

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 19/02/2022.

ANÁLISE ESTATÍSTICA BASEADA NA CONSTRUÇÃO DE
ÁRVORES DE CLASSIFICAÇÃO E REGRESSÃO DE MICRORNAS
DE PACIENTES COM ADENOCARCINOMA PULMONAR

Bethina da Rocha Camargo

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro – 2020

**ANÁLISE ESTATÍSTICA BASEADA NA CONSTRUÇÃO DE
ÁRVORES DE CLASSIFICAÇÃO E REGRESSÃO DE MICRORNAS
DE PACIENTES COM ADENOCARCINOMA PULMONAR**

Bethina da Rocha Camargo

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

Coorientadora: Prof.^a Dr.^a **Patrícia Pintor dos Reis**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro – 2020

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Camargo, Bethina da Rocha.

Análise estatística baseada na construção de árvores de classificação e regressão de microRNAs de pacientes com adenocarcinoma pulmonar / Bethina da Rocha Camargo. - Botucatu, 2020

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Rogério Antonio de Oliveira

Coorientador: Patrícia Pintor dos Reis

Capes: 10202080

1. Pulmões - Câncer. 2. Análise de sobrevivência (Biometria). 3. Adenocarcinoma. 4. MicroRNAs.

Palavras-chave: Análise de sobrevivência; Bioestatística; Biomarcadores; Câncer de pulmão; Genética.

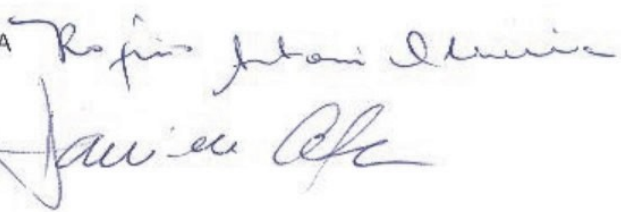
ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE BETHINA DA ROCHA CAMARGO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 19 dias do mês de fevereiro do ano de 2020, às 14:00 horas, no(a) Departamento de Bioestatística, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. ROGERIO ANTONIO DE OLIVEIRA - Orientador(a) do(a) Depto. de Bioestatística / IB/Botucatu - Unesp, Profa. Dra. HILDETE PRISCO PINHEIRO do(a) Departamento de Estatística / Universidade Estadual de Campinas - UNICAMP (Por meio de videoconferência), Profa. Dra. DANIELE CRISTINA CATANEO do(a) Departamento de Cirurgia e Ortopedia / Faculdade de Medicina de Botucatu - UNESP, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE MESTRADO de BETHINA DA ROCHA CAMARGO, intitulada **ANÁLISE ESTATÍSTICA BASEADA NA CONSTRUÇÃO DE ÁRVORES DE CLASSIFICAÇÃO E REGRESSÃO DE MICRORNAS DE PACIENTES COM ADENOCARCINOMA PULMONAR**. Após a exposição, a discente foi arguida oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: APROVADA. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof. Dr. ROGERIO ANTONIO DE OLIVEIRA

Profa. Dra. HILDETE PRISCO PINHEIRO

Profa. Dra. DANIELE CRISTINA CATANEO



Mas quando você orar, vá para seu quarto, feche a porta e ore a seu Pai, que está no secreto. Então seu Pai, que vê no secreto, o recompensará.

Mateus 6:6

Dedicatória

Dedico primeiramente a Deus que me guiou durante esta trajetória, a meus pais Jarbas e Marlene, meu irmão Benedito e meu namorado Felipe. Meu carinho e admiração por vocês é eterna, amo vocês.

Agradecimentos

Primeiramente agradeço a Deus, por ser essencial em minha vida, autor do meu destino, meu guia durante esta caminhada e socorro nos momentos difíceis.

Em especial ao meu pai Jarbas, minha mãe Marlene, meu irmão Benedito e meu namorado Felipe que dividem comigo todos os seus dias, são meus maiores exemplos de vida, inspiração e motivação. A vocês o meu muito obrigada, por me apoiarem e estarem ao meu lado ajudando na realização deste sonho.

Ao meu orientador Prof. Dr. Rogério Antonio de Oliveira e coorientadora Prof.^a Dr.^a Patrícia Pintor dos Reis pelo aprendizado, auxílio, motivação, dedicação e confiança.

À banca examinadora da dissertação Prof.^a Dr.^a Hildete P. Pinheiro e Prof.^a Dr.^a Daniele C. Cataneo pelas sugestões e comentários relevantes.

A todos os professores do departamento de Bioestatística UNESP/IBB. Em especial a Prof.^a Dr.^a Liciane e Prof.^a Dr.^a Miriam pelas contribuições na qualificação.

Em especial aos meus amigos Elizabete, Juliana, Gustavo, Eduardo e também, a todos os colegas de departamento pelos excelentes momentos que compartilhamos.

Aos Funcionários Arthur, Júnior e Luiz por sempre estarem dispostos a ajudar.

Ao Departamento de Bioestatística do IBB pelos ambientes adequados que possibilitaram a realização deste trabalho e ao Programa de Pós-Graduação em Biometria.

Agradeço apoio da Coordenação de Aperfeiçoamento Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Sumário

	Página
LISTA DE FIGURAS	ix
LISTA DE TABELAS	xi
RESUMO	xii
ABSTRACT	xiii
1 INTRODUÇÃO	1
1.1 Câncer de pulmão	2
1.2 MicroRNAs e o câncer	4
1.3 MicroRNAs e alvos moleculares terapêuticos	6
1.4 Árvore de classificação e regressão	7
2 OBJETIVOS	9
3 TESTES E MODELOS ESTATÍSTICOS	11
3.1 Teste de Wilcoxon-Mann-Whitney	11
3.2 Regressão logística	12
3.2.1 Teste qui-quadrado de Pearson	15
3.2.2 Estatística C	17
3.3 Análise de sobrevivência	17
3.3.1 Funções associadas ao tempo de sobrevivência	19
3.3.2 Relações entre as funções	20

	viii
3.3.3	Estimador de Kaplan Meier 21
3.3.4	Teste de log-rank 22
3.4	Construção de árvores de decisão 24
3.4.1	O método CART 26
3.4.2	Árvore de classificação 26
3.4.3	Árvore de regressão 28
3.4.4	Crescimento e poda 29
4	ANÁLISE DE DADOS E RESULTADOS 31
4.1	Análise exploratória 31
4.2	Teste de Wilcoxon-Mann-Whitney 35
4.3	Regressão logística 41
4.4	Análise do tempo de seguimento do paciente 43
4.5	Árvores de classificação e regressão 48
4.5.1	Árvore de classificação 48
4.5.2	Árvore de sobrevivência 54
4.6	Análise biológica 60
5	CONCLUSÕES 64
	REFERÊNCIAS BIBLIOGRÁFICAS 66

Lista de Figuras

	Página
1 Anatomia e função dos pulmões humano.	1
2 Ilustração de uma árvore de decisão.	25
3 Algoritmo de divisão do CART.	27
4 Distribuição da idade e dos dias até o último acompanhamento.	32
5 Boxplot da expressão dos microRNAs por tecido normal e tumoral para o teste de Wilcoxon-Mann-Whitney.	37
6 Boxplot da expressão dos microRNAs por tecido normal e tumoral para o teste de Wilcoxon-Mann-Whitney.	38
7 Boxplot da expressão dos microRNAs por tecido normal e tumoral para o teste de Wilcoxon-Mann-Whitney.	39
8 Boxplot da expressão dos microRNAs por tecido normal e tumoral para o teste de Wilcoxon-Mann-Whitney.	40
9 Boxplot da expressão do microRNA por tecido normal e tumoral para o teste de Wilcoxon-Mann-Whitney.	41
10 Boxplot da expressão dos microRNAs por tecido normal e tumoral para a regressão logística.	42
11 Boxplot da expressão dos microRNAs por tecido normal e tumoral para a regressão logística.	43
12 Curvas de sobrevivência estimadas por meio de Kaplan Meier para os grupos de estágio do tumor.	44
13 Curvas de sobrevivência estimadas por meio de Kaplan Meier para as etnias.	45

14	Curvas de sobrevivência por meio do estimador de Kaplan Meier para o miR-34b-3p e miR-185-3p.	46
15	Curvas de sobrevivência por meio do estimador de Kaplan Meier para o miR-1247-3p e miR-4661-5p.	47
16	Boxplot da expressão do microRNA por tecido normal e tumoral para a análise do tempo de seguimento do paciente.. . . .	48
17	Árvore de classificação sem poda, em que, as cores azul e cinza representam tecido tumoral e normal, respectivamente.	49
18	Gráfico do custo-complexidade, erro de validação cruzada e número de divisões para árvore de classificação.	51
19	Gráfico do número de divisões pelo erro de validação cruzada para árvore de classificação.	52
20	Árvore de classificação pós poda, em que, as cores azul e cinza representam tecido tumoral e normal, respectivamente.	53
21	Boxplot da expressão dos microRNAs por tecido normal e tumoral para a árvore de classificação pós poda.	54
22	Árvore de sobrevivência sem poda.	55
23	Gráfico do custo-complexidade, erro de validação cruzada e número de divisões para árvore de sobrevivência.	57
24	Gráfico do número de divisões pelo erro de validação cruzada para árvore de sobrevivência.	57
25	Árvore de sobrevivência pós poda. RR: Risco relativo.	58
26	Boxplot da expressão dos microRNAs por tecido normal e tumoral para a árvore de sobrevivência pós poda.	59
27	Boxplot da expressão do microRNA por tecido normal e tumoral para a árvore de sobrevivência pós poda.	60
28	Genes para a via molecular metabolismo celular do nicotinato e nicotinamida.	62

Lista de Tabelas

	Página
1 Tabela de contingência para o tempo t_j	23
2 Análises descritivas das variáveis qualitativas dos dados demográficos e características clínico-patológicas.	33
3 Análises descritivas das variáveis quantitativas dos dados demográficos e características clínico-patológicas.	33
4 Análises descritivas das variáveis qualitativas de acordo com o tecido e valor p	34
5 Análises descritivas das variáveis quantitativas de acordo com o tecido e valor p	35
6 Expressão dos microRNAs significativos por tecido normal ou tumoral e valor p	36
7 Estimativas dos parâmetros do modelo de regressão logística.	42
8 Resumo do modelo para a poda da árvore de classificação.	50
9 Resumo do modelo para a poda da árvore de sobrevivência.	56
10 Vias moleculares identificadas a partir dos genes-alvo de microRNAs associados a árvore de classificação.	61
11 Vias moleculares identificadas a partir dos genes-alvo de microRNAs associados a árvore de sobrevivência.	63

ANÁLISE ESTATÍSTICA BASEADA NA CONSTRUÇÃO DE ÁRVORES DE CLASSIFICAÇÃO E REGRESSÃO DE MICRORNAS DE PACIENTES COM ADENOCARCINOMA PULMONAR

Autora: BETHINA DA ROCHA CAMARGO

Orientador: Prof. Dr. ROGÉRIO ANTONIO DE OLIVEIRA

Coorientadora: Prof.^a Dr.^a PATRÍCIA PINTOR DOS REIS

RESUMO

O adenocarcinoma pulmonar é um problema mundial na saúde pública e representa uma das maiores causas de morte por câncer no mundo. Os microRNAs são grandes reguladores e têm sido propostos como biomarcadores em diversos tipos de cânceres. O objetivo desta pesquisa é encontrar possíveis microRNAs para melhorar a classificação dos tecidos (normal ou tumoral) e a sobrevivência dos pacientes. Utilizou-se o banco de dados do projeto Atlas do Genoma do Câncer (TCGA) para pacientes com adenocarcinoma pulmonar (LUAD). A análise estatística empregada foi baseada na construção de árvores de classificação, que encontrou o miR-21-5p, miR-133a-3p, miR-1287-3p e let-7g-3p estatisticamente significativos, e árvore de regressão para a sobrevivência dos pacientes, que encontrou o miR-887-3p, miR-1271-5p, miR-128-1-5p, miR-493-3p e miR-4999-5p estatisticamente significativos.

**STATISTICAL ANALYSIS BASED ON THE CONSTRUCTION OF
CLASSIFICATION TREES AND REGRESSION OF MICRORNAS
FROM PATIENTS WITH PULMONARY ADENOCARCINOMA**

Author: BETHINA DA ROCHA CAMARGO

Adviser: Prof. Dr. ROGÉRIO ANTONIO DE OLIVEIRA

Coadviser: Prof.^a Dr.^a PATRÍCIA PINTOR DOS REIS

ABSTRACT

Pulmonary adenocarcinoma is a worldwide public health problem and represents one of the biggest causes of cancer death in the world. MicroRNAs are strong regulators and have been proposed as biomarkers in several types of cancer. The objective of this research is to find possible microRNAs to improve the classification of tissues (normal or tumor) and patient survival. We used the database of the Atlas of Cancer Genome project (TCGA) for patients with pulmonary adenocarcinoma (LUAD). A statistical analysis employed was based on the construction of classification trees, which found miR-21-5p, miR-133a-3p, miR-1287-3p and let-7g-3p, with statistical classification, and regression tree for the situation of patients, who found miR-887-3p, miR-1271-5p, miR-128-1-5p, miR-493-3p and miR-4999-5p are statistically used.

1 INTRODUÇÃO

O pulmão é o órgão do sistema respiratório responsável pelas trocas gasosas entre o sangue e o ambiente. Sua principal função é oxigenar o sangue e eliminar o dióxido de carbono (CO_2) do corpo. Os seres humanos possuem dois pulmões que estão localizados na cavidade torácica e são cobertos por uma membrana protetora chama pleura. Cada pulmão, direito ou esquerdo, é dividido em partes ou lobos, o pulmão direito é maior e possui três lobos e o esquerdo possui apenas dois.

A Figura 1 ilustra a anatomia e a função dos pulmões humano, pode-se ressaltar que o órgão tem o formato piramidal, cor rosada e consistência esponjosa.

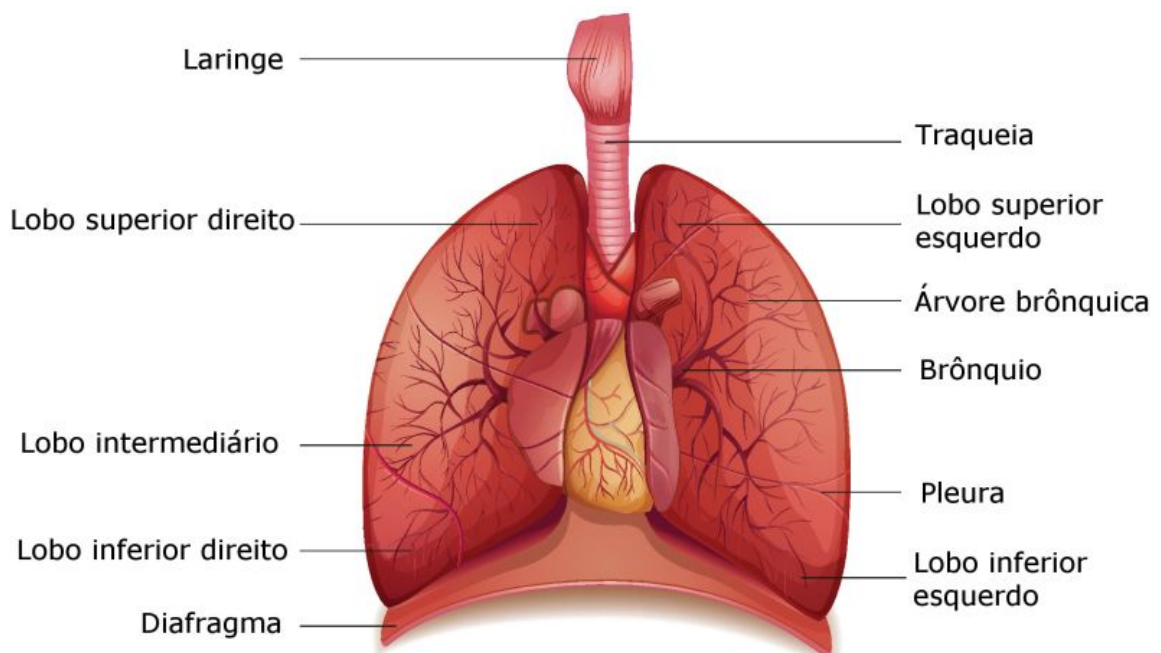


Figura 1: Anatomia e função dos pulmões humano.

Fonte: <https://www.infoescola.com/anatomia-humana/pulmoes/>

1.1 Câncer de pulmão

O câncer de pulmão é um dos cânceres mais populares e líder em mortalidade, com aproximadamente 1,8 milhões de casos novos e cerca de 1,6 milhões de óbitos por ano, mundialmente (Globocan, 2018). No Brasil, as estimativas de incidência para os anos de 2018 e 2019 apontaram o câncer de pulmão como o segundo mais frequente, com ocorrência de aproximadamente 31 mil casos novos, sendo a maioria em homens (18.740 homens e 12.530 mulheres). No sudeste do Brasil a incidência de carcinomas de pulmão em 2018 foi a terceira mais frequente com 19,22 casos a cada cem mil homens e 12,72 casos a cada cem mil mulheres (INCA, 2018).

O carcinoma pulmonar é geralmente diagnosticado em estágio avançado da doença, principalmente porque os sintomas não são percebidos nas fases iniciais de seu desenvolvimento, o diagnóstico tardio é apontado como responsável pela baixa sobrevivência dos pacientes, de 10% a 15% dos pacientes diagnosticados (INCA, 2018).

O tabagismo está relacionado como a principal causa de câncer de pulmão, na maioria da população os casos tabaco-relacionados representam cerca de 85% dos casos desse câncer (Ferlay et al., 2013). Além disso, há outros fatores de risco para este carcinoma, como a radiação e exposição à carcinógenos ocupacionais, como agentes químicos, e ambientais como, a poluição do ar, por exemplo (Schmid et al., 2010). Desta forma, esta doença pode ocorrer em indivíduos fumantes, ex-fumantes e não fumantes.

O câncer pulmonar é uma preocupante doença da saúde pública, assim deve ser priorizado a prevenção da doença e o desenvolvimento de melhores estratégias de tratamento.

Para decisões referentes a possíveis tratamentos do câncer de pulmão deve-se atentar aos subtipos histológicos¹ envolvidos neste tipo de carcinoma, que pode ser dividido em dois principais grupos: o carcinoma pulmonar de células pequenas (SCC, em inglês *small cell carcinoma*) e o carcinoma pulmonar de células não pe-

¹Estudos da estrutura microscópica, composição e função dos tecidos vivos.

quenas (NSCLC, em inglês *non-small cell lung carcinoma*). O primeiro ocorre em 15% dos pacientes, caracteriza-se por tumores mais agressivos e está associado a um prognóstico pobre com baixa sobrevivência. O segundo, é o tipo histológico mais comum, com 85% dos casos, é subdividido em vários subtipos histológicos, sendo o adenocarcinoma (AD, em inglês *adenocarcinoma*) o mais frequentemente diagnosticado (Herbst et al., 2008).

O AD é um problema mundial na saúde pública, representa um dos grandes motivos para as causas de morte por câncer no mundo, forma-se na periferia dos pulmões e é o subtipo mais encontrado de câncer de pulmão entre os não-fumantes, principalmente em mulheres (Chang et al., 2015).

Os pacientes diagnosticados com NSCLC podem fazer o tratamento da doença por meio de cirurgia, quimioterapia, radioterapia e, atualmente por meio de terapias com alvos moleculares, que nos últimos anos obtiveram vários marcadores moleculares e modelos propostos pelo mundo. Entretanto, devido à heterogeneidade observada nos pacientes com AD, o desenvolvimento de tratamentos aprimorados, como, por exemplo biomarcadores para o prognóstico da doença e para a prevenção do aparecimento da doença, ainda é um grande problema devido à complexidade biologia do câncer (Souza, 2016).

Estudos têm sido realizados com a intenção de identificar novas mutações ou alterações no carcinoma pulmonar e tais estudos têm identificado mutações em NSCLC. As alterações moleculares responsáveis pelo processo de formação do carcinoma pulmonar, especialmente o subtipo AD, ainda não são muito conhecidas. Considerando, então, que as terapias com alvos moleculares têm beneficiado poucos pacientes, ainda é necessária a identificação de novos alvos moleculares relacionados a este carcinoma (Chang et al., 2015).

O estudo sobre os microRNAs serve como base para o esclarecimento das causas de doenças e dos mecanismos que atuam sobre o organismo, também tem propiciado utilização para diagnóstico, prognóstico e terapia do câncer (Chamorro et al., 2016).

Nos últimos anos, confirmou-se que a desregulação na expressão dos microRNAs está relacionada à várias enfermidades em humanos, como problemas cardíacos e em cânceres.

A colaboração entre o Instituto Nacional do Câncer (NCI, em inglês *National Cancer Institute*) e o Instituto Nacional de Pesquisa do Genoma Humano (NHGRI, em inglês *National Human Genome Research Institute*) gerou a caracterização do perfil mutacional do pulmão, inclusive da expressão dos microRNAs. Nomeado de Atlas do Genoma do Câncer (TCGA, em inglês *The Cancer Genome Atlas*), é um banco de dados *online*, está publicamente disponível e tem sido vastamente utilizado pela comunidade de pesquisa, para realização de estudos sobre a desregularização da expressão dos microRNAs em vários tipos de cânceres.

1.2 MicroRNAs e o câncer

O primeiro microRNA encontrado em pesquisas foi por Lee et al. (1993). O microRNA lin-4, foi relacionado com o tempo de progresso do *nematóide Caenorhabditis elegans*, por meio da modulação da proteína lin-14.

Em seguida, definiu-se que sequências de pequenos RNAs, não codificadores, de 19 a 22 nucleotídeos de extensão estavam, também, expressas em outros tipos de organismos, como em *Homo sapiens*. Estas sequências foram nomeadas de microRNAs, são encontradas em diferentes espécies e cada espécie apresenta sua particularidade a respeito da expressão do microRNA (Iorio & Croce, 2009).

A nomenclatura do microRNA é composta pela espécie e o número do microRNA, nesta ordem, de acordo com a sua descoberta. Assim, o hsa-miR-21 refere-se ao microRNA 21 em *Homo sapiens*.

Encontrar a expressão dos microRNAs em cânceres humanos não criou expectativas para possíveis tratamentos até Calin et al. (2002) clonarem os miR-15a e miR-16-1, que inibiam a expressão de um gene anti apoptótico, em pacientes com leucemia linfocítica crônica. Os microRNAs agiam como supressores tumorais e estavam com a expressão diminuída em 70%.

Após este fato, outros estudos surgiram descrevendo diferentes perfis da expressão dos microRNAs em outras categorias de câncer, entre eles, o câncer de pulmão, cólon, próstata, entre outros. Além disso, também pode haver outras doenças relacionadas a diminuição ou aumento da expressão dos microRNAs. É importante o reconhecimento da função regulatória destes, para assim, obter clareza em seu papel no desenvolvimento das doenças (Liu et al., 2012).

Evidenciou-se a expressão alterada dos microRNAs ao comparar tecidos normais e tumorais e, também, em alguns cânceres específicos os microRNAs mostraram expressões diferentes nas fases da doença, ou seja, desde a formação do câncer até a metástase (Medina & Slack, 2008).

As expressões dos microRNAs podem ser alteradas em relação as variações do câncer, condições clínicas, estágio e perfil genético do paciente. Percebe-se, de um modo geral, que há uma redução significativa da expressão de microRNAs na presença do câncer, ainda assim, existem muitos microRNAs que podem estar superexpressos, ou seja, com a expressão do microRNA aumentada em relação as expressões consideradas normais. Cada tecido possui um valor referência para a expressão dos microRNA, logo a expressão do microRNA é capaz de estar aumentada ou diminuída em um tipo de câncer e o oposto em outro (Lee & Dutta, 2009).

Em alguns casos, o aparecimento e crescimento de tumores está relacionado a superexpressão dos microRNAs, tem-se como exemplos o miR-221 em glioblastomas², miR-372 e miR-373 em neoplasias do testículo e miR-221, miR-222 e miR-146 em tumores de tireoide (Lee & Dutta, 2009).

A literatura mostra que os microRNAs podem ser responsáveis pelos processos de invasão e migração de tecidos tumorais, caracterizando tumores malignos. Pesquisas encontraram superexpressão, do miR-21 em cânceres. O miR-21 é caracterizado por promover a mobilidade e invasão de tecidos tumorais, pois liga-se e bloqueia os genes supressores de tumor (Lee & Dutta, 2009).

Um estudo clínico sobre o papel do let-7 (let-7a ao let-7i) em câncer de

²Tipo mais comum e agressivo de tumor maligno cerebral que acomete os seres humanos.

pulmão, mostrou um diagnóstico ruim quando este está subexpresso, ou seja, com a expressão do microRNAs diminuída em relação as expressões consideradas normais, e a introdução do let-7 como um supressor tumoral (Takamizawa et al., 2004).

Outras pesquisas discutiram a importância da expressão dos microRNAs para o processo de formação do câncer e para o desenvolvimento do câncer de mama, já que os hormônios podem apresentar expressões diferentes nestas fases da doença (Nakano et al., 2013). Diferentes expressões de microRNAs têm sido retratadas ao comparar tecidos tumorais e normais em vários os tipos de cânceres (Yin et al., 2014).

Tsunoda et al. (2011) em sua pesquisa constatou que a inibição da morte celular programada está ligada com a superexpressão do miR-221 e miR-222 em pacientes com câncer colorretal.

Com a grande quantidade de informações a respeito dos microRNAs, surgiu a necessidade de organizar as informações em banco de dados. Hoje, existem vários bancos de dados virtuais que fornecem informações sobre os microRNAs. O miR-base, foi criado em 2002, representa o principal depósito online que contém todas as informações sobre os microRNAs publicados pelo mundo.

1.3 MicroRNAs e alvos moleculares terapêuticos

No decorrer dos anos, tem-se percebido crescimento nas pesquisas e no desenvolvimento de técnicas de biologia molecular, mostrando uma tendência em cada vez mais entender e desenvolver terapias-alvos específicas, com o objetivo de bloquear vias moleculares desreguladas nos tecidos tumorais.

Desde que os microRNAs foram descobertos, ampliou-se as possibilidades de tê-los como biomarcadores para o diagnóstico e prognóstico de cânceres, principalmente nas fases iniciais, em que não é possível detectar a doença com exames convencionais. Esses tratamentos têm como objetivo atuar seletivamente nas células tumorais que contém alterações moleculares (Cinegaglia, 2015).

A maioria dos estudos utilizam do soro para a identificação dos níveis de

expressão dos microRNAs, entretanto, outros estudos relatam que a identificação destes níveis pode ser realizada por meio da urina, saliva, suco pancreático ou fluido cístico. Encontrar os microRNAs nestes tipos de fluidos corporais são fundamentais para detecção de possíveis biomarcadores e representam uma forma não-invasiva para o diagnóstico da doença, principalmente em pacientes com a saúde mais fraca (Wang et al., 2014).

Algumas pesquisas a respeito de terapias baseadas em microRNAs possuem a intenção de restaurar ou inibir a função dos microRNAs usando microRNAs *mimics* ou *antagomirs*, que são como uma imitação dos microRNAs. Estudos *in vitro* e *in vivo* tem mostrado a eficácia do uso de microRNAs *mimics* contra o desenvolvimento tumoral (Ulivi et al., 2013). Porém pesquisas estão sendo desenvolvidas para definir a segurança na utilização dos microRNAs para este fim. Primeiramente são identificados os microRNAs relacionados ao tipo de câncer e posteriormente estudadas as necessidades biológicas para a ação.

A compreensão sobre os processos associados a formação do câncer humano, a caracterização do perfil genético, as alterações na expressão e mutações de microRNAs em adenocarcinoma, poderá contribuir para o desenvolvimento de biomarcadores com potencial aplicações clínicas. Biomarcadores moleculares serão úteis na evolução de estratégias para identificação precoce de doença, diagnóstico, prognóstico e terapêuticas mais precisas para o tratamento de pacientes.

1.4 Árvore de classificação e regressão

A aplicação de técnicas com segmentação ou agrupamentos na construção de árvores foi desenvolvida para solucionar os problemas complexos que envolvem conjunto de dados de dimensão elevada, composto por muitas variáveis qualitativas e quantitativas.

O método de árvores de classificação e regressão apareceu inicialmente no trabalho de Morgan & Sonquist (1963). Posteriormente nos trabalhos de Hunt et al. (1966), Messenger & Mandell (1972) e Morgan & Messenger (1973), que apresenta-

ram este método para problemas de classificação. Entretanto, as modificações tiveram uma grande contribuição para a popularização de árvores binárias em problemas de classificação por meio dos trabalhos realizados por Quinlan (1986) e também nos seus trabalhos dos anos 1979 e 1983, e na mesma época por Breiman et al. (1984).

Os problemas de regressão utilizaram a árvore primeiramente no trabalho de Morgan & Sonquist (1963) com o programa AID (*Automatic Interaction Detection*) e posteriormente apareceram generalizações como em Sonquist (1970), Sonquist et al. (1973) e em Van Eck (1980).

O algoritmo para árvores de classificação e regressão (CART, em inglês *Classification and Regression Trees*) foi apresentado por Breiman et al. (1984) e hoje está implementado em muitos programas estatísticos, como o *software* R. O algoritmo é uma técnica não-paramétrica que gera árvores de classificação ou de regressão, de acordo variável, qualitativa ou quantitativa, respectivamente.

O CART apresenta algumas vantagens em sua utilização, como: capacidade de encontrar relações entre os dados; os resultados são gerados no formato de árvore de decisão, implicando em facilidade, legibilidade e simplicidade com a interpretação dos resultados (Fonseca, 1994). As árvores de decisão que o algoritmo CART gera possuem respostas binárias e podem ser percorridas do nó pai até os nós finais, respondendo a perguntas com respostas de “sim” ou “não”.

Com o intuito de encontrar microRNAs como potenciais biomarcadores utiliza-se como método estatístico a construção de árvores de classificação e regressão, por ser uma técnica que gera resultados de fácil interpretação conforme descrito no Capítulo 3.

5 CONCLUSÕES

O banco de dados analisado neste trabalho continha dados demográficos, fatores de risco, características clínico-patológicas e 709 variáveis de microRNAs, para 508 amostras histopatológicas de tecidos normais e 45 tumorais.

O trabalho teve como objetivos específicos analisar o banco de dados por meio de quatro métodos estatísticos: teste de Mann-Whitney, regressão logística, análise de sobrevivência e árvores de classificação e regressão, o qual, foi realizado por meio do algoritmo CART. Depois de realizar as análises pode-se tirar algumas conclusões a respeito de cada método, como:

- Comparando os testes de regressão logística, que testa o efeito da probabilidade do tecido ser tumoral, e teste de Wilcoxon-Mann-Whitney, que compara as medianas em relação ao tecido normal as variáveis clinico-histopatológicas não apresentaram significância estatística. O teste Mann-Whitney apresentou 25 microRNAs estatisticamente significativos, a regressão logística apresentou 18 microRNAs, e destes tem-se 13 microRNAs concordaram (miR-143-3p, miR-144-5p, miR-20a-3p, miR-21-5p, miR-28-5p, miR-30a-5p, miR-320d, miR-33b-5p, miR-409-5p, miR-429, miR-486-5p, miR-490-3p e miR-629-3p).
- Na análise do tempo de seguimento do paciente encontrou-se as variáveis estagio do tumor, etnia, miR-34b-3p, miR-185-3p, miR-200b-5p, miR-331-5p, miR-1247-3p, miR-4661-5p e miR-5187-5p.
- Para a árvore de classificação restaram apenas 4 microRNAs: miR-21-5p, miR-133a-3p, miR-1287-3p e let-7g-3p. O miR-21-5p também aparece tanto no teste de Wilcoxon-Mann-Whitney quanto na regressão logística.

- A árvore de regressão para a sobrevivência pós poda apresentou 5 microRNAs: miR-887-3p, miR-1271-5p, miR-128-1-5p, miR-493-3p e miR-4999-5p. Na regressão logística apareceu o miR-493-5p, da família do miR-493.

Para árvore de sobrevivência sem poda apareceu o miR-4661-5p, que também apareceu no teste de Wilcoxon-Mann-Whitney e análise do tempo de seguimento, o miR-378a-3p, apareceu no teste de Wilcoxon-Mann-Whitney, e a família do miR-20, que apareceu o miR-20b-5p na árvore de sobrevivência sem poda e o miR-20a-3p no teste de Wilcoxon-Mann-Whitney e na regressão logística.

Nesse trabalho, foi encontrado-se o miR-21-5p superexpresso no tecido tumoral, da família dos miR-21. O miR-21 promove a mobilidade e a invasão de tecidos tumorais, ao se ligar e bloquear genes supressores (Lee & Dutta, 2009).

Takamizawa et al. (2004) mostram a influência das famílias do let-7 associadas na diminuição deste microRNAs com um pior prognóstico. E a introdução dele indica a potencialidade de supressor tumoral. No trabalho, o let-7g-3p também foi estatisticamente significativo no processo de discriminação dos tecidos histologicamente normais ou tumorais dos pacientes com o adenocarcinoma pulmonar, conforme resultado obtido pela construção da árvore de classificação.

A análise biológica foi feita com a intenção de averiguar se as vias moleculares que estes microRNAs regulam, estavam ligadas ao surgimento do câncer, pode-se observar que há ligação, ou seja, os microRNAs encontrados neste trabalho possuem ligações e podem ser potenciais biomarcadores para o adenocarcinoma pulmonar.

Baseando-se nos resultados encontrados nas análises estatísticas e nas análises biológicas, realizadas com o intuito de responder o objetivo principal do trabalho, sobre os microRNAs relacionados ao câncer pulmonar, pode-se concluir que os microRNAs encontrados possuem potencial para serem biomarcadores do adenocarcinoma pulmonar e, após a investigação mais profundas destes microRNAs, estes poderiam ser futuramente utilizados em terapêuticas biológicas no tratamento do câncer pulmonar.

REFERÊNCIAS BIBLIOGRÁFICAS

ARGENTON, J. L. P.; ET AL. Árvore de regressão para dados censurados e correlacionados, 2013.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. Classification and regression trees. *Wadsworth Int. Group*, v.37, n.15, p.237–251, 1984.

CALIN, G. A.; DUMITRU, C. D.; SHIMIZU, M.; BICHI, R.; ZUPO; ET AL. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the national academy of sciences*, v.99, n.24, p.15524–15529, 2002.

CHAMORRO, A. C.; ALBUQUERQUE, E. P. A.; DELLA ROSA, V. A. MICRO-RNAs E CÂNCER: ABORDAGENS E PERSPECTIVAS/MICRORNAs AND CANCER: APPROACHES AND PERSPECTIVES. *Revista de Pesquisa em Saúde*, v.16, n.2, 2016.

CHANG, J. T.-H.; LEE, Y. M.; HUANG, R. S. The impact of the Cancer Genome Atlas on lung cancer. *Translational Research*, v.166, n.6, p.568–585, 2015.

CHEN, J.; BARDES, E. E.; ARONOW, B. J.; JEGGA, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research*, v.37, n.suppl_2, p.W305–W311, 2009.

CIAMPI, A.; ZIGHED, D. A.; CLECH, J. Trees and induction graphs for multivariate response. In: Springer, 2000. *European Conference on Principles of Data Mining and Knowledge Discovery*; resumos. , 2000. 359–366.

CINEGAGLIA, N. D. C. MicroRNoma do carcinoma de pulmão de células não pequenas, 2015.

COLOSIMO, E.; GIOLO, S. Análise de Sobrevivência Aplicada. 1ª edição. **São Paulo: Editora Edgard Blücher**, 2006.

CORREA, A. P. A.; QUEIROZ, E.; TREVISAN, N. Teste Qui-Quadrado. **Universidade Federal do Paraná. Disponível em: http://www.leg.ufpr.br/lib/exe/fetch.php/disciplinas:ce001:teste_do_quiquadrado.pdf** Acesso em, 2006.

DIAS FILHO, J. M.; CORRAR, L.; PAULO, E. Análise multivariada. **São Paulo: Atlas**, 2007.

FERLAY, J.; SOERJOMATARAM, I.; ERVIK, M.; DIKSHIT, R.; ESER, S.; MATHERS, C.; REBELO, M.; PARKIN, D. M.; FORMAN; ET AL. GLOBOCAN 2012 v1. 0, cancer incidence and mortality worldwide: IARC CancerBase No. 11. **Lyon, France: International agency for research on cancer**, v.2016, 2013.

FIGUEIRA, C. V. Modelos de regressão logística. Porto Alegre, 2006. 138p. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul.

FONSECA, J. Indução de árvores de decisão, 1994. Tese (Doutorado) - Dissertação de Mestrado, Universidade Nova de Lisboa, Lisboa.

GARSON, D. PA 765 Statnotes: An online textbook. **<http://www2.chass.ncsu.edu/garson/pa765/statnote.htm>** [Accessed 28.2. 04], 2004.

GLOBOCAN. Lung Cancer. Estimated Incidence, Mortality and Prevalence Worldwide in 2012. <http://gco.iarc.fr/today/fact-sheets-populations>, 2018. Accessed: 2019-28-02.

GUO, C.; YO, S.; JANG, W. Evaluating predictive accuracy of survival models with PROC PHREG. In: SAS, 2018. ; resumos. , 2018.

HARRELL JR, F. The PHGLM procedure. SUGI Supplemental Library User's Guide. **SAS (Statistical Analysis System) Institute Inc., Version**, v.5, 1986.

HERBST, R. S.; HEYMACH, J. V.; LIPPMAN, S. M. Lung Cancer. **New England Journal of Medicine**, v.359, n.13, p.1367–1380, 2008, PMID: 18815398.

HOSMER JR, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. John Wiley & Sons, 2013. 398v.

HOTHORN, T.; ZEILEIS, A. partykit: A Modular Toolkit for Recursive Partytioning in R. **Journal of Machine Learning Research**, v.16, n.118, p.3905–3909, 2015.

HUNT, E. B.; MARIN, J.; STONE, P. J. Experiments in induction, 1966.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2018: Incidência de Câncer no Brasil. Rio de Janeiro. <https://www.inca.gov.br/publicacoes/livros/estimativa-2018-incidencia-de-cancer-no-brasil>, 2018. Accessed: 2019-28-02.

IORIO, M. V.; CROCE, C. M. MicroRNAs in cancer: small molecules with a huge impact. **Journal of clinical oncology**, v.27, n.34, p.5848, 2009.

KALBFLEISCH, J. D.; PRENTICE, R. L. **The statistical analysis of failure time data**. John Wiley & Sons, 2011. 360v.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, v.53, p.457–481, 1958.

KASSAMBARA, A.; KOSINSKI, M. **survminer: Drawing Survival Curves using 'ggplot2'**, 2019. R package version 0.4.4.

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**, v.75, n.5, p.843–854, 1993.

LEE, Y. S.; DUTTA, A. MicroRNAs in cancer. **Annual Review of Pathological Mechanical Disease**, v.4, p.199–227, 2009.

LIU, B.; LI, J.; CAIRNS, M. J. Identifying miRNAs, targets and functions. **Briefings in bioinformatics**, v.15, n.1, p.1–19, 2012.

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**, p.50–60, 1947.

MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemother Rep**, v.50, p.163–170, 1966.

MEDINA, P. P.; SLACK, F. J. microRNAs and cancer: an overview. **Cell cycle**, v.7, n.16, p.2485–2492, 2008.

MESSENGER, R.; MANDELL, L. A modal search technique for predictive nominal scale multivariate analysis. **Journal of the American statistical association**, v.67, n.340, p.768–772, 1972.

MILBORROW, S. **rpart.plot: Plot “rpart”Models: An Enhanced Version of “plot.rpart”**, 2019. R package version 3.0.7.

MORGAN, J. N.; MESSENGER, R. C. THAID, a sequential analysis program for the analysis of nominal scale dependent variables, 1973.

MORGAN, J. N.; SONQUIST, J. A. Problems in the analysis of survey data, and a proposal. **Journal of the American statistical association**, v.58, 1963.

NAKANO, K.; MIKI, Y.; HATA, S.; EBATA, A.; TAKAGI, K.; MCNAMARA, K. M.; SAKURAI, M.; MASUDA, M.; HIRAKAWA, H.; ISHIDA, T.; ET AL. Identification of androgen-responsive microRNAs and androgen-related genes in breast cancer. **Anticancer research**, v.33, n.11, p.4811–4819, 2013.

NIKITOVIC, D.; BERDIAKI, A.; SPYRIDAKI, I.; KRASANAKIS, T.; TSATSAKIS, A.; TZANAKAKIS, G. N. Proteoglycans—biomarkers and targets in cancer therapy. **Frontiers in endocrinology**, v.9, p.69, 2018.

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**, v.50, 1900.

QUINLAN, J. R. Induction of decision trees. **Machine learning**, p.81–106, 1986.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018.

SCHMID, K.; KUWERT, T.; DREXLER, H. Radon in indoor spaces: an underestimated risk factor for lung cancer in environmental medicine. **Deutsches Arzteblatt international**, v.107, n.11, p.181, 2010.

SIEGEL, S.; CASTELLAN JR, N. J. **Estatística não-paramétrica para ciências do comportamento**. Artmed Editora, 1975.

SOBRAL, A. P. B. Previsão de carga horária—Uma nova abordagem por Árvore de Decisão, 2003. Tese (Doutorado) - PUC-Rio.

SONQUIST, J.; BAKER, E.; MORGAN, J. Searching for Structure (Institute for Social Research, University of Michigan). **Ann Arbor**, 1973.

SONQUIST, J. A. Multivariate model building, 1970.

SOUZA, C. D. P. Perfil de expressão de microRNAs e seus alvos moleculares em carcinoma pulmonar, 2016.

TAKAMIZAWA, J.; KONISHI, H.; YANAGISAWA, K.; TOMIDA, S.; OSADA, H.; ENDOH, H.; HARANO, T.; YATABE, Y.; NAGINO; ET AL. Reduced expression

of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. **Cancer research**, v.64, n.11, p.3753–3756, 2004.

THERNEAU, T.; ATKINSON, B. **rpart: Recursive Partitioning and Regression Trees**, 2018. R package version 4.1-13.

THERNEAU, T. M. **A Package for Survival Analysis in S**, 2015. Version 2.38.

TIMOFEEV, R. Classification and regression trees (CART). **Theory and Applications**, 2004.

TSUNODA, T.; TAKASHIMA, Y.; YOSHIDA, Y.; DOI, K.; TANAKA, Y.; FUJIMOTO, T.; MACHIDA, T.; OTA, T.; KOYANAGI, M.; KUROKI, M.; ET AL. Oncogenic KRAS regulates miR-200c and miR-221/222 in a 3D-specific manner in colorectal cancer cells. **Anticancer research**, v.31, n.7, p.2453–2459, 2011.

ULIVI, P.; FOSCHI, G.; MENGOZZI, M.; SCARPI, E.; SILVESTRINI, R.; AMADORI, D.; ZOLI, W. Peripheral blood miR-328 expression as a potential biomarker for the early diagnosis of NSCLC. **International journal of molecular sciences**, v.14, n.5, p.10332–10342, 2013.

VAN ECK, N. A. Statistical analysis and data management highlights of OSIRIS IV. **The American Statistician**, v.34, n.2, p.119–121, 1980.

WANG, J.; ZHANG, K.-Y.; LIU, S.-M.; SEN, S. Tumor-associated circulating microRNAs as biomarkers of cancer. **Molecules**, v.19, n.2, p.1912–1938, 2014.

WILCOXON, F. Individual comparisons by ranking methods. *Biom Bull*, 1945.

YIN, J.-Y.; DENG, Z.-Q.; LIU, F.-Q.; QIAN, J.; LIN; OTHER. Association between mir-24 and mir-378 in formalin-fixed paraffin-embedded tissues of breast cancer. **International journal of clinical and experimental pathology**, v.7, 2014.

YOUNG, L. On randomness in ordered sequences. **The Annals of Mathematical Statistics**, v.12, n.3, p.293–300, 1941.