrevivência dos pacientes com câncer da cavidade oral. Para o câncer de laringe, nove miRNAs e uma assinatura combinada de 16 miRNAs foram associadas à sobrevivência dos pacientes. Esses resultados destacam o potencial desses miRNAs e seus alvos como biomarcadores no diagnóstico, prognóstico e compreensão do processo biológico do CCP.

**Palavras-chave:** Cavidade oral, Laringe, Redes de interação, Árvore de decisão, Análise de sobrevivência.

# Abstract

Head and neck cancer (HNC) presents a low survival rate, mainly due to late diagnosis, lack of biological understanding, and a shortage of reliable biomarkers. In this context, microRNAs (miRNAs or miRs) and their targets have emerged as relevant biomarkers for the formation, progression, and prognosis of various types of cancers, including HNC, and have been the subject of constant investigation. This study aimed to identify miRNAs and their target genes as potential biomarkers for the diagnosis, prognosis, and treatment of HNC, using various bioinformatics tools. Analysis of miRNA-Seq and mRNA-Seq expressions in HNC patients was conducted based on data from The Cancer Genome Atlas (TCGA) project, along with corresponding clinicopathological information. The analyses included differential expression, principal components, clustering, construction of miRNA-mRNA interaction networks, protein co-expression networks, functional enrichment analyses, tree-based classification, as well as survival analysis. In the context of the oral cavity, 55 differentially expressed miRNAs (DEmiRNAs) associated with 103 differentially expressed mRNAs (DEmRNAs) were identified, and in the larynx, 40 DEmiRNAs associated with 63 DEmRNAs were identified. Transcriptional alterations were observed, with emphasis on the COL1A1 gene, regulated mainly by miR-133a-3p, miR-29c-3p, and miR-133b, indicating its importance in protein digestion and absorption. Decision trees based on classification highlighted the relevance of miR-101-3p, miR-455-3p, miR-99a-5p, and miR-30a-5p as potential biomarkers in the diagnosis of oral cavity cancer, distinguishing the primary tumor from adjacent normal tissue. In laryngeal cancer, miR-338-5p, miR-1910-5p, miR-99a-5p, miR-196b-5p, and miR-139-3p emerged as biomarker candidates regardless of disease staging. Univariate Cox proportional hazards analysis identified five miRNAs and a combined signature of 24 miRNAs associated with the survival of oral cavity cancer patients. For laryngeal cancer, nine miRNAs and a combined signature of 16 miRNAs were associated with patient survival. These results highlight the potential of these miRNAs and their targets as biomarkers in the diagnosis, prognosis, and understanding of the biological process of HNC.

**Keywords:** Oral cavity, Larynx, Interaction networks, Decision tree, Survival analysis.

# Lista de figuras

Figura 1 – Etapas da carcinogênese: iniciação, promoção, progressão e metástase. Figura criada em <Biorender.com>. . . . .	4
Figura 2 – Ilustração da heterogeneidade tumoral (intertumoral e intratumoral). Figura criada em <Biorender.com>. . . . .	5
Figura 3 – Ilustração da anatomia da cabeça e pescoço. Figura criada em <Biorender.com>. . . . .	7
Figura 4 – Biogênese e mecanismo dos miRNAs. Figura criada em <Biorender.com>. . . . .	9
Figura 5 – Nomenclatura dos miRNAs coletados no miRBase. . . . .	10
Figura 6 – Ilustração da árvore de decisão. . . . .	27
Figura 7 – Ilustração da floresta aleatória. Figura adaptada de Khan <i>et al.</i> (2021). . . . .	29
Figura 8 – Ilustração da curva ROC juntamente com a Área Sob a Curva ROC. . . . .	32
Figura 9 – Ilustração esquemática dos métodos de validação. (A) Método Holdout e (B) validação cruzada <i>k-fold</i> . . . . .	33
Figura 10 – Fluxograma das etapas utilizadas no estudo. . . . .	39
Figura 11 – (A) Gráfico de barras referentes ao tecido normal adjacente e tumor primário nos sítios anatômicos do CCP. (B) Gráfico de barras referentes ao órgão de origem do câncer da cavidade oral. . . . .	53
Figura 12 – Gráficos bivariados das características clinicopatológicas. (A) <i>Boxplots</i> das idades (anos) dos pacientes por estadiamento patológico. (B) Histograma das idades dos pacientes estratificado por sexo. (C) Histograma dos tempos de censura e falha. (D) Histograma das idades (anos) por estado vital dos pacientes nos sítios anatômicos do câncer de cabeça e pescoço. . . . .	54
Figura 13 – Distribuição da expressão de oito miRNAs e oito mRNAs nos sítios anatômicos da cavidade oral (A) e laringe (B), respectivamente. . . . .	55
Figura 14 – <i>Boxplots</i> da expressão dos miRNAs para os primeiros 112 amostras da cavidade oral (a, c, e) e laringe (b, d, f), respectivamente. . . . .	56
Figura 15 – Expressão Diferencial de miRNAs em comparação entre tecidos tumorais primários e normais adjacentes relacionados ao câncer de cavidade oral. (A) O gráfico MA da razão logarítmica (M) versus a média (A) das diferenças entre os dois grupos para cada miRNA. (B) Gráfico de Vulcão dos DE miRNAs. (C) Gráfico de Barras com os 10 DE miRNAs mais desregulados. (D) Gráfico de pirulitos divergentes com os 10 DE miRNAs em relação ao risco relativo. . . . .	58
Figura 16 – Gráfico de barras dos 10 DE miRNAs mais desregulados no câncer da cavidade oral pelo método de edgeR (A) e o teste de Wilcoxon-Mann-Whitney (B), respectivamente. (C) Diagrama de Venn mostrando a sobreposição dos resultados obtidos pelos três métodos (DESeq2, edgeR e o teste de Wilcoxon-Mann-Whitney). . . . .	59

Figura 17 – Gráficos de violino ilustrando a expressão dos cinco principais DEmiRNAs superexpressos (A) e dos cinco DEmiRNAs subexpressos (B) ao comparar o tumor primário com o tecido normal adjacente ao tumor, relacionados ao câncer na cavidade oral. . . . .	60
Figura 18 – Análise de componentes principais com base na expressão dos DEmiRNA de cada amostra associados ao câncer da cavidade oral. (A) <i>Scree plot</i> dos autovalores para cada componente principal. (B) <i>Scree plot</i> da variação explicada dos componentes principais. (C) Gráficos de barras para duas dimensões com base na contribuição dos miRNAs. (D) PCA mostra o padrão de expressão dos miRNAs na estratificação do tumor primário e tecido adjacente. . . . .	61
Figura 19 – Mapa de calor de agrupamento hierárquico da expressão de miRNA entre as amostras dos dois grupos (tecido tumoral e normal adjacente ao tumor) no câncer da cavidade oral usando a distância Euclideana e o método de ligação completa. . . . .	62
Figura 20 – Dendrograma circular hierárquico da separação entre grupos de DEmiRNAs superexpressos e subexpressos no câncer da cavidade oral usando a distância Euclideana e o método de ligação completa. . . . .	63
Figura 21 – Matriz de correlação linear de Spearman entre os DEmiRNAs associados ao câncer da cavidade oral. . . . .	64
Figura 22 – DEmiRNAs na comparação do tumor primário e normal adjacente relacionados ao câncer da laringe. (A) Gráfico MA da razão logarítmica (M) versus a média (A) das diferenças entre os dois grupos para cada miRNA. (B) Gráfico de Vulcão dos DEmiRNAs. (C) Gráfico de Barras com os 10 DEmiRNAs mais desregulados. (D) Gráfico de pirulitos divergentes com os 10 DEmiRNAs em relação ao risco relativo. . . . .	65
Figura 23 – Gráficos de violino da expressão dos 10 primeiros DEmiRNAs (superexpresso (A) e subexpresso (B)) na comparação do tumor primário com o tecido normal adjacente ao tumor associados ao câncer da laringe. . . . .	66
Figura 24 – Gráfico de Barras dos 10 DEmiRNAs mais desregulados no câncer da laringe pelo método de edgeR (A) e Wilcoxon-Mann-Whitney (B), respectivamente. (C) Diagrama de Venn da sobreposição dos resultados obtidos pelos três métodos. . . . .	67
Figura 25 – Análise de componentes principais com base na expressão dos DEmiRNA de cada amostra associados ao câncer da laringe. (A) <i>Scree-plot</i> dos autovalores para cada componente principal. (B) <i>Scree-plot</i> da variação explicada dos componentes principais. (C) Gráficos de barras para duas dimensões com base na contribuição dos miRNAs. (D) PCA mostra o padrão de expressão dos miRNAs na estratificação do tumor primário e tecido normal adjacente. . . . .	68

Figura 26 – Mapa de calor de agrupamento hierárquico da expressão de miRNA entre as amostras dos dois grupos (tumor primário e normal adjacente ao tumor) no câncer da laringe oral usando a distância Euclideana e o método de ligação completa. . . . .	69
Figura 27 – Dendrograma circular hierárquico da separação entre grupos de DEmiRNAs superexpressos e subexpressos no câncer da laringe usando a distância Euclideana e o método de ligação completa. . . . .	70
Figura 28 – Matriz de correlação linear de Spearman entre DEmiRNAs no câncer da laringe. . . . .	71
Figura 29 – Diagrama de Venn dos DEmiRNAs validados com outras fontes externas para câncer da cavidade oral (A), laringe (B), tipos de cânceres (C) e câncer de cabeça e pescoço (D), respectivamente. . . . .	73
Figura 30 – Diagrama de Venn mostra a sobreposição entre (A) Vias KEGG e (B) Anotação GO em câncer da cavidade oral e laringe, respectivamente. . . . .	74
Figura 31 – Gráficos de bolhas mostrando as trinta principais vias KEGG e os termos GO associados aos miRNAs desregulados em pacientes com câncer oral (A, B) e laringe (C, D), respectivamente. . . . .	75
Figura 32 – Diagrama de Venn mostra a sobreposição dos DEmiRNAs entre o câncer da cavidade oral e da laringe. . . . .	76
Figura 33 – Análise de Enriquecimento de vias KEGG em DEmiRNAs sobrepostos entre o câncer da cavidade oral e da laringe. . . . .	76
Figura 34 – (A, B) Diagrama de Venn mostrando a sobreposição dos DEmiRNAs entre estadiamento inicial e avançado do câncer da cavidade oral e laringe, respectivamente. (C) Gráficos de violino da expressão dos cinco DEmiRNAs nos estadiamentos avançados do câncer de cavidade oral. (D) Gráfico de violino da expressão dos DEmiRNAs no estadiamento inicial e avançado do câncer de laringe. . . . .	78
Figura 35 – <i>MA-Plot</i> (A) e <i>volcano-Plot</i> (B) dos DEmiRNAs entre o tumor primário e tecido de normal adjacente relacionados ao câncer da cavidade oral (A, B) e laringe (C, D), respectivamente. . . . .	79
Figura 36 – Diagrama de Venn mostrando a sobreposição entre DEmiRNAs e genes alvo dos DEmiRNAs (mirDIP e miRTarBase) associados ao câncer da cavidade oral (A) e da laringe (B), respectivamente. . . . .	80
Figura 37 – (A, C) Gráfico de barras dos 10 DEmiRNAs mais desregulados (valores FDR muito baixos) e influenciado pelo los valores $\log_2(FC)$ para o câncer da cavidade oral e laringe, respectivamente. (B, C) Análise de componentes principais dos DEmiRNAs relacionados ao câncer da cavidade oral e laringe. . . . .	81

Figura 38 – <i>Heatmap</i> e dendrograma circular dos DEmRNAs entre as amostras dos dois grupos (tecido do tumor primário e normal adjacente) no câncer da cavidade oral (A,B) e laringe (C, D) usando a distância Euclideana e o método de ligação completa para as colunas (amostras) e Ward.D2 para as linhas (DEmRNAs). A seta vermelha em (B) significa que o gene ADD2 é superexpresso e que foi agrupado incorretamente. . . . .	82
Figura 39 – Rede regulatória entre DEmiRNAs e DEmRNAs associadas ao câncer da cavidade oral (A) e laringe (B), respectivamente. . . . .	83
Figura 40 – Rede regulatória entre DEmiRNAs e DEmRNAs do estadiamento avançado do câncer da cavidade oral (A), inicial e avançado do câncer da laringe (B), respectivamente. . . . .	84
Figura 41 – Anotação funcional e análise de enriquecimento dos DEmRNAs. (A, B, C, D) Rede interativa de análise de enriquecimento e gráfico de bolhas exibindo as vias KEGG dos DEmRNAs no câncer da cavidade oral e laringe, respectivamente.(E, F) Gráfico de bolhas de enriquecimento do termo Ontologia Genética para cada um dos termos de instância (processo biológico, componente celular e função molecular) no câncer da cavidade oral e laringe, respectivamente. (G, H) Gráfico de barras da análise de enriquecimento da Ontologia de Doenças no câncer da cavidade oral e laringe, respectivamente.	87
Figura 42 – (A, E) Rede de interação proteína-proteína (PPI) associada ao câncer da cavidade oral e laringe, respectivamente. (B, F) <i>Hub</i> genes referentes ao câncer da cavidade oral e laringe, respectivamente. A cor vermelha a amarela representa o grau de conectividade de cima para baixo baseada no algoritmo de <i>Maximal Clique Centrality</i> (MCC). (C, G) Rede regulatória entre DEmiRNAs e <i>hub</i> genes referentes ao câncer da cavidade oral e laringe, respectivamente. (D, H) Enriquecimento da via KEGG dos <i>hubs</i> referentes ao câncer da cavidade oral e laringe, respectivamente. . . . .	90
Figura 43 – <i>Heatmap</i> baseada na árvore de decisão para classificação de tecidos do câncer da cavidade oral, considerando todos os estadiamentos patológicos. As cores do mapa apresentam o valor relativo da expressão dos miRNAs em cada amostra.	92
Figura 44 – Classificação dos tecidos utilizando CTree e floresta aleatória, abrangendo todos os estadiamentos patológicos (A, B), iniciais (C, D) e avançados (E, F) do câncer da cavidade oral. . . . .	95
Figura 45 – <i>Heatmap</i> baseada na árvore de decisão para classificação de tecidos do câncer da laringe, considerando todos os estadiamentos patológicos. As cores do mapa apresentam o valor relativo da expressão dos miRNAs em cada amostra.	96
Figura 46 – Classificação dos tecidos utilizando CTree e floresta aleatória, abrangendo todos os estadiamentos patológicos (A, B) e avançados (C, D) do câncer da laringe. . . . .	98

Figura 47 – Curvas de sobrevivência estimada pelo método de Kaplan-Meier de acordo com o câncer da cavidade oral e laringe. . . . .	99
Figura 48 – Curva de sobrevivência estimada pelo método de Kaplan-Meier, teste de Log-Rank e modelo de Cox univariado estratificado por estadiamento patológico no conjunto treinamento do câncer da cavidade oral. . . . .	102
Figura 49 – Curvas de sobrevivência estimada pelo método de Kaplan-Meier, teste de Log-Rank e modelo de Cox univariado de acordo com a expressão de miRNAs em diferentes estadiamentos patológicos (A) inicial (B) e avançado (C) do câncer da cavidade oral no conjunto de treinamento. . . . .	104
Figura 50 – Gráfico dos riscos do modelo multivariado de riscos proporcionais de Cox dos miRNAs associados com a sobrevivência dos pacientes com câncer oral considerando todos os estadiamentos (I - IV) e avançado (III - IV) no conjunto treinamento. Cada miRNA encontra-se categorizada (1: expressão aumentada e 0: expressão diminuída). . . . .	106
Figura 51 – Árvore de sobrevivência podada obtida pelo CART <sup>®</sup> e gráfico de barras da importância de miRNAs obtidos pela floresta aleatória na previsão da sobrevivência dos pacientes considerando todos os estadiamentos (A, B), inicial (C, D) e avançado (E, F) do câncer da cavidade oral no conjunto treinamento. . . . .	108
Figura 52 – Curvas de sobrevivências estimada pelo método de Kaplan-Meier para sobrevivência dos pacientes estratificado por estadiamento patológico no câncer da laringe. . . . .	109
Figura 53 – Curvas de sobrevivência estimada pelo método de Kaplan-Meier, teste de Log-Rank e modelo de Cox univariado para sobrevivência dos pacientes de acordo com a expressão dos miRNAs considerando todos os estadiamentos (A) e o estadiamento avançado (B) do câncer da laringe. . . . .	112
Figura 54 – Gráfico de riscos do modelo multivariado de riscos proporcionais de Cox dos miRNAs associados com a sobrevivência dos pacientes considerando todos os estadiamentos do câncer da laringe. Cada miRNA encontra-se categorizada (1: expressão aumentada e 0: expressão diminuída). . . . .	112
Figura 55 – Árvore de sobrevivência podada obtido pelo CART <sup>®</sup> e fator de importância de miRNAs obtidas via floresta aleatória na previsão da sobrevivência dos pacientes considerando todos os estadiamentos (A, B) e avançado (C, D) no conjunto treinamento do câncer da laringe, respectivamente. . . . .	113

# Lista de tabelas

Tabela 1	– Tipos de erros de um teste de hipóteses. . . . .	19
Tabela 2	– Erros em contrastes de hipóteses múltiplas. . . . .	20
Tabela 3	– Matriz de confusão comparando valores preditos e valores observados. . . . .	30
Tabela 4	– Métricas de desempenho na classificação. . . . .	31
Tabela 5	– Descrição das características clinicopatológicas dos 528 pacientes diagnosticados com câncer de cabeça e pescoço. . . . .	40
Tabela 6	– Matriz de expressão do miRNA-Seq. Cada linha corresponde a um miRNA específico e cada coluna uma amostra referente ao tumor primário (exemplo, TCGA-4P-AA8J-01) e tecido sólido normal ou adjacente ao tumor (exemplo, TCGA-CV-6933-11) com seus respectivos níveis de expressão associados ao câncer de cabeça e pescoço. . . . .	41
Tabela 7	– Matriz de expressão de genes transcritos codificadores (mRNA-Seq). Cada linha corresponde a um mRNA específico e cada coluna uma amostra referente ao tumor primário (exemplo, TCGA-4P-AA8J-01) e tecido histologicamente normal adjacente ao tumor (exemplo, TCGA-H7-A6C5-11) com seus respectivos níveis de expressão associados ao câncer de cabeça e pescoço. . . . .	41
Tabela 8	– Conjuntos de dados destinados à validação dos DEmiRNAs presentes no banco de dados dbDMEC (versão 3.0) (XU <i>et al.</i> , 2022). . . . .	46
Tabela 9	– Frequência absoluta ( $f_a$ ) e percentual ( $f_p$ ) das características clinicopatológicas dos pacientes com câncer de cabeça e pescoço (CCP) em cada sítio anatômico da doença. . . . .	52
Tabela 10	– Análise das curvas de ROC na discriminação do tumor primário e do tecido normal adjacente ao tumor no câncer da cavidade oral e laringe. . . . .	91
Tabela 11	– Matriz de confusão e avaliação de desempenho dos modelos baseados nas árvores de decisão (CTree e floresta aleatória) utilizando <i>Five-Fold Cross-Validation</i> na classificação binária de tumor primário (1) e tecido normal adjacente (0) em relação ao câncer de cavidade oral com diferentes técnicas de balanceamento. . . . .	93
Tabela 12	– Matriz de confusão e avaliação de desempenho dos modelos baseados nas árvores de decisão (CTree e floresta aleatória) utilizando <i>Five-Fold Cross-Validation</i> na classificação binária de tumor primário (1) e tecido normal adjacente (0) em relação ao câncer da laringe com diferentes técnicas de balanceamento. . . . .	97
Tabela 13	– Comparação do desempenho de modelos multivariados de Cox e árvores de sobrevivência e floresta de sobrevivência aleatória no câncer da cavidade oral e laringe, considerando os estadiamentos da doença. . . . .	100

Tabela 14 – Estimativas da regressão de Cox univariado e multivariado das características clinicopatológicas do câncer da cavidade oral no conjunto treinamento. . . . .	101
Tabela 15 – Estimativas de regressão de Cox univariada e multivariada das características clinicopatológicas do câncer da laringe. . . . .	109
Tabela 16 – Distribuição das características clínico-patológicas em relação ao tecido normal e tumoral nos sítios anatômicos do CCP. . . . .	142
Tabela 17 – Estimativas dos parâmetros do modelo Binomial Negativa da expressão dos DEmiRNAs na comparação dos tecidos (tumoral vs. normal) na cavidade oral. . . . .	143
Tabela 18 – Estimativas dos coeficientes do modelo Binomial Negativa da expressão dos DEmiRNAs em amostras de tecido tumoral primário em comparação com os tecidos normais adjacentes ao tumor dos pacientes com câncer da laringe. . . . .	144
Tabela 19 – Estimativas dos coeficientes do modelo pelo edgeR em amostras de tecido tumoral primário em comparação com os tecidos normais adjacentes ao tumor dos pacientes com câncer da laringe. . . . .	145
Tabela 20 – Estatísticas descritivas da expressão dos miRNAs identificados na cavidade oral comparando os tecidos usando o teste de Wilcoxon-Mann-Whitney. . . . .	146
Tabela 21 – Estatísticas descritivas da expressão dos miRNAs identificados na laringe comparando os tecidos (tumor primário vs. normal adjacente) usando o teste de Wilcoxon-Mann-Whitney. . . . .	147
Tabela 22 – Estimativas dos coeficientes do modelo pelo Deseq2 em amostras de tecido tumoral primário em comparação com os tecidos normais adjacentes ao tumor dos pacientes com câncer da laringe. . . . .	148
Tabela 23 – DEmiRNAs identificados com o teste WMW nos estadiamento iniciais e avançados e áreas sob a curva (AUC), especificidade (TNR) e sensibilidade (TPR) na capacidade de discriminação do tumor primário (T) e tecido normal adjacente ao tumor (NAT) na cavidade oral. . . . .	149
Tabela 24 – DEmiRNAs identificados com o teste WMW nos estadiamentos iniciais e avançados e áreas sob a curva (AUC), especificidade (TNR) e sensibilidade (TPR) na capacidade de discriminação do tumor primário (T) e tecido normal adjacente ao tumor (NAT) na laringe. . . . .	150

# Lista de abreviaturas e siglas

DNA	Ácido Desoxirribonucleico
RNA	Ácido ribonucleico
RNA-Seq	Sequenciamento de RNA
ncRNA	RNA não codificante
miRNA	microRNA
miRNA-Seq	Sequenciamento de microRNA
mRNA	RNA mensageiro
mRNA-Seq	Sequenciamento de mRNA
NGS	Sequenciamento da Próxima Geração
AJCC	<i>American Joint Committee on Cancer</i>
UTCC	<i>Union for International Cancer Control</i>
TNM	Tumor, linfonodos, metástase
TCGA	Genoma Atlas do Câncer
mirDIP	<i>microRNA Data Integration Portal</i>
CCP	Câncer de cabeça e pescoço
CCP	Carcinoma de células escamosas de cabeça e pescoço
CCE	Carcinoma de células escamosas
CCEF	Carcinoma de células escamosas fusiformes
CCEB	Carcinoma de células escamosas basalóide
CCEQ	Carcinoma de células escamosas queratinizante
CCENQ	Carcinoma de células escamosas não queratinizado
CART®	Árvores de Classificação e Regressão
MLG	Modelos Lineares Generalizados

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>1.1</b>	<b>Objetivos</b>	<b>2</b>
<b>1.1.1</b>	<b>Objetivo geral</b>	<b>2</b>
<b>1.1.2</b>	<b>Objetivos específicos</b>	<b>3</b>
<b>2</b>	<b>ASPECTOS BIOLÓGICOS DO CÂNCER DE CABEÇA E PESCOÇO E SUA RELAÇÃO COM O TRANSCRIPTOMA</b>	<b>4</b>
<b>2.1</b>	<b>Biologia do câncer</b>	<b>4</b>
<b>2.2</b>	<b>Cânceres de cabeça e pescoço</b>	<b>6</b>
<b>2.3</b>	<b>MiRNAs na regulação da expressão gênica</b>	<b>8</b>
<b>2.4</b>	<b>MiRNAs e genes envolvidos no câncer de cabeça e pescoço</b>	<b>10</b>
<b>3</b>	<b>FERRAMENTAS DA BIOINFORMÁTICA APLICADA À GENÔMICA</b>	<b>12</b>
<b>3.1</b>	<b>Análise de expressão diferencial e predição de alvos</b>	<b>12</b>
<b>3.2</b>	<b>Redes regulatórias e análise de enriquecimento funcional</b>	<b>12</b>
<b>4</b>	<b>MÉTODOS ESTATÍSTICOS APLICADOS À EXPRESSÃO GÊNICA</b>	<b>14</b>
<b>4.1</b>	<b>Modelos Lineares Generalizados</b>	<b>14</b>
<b>4.1.1</b>	<b>Modelo de regressão de Poisson</b>	<b>15</b>
<b>4.1.2</b>	<b>Modelo de regressão Binomial Negativa</b>	<b>15</b>
<b>4.2</b>	<b>Teste de múltiplas hipóteses</b>	<b>19</b>
<b>4.3</b>	<b>Testes não paramétricos</b>	<b>21</b>
<b>4.3.1</b>	<b>Teste não paramétrico de Wilcoxon-Mann-Whitney (WMW)</b>	<b>21</b>
<b>4.3.2</b>	<b>Teste Qui-quadrado de Pearson</b>	<b>22</b>
<b>4.4</b>	<b>Análise multivariada dos dados</b>	<b>22</b>
<b>4.4.1</b>	<b>Análise de componentes principais</b>	<b>23</b>
<b>4.4.2</b>	<b>Análise de agrupamento – <i>Clustering</i></b>	<b>25</b>
<b>4.5</b>	<b>Árvores de decisão e floresta aleatória</b>	<b>26</b>
<b>4.5.1</b>	<b>Técnicas de reamostragem para dados desbalanceados</b>	<b>30</b>
<b>4.5.2</b>	<b>Métricas de desempenho das árvores de decisão</b>	<b>30</b>
<b>4.6</b>	<b>Análise de sobrevivência</b>	<b>33</b>
<b>4.6.1</b>	<b>Estimador de Kaplan-Meier</b>	<b>35</b>
<b>4.6.2</b>	<b>Modelo de riscos proporcionais de Cox</b>	<b>35</b>
<b>4.6.3</b>	<b>Árvores de regressão para dados censurados</b>	<b>37</b>
<b>4.6.4</b>	<b>Métrica de desempenho para dados censurados</b>	<b>38</b>

5	BANCO DE DADOS E MÉTODOS DE ANÁLISE . . . . .	39
5.1	Conjunto de dados do TCGA . . . . .	40
5.2	Análise de expressão diferencial dos miRNAs e mRNAs . . . . .	42
5.3	Análise de componentes principais e <i>Clustering</i> . . . . .	44
5.4	Validação da expressão dos miRNAs com fontes externas . . . . .	45
5.5	Análise de enriquecimento e anotação funcional de miRNAs . . . . .	46
5.6	Predição de alvos, redes de interação miRNA-mRNA e análises de enriquecimento GO e KEGG de mRNAs . . . . .	46
5.7	Rede de interação proteína-proteína, identificação de <i>hub</i> genes, rede regulatória miRNA- <i>hub</i> e análise de enriquecimento de vias KEGG . . .	47
5.8	Algoritmos de classificação baseadas em árvores de decisão e florestas aleatórias . . . . .	48
5.9	Análise de sobrevivência . . . . .	49
6	RESULTADOS . . . . .	51
6.1	Análise preliminar dos dados . . . . .	51
6.1.1	Características clinicopatológicas . . . . .	51
6.1.2	Perfil de expressão de miRNAs e mRNAs . . . . .	55
6.2	Análise da expressão diferencial dos miRNAs . . . . .	57
6.2.1	Identificação de DEmiRNAs no câncer da cavidade oral . . . . .	57
6.2.2	Identificação de DEmiRNAs no câncer da laringe . . . . .	64
6.2.3	Validação de miRNAs candidatos com outras fontes de dados . . . . .	71
6.2.4	Análise de enriquecimento e anotação ontológica de miRNAs . . . . .	74
6.2.5	Identificação de DEmiRNAs comumente expressos no câncer da cavidade oral e laringe . . . . .	75
6.2.6	Identificação de DEmiRNAs associados ao estadiamento do câncer da cavidade oral e laringe . . . . .	77
6.3	Identificação DEmRNAs, predição de alvos, redes de interação miRNA-mRNA, redes de proteína-proteína e análise funcional . . . . .	79
6.3.1	Rede de interação regulatória miRNA-mRNA . . . . .	83
6.3.2	Análises de enriquecimento e de vias dos DEmRNAs . . . . .	85
6.3.3	Rede de interação proteína-proteína e redes regulatórias miRNA- <i>hub</i> . .	88
6.4	Discriminação de tecidos por meio de curvas ROC: avaliando a expressão individual de miRNAs . . . . .	91
6.4.1	Classificadores baseados em árvore de decisão no diagnóstico de câncer da cavidade oral e laringe usando a expressão de miRNAs . . . . .	92
6.5	MiRNAs como biomarcadores de sobrevivência para câncer da cavidade oral e laringe . . . . .	99
6.5.1	MiRNAs como preditores de sobrevivência em pacientes com câncer da cavidade oral . . . . .	101

6.5.2	MiRNAs como preditores de sobrevivência em pacientes com câncer da laringe . . . . .	109
7	DISCUSSÃO . . . . .	114
8	CONCLUSÃO . . . . .	120
8.1	Limitações do estudo . . . . .	121
8.2	Trabalhos futuros . . . . .	121
	REFERÊNCIAS . . . . .	123
	 <b>APÊNDICES</b>	 <b>141</b>
	<b>APÊNDICE A</b> <b>VARIÁVEIS CLÍNICAS, MIRNA-SEQ E MRNA-SEQ . .</b>	<b>142</b>
A.1	Análise bivariada das características clinicopatológicas . . . . .	142
A.2	Identificação de DEmiRNAs pelo DESeq2: oral e laringe . . . . .	143
A.3	Identificação de DEmiRNAs com edgeR . . . . .	145
A.4	Identificação de DEmiRNAs pelo WMW: oral e laringe . . . . .	146
A.5	Identificação de DEmRNAs com DESeq2: oral e laringe . . . . .	148
A.6	Identificação dos DEmiRNAs nos estadiamento iniciais e avançados do câncer da cavidade oral e laringe . . . . .	149

# 1 Introdução

O câncer de cabeça e pescoço (CCP), também conhecido como carcinoma de células escamosas de cabeça e pescoço (CCECP), é considerado um dos tipos de câncer mais agressivos, e complexo no diagnóstico, pois frequentemente é detectado em estadiamentos avançados (MODY *et al.*, 2021; INCA, 2022). Esta doença está associada a altas taxas de incidência e mortalidade, levando a óbito mais de 400 mil pessoas anualmente (SUNG *et al.*, 2021) e prevê-se que até 2030, a incidência possa aumentar em até 30% (JOHNSON *et al.*, 2020). A maioria dos diagnósticos de CCP ocorre nos sítios anatômicos da cavidade oral e laringe, principalmente em homens com mais de 40 anos e em grande parte dos casos, associado ao consumo prolongado de tabaco e a ingestão de álcool (NETWORK *et al.*, 2015; SHAW; BEASLEY, 2016; GORMLEY *et al.*, 2022). No entanto, apesar dos avanços nas estratégias terapêuticas, a taxa de sobrevivência dos pacientes com CCP ainda é baixa, devido ao insucesso terapêutico que muitas vezes decorre de diagnósticos tardios, compreensão biológica limitada, dificuldade na avaliação da progressão do cancro devido à localização anatômica, diversidade clinicopatológica, heterogeneidade observada na resposta ao tratamento e à falta de biomarcadores, como os miRNAs clinicamente validados para o diagnóstico precoce (ALSAHAFI *et al.*, 2019; PASI *et al.*, 2020; VAHABI; BLANDINO; AGOSTINO, 2021).

Atualmente, com o avanço das técnicas de sequenciamento de próxima geração (NGS) e o estabelecimento de repositórios de dados públicos, como o *Genomic Data Commons* (GDC), cBioPortal, *UCSC Xena*, *Human Protein Atlas*, *Pan-Cancer Atlas* e *Gene Expression Omnibus* (GEO), têm aumentado significativamente a quantidade de dados genômicos, transcriptômicos e clínicos de diversos tipos de câncer (NETWORK *et al.*, 2015). Como consequência, métodos computacionais e estatísticos, como aprendizado de máquina para mineração e modelagem preditiva generalizável, têm sido amplamente adotados na bioinformática para identificar e caracterizar o RNA-Seq para o diagnóstico e prognóstico do câncer. Além disso, as ferramentas de análise de enriquecimento funcional estão se tornando cada vez mais importantes na decodificação da expressão dos miRNAs e mRNAs que estão relacionados aos processos biológicos e vias de sinalização (CHEN *et al.*, 2019; GARCIA-MORENO; CARMONA-SAEZ, 2020; BUITRAGO; LOVERING; CAPORALI, 2023).

Geralmente, em estudos da expressão do RNA-Seq, empregam-se métodos de expressão diferencial para reduzir a complexidade dos dados e identificar genes expressos diferencialmente

no câncer (MCDERMAID *et al.*, 2019). Esses métodos são complementados por métodos não supervisionados, como técnicas de agrupamento e análise de componentes principais (PCA), para identificar padrões na expressão gênica. Além disso, métodos supervisionados, como regressão linear, regressão logística, árvores de decisão, floresta aleatória, máquinas de vetores de suporte (SVM), redes neurais, classificador Naive Bayes, k-vizinhos mais próximos, redes Bayesianas, *Lasso*, *Ridge* entre outras são utilizados para construir modelos preditivos. No entanto, as árvores de decisão são um diferencial em relação aos outros algoritmos, pois não requerem especificações de parâmetros complexos. Outro método amplamente utilizado é a análise de sobrevivência que visa examinar o tempo até a ocorrência de um evento de interesse relacionado à doença, sendo complementada por uma série de métodos estatísticos e modelos mais complexos de aprendizado de máquina, como a floresta de sobrevivência aleatória (RSF), utilizado para construir o modelo de sobrevivência com dados genômicos de alta dimensão (MIRZA *et al.*, 2019; LEE; LIM, 2019).

Alguns estudos bem-sucedidos na aplicação desses métodos no diagnóstico e prognóstico do CCP, utilizando a expressão de miRNAs, foram publicados em Nair *et al.* (2015), Cecco *et al.* (2017), Lopez *et al.* (2018), Lu *et al.* (2019), Kumar, Patnaik e Dixit (2020), Panvongsa *et al.* (2021), Piotrowski *et al.* (2021) e Qi *et al.* (2021). No entanto, mesmo que existam estudos promissores sobre miRNAs e mRNAs no CCP, ainda não foram identificados biomarcadores específicos que possam ser utilizados na prática clínica devido à complexidade anatômica desta doença.

Desse modo, verifica-se que existe ainda uma grande necessidade de adotar abordagens integradas de bioinformática para atingir uma compreensão mais abrangente e detalhada da complexidade biológica dos miRNAs e mRNAs associadas ao CCP, o que poderia resultar em avanços significativos no diagnóstico precoce, prognósticos mais precisos e no desenvolvimento de tratamentos personalizados mais eficazes e menos prejudiciais, proporcionando qualidade de vida e possibilidade de cura para os pacientes, classificados com base em sua localização anatômica.

## 8 Conclusão

Com base nos resultados obtidos, foram identificados 58 DEmiRNAs no contexto do câncer da cavidade oral, destacando-se como os mais significantes: miR-455-3p, miR-196b-5p, miR-196a-5p, miR-503-5p, miR-210-3p, miR-101-3p, miR-381-3p, miR-30a-3p, miR-30a-5p e miR-378a-5p. Paralelamente, foram identificadas 1423 DEmRNAs, sendo CA9, LAMC2, INHBA, IL11, TGFBI, PRH2, PRB3, KLK1, NOSTRIN e LPO os mais notáveis neste cenário. Na análise do câncer de laringe, observou-se 44 DEmiRNAs, com miR-196b-5p, miR-196a-5p, miR-615-3p, miR-4652-5p, miR-455-3p, miR-338-5p, miR-139-3p, miR-378c, miR-378d e miR-99a-5p como os mais significativos. Adicionalmente, foram identificadas 1507 DEmRNAs, sendo ESM1, ZIC5, COL4A1, COL4A2, MMP11, EMP1, UNL3, SLC35C1, KAT2B e FCHO2 os mais proeminentes.

De maneira significativa, 29 miRNAs foram compartilhados entre os dois tipos de câncer, indicando um enriquecimento em várias vias KEGG, como carcinogênese viral, vias do câncer, câncer de próstata, desregulação transcricional e câncer de bexiga, sugerindo interconexões entre esses processos patológicos. Identificou-se também dez genes centrais, sendo COL1A1 e COL3A1 os mais importantes em ambos os tipos de cânceres, indicando que o miR-29c-3p é o principal regulador dos genes associados principalmente às vias enriquecidas, como digestão e absorção de proteínas. Adicionalmente, cinco miRNAs (miR-1269a, miR-767-5p, miR-381-5p, miR-133a-3p e miR-133b) estiveram especificamente ligados ao estadiamento avançado do câncer de cavidade oral, regulando genes como COL1A1, FBN2, FOXL2, NRIP3, ONECUT2 e SERPINH1. No câncer de laringe, o miR-105-5p associou-se ao estadiamento inicial, regulando exclusivamente o gene CLIC5.

Na avaliação dos algoritmos de classificação, o RFT superou o CTree para ambos os cânceres. Os miR-101-3p, miR-455-3p, miR-99a-5p e miR-30a-5p foram identificados como potenciais biomarcadores no diagnóstico do câncer de cavidade oral, distinguindo o tumor primário do tecido normal adjacente. No câncer de laringe, miR-99a-5p, miR-139-3p e miR-196b-5p emergiram como candidatos a biomarcadores independentemente do estadiamento da doença.

Em suma, o modelo de riscos proporcionais de Cox foi eficaz na identificação de biomarcadores prognósticos em dois tipos de câncer. No câncer de cavidade oral, a idade e o estadiamento avançado surgiram como fatores de risco para óbito. Além disso, observou-se que a superexpressão de miR-210-3p e miR-101-3p estava correlacionada a uma menor taxa de

sobrevivência, enquanto miR-503-3p estava associada a uma maior taxa de sobrevivência. Em estadiamento avançados, miR-133b foi vinculada a uma menor taxa de sobrevivência, enquanto miR-133a-3p foi associada a uma maior taxa de sobrevivência. No câncer de laringe, somente pacientes do sexo feminino e miR-1-3p apresentaram um prognóstico desfavorável.

Desta forma, os resultados obtidos neste estudo representam uma valiosa contribuição para o avanço do entendimento da biologia molecular dos cânceres de cavidade oral e laringe. Essas descobertas não apenas promovem novas perspectivas no diagnóstico, prognóstico e tratamento da doença, mas também abrem portas para o desenvolvimento de estratégias terapêuticas inovadoras e aprimoramentos na abordagem clínica dos pacientes afetados por cânceres de cabeça e pescoço. Contudo, é crucial ressaltar a necessidade de validar esses miRNAs por meio de estudos experimentais, a fim de avaliar sua relevância clínica.

## 8.1 Limitações do estudo

Limitações deste estudo não invalidam os resultados obtidos, mas fornecem informações sobre os desafios e as restrições enfrentadas pelos pesquisadores. As limitações incluem dificuldades na obtenção conjunta de dados de expressão de miRNAs maduros e mRNAs do TCGA, bem como a disponibilidade limitada de dados em comparação com outras plataformas, como o GEO e o *ArrayExpress*. Cada plataforma possui limitações distintas em termos de amostras, tipos de dados e abrangência de genes, o que pode restringir a quantidade de dados disponíveis para análise, dependendo dos objetivos do estudo.

Embora o TCGA forneça algumas informações clínicas, estas podem não ser suficientes para análises mais sofisticadas ou relacionadas a doenças específicas, sendo a qualidade dos dados clínicos crucial para a interpretação dos resultados. Em estudos retrospectivos, a obtenção de dados clínicos completos pode ser desafiadora, com informações incompletas identificadas em variáveis como tempo de seguimento, idade, estadiamento do tumor e outros.

A variação na expressão de miRNAs e mRNAs entre os indivíduos também dificultou a identificação de padrões de expressão relevantes para determinar a condição clínica da doença. Essa variação pode ser influenciada por fatores como variações genéticas individuais, fatores ambientais e a heterogeneidade tumoral.

## 8.2 Trabalhos futuros

Com base nos resultados desta tese, algumas direções promissoras são sugeridas para orientar futuros desenvolvimentos na pesquisa:

- Recomenda-se a realização de análises de classificação e sobrevivência considerando os genes alvo dos miRNAs;

- Recomenda-se aprimorar a análise de expressão diferencial por meio da integração de dados transcriptômicos de diversas plataformas, juntamente com os dados clínicos, utilizando diferentes métodos de redução de dimensionalidade. Para a classificação de tecidos, sugere-se a adoção de modelos alternativos de classificação binária que levem em consideração o desbalanceamento dos dados, empregando funções de ligação assimétricas. Esses modelos devem ser comparados com os algoritmos de classificação do aprendizado de máquina, mediante técnicas como reamostragem e validação cruzada;
- No contexto da análise de sobrevivência, é recomendável comparar métodos paramétricos de sobrevivência com algoritmos de sobrevivência utilizados no aprendizado de máquina, tais como máquinas de vetores de suporte e redes neurais. Novamente, a utilização de técnicas como reamostragem e validação cruzada é essencial nesse processo. Essas análises, tanto para classificação quanto para sobrevivência, podem ser conduzidas sob as abordagens frequentista ou Bayesiana envolvendo a bioinformática, proporcionando uma compreensão abrangente e robusta dos resultados obtidos.

# REFERÊNCIAS

- ABDI, H. *et al.* Bonferroni and šidák corrections for multiple comparisons. *Encyclopedia of Measurement and Statistics*, Sage Thousand Oaks, CA, v. 3, p. 103–107, 2007. 19
- ADELSTEIN, D. *et al.* Nccn guidelines insights: head and neck cancers, version 2.2017. *Journal of the National Comprehensive Cancer Network*, Harborside Press, LLC, v. 15, n. 6, p. 761–770, 2017. 8
- AHMAD, A.; IMRAN, M.; AHSAN, H. Biomarkers as biomedical bioindicators: Approaches and techniques for the detection, analysis, and validation of novel biomarkers of diseases. *Pharmaceutics*, MDPI, v. 15, n. 6, p. 1630, 2023. 8
- AHMAD, J. *et al.* Microrna in carcinogenesis & cancer diagnostics: a new paradigm. *The Indian Journal of Medical Research*, Wolters Kluwer–Medknow Publications, v. 137, n. 4, p. 680, 2013. 11
- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Ieee, v. 19, n. 6, p. 716–723, 1974. 36
- AL-HAMEED, K. A. A. Spearman’s correlation coefficient in statistical analysis. *International Journal of Nonlinear Analysis and Applications*, Semnan University, v. 13, n. 1, p. 3249–3255, 2022. 45
- ALFARO, E.; GÁMEZ, M.; GARCÍA, N. *Ensemble classification methods with applications in R*. [S.l.]: John Wiley & Sons, 2018. 26
- ALSAHAFI, E. *et al.* Clinical update on head and neck cancer: molecular biology and ongoing challenges. *Cell Death & Disease*, Nature Publishing Group UK London, v. 10, n. 8, p. 540, 2019. 1, 8
- ALWANI, A.; BAJ-KRZYWORZEKA, M. Mikrorna jako cel terapeutyczny w chorobach nowotworowych. *Postępy Biochemii*, v. 67, n. 3, 2021. 11
- AMBROS, V. *et al.* A uniform system for microrna annotation. *RNA*, Cold Spring Harbor Lab, v. 9, n. 3, p. 277–279, 2003. 10
- AMIN, M. B. *et al.* *AJCC cancer staging manual*. [S.l.]: Springer, 2017. v. 1024. 8
- ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. *Nature Precedings*, Nature Publishing Group, p. 1–1, 2010. 15, 16, 17, 19
- ANDERSON, G. *et al.* An updated review on head and neck cancer treatment with radiation therapy. *Cancers*, MDPI, v. 13, n. 19, p. 4912, 2021. 8
- ASSIS, A. F. *et al.* What is the transcriptome and how it is evaluated. In: *Transcriptomics in Health and Disease*. [S.l.]: Springer, 2022. p. 3–50. 12
- ATHANASIOS, A. *et al.* Protein-protein interaction (ppi) network: recent advances in drug discovery. *Current Drug Metabolism*, Bentham Science Publishers, v. 18, n. 1, p. 5–10, 2017. 13

- AUPÉRIN, A. Epidemiology of head and neck cancers: an update. *Current Opinion in Oncology*, LWW, v. 32, n. 3, p. 178–186, 2020. 7
- AZAR, M. R. M. H. *et al.* Altered expression levels of mir-212, mir-133b and mir-27a in tongue squamous cell carcinoma (tscc) with clinicopathological considerations. *Gene Reports*, Elsevier, v. 19, p. 100622, 2020. 115
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 57, n. 1, p. 289–300, 1995. 19, 20, 21
- BERTSIMAS, D. *et al.* Optimal survival trees. *Machine Learning*, Springer, v. 111, n. 8, p. 2951–3023, 2022. 37
- BIAN, T. *et al.* Identification of clic5 as a prognostic biomarker and correlated immunomodulator for lung adenocarcinoma. *Combinatorial Chemistry & High Throughput Screening*, Bentham Science Publishers, v. 26, n. 14, p. 2452–2468, 2023. 116
- BLIGHE, K. *et al.* Package ‘pcatools.’. *REAGENT or RESOURCE SOURCE IDENTIFIER* Deposited data GitHub repository with code and data for analysis GitHub [https://github.com/CharlotteEPage/NFI\\_Disease\\_Ecology](https://github.com/CharlotteEPage/NFI_Disease_Ecology) Raw data tables Mendeley <https://data.mendeley.com/datasets/y869bhzmzr/1> Software and algorithms R version, v. 4, n. 1, p. 9–11, 2019. 44
- BORKOVEC, M.; MADIN, N. ggparty:“ggplot” visualizations for the “partykit” package. *R package version*, v. 1, n. 0, 2019. 48
- BOZEC, A. *et al.* Current role of primary surgical treatment in patients with head and neck squamous cell carcinoma. *Current Opinion in Oncology*, LWW, v. 31, n. 3, p. 138–145, 2019. 8
- BRAY, F. *et al.* Erratum: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, v. 70, n. 4, p. 313, 2020. 6
- BREIMAN, L. Random forests. *Machine Learning*, Springer, v. 45, n. 1, p. 5–32, 2001. 26, 28, 29, 37
- BRIEMAN, L. *et al.* *Classification and regression tree analysis*. [S.l.]: CRC Press: Boca Raton, FL, USA, 1984. 27, 28
- BRIERLEY, J. D.; GOSPODAROWICZ, M. K.; WITTEKIND, C. *TNM classification of malignant tumours*. [S.l.]: John Wiley & Sons, 2017. 8
- BROSEGHINI, E. *et al.* Diagnostic and prognostic value of micrnas in patients with laryngeal cancer: A systematic review. *Non-coding RNA*, MDPI, v. 9, n. 1, p. 9, 2023. 115
- BUITRAGO, D. L.; LOVERING, R. C.; CAPORALI, A. Insights into online micrna bioinformatics tools. *Non-coding RNA*, MDPI, v. 9, n. 2, p. 18, 2023. 1
- BUJA, A.; EYUBOGLU, N. Remarks on parallel analysis. *Multivariate Behavioral Research*, Taylor & Francis, v. 27, n. 4, p. 509–540, 1992. 24
- CALIFF, R. M. Biomarker definitions and their applications. *Experimental Biology and Medicine*, SAGE Publications Sage UK: London, England, v. 243, n. 3, p. 213–221, 2018. 8

- CANNING, M. *et al.* Heterogeneity of the head and neck squamous cell carcinoma immune landscape and its impact on immunotherapy. *Frontiers in Cell and Developmental Biology*, Frontiers Media SA, v. 7, p. 52, 2019. 8
- CARLBERG, C.; VELLEUER, E. Introduction to cancer. In: \_\_\_\_\_. *Cancer Biology: How Science Works*. Cham: Springer International Publishing, 2021. p. 1–16. ISBN 978-3-030-75699-4. 4
- CARLSON, M. *et al.* org. hs. eg. db: Genome wide annotation for human. *R package version*, v. 3, n. 2, p. 3, 2019. 47
- CARVALHO, P. R.; MUNITA, C. S.; LAPOLLI, A. L. Validity studies among hierarchical methods of cluster analysis using cophenetic correlation coefficient. *Brazilian Journal of Radiation Sciences*, v. 7, n. 2A, 2019. 26
- CASELLA, G.; BERGER, R. L. *Inferência estatística*. [S.l.]: Cengage Learning, 2010. 15, 16, 19
- CATTELL, R. B. The scree test for the number of factors. *Multivariate Behavioral Research*, Taylor & Francis, v. 1, n. 2, p. 245–276, 1966. 24
- CECCO, L. D. *et al.* Integrative mirna-gene expression analysis enables refinement of associated biology and prediction of response to cetuximab in head and neck squamous cell cancer. *Genes*, MDPI, v. 8, n. 1, p. 35, 2017. 2, 10
- CHARRAD, M. *et al.* Nbclust: an r package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, v. 61, p. 1–36, 2014. 45
- CHAWLA, N. V. *et al.* Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002. 30
- CHEN, L. *et al.* Trends in the development of mirna bioinformatics tools. *Briefings in Bioinformatics*, Oxford University Press, v. 20, n. 5, p. 1836–1852, 2019. 1
- \_\_\_\_\_. Integrated data analysis implicates plau, serpine1, spp1, and mmp1 as prognostic factors in head and neck squamous cell carcinoma. *Journal of Postgraduate Medicine, Education and Research*, v. 55, n. 4, p. 160–170, 2021. 116
- CHEN, R.; ZHANG, Y.; ZHANG, X. Mir-1254 functions as a tumor suppressor in oral squamous cell carcinoma by targeting cd36. *Technology in Cancer Research & Treatment*, SAGE Publications Sage CA: Los Angeles, CA, v. 18, p. 1533033819859447, 2019. 115
- CHEN, Y.; WANG, X. mirdb: an online database for prediction of functional microrna targets. *Nucleic Acids Research*, Oxford University Press, v. 48, n. D1, p. D127–D131, 2020. 12
- CHEN, Y. *et al.* mir-30a-3p inhibits renal cancer cell invasion and metastasis through targeting atg12. *Translational Andrology and Urology*, AME Publications, v. 9, n. 2, p. 646, 2020. 119
- CHIN, C.-H. *et al.* cytohubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, BioMed Central, v. 8, n. 4, p. 1–7, 2014. 47
- CHOW, L. Q. Head and neck cancer. *New England Journal of Medicine*, Mass Medical Soc, v. 382, n. 1, p. 60–72, 2020. 6, 8
- CIAMPI, A. *et al.* Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. *Computational Statistics & Data Analysis*, Elsevier, v. 4, n. 3, p. 185–204, 1986. 37

- COLAPRICO, A. *et al.* Tcgabioblinks: an r/bioconductor package for integrative analysis of tcga data. *Nucleic Acids Research*, Oxford University Press, v. 44, n. 8, p. e71–e71, 2016. 40
- COLEVAS, A. D. *et al.* Nccn guidelines insights: head and neck cancers, version 1.2018. *Journal of the National Comprehensive Cancer Network*, Harborside Press, LLC, v. 16, n. 5, p. 479–490, 2018. 8
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. [S.l.]: Editora Blucher, 2006. 33, 34, 35, 36
- CONRAD, O. *et al.* Tumor-suppressive and immunomodulating activity of mir-30a-3p and mir-30e-3p in hnscc cells and tumoroids. *International Journal of Molecular Sciences*, MDPI, v. 24, n. 13, p. 11178, 2023. 119
- CORMACK, R. M. A review of classification. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 134, n. 3, p. 321–353, 1971. 26
- COUSSEAU, C. P. V. *et al.* mir-99a-5p as a biomarker for lymph node metastasis prediction in oral squamous cell carcinoma patients. *Head & Neck*, Wiley Online Library, v. 45, n. 10, p. 2489–2497, 2023. 118
- COWLAND, J. B.; HOTHER, C.; GRØNBAEK, K. Micrnas and cancer. *Apmis*, Wiley Online Library, v. 115, n. 10, p. 1090–1106, 2007. 9
- COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 34, n. 2, p. 187–202, 1972. 35, 36
- CRIMI, S. *et al.* Droplet digital pcr analysis of liquid biopsy samples unveils the diagnostic role of hsa-mir-133a-3p and hsa-mir-375-3p in oral cancer. *Biology*, MDPI, v. 9, n. 11, p. 379, 2020. 11, 116
- CRISPIM, D. L. *et al.* Comparação de métodos de agrupamentos hierárquicos aglomerativos em indicadores de sustentabilidade em municípios do estado do pará. *Research, Society and Development*, v. 9, n. 2, p. e60922067–e60922067, 2020. 26
- DAI, S. *et al.* The important role of mir-1-3p in cancers. *Journal of Translational Medicine*, Springer, v. 21, n. 1, p. 769, 2023. 116
- DALMAN, M. R. *et al.* Fold change and p-value cutoffs significantly alter microarray interpretations. In: SPRINGER. *BMC Bioinformatics*. [S.l.], 2012. v. 13, p. 1–4. 18
- DENISSOFF, A. *et al.* Exposure to alcohol and overall survival in head and neck cancer: A regional cohort study. *Head & Neck*, Wiley Online Library, v. 44, n. 10, p. 2109–2117, 2022. 7
- D'HAESELEER, P. How does gene expression clustering work? *Nature Biotechnology*, Nature Publishing Group US New York, v. 23, n. 12, p. 1499–1501, 2005. 26
- DÍEZ, J.; LÓPEZ-NOVOA, J. Introducción a la biología molecular. *Nefrología*, Elsevier, v. 13, n. 1, p. 14–22, 1993. 9
- DILLIES, M.-A. *et al.* A comprehensive evaluation of normalization methods for illumina high-throughput rna sequencing data analysis. *Briefings in Bioinformatics*, Oxford University Press, v. 14, n. 6, p. 671–683, 2013. 17

- DONG, X. *et al.* Plasma mir-1247-5p, mir-301b-3p and mir-105-5p as potential biomarkers for early diagnosis of non-small cell lung cancer. *Thoracic Cancer*, Wiley Online Library, v. 12, n. 4, p. 539–548, 2021. 116
- D'SOUZA, W.; KUMAR, A. micrnas in oral cancer: moving from bench to bed as next generation medicine. *Oral Oncology*, Elsevier, v. 111, p. 104916, 2020. 11
- FALZONE, L. *et al.* Identification of novel micrnas and their diagnostic and prognostic significance in oral cancer. *Cancers*, MDPI, v. 11, n. 5, p. 610, 2019. 46
- FAN, J.; UPADHYE, S.; WORSTER, A. Understanding receiver operating characteristic (roc) curves. *Canadian Journal of Emergency Medicine*, Cambridge University Press, v. 8, n. 1, p. 19–20, 2006. 31
- FANG, R. *et al.* Downregulation of mir-29c-3p is associated with a poor prognosis in patients with laryngeal squamous cell carcinoma. *Diagnostic Pathology*, Springer, v. 14, p. 1–7, 2019. 115
- FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006. 30, 31
- FEI, Y. *et al.* Identification of mirna-mrna crosstalk in laryngeal squamous cell carcinoma. *Molecular Medicine Reports*, Spandidos Publications, v. 16, n. 4, p. 4179–4186, 2017. 115
- FEKETE, J. T.; WELKER, Á.; GYÓRFFY, B. Mirna expression signatures of therapy response in squamous cell carcinomas. *Cancers*, MDPI, v. 13, n. 1, p. 63, 2020. 114
- FILETTI, V. *et al.* Diagnostic and prognostic value of three micrnas in environmental asbestiform fibers-associated malignant mesothelioma. *Journal of Personalized Medicine*, MDPI, v. 11, n. 11, p. 1205, 2021. 118
- FILHO, A. D. P. C. Uso de big data em saúde no brasil: perspectivas para um futuro próximo. *Epidemiologia e Serviços de Saúde*, SciELO Brasil, v. 24, p. 325–332, 2015. 19, 20
- FISHER, R. A. On the interpretation of  $\chi^2$  from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, JSTOR, v. 85, n. 1, p. 87–94, 1922. 22
- FOULKES, A. S. *Applied statistical genetics with R*. [S.l.]: Springer-Verlag, 2009. 19
- GAI, C. *et al.* Salivary extracellular vesicle-associated mirnas as potential biomarkers in oral squamous cell carcinoma. *BMC cancer*, BioMed Central, v. 18, n. 1, p. 1–11, 2018. 115
- GALBIATTI, A. L. S. *et al.* Câncer de cabeça e pescoço: causas, prevenção e tratamento. *Brazilian Journal of Otorhinolaryngology*, SciELO Brasil, v. 79, p. 239–247, 2013. 8
- GALILI, T. dendextend: an r package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, Oxford University Press, v. 31, n. 22, p. 3718–3720, 2015. 45
- GAO, C.-H.; YU, G.; CAI, P. ggvenndiagram: an intuitive, easy-to-use, and highly customizable r package to generate venn diagram. *Frontiers in Genetics*, Frontiers, p. 1598, 2021. 47
- GAO, T. *et al.* mir-30a-3p inhibits the proliferation of laryngeal cancer cells by targeting dnmt3a through regulating dna methylation of pten. *Oncologie*, De Gruyter, n. 0, 2023. 119

GARCIA-MORENO, A.; CARMONA-SAEZ, P. Computational methods and software tools for functional analysis of mirna data. *Biomolecules*, MDPI, v. 10, n. 9, p. 1252, 2020. 1

GARRIDO-CANO, I. *et al.* Circulating mir-99a-5p expression in plasma: A potential biomarker for early diagnosis of breast cancer. *International Journal of Molecular Sciences*, MDPI, v. 21, n. 19, p. 7427, 2020. 118

GAUCH, H. G.; JR, H. G. G. *Multivariate analysis in community ecology*. [S.l.]: Cambridge University Press, 1982. 26

GE, S. X.; JUNG, D.; YAO, R. Shinygo: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, Oxford University Press, v. 36, n. 8, p. 2628–2629, 2020. 48

GENTLEMAN, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, Springer, v. 5, n. 10, p. 1–16, 2004. 16, 40, 43

GIBBONS, J. D.; CHAKRABORTI, S. *Nonparametric statistical inference*. [S.l.]: CRC press, 2014. 21

GLUSMAN, G. *et al.* Optimal scaling of digital transcriptomes. *PloS One*, Public Library of Science San Francisco, USA, v. 8, n. 11, p. e77885, 2013. 17

GODARD, P.; EYLL, J. V. Pathway analysis from lists of micrnas: common pitfalls and alternative strategy. *Nucleic acids Research*, Oxford University Press, v. 43, n. 7, p. 3490–3497, 2015. 13

GOŁĄBEK, K. *et al.* mir-125b-5p, mir-155-3p, and mir-214-5p and target e2f2 gene in oral squamous cell carcinoma. *International Journal of Molecular Sciences*, MDPI, v. 24, n. 7, p. 6320, 2023. 115

GORMLEY, M. *et al.* Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. *British Dental Journal*, Nature Publishing Group UK London, v. 233, n. 9, p. 780–786, 2022. 1

GOYAL, N. *et al.* Risk factors for head and neck cancer in more and less developed countries: Analysis from the inhance consortium. *Oral Diseases*, Wiley Online Library, v. 29, n. 4, p. 1565–1578, 2023. 7

GRAMBSCH, P. M.; THERNEAU, T. M. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, Oxford University Press, v. 81, n. 3, p. 515–526, 1994. 36

GRANDIS, E.-N. A. C. J.; WHO, J. T. T. S. P. *WHO classification of head and neck tumours*. [S.l.]: IARC. Lyon, 2017. 6

GRIFFITHS-JONES, S. *et al.* mirbase: microrna sequences, targets and gene nomenclature. *Nucleic Acids Research*, Oxford University Press, v. 34, n. suppl\_1, p. D140–D144, 2006. 10

GU, Z. Complex heatmap visualization. *Imeta*, Wiley Online Library, v. 1, n. 3, p. e43, 2022. 45

GUO, M. *et al.* Epigenetic heterogeneity in cancer. *Biomarker Research*, BioMed Central, v. 7, n. 1, p. 1–19, 2019. 5

HAIR, J. F. *et al.* *Análisis multivariante*. [S.l.: s.n.], 2004. 23

\_\_\_\_\_. *Análise multivariada de dados*. [S.l.]: Bookman editora, 2009. 25, 26

- HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. [S.l.]: Morgan Kaufmann, 2012. 26, 27
- HARDIN, J. W. *et al.* *Generalized linear models and extensions*. [S.l.]: Stata Press, 2007. 16
- HARRELL, F. E. *et al.* Evaluating the yield of medical tests. *Jama*, American Medical Association, v. 247, n. 18, p. 2543–2546, 1982. 38
- HE, B. *et al.* Mir-133a-3p inhibits oral squamous cell carcinoma (oscc) proliferation and invasion by suppressing col1a1. *Journal of Cellular Biochemistry*, Wiley Online Library, v. 119, n. 1, p. 338–346, 2018. 116, 117
- HE, X.; ZHANG, J. Why do hubs tend to be essential in protein networks? *PLoS Genetics*, Public Library of Science San Francisco, USA, v. 2, n. 6, p. e88, 2006. 13
- HESKETH, R. *Introduction to Cancer Biology*. [S.l.]: Cambridge University Press, 2023. 4
- HILBE, J. M. *Modeling count data*. [S.l.]: Cambridge University Press, 2014. 16
- HONGYU, K.; SANDANIELO, V. L. M.; JUNIOR, G. J. de O. Análise de componentes principais: resumo teórico, aplicação e interpretação. *ES Engineering and Science*, v. 5, n. 1, p. 83–90, 2016. 23
- HORN, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrika*, Springer, v. 30, p. 179–185, 1965. 24
- HOTHORN, T. *et al.* Package 'coin'. 2017. Disponível em: <<http://coin.r-forge.r-project.org>>. Acesso em: 24 junho 2021. 44
- HOTHORN, T.; HORNIK, K.; ZEILEIS, A. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 15, n. 3, p. 651–674, 2006. 28
- HOTHORN, T.; ZEILEIS, A.; HOTHORN, M. T. Package 'partykit'. 2023. 48, 50
- HOU, L. *et al.* Collagen type 1 alpha 1 chain is a novel predictive biomarker of poor progression-free survival and chemoresistance in metastatic lung cancer. *Journal of Cancer*, Ivyspring International Publisher, v. 12, n. 19, p. 5723, 2021. 117
- HUBER, W. *et al.* Orchestrating high-throughput genomic analysis with bioconductor. *Nature methods*, Nature Publishing Group, v. 12, n. 2, p. 115–121, 2015. 40, 43
- HUSSON, F. *et al.* Package 'factominer'. *An R package*, v. 96, p. 698, 2016. 44
- INCA. *Brasil - estimativa dos casos novos: Estimativas para o ano de 2023*. Rio de Janeiro: [s.n.], 2022. Disponível em: <<https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/estimativa/estado-capital/brasil>>. 1, 6
- IRIMIE-AGHIORGHIESEI, A. I. *et al.* Prognostic value of mir-21: an updated meta-analysis in head and neck squamous cell carcinoma (hnscc). *Journal of Clinical Medicine*, MDPI, v. 8, n. 12, p. 2041, 2019. 11
- ISHWARAN, H. *et al.* Random survival forests. 2008. 37, 38

- ISHWARAN, H.; KOGALUR, U. B.; KOGALUR, M. U. B. Package ‘randomforestsrc’. *Breast*, v. 6, n. 1, p. 854, 2023. 50
- JAMAL-HANJANI, M. *et al.* Translational implications of tumor heterogeneity. *Clinical Cancer Research*, AACR, v. 21, n. 6, p. 1258–1266, 2015. 5, 8
- JANITZA, S.; HORNUNG, R. On the overestimation of random forest’s out-of-bag error. *PloS one*, Public Library of Science San Francisco, CA USA, v. 13, n. 8, p. e0201904, 2018. 28, 29
- JASKOWIAK, P. A.; COSTA, I. G.; CAMPELLO, R. J. Clustering of rna-seq samples: Comparison study on cancer data. *Methods*, Elsevier, v. 132, p. 42–49, 2018. 26
- JIANG, L.-h.; ZHANG, H.-d.; TANG, J.-h. Mir-30a: A novel biomarker and potential therapeutic target for cancer. *Journal of Oncology*, Hindawi, v. 2018, 2018. 118
- JOHNSON, D. E. *et al.* Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers*, Nature Publishing Group, v. 6, n. 1, p. 1–22, 2020. 1, 6
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. [S.l.]: Pearson London, UK, 2014. 23, 25, 26
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, Springer, v. 23, n. 3, p. 187–200, 1958. 24
- KALFERT, D. *et al.* Multifunctional roles of mir-34a in cancer: a review with the emphasis on head and neck squamous cell carcinoma and thyroid cancer with clinical implications. *Diagnostics*, MDPI, v. 10, n. 8, p. 563, 2020. 11
- KANTIDAKIS, G. *et al.* Survival prediction models since liver transplantation-comparisons between cox models and machine learning techniques. *BMC Medical Research Methodology*, Springer, v. 20, p. 1–14, 2020. 38
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. 35
- KASSAMBARA, A. *et al.* Drawing survival curves using ‘ggplot2’. *R package version 0.3*, v. 1, 2017. 49
- KE, J. *et al.* Mir-1-3p suppresses cell proliferation and invasion and targets stc2 in gastric cancer. *European Review for Medical & Pharmacological Sciences*, v. 23, n. 20, 2019. 116
- KHALID, S. *et al.* Predicting risk through artificial intelligence based on machine learning algorithms: A case of pakistani nonfinancial firms. *Complexity*, Hindawi, v. 2022, 2022. 27
- KHAN, M. Y. *et al.* Automated prediction of good dictionary examples (gdex): a comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques. *Complexity*, Hindawi Limited, v. 2021, p. 1–18, 2021. ix, 26, 28, 29
- KINGSFORD, C.; SALZBERG, S. L. What are decision trees? *Nature Biotechnology*, Nature Publishing Group, v. 26, n. 9, p. 1011–1013, 2008. 26
- KLEIN, J. D.; GRANDIS, J. R. The molecular pathogenesis of head and neck cancer. *Cancer Biology & Therapy*, Taylor & Francis, v. 9, n. 1, p. 1–7, 2010. 4

KLEINBAUM, D. G.; KLEIN, M. *et al.* *Survival analysis: a self-learning text*. [S.l.]: Springer, 2012. v. 3. 33, 34, 35

KONG, F. *et al.* Mir-381-3p suppresses biological characteristics of cancer in head-neck squamous cell carcinoma cells by targeting nuclear autoantigenic sperm protein (nasp). *Bioscience, Biotechnology, and Biochemistry*, Taylor & Francis, v. 84, n. 4, p. 703–713, 2020. 119

KONINGS, H. *et al.* A literature review of the potential diagnostic biomarkers of head and neck neoplasms. *Frontiers in Oncology*, Frontiers Media SA, v. 10, p. 1020, 2020. 7

KOTSIANTIS, S.; PINTELAS, P. Mixture of expert agents for handling imbalanced data sets. *Annals of Mathematics, Computing & Teleinformatics*, v. 1, n. 1, p. 46–55, 2003. 30

KOWALSKI, L. P. *et al.* Survival trends of patients with oral and oropharyngeal cancer treated at a cancer center in são paulo, brazil. *Clinics*, SciELO Brasil, v. 75, 2020. 7

KOZOMARA, A.; BIRGAOANU, M.; GRIFFITHS-JONES, S. mirbase: from microRNA sequences to function. *Nucleic Acids Research*, Oxford University Press, v. 47, n. D1, p. D155–D162, 2019. 10

KRZYWINSKI, M.; ALTMAN, N. Points of significance: Nonparametric tests. *Nature Methods*, v. 11, n. 5, p. 467–468, 2014. 21

KUHN, D. E. *et al.* Experimental validation of mirna targets. *Methods*, Elsevier, v. 44, n. 1, p. 47–54, 2008. 13

KUHN, M. *et al.* Package ‘caret’. *The R Journal*, v. 223, n. 7, 2020. 48

KUMAR, S.; PATNAIK, S.; DIXIT, A. Predictive models for stage and risk classification in head and neck squamous cell carcinoma (hnscc). *PeerJ*, PeerJ Inc., v. 8, p. e9656, 2020. 2

KUMAR, S.; VIJAYAN, M.; REDDY, P. H. MicroRNA-455-3p as a potential peripheral biomarker for alzheimer’s disease. *Human Molecular Genetics*, Oxford University Press, v. 26, n. 19, p. 3808–3822, 2017. 118

KUMAR, V. E. *et al.* Targeting epigenetic modifiers of tumor plasticity and cancer stem cell behavior. *Cells*, MDPI, v. 11, n. 9, p. 1403, 2022. 5

LANGMEAD, B.; HANSEN, K. D.; LEEK, J. T. Cloud-scale rna-sequencing differential expression analysis with myrna. *Genome Biology*, Springer, v. 11, n. 8, p. 1–11, 2010. 15

LE, T. T.; MOORE, J. H. treeheatr: an r package for interpretable decision tree visualizations. *Bioinformatics*, Oxford University Press, v. 37, n. 2, p. 282–284, 2021. 48

LEBLANC, M.; CROWLEY, J. Relative risk trees for censored survival data. *Biometrics*, JSTOR, p. 411–425, 1992. 37

\_\_\_\_\_. Survival trees by goodness of split. *Journal of the American Statistical Association*, Taylor & Francis, v. 88, n. 422, p. 457–467, 1993. 37

LEE, H. M. *et al.* Detection of increased serum mir-122-5p and mir-455-3p levels before the clinical diagnosis of liver cancer in people with type 2 diabetes. *Scientific Reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 23756, 2021. 118

- LEE, R. C.; FEINBAUM, R. L.; AMBROS, V. The c. elegans heterochronic gene lin-4 encodes small rnas with antisense complementarity to lin-14. *Cell*, Elsevier, v. 75, n. 5, p. 843–854, 1993. 9
- LEE, S.; LIM, H. Review of statistical methods for survival analysis using genomic data. *Genomics & Informatics*, Korea Genome Organization, v. 17, n. 4, 2019. 2, 34
- LEI, Y. *et al.* Applications of single-cell sequencing in cancer research: progress and perspectives. *Journal of Hematology & Oncology*, Springer, v. 14, n. 1, p. 91, 2021. 5
- LEMESHOW, S.; MAY, S.; HOSMER, D. W. *Applied survival analysis: regression modeling of time-to-event data*. [S.l.]: John Wiley & Sons, 2011. 33
- LI, G. *et al.* High expression of collagen 1a2 promotes the proliferation and metastasis of esophageal cancer cells. *Annals of Translational Medicine*, AME Publications, v. 8, n. 24, 2020. 117
- LI, J. *et al.* The clinical significance of collagen family gene expression in esophageal squamous cell carcinoma. *PeerJ*, PeerJ Inc., v. 7, p. e7705, 2019. 117
- LI, R. *et al.* Gdcrnatools: an r/bioconductor package for integrative analysis of lncrna, mirna and mrna data in gdc. *Bioinformatics*, Oxford University Press, v. 34, n. 14, p. 2515–2517, 2018. 40, 43
- LI, W. Volcano plots in analyzing differential expressions with mrna microarrays. *Journal of Bioinformatics and Computational Biology*, World Scientific, v. 10, n. 06, p. 1231003, 2012. 18
- LI, X. *et al.* Integrated analysis of microrna (mirna) and mrna profiles reveals reduced correlation between microrna and target gene in cancer. *BioMed Research International*, Hindawi, v. 2018, 2018. 45
- \_\_\_\_\_. mir-34a-5p functions as a tumor suppressor in head and neck squamous cell cancer progression by targeting flotillin-2. *International Journal of Biological Sciences*, Ivyspring International Publisher, v. 17, n. 15, p. 4327, 2021. 11
- LI, Y. *et al.* Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biology*, Springer, v. 23, n. 1, p. 79, 2022. 44, 114, 117
- LI, Z. *et al.* Identifying key microrna signatures for neurodegenerative diseases with machine learning methods. *Frontiers in Genetics*, Frontiers, v. 13, p. 880997, 2022. 117
- LING, C. X.; LI, C. Data mining for direct marketing: Problems and solutions. In: *Kdd*. [S.l.: s.n.], 1998. v. 98, p. 73–79. 30
- LIU, J.; DANG, H.; WANG, X. W. The significance of intertumor and intratumor heterogeneity in liver cancer. *Experimental & Molecular Medicine*, Nature Publishing Group, v. 50, n. 1, p. e416–e416, 2018. 5
- LIU, S.; LIAO, G.; LI, G. Regulatory effects of coll1a1 on apoptosis induced by radiation in cervical cancer cells. *Cancer Cell International*, Springer, v. 17, p. 1–9, 2017. 117

- LOPEZ, Y. O. N. *et al.* Characteristic mirna expression signature and random forest survival analysis identify potential cancer-driving mirnas in a broad range of head and neck squamous cell carcinoma subtypes. *Reports of Practical Oncology and Radiotherapy*, v. 23, n. 1, p. 6–20, 2018. [2](#), [10](#), [114](#)
- LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, BioMed Central, v. 15, n. 12, p. 1–21, 2014. [16](#), [17](#), [18](#), [19](#), [43](#)
- LU, L. *et al.* A novel seven-mirna prognostic model to predict overall survival in head and neck squamous cell carcinoma patients. *Molecular Medicine Reports*, Spandidos Publications, v. 20, n. 5, p. 4340–4348, 2019. [2](#), [10](#)
- LU, Y. *et al.* Hsa-mir-301a-3p acts as an oncogene in laryngeal squamous cell carcinoma via target regulation of smad4. *Journal of Cancer*, Ivyspring International Publisher, v. 6, n. 12, p. 1260, 2015. [115](#)
- LUNARDON, N.; MENARDI, G.; TORELLI, N. Rose: a package for binary imbalanced learning. *R Journal*, v. 6, n. 1, 2014. [30](#), [48](#)
- LUO, K. *et al.* Identification of critical mirnas as novel diagnostic markers for laryngeal squamous cell carcinoma. *Disease Markers*, Hindawi, v. 2022, 2022. [46](#), [115](#)
- LYDIATT, W. M. *et al.* Head and neck cancers—major changes in the american joint committee on cancer eighth edition cancer staging manual. *CA: a Cancer Journal for Clinicians*, Wiley Online Library, v. 67, n. 2, p. 122–137, 2017. [8](#)
- MA, J. *et al.* Prediction of the mechanism of mirnas in laryngeal squamous cell carcinoma based on the mirna-mrna regulatory network. *PeerJ*, PeerJ Inc., v. 9, p. e12075, 2021. [116](#)
- MAIMON, O. Z.; ROKACH, L. *Data mining with decision trees: theory and applications*. [S.l.]: World scientific, 2014. v. 81. [33](#)
- MALENOVÁ, G.; ROWSON, D.; BOEVA, V. Exploring pathway-based group lasso exploring pathway-based group lasso for cancer survival analysis: a special case of multi-task learning. *Frontiers in Genetics*, Frontiers, p. 2276, 2021. [33](#)
- MANLY, B. F.; ALBERTO, J. A. N. *Multivariate statistical methods: a primer*. [S.l.]: Chapman and Hall/CRC, 2016. [23](#)
- MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, JSTOR, p. 50–60, 1947. [21](#)
- MANTEL, N. *et al.* Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, v. 50, n. 3, p. 163–170, 1966. [35](#)
- MARIONI, J. C. *et al.* Rna-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, Cold Spring Harbor Lab, v. 18, n. 9, p. 1509–1517, 2008. [17](#)
- MARTINEZ, B.; PELOW, P. V. Micrnas as potential biomarkers for diagnosis of attention deficit hyperactivity disorder. *Neural Regeneration Research*, Medknow, 2023. [118](#)

- MARZBAN, C. The roc curve and the area under it as performance measures. *Weather and Forecasting*, American Meteorological Society, v. 19, n. 6, p. 1106–1114, 2004. 31
- MCDERMAID, A. *et al.* Interpretation of differential gene expression results of rna-seq data: review and integration. *Briefings in Bioinformatics*, Oxford University Press, v. 20, n. 6, p. 2044–2054, 2019. 2, 12
- MCHUGH, M. L. The chi-square test of independence. *Biochemia Medica*, Medicinska naklada, v. 23, n. 2, p. 143–149, 2013. 22
- MENARD, S. *Applied logistic regression analysis*. [S.l.]: Sage, 2002. 14
- MINGOTI, S. A. Análise de dados através de métodos estatística multivariada: uma abordagem aplicada. In: *Análise de dados através de métodos estatística multivariada: uma abordagem aplicada*. [S.l.: s.n.], 2007. p. 295–295. 23, 26
- MIRZA, B. *et al.* Machine learning and integrative analysis of biomedical big data. *Genes*, MDPI, v. 10, n. 2, p. 87, 2019. 2
- MODY, M. D. *et al.* Head and neck cancer. *The Lancet*, Elsevier, v. 398, n. 10318, p. 2289–2299, 2021. 1, 6
- MONARD, M. C.; BARANAUSKAS, J. A. Indução de regras e árvores de decisão. *Sistemas Inteligentes-fundamentos e Aplicações*, Editora Manole Ltda, v. 1, p. 115–139, 2003. 28
- MONROY, L. G. D.; RIVERA, M. A. M.; DÁVILA, L. R. L. *Análisis estadístico de datos categóricos*. [S.l.]: Universidad Nacional de Colombia, 2018. 15
- NAIR, J. *et al.* Gene and mirna expression changes in squamous cell carcinoma of larynx and hypopharynx. *Genes & cancer*, Impact Journals, LLC, v. 6, n. 7-8, p. 328, 2015. 2
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972. 14
- NETWORK, C. G. A. *et al.* Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, Nature Publishing Group, v. 517, n. 7536, p. 576, 2015. 1, 11
- NIGRO, C. L. *et al.* Head and neck cancer: improving outcomes with a multidisciplinary approach. *Cancer Management and Research*, Dove Press, v. 9, p. 363, 2017. 8
- O'BRIEN, J. *et al.* Overview of microrna biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology*, Frontiers Media SA, v. 9, p. 402, 2018. 12
- OSHIMA, S. *et al.* Identification of tumor suppressive genes regulated by mir-31-5p and mir-31-3p in head and neck squamous cell carcinoma. *International Journal of Molecular Sciences*, MDPI, v. 22, n. 12, p. 6199, 2021. 115
- OTASEK, D. *et al.* Cytoscape automation: empowering workflow-based network analysis. *Genome Biology*, Springer, v. 20, p. 1–15, 2019. 47
- OUELLETTE, J. N. *et al.* Navigating the collagen jungle: the biomedical potential of fiber organization in cancer. *Bioengineering*, MDPI, v. 8, n. 2, p. 17, 2021. 116

- PANVONGSA, W. *et al.* Plasma extracellular vesicle microrna-491-5p as diagnostic and prognostic marker for head and neck squamous cell carcinoma. *Cancer Science*, Wiley Online Library, v. 112, n. 10, p. 4257–4269, 2021. 2, 10, 114
- PASI, F. *et al.* Radiation-induced circulating mirna expression in blood of head and neck cancer patients. *Radiation and Environmental Biophysics*, Springer, p. 1–8, 2020. 1
- PEARSON, K. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Taylor & Francis, v. 50, n. 302, p. 157–175, 1900. 22
- PECORINO, L. *Molecular biology of cancer: mechanisms, targets, and therapeutics*. [S.l.]: Oxford University Press, 2021. 4
- PENG, Y.; CROCE, C. M. The role of micrnas in human cancer. *Signal Transduction and Targeted Therapy*, Nature Publishing Group, v. 1, n. 1, p. 1–9, 2016. 8
- PETRONACCI, C. M. C. *et al.* mirnas expression of oral squamous cell carcinoma patients: Validation of two putative biomarkers. *Medicine*, Wolters Kluwer Health, v. 98, n. 13, 2019. 115
- PIOTROWSKI, I. *et al.* mirnas as biomarkers for diagnosing and predicting survival of head and neck squamous cell carcinoma patients. *Cancers*, Multidisciplinary Digital Publishing Institute, v. 13, n. 16, p. 3980, 2021. 2, 10, 115
- PORCHERI, C.; MITSIADIS, T. A. Notch in head and neck cancer. In: *Notch Signaling in Embryology and Cancer*. [S.l.]: Springer, 2020. p. 81–103. 6
- PROVINCE, M. A. 30 sequential methods of analysis for genome scans. Elsevier, 2001. 20
- PURAM, S. V.; ROCCO, J. W. Molecular aspects of head and neck cancer therapy. *Hematology/Oncology Clinics*, Elsevier, v. 29, n. 6, p. 971–992, 2015. 7
- QI, C.-L. *et al.* Integrated analysis of deregulation microrna expression in head and neck squamous cell carcinoma. *Medicine*, Wolters Kluwer Health, v. 100, n. 6, 2021. 2, 10, 115
- QIAN, H. *et al.* Identification of mirna biomarkers for stomach adenocarcinoma. *BMC Bioinformatics*, Springer, v. 23, n. 1, p. 181, 2022. 119
- QUINLAN, J. R. Induction of decision trees. *Machine Learning*, Springer, v. 1, n. 1, p. 81–106, 1986. 27
- \_\_\_\_\_. Program for machine learning. *C4.5*, Morgan Kaufmann Pub, 1993. 27
- QUIRICO, L.; ORSO, F. The power of micrnas as diagnostic and prognostic biomarkers in liquid biopsies. *Cancer Drug Resistance*, OAE Publishing Inc, v. 3, n. 2, p. 117, 2020. 5
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. 16, 40, 43, 44
- RANA, R.; SINGHAL, R. *et al.* Chi-square test and its application in hypothesis testing. *Journal of the Practice of Cardiovascular Sciences*, Medknow Publications, v. 1, n. 1, p. 69, 2015. 22
- REINHART, B. J. *et al.* The 21-nucleotide let-7 rna regulates developmental timing in *caenorhabditis elegans*. *Nature*, Nature Publishing Group, v. 403, n. 6772, p. 901–906, 2000. 10

- RENCHER, A. C. *Methods of Multivariate Analysis, 2nd edn.* New York: A John Wiley & Sons, 2002. v. 2. 23, 25, 26
- RIPLEY, B. *et al.* Package ‘mass’. *Cran R*, v. 538, p. 113–120, 2013. 49
- RITCHIE, M. E. *et al.* Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research*, Oxford Academic, v. 43, n. 7, p. e47–e47, 2015. 46
- ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, Oxford University Press, v. 26, n. 1, p. 139–140, 2010. 16, 17, 43, 44
- ROBINSON, M. D.; SMYTH, G. K. Small-sample estimation of negative binomial dispersion, with applications to sage data. *Biostatistics*, Oxford University Press, v. 9, n. 2, p. 321–332, 2008. 15
- RODRIGUES, R. M. *et al.* How pathological criteria can impact prognosis of tongue and floor of the mouth squamous cell carcinoma. *Journal of Applied Oral Science*, SciELO Brasil, v. 28, 2019. 7
- ROSA, G. J. d. M.; ROCHA, L. B. d.; FURLAN, L. R. Estudos de expressão gênica utilizando-se microarrays: delineamento, análise, e aplicações na pesquisa zootécnica. *Revista Brasileira de Zootecnia*, SciELO Brasil, v. 36, p. 186–209, 2007. 20
- SAITO, R. *et al.* mir-99a-5p as possible diagnostic and prognostic marker in patients with gastric cancer. *Journal of Surgical Research*, Elsevier, v. 250, p. 193–199, 2020. 118
- SALAHUDDIN, S. *et al.* Human immunodeficiency virus is associated with poor overall survival among patients with head and neck cancer. *Clinical Infectious Diseases*, Oxford University Press US, v. 76, n. 8, p. 1449–1458, 2023. 7
- SALINAS-RODRÍGUEZ, A.; MANRIQUE-ESPINOZA, B.; SOSA-RUBÍ, S. G. Análisis estadístico para datos de conteo: aplicaciones para el uso de los servicios de salud. *Salud pública de México*, SciELO Public Health, v. 51, p. 397–406, 2009. 15
- SCHOENFELD, D. Partial residuals for the proportional hazards regression model. *Biometrika*, Oxford University Press, v. 69, n. 1, p. 239–241, 1982. 37
- SHAW, R.; BEASLEY, N. Aetiology and risk factors for head and neck cancer: United kingdom national multidisciplinary guidelines. *The Journal of Laryngology & Otology*, Cambridge University Press, v. 130, n. S2, p. S9–S12, 2016. 1, 6
- SHEN, Y. *et al.* Col3a1: Potential prognostic predictor for head and neck cancer based on immune-microenvironment alternative splicing. *Cancer Medicine*, Wiley Online Library, v. 12, n. 4, p. 4882–4894, 2023. 117
- SHERAFATIAN, M.; ARJMAND, F. Decision tree-based classifiers for lung cancer diagnosis and subtyping using tcga mirna expression data. *Oncology Letters*, Spandidos Publications, v. 18, n. 2, p. 2125–2131, 2019. 117
- SIDDIQUI, I. A. *et al.* Resveratrol nanoformulation for cancer prevention and therapy. *Annals of the New York Academy of Sciences*, Wiley Online Library, v. 1348, n. 1, p. 20–31, 2015. 4
- SIEGEL, R. L. *et al.* Cancer statistics, 2021. *Ca Cancer J Clin*, v. 71, n. 1, p. 7–33, 2021. 6

- SIEGMUND, K. D.; LAIRD, P. W.; LAIRD-OFFRINGA, I. A. A comparison of cluster analysis methods using dna methylation data. *Bioinformatics*, Oxford University Press, v. 20, n. 12, p. 1896–1904, 2004. 26
- SMYTH, G. K.; YANG, Y. H.; SPEED, T. Statistical issues in cdna microarray data analysis. *Functional genomics: methods and protocols*, Springer, p. 111–136, 2003. 18
- SONESON, C.; DELORENZI, M. A comparison of methods for differential expression analysis of rna-seq data. *BMC Bioinformatics*, BioMed Central, v. 14, n. 1, p. 1–18, 2013. 12
- SUNG, H. *et al.* Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, Wiley Online Library, 2021. 1, 6
- TAHERKHANI, A. *et al.* Pathogenesis and prognosis of primary oral squamous cell carcinoma based on micrnas target genes: a systems biology approach. *Genomics & Informatics*, Korea Genome Organization, v. 20, n. 3, 2022. 115
- TANG, M. *et al.* Col3a1 and its related molecules as potential biomarkers in the development of human ewing’s sarcoma. *BioMed Research International*, Hindawi, v. 2021, 2021. 117
- TASTSOGLU, S. *et al.* Diana-mirpath v4. 0: expanding target-based mirna functional analysis in cell-type and tissue contexts. *Nucleic Acids Research*, Oxford University Press, p. gkad431, 2023. 46
- THABTAH, F. *et al.* Data imbalance in classification: Experimental evaluation. *Information Sciences*, Elsevier, v. 513, p. 429–441, 2020. 30
- THERNEAU, T. *et al.* rpart: Recursive partitioning and regression trees. 2022. 50
- THERNEAU, T. M.; LUMLEY, T. Survival analysis. *R Top Doc*, v. 128, n. 10, p. 28–33, 2015. 49
- THOMAIDOU, A. C. *et al.* Promising biomarkers in head and neck cancer: The most clinically important mirnas. *International Journal of Molecular Sciences*, MDPI, v. 23, n. 15, p. 8257, 2022. 11
- TORGO, L.; TORGO, M. L. Package ‘dmwr’. *Comprehensive R archive network*, Citeseer, 2013. 48
- TURASHVILI, G. *et al.* Novel prognostic and predictive microrna targets for triple-negative breast cancer. *The FASEB Journal*, Wiley Online Library, v. 32, n. 11, p. 5937–5954, 2018. 119
- TURKMAN, M. A. A.; SILVA, G. L. Modelos lineares generalizados-da teoria à prática. *Sociedade Portuguesa de Estatística*, Lisboa, 2000. 14, 15
- UKEY, S. *et al.* Study of microrna (mir-221-3p, mir-133a-3p, and mir-9-5p) expressions in oral submucous fibrosis and squamous cell carcinoma. *Indian Journal of Clinical Biochemistry*, Springer, v. 38, n. 1, p. 73–82, 2023. 115
- VAHABI, M.; BLANDINO, G.; AGOSTINO, S. D. Micrnas in head and neck squamous cell carcinoma: a possible challenge as biomarkers, determinants for the choice of therapy and targets for personalized molecular therapies. *Translational Cancer Research*, AME Publications, v. 10, n. 6, p. 3090, 2021. 1, 11

- VERHOEVEN, K. J.; SIMONSEN, K. L.; MCINTYRE, L. M. Implementing false discovery rate control: increasing your power. *Oikos*, Wiley Online Library, v. 108, n. 3, p. 643–647, 2005. 19
- VILLE, B. D. Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, Wiley Online Library, v. 5, n. 6, p. 448–455, 2013. 26
- WALAYAT, A.; YANG, M.; XIAO, D. Therapeutic implication of mirna in human disease. *Antisense Therapy*, IntechOpen, 2018. 9
- WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, v. 54, n. 3, p. 426–482, 1943. 18
- WALTERS, S. J. *What is a Cox model?* [S.l.]: Citeseer, 1999. 35
- WANG, C. *et al.* Mir-29c inhibits the metastasis of oral squamous cell carcinoma and promotes its cell cycle arrest by targeting serpinh1. *Annals of Translational Medicine*, AME Publications, v. 9, n. 18, 2021. 116
- WANG, H. *et al.* mir-101-3p and mir-199b-5p promote cell apoptosis in oral cancer by targeting bicc1. *Molecular and Cellular Probes*, Elsevier, v. 52, p. 101567, 2020. 118
- WANG, Y. *et al.* Mir-34a suppresses hnscc growth through modulating cell cycle arrest and senescence. *Neoplasma*, v. 64, n. 4, p. 543–553, 2017. 11
- WANG, Z. *et al.* Microrna-1-3p inhibits the proliferation and migration of oral squamous cell carcinoma cells by targeting dkk1. *Biochemistry and Cell Biology*, NRC Research Press, v. 96, n. 3, p. 355–364, 2018. 115
- WEI, D. *et al.* mir-29c-3p represses the angiogenesis of esophageal squamous cell carcinoma by targeting serpinh1 to regulate the wnt signaling pathway. *Acta Cirúrgica Brasileira*, SciELO Brasil, v. 38, p. e385223, 2023. 116
- WEI, M. *et al.* Predicting therapeutic responses in head and neck squamous cell carcinoma from tp53 mutation detected by cell-free dna. *Translational Cancer Research*, AME Publications, v. 12, n. 12, p. 3604, 2023. 11
- WEI, T. *et al.* Package ‘corrplot’. *Statistician*, v. 56, n. 316, p. e24, 2017. 45
- WEISS, C. N.; ITO, K. A macro view of micrnas: the discovery of micrnas and their role in hematopoiesis and hematologic disease. *International Review of Cell and Molecular Biology*, Elsevier, v. 334, p. 99–175, 2017. 9
- WICKHAM, H.; CHANG, W.; WICKHAM, M. H. Package ‘ggplot2’. *Create Elegant Data Visualisations using the Grammar of Graphics*, v. 2, n. 1, p. 1–189, 2016. 43, 47, 49, 50
- WIENKE, A. *Frailty models in survival analysis*. [S.l.]: Chapman and Hall/CRC, 2010. 33
- WILCOXON, F. *Individual comparisons by ranking methods*. *Biom Bull 1 (6): 80–83*. [S.l.]: JSTOR, 1945. 21
- WILLIAMS, G. *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. [S.l.]: Springer Science & Business Media, 2011. 50

- WONG, N. W. *et al.* Oncomir: an online resource for exploring pan-cancer microRNA dysregulation. *Bioinformatics*, Oxford University Press, v. 34, n. 4, p. 713–715, 2018. 45
- WU, F. *et al.* A prognostic model for oral squamous cell carcinoma using 7 genes related to tumor mutational burden. *BMC Oral Health*, Springer, v. 22, n. 1, p. 152, 2022. 117
- WU, T. *et al.* clusterprofiler 4.0: A universal enrichment tool for interpreting omics data. *The innovation*, Elsevier, v. 2, n. 3, 2021. 47
- XU, F. *et al.* dbdemic 3.0: functional exploration of differentially expressed mirnas in cancers of human and model organisms. *Genomics, Proteomics and Bioinformatics*, Oxford University Press, v. 20, n. 3, p. 446–454, 2022. xiv, 45, 46
- XU, J. *et al.* mir-30a-3p/igf2bp1 affect the proliferation and migration of lung adenocarcinoma cells. *J Clin Images Med Case Rep*, v. 4, n. 10, p. 2628, 2023. 119
- YAKA, H. *et al.* A comparison of machine learning algorithms for estimation of higher heating values of biomass and fossil fuels from ultimate analysis. *Fuel*, Elsevier, v. 320, p. 123971, 2022. 32
- YANG, C.-C. *et al.* Differential impact of age on survival in head and neck cancer according to classic cox regression and decision tree analysis. *Clinical Otolaryngology*, Wiley Online Library, v. 44, n. 3, p. 244–253, 2019. 7
- YANG, X. *et al.* Diagnostic value of strand-specific mi rna-101-3p and mi rna-101-5p for hepatocellular carcinoma and a bioinformatic analysis of their possible mechanism of action. *FEBS Open Bio*, Wiley Online Library, v. 8, n. 1, p. 64–84, 2018. 118
- YAO, L. *et al.* Analysis of transcriptome expression profiling data in oral leukoplakia and early and late-stage oral squamous cell carcinoma. *Oncology Letters*, Spandidos Publications, v. 25, n. 4, p. 1–11, 2023. 116
- YEUNG, K. Y.; RUZZO, W. L. Principal component analysis for clustering gene expression data. *Bioinformatics*, Oxford University Press, v. 17, n. 9, p. 763–774, 2001. 23
- YIN, Y. *et al.* mir-133a-3p suppresses cell proliferation, migration, and invasion and promotes apoptosis in esophageal squamous cell carcinoma. *Journal of Cellular Physiology*, Wiley Online Library, v. 234, n. 8, p. 12757–12770, 2019. 116
- YOON, A. J. *et al.* Prognostic value of mir-375 and mir-214-3p in early stage oral squamous cell carcinoma. *American Journal of Translational Research*, e-Century Publishing Corporation, v. 6, n. 5, p. 580, 2014. 115
- YU, C. *et al.* Identification of key genes and signaling pathways associated with the progression of gastric cancer. *Pathology & Oncology Research*, Springer, v. 26, p. 1903–1919, 2020. 117
- YU, N. *et al.* Identification of tumor suppressor mirnas by integrative mirna and mrna sequencing of matched tumor–normal samples in lung adenocarcinoma. *Molecular oncology*, Wiley Online Library, v. 13, n. 6, p. 1356–1368, 2019. 119
- YU, Y.-Z. *et al.* mir-381-3p suppresses breast cancer progression by inhibition of epithelial–mesenchymal transition. *World Journal of Surgical Oncology*, Springer, v. 19, p. 1–11, 2021. 119

- ZEILEIS, A.; HOTHORN, T.; HORNIK, K. Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 17, n. 2, p. 492–514, 2008. 28
- ZHANG, G. *et al.* mir-1-3p inhibits osteosarcoma cell proliferation and cell cycle progression while promoting cell apoptosis by targeting cdk14 to inactivate wnt/beta-catenin signaling. *Molecular Biotechnology*, Springer, p. 1–14, 2023. 116
- ZHANG, J. M. *et al.* An interpretable framework for clustering single-cell rna-seq datasets. *BMC Bioinformatics*, BioMed Central, v. 19, n. 1, p. 1–12, 2018. 26
- ZHANG, L. *et al.* Functional analysis of mir-767-5p during the progression of hepatocellular carcinoma and the clinical relevance of its dysregulation. *Histochemistry and Cell Biology*, Springer, v. 154, p. 231–243, 2020. 117
- ZHAO, X.; CUI, L. A robust six-mirna prognostic signature for head and neck squamous cell carcinoma. *Journal of Cellular Physiology*, Wiley Online Library, v. 235, n. 11, p. 8799–8811, 2020. 115
- ZHAO, X. *et al.* mir-101-3p negatively regulates inflammation in systemic lupus erythematosus via mapk1 targeting and inhibition of the nf- $\kappa$ b pathway. *Molecular Medicine Reports*, Spandidos Publications, v. 23, n. 5, p. 1–13, 2021. 118
- ZHUANG, Z. *et al.* Microrna-204-5p is a tumor suppressor and potential therapeutic target in head and neck squamous cell carcinoma. *Theranostics*, Ivyspring International Publisher, v. 10, n. 3, p. 1433, 2020. 119