

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-graduação em Bioinformática

Ana Paula de Abreu

**PROSPECÇÃO DE BIOMARCADORES PARA CÂNCER DE MAMA
SUBTIPO LUMINAL A EM ESTÁGIO INICIAL USANDO ÁLGEBRA
LINEAR.**

Belo Horizonte

2019

Ana Paula de Abreu

**PROSPECÇÃO DE BIOMARCADORES PARA CÂNCER DE MAMA
SUBTIPO LUMINAL A EM ESTÁGIO INICIAL USANDO ÁLGEBRA
LINEAR.**

Versão final

Dissertação apresentada ao Programa de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre.

Orientador Prof. Dr. Marcos Augusto dos Santos

Belo Horizonte

2019

43

Abreu, Ana Paula de.

Prospecção de biomarcadores para câncer de mama subtipo luminal A em estágio inicial usando álgebra linear [manuscrito] / Ana Paula de Abreu. – 2019.

40 f. : il. ; 29,5 cm.

Orientador Prof.(a). Dr. Marcos Augusto dos Santos. Coorientador: José Miguel Ortega.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Neoplasias da Mama. 3. Biomarcadores. 4. Mineração de Dados. 5. MicroRNAs. I. Santos, Marcos Augusto dos. II. Ortega, José Miguel. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004



ATA DA DEFESA DE DISSERTAÇÃO

Ana Paula de Abreu

66/2019
entrada
2º/2017
CPF:
064.694.706-02

Às quatorze horas do dia **22 de julho de 2019**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Prospecção de Biomarcadores para Câncer de Mama Subtipo Luminal A em estágio inicial usando Álgebra Linear**", requisito para obtenção do grau de Mestre em **Bioinformática**. Abrindo a sessão, o Presidente da Comissão, **Dr. Marcos Augusto dos Santos**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Dr. Marcos Augusto dos Santos	UFMG	274525106-4	Aprovada
Dr. José Miguel Ortega	UFMG	059501268-07	Aprovada
Dr. Bráulio Roberto Gonçalves Marinho Couto	UniBH	600783406-7	Aprovada
Dra. Francielly Moraes Rodrigues da Costa	UFMG	060178996-27	Aprovada
Dra. Susanne Facchin	IFMG	056899046-35	Aprovada

Pelas indicações, a candidata foi considerada: Aprovada
O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 22 de julho de 2019.

Dr. Marcos Augusto dos Santos - Orientador [Assinatura]

Dr. José Miguel Ortega - Coorientador [Assinatura]

Dr. Bráulio Roberto Gonçalves Marinho Couto [Assinatura]

Dra. Francielly Moraes Rodrigues da Costa [Assinatura]

Dra. Susanne Facchin [Assinatura]



"Prospecção de Biomarcadores para Câncer de Mama Subtipo Luminal A em estágio inicial usando Álgebra Linear"

Ana Paula de Abreu

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Prof. Marcos Augusto dos Santos - Orientador
UFMG

Prof. José Miguel Ortega - Coorientador
UFMG

Prof. Bráulio Roberto Gonçalves Marinho Couto
UniBH

Francielly Moraes Rodrigues da Costa
UFMG

Susanne Facchin
IFMG

Belo Horizonte, 22 de julho de 2019.

A Deus por tornar esse sonho possível.
“Lâmpada para os meus pés é a tua palavra e, luz
para os meus caminhos.” (SL. 119:105, Almeida
Revista e Atualizada).”

AGRADECIMENTOS

A Deus por sempre estar presente e provendo tudo o que foi necessário para a conclusão desse trabalho.

A minha família por permitir que meus sonhos fossem possíveis e por prover os meios para que sempre sejam concretizados. Pai, a você por não deixar que faltasse estímulo, perseverança, compreensão e companheirismo. Mãe, você por não deixar que faltasse amor, carinho, abdições e orações.

A meu orientador, Professor Doutor Marcos Augustos dos Santos por toda a paciência, empenho com que sempre me orientou neste trabalho e pelas boas conversas. Muito obrigada por me ter corrigido quando necessário sem nunca me desmotivar e pela amizade construída nessa caminhada.

A meu coorientador, Professor Doutor José Miguel Ortega pela disposição de sempre prover os recursos quando necessários.

Ao meu colega Lúcio Pacorri pelo auxílio e disponibilidade em sanar todas as dúvidas com prontidão e afinco.

Ao meu tio Antônio Carlos pelo carinho e prontidão em sempre ajudar. A Sabrina pela amizade inestimável e por sempre me motivar a continuar.

A funcionários da Secretaria do Programa de Pós-Graduação em Bioinformática Tiago e Sheila, pelo empenho em suas profissões e zelo conosco alunos.

RESUMO

O câncer de mama não é uma doença única, e sim uma doença heterogênea, com múltiplas formas histopatológicas, diferentes respostas à terapêutica e evoluções variáveis. Com efeito, sob o rótulo de câncer de mama parecem estar agrupados distintos tipos de neoplasia com origem nas células da mama.

Com o advento e o progresso da biotecnologia, desde 2001, uma nova classificação para os carcinomas mamários foi implementada à rotina de oncologistas e pacientes. Hoje é possível classificar molecularmente os diferentes tipos de câncer de mama por perfis de expressão gênica. Em casos em que o paciente tenha recebido um diagnóstico positivo para câncer de mama, há a necessidade de confirmação do resultado por método invasivo (biópsia) do tecido para garantir a confirmação diagnóstica oportuna e com qualidade. No entanto, a biópsia de tecido mamário, que hoje é adotada como padrão clássico de confirmação para câncer mamário, trata-se de um método invasivo e atualmente muitas das pesquisas realizadas sobre o câncer tem sido focado na busca por biomarcadores com o intuito de otimizar uma metodologia diagnóstica não invasiva e com elevada acurácia e que possa ser utilizada precocemente. Dentre os marcadores utilizados nessas metodologias modernas, destaca-se os miRNAs.

MicroRNAs (miRNAs) são uma classe de RNAs não-codificantes curtos que abrem um novo caminho para entender as respostas imunes e a progressão do câncer. Apesar das funções biológicas dos miRNAs se tornarem cada vez mais aparentes, a função dos miRNAs na regulação da resistência ao câncer de mama permanece sob investigação.

Essa dissertação tem por finalidade prospectar possíveis biomarcadores para câncer de mama usando novas ferramentas para agregar aos métodos de mineração de dados para detecção precoce do câncer de mama por meio de biópsia líquida usando perfis de conjuntos de dados através de um modelo de regressão. Este trabalho procura identificar possíveis biomarcadores que serão validados em uma estratégia de classificação. A partir de uma triagem obtida pela aplicação de uma metodologia que tem se mostrado efetiva. Portanto, tais marcadores são confrontados com o que existe sobre eles na literatura.

Os achados suportam a hipótese de que os perfis de expressão de miRNA circulante podem atuar como uma janela direta sobre a atividade tumoral e nos leva a questionar o mecanismo pelo qual eles entram na corrente sanguínea, além de seu papel funcional, se algum, na circulação periférica.

Palavras-Chave: Câncer de Mama, Biomarcadores, miRNAs, Mineração de Dados.

ABSTRACT

Breast cancer is not a single disease, but a heterogeneous disease with multiple histopathological forms, different responses to therapy and varying outcomes. In fact, under the label of breast cancer, different types of breast cell neoplasia appear to be grouped.

With the advent and progress of biotechnology since 2001, a new classification for breast carcinomas has been implemented in the routine of oncologists and patients. Today it is possible to molecularly classify the different types of breast cancer by gene expression profiles. In cases where the patient has been positively diagnosed with breast cancer, confirmation of the result by invasive tissue biopsy is necessary to ensure timely and quality diagnostic confirmation. However, breast tissue biopsy, which is now adopted as the classic confirmatory standard for breast cancer, is an invasive method and currently much of the cancer research has been focused on finding biomarkers to optimize a non-invasive diagnostic methodology with high accuracy that can be used early. Among the markers used in these modern methodologies, miRNAs stand out.

MicroRNAs (miRNAs) are a class of short noncoding RNAs that open a new pathway to understanding immune responses and cancer progression. Although the biological functions of miRNAs are becoming increasingly apparent, the role of miRNAs in regulating breast cancer resistance remains under investigation.

This dissertation aims to prospect possible biomarkers for breast cancer using new tools to aggregate data mining methods for early detection of breast cancer by liquid biopsy using data set profiles using a regression model. This paper seeks to identify possible biomarkers that will be validated in a classification strategy. From a screening obtained by applying a methodology that has been shown to be effective. Therefore, such markers are confronted with what exists about them in the literature.

The findings support the hypothesis that circulating miRNA expression profiles may act as a direct window on tumor activity and lead us to question the mechanism by which they enter the bloodstream, and their functional role, if any, in the circulation. peripheral.

Keywords: Breast Cancer, Biomarkers, miRNAs, Data Mining.

LISTA DE SÍMBOLOS E ABREVIATURAS

AKT	Protein kinase B serine/threonine-specific protein kinase
BCRP	Breast cancer resistance protein
CDK4	cyclin-dependent kinases
CYLD	Protein Ubiquitin carboxyl-terminal hydrolase
DNA	Deoxyribonucleic acid
DOX	Doxorubicin
FOXF2	Proteína Forkhead box F2
<i>GEO</i>	Gene Expression Omnibus
GRB2	Growth factor receptor-bound protein 2
HER- 2	Human Epidermal growth factor Receptor 2
HMGA1	Protein High mobility group AT-hook 1
hsa-mir	homo sapiens micro Ribonucleic acid
INCA	Instituto Nacional de Câncer José Alencar Gomes da Silva
Lum A	Luminal A
Lum B	Luminal B
MAPK	Mitogen Activated Protein Kinases - Proteíno-quinases ativadas por mitógenos
MCF-7	Células epiteliais de câncer de mama humano positivas para receptor de estrogênio derivadas de um carcinoma in situ (Michigan Cancer Foundation-7)
MDA-MB-231	Células triplo-negativas para receptor de estrogênio, receptor de progesterona e receptor do fator de crescimento epidermal humano 2 derivadas de um carcinoma metastático
miRNA (miR)	micro Ribonucleic acid
NCBI	National Center for Biotechnology Information
NF-KB	Fator nuclear kappa B
PI3K	Phosphoinositide 3-kinases
PIK3R2	subunidade reguladora da fosfoinositida-3-quinase 2
PTEN	Protein Phosphatase and tensin homolog
Ras	Protein RAt Sarcoma vírus, ou vírus do sarcoma de rato
RNA	Ribonucleic acid
SLN	metástase em linfonodos não-sentinela
SNAI1	Zinc finger protein codificada pelo gene SNAI1
SOX4	Transcription fator SOX (SRY-related HMG-box)
SVD	Singular Value Decomposition
TNBC	Triple Negative Breast Cancer
VEGF	Fator de Crescimento do Endotélio Vascular
VEGFA	Fator de Crescimento do Endotélio Vascular A

Sumário


1. INTRODUÇÃO.....	11
2. OBJETIVOS.....	14
3. MATERIAIS E MÉTODOS.....	15
3.1. Coleta e Geração de Dados.....	15
3.2. Modelo 1 (SVD)	15
3.3. Modelo de Regressão Logística Modificada.....	17
4. RESULTADOS E DISCUSSÃO.....	20
4.1. Modelo 1 (SVD)	20
4.2. Modelo de Regressão Logística Modificada.....	21
4.3 Avaliação dos miRNAs na literatura.....	25
4.3.1 Alfas Negativos.....	25
4.3.2 Alfas Positivos.....	26
5. CONCLUSÕES.....	29
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	32

1. INTRODUÇÃO

Nenhuma doença deveria assustar as mulheres tanto quanto o câncer de mama. Embora a probabilidade de uma mulher de 60 anos sofrer um ataque cardíaco ou doença cerebral seja muito maior, o fantasma do nódulo maligno no seio tem efeito desproporcional devastador por interferência à sexualidade e ao medo da morte (VARELLA, 2019).

Segundo o INCA, após os cânceres de pele não melanoma, o câncer de mama é o tipo mais frequente e a primeira causa de morte entre as mulheres. O câncer de mama tem letalidade relativamente baixa, uma vez que a taxa de mortalidade é menor que um terço da taxa de incidência. Nas estimativas divulgadas pelo INCA para cada ano do biênio 2018-2019, espera-se 59.700 novos casos de câncer de mama para o Brasil.

O câncer de mama não é uma doença única, e sim uma doença heterogênea, com múltiplas formas histopatológicas, diferentes respostas à terapêutica e evoluções variáveis (BARROS et al., 2015). Com efeito, sob o rótulo de câncer de mama parecem estar agrupados distintos tipos de neoplasia com origens celulares diversas, que se desenvolvem localizados nas mamas.

Localização Primária	Casos	%			Localização Primária	Casos	%
					Mama Feminina	59.700	29,5%
Próstata	68.220	31,7%			Traqueia, Brônquio e Pulmão	18.740	8,7%
Traqueia, Brônquio e Pulmão	18.740	8,7%			Cólon e Reto	17.380	8,1%
Cólon e Reto	17.380	8,1%			Colo do Útero	16.370	8,1%
Estômago	13.540	6,3%			Traqueia, Brônquio e Pulmão	12.530	6,2%
Cavidade Oral	11.200	5,2%			Glândula Tireoide	8.040	4,0%
Esôfago	8.240	3,8%			Estômago	7.750	3,8%
Bexiga	6.690	3,1%			Corpo do Útero	6.600	3,3%
Laringe	6.390	3,0%			Ovário	6.150	3,0%
Leucemias	5.940	2,8%			Sistema Nervoso Central	5.510	2,7%
Sistema Nervoso Central	5.810	2,7%	Leucemias	4.860	2,4%		

*Números arredondados para múltiplos de 10.

Figura 1: Distribuição proporcional dos dez tipos de câncer mais incidentes estimados para 2018 por sexo, exceto pele não melanoma*. Fonte: INCA, 2017

Com o advento e o progresso da biotecnologia, desde 2001 uma nova classificação para os carcinomas mamários foi implementada na rotina de oncologistas e pacientes, hoje é possível classificar molecularmente por perfis de expressão gênica, em: luminal A, luminal B, HER-2 negativo, luminal HER-2 positivo e triplo negativo.

Hoje a conduta médica adota para a triagem de casos assintomáticos é a realização da mamografia bienal para as mulheres na faixa etária de 50 a 69 anos. Segundo o Ministério

da Saúde, em 2015, no Brasil, foram gastos 2,5 milhões de reais em mamografias. Em casos em que alguma alteração morfológica é detectada neste exame de imagem, há a necessidade de confirmação do resultado por método invasivo (biópsia do tecido mamário) do tecido para garantir a confirmação diagnóstica oportuna e com qualidade.

Atualmente muitas das pesquisas realizadas sobre o câncer tem focado na busca por biomarcadores com o intuito de um diagnóstico precoce e não invasivo. Um estudo piloto de miRNAs circulantes realizado em 2010 por ZHAO e colaboradores., obteve resultados significativos com potencial para possíveis biomarcadores de câncer de mama em estágio inicial entre pacientes de diferentes etnias (caucasianos Americanos (CAs) e afro-americanos (AAs)). Descobriram que 31 miRNAs foram expressos diferencialmente entre os casos de CA e os controles de CA e 18 miRNAs foram expressos de maneira significativa entre os casos de e os controles de AA. Usando miRNAs expressos diferencialmente, casos de câncer de mama e controles saudáveis podem ser discriminados com especificidade e sensibilidade razoáveis. Curiosamente, existem apenas dois miRNAs diferencialmente expressos que se sobrepõem entre os sujeitos do estudo CA e AA, sugerindo possíveis diferenças raciais na expressão do miRNA circulante. Além do mais, foi realizada uma análise exploratória denominada *via in silico* (uso do algoritmo TargetScan) onde previram que esses miRNAs circulantes diferencialmente expressos podem afetar vias críticas conducentes à formação e progressão do câncer de mama, um mecanismo potencialmente importante que merece mais investigações (ZHAO et al., 2010).

Um biomarcador ideal deve ser amostrado de forma não invasiva, de fácil acesso (baixo custo e rapidez) e com alta sensibilidade e especificidade para a detecção precoce de tumores em quase todos os pacientes e ausentes ou minimamente presente em indivíduos saudáveis, livres de tumor. Uma das formas de obtenção seria através da biópsia líquida (sangue total, soro e plasma).

Um dos problemas enfrentados para a validação de miRNAs para fins de diagnóstico seria quanto as qualidades (sensibilidade, especificidade e reprodutibilidade) para uso isolado. O papel funcional dos miRNAs na circulação ainda não foi completamente elucidado; uma vez na circulação, o transporte de miRNA não é uniforme. Segundo OGAWA e colaboradores, alguns são encapsulados em microvesículas, corpos apoptóticos, exossomos ou partículas de lipoproteínas de alta densidade (HDL), enquanto outros estão em combinação com proteínas da família Argonaute (AGO).

MicroRNAs (miRNAs) são uma classe de RNAs não-codificantes curtos que abrem um novo caminho para entender as respostas imunes e a progressão do câncer. Apesar das

funções biológicas dos miRNAs se tornarem cada vez mais aparentes, a função dos miRNAs na regulação da resistência ao câncer de mama permanece sob investigação.

Os miRNAs, após secreção seletiva por células tumorais em microvesículas circulantes, estão presentes no plasma humano em uma forma notavelmente estável e protegida da atividade da RNase endógena (MITCHELL et al., 2008). Outra hipótese defendida, sugere que miRNAs possam ser detectáveis na circulação decorrente o vazamento passivo de células apoptóticas e necróticas (OGAWA et al., 2010).

Em um trabalho anterior Alysson dos Santos apresentou métodos eficientes de árvore *Full Logistic Regression* (FLR), *Reduced Logistic Regression* (RLR) e *Ad Hoc Logistic Regression* (ALR). para classificar o câncer de mama em fase inicial com resultados notáveis. Esse trabalho não apenas possibilitou a detecção do câncer de mama precoce em tecidos a partir do seu perfil miRNA, mas também foi possível classificá-los nos quatro subtipos moleculares possíveis (TNBC, Luminal A, Luminal B e HER-2). (SANTOS, 2017)

2. OBJETIVOS

Essa dissertação tem por finalidade prospectar possíveis biomarcadores para câncer de mama usando novas ferramentas para agregar aos métodos de mineração de dados para detecção precoce do câncer de mama por meio de biópsia líquida usando perfis de conjuntos de dados através de um modelo de regressão. Mais especificamente, a partir de única base de dados com os perfis de pacientes com câncer de mama em estágio inicial que está disponível no repositório <https://www.ncbi.nlm.nih.gov/> publicado em 24 de abril de 2013 com o registro da *Query DataSets* for GSE46355. Este trabalho procura identificar possíveis biomarcadores que serão validados em uma estratégia de classificação. A partir de uma triagem obtida pela aplicação de uma metodologia de regressão logística modificada que tem se mostrado efetivo (COSTA, F. M. R, 2017), pois o subconjunto selecionado é capaz de classificar os pacientes de forma eficiente. Portanto, tais marcadores são confrontados com o que existe sobre eles na literatura. Apresentaremos:

- A classificação do perfil de miRNAs GSE46355 utilizando todos os miRNAs.
- Possíveis miRNAs com potencial para biomarcadores para o câncer de mama em estágio inicial.
- Comparar nossos resultados com algumas referências em literatura.

3. MATERIAIS E MÉTODOS

3.1. Coleta e Geração de Dados

O conjunto de dados utilizado para demonstrar a utilidade da metodologia proposta abaixo foi obtido do repositório <https://www.ncbi.nlm.nih.gov/> publicado em 24 de abril de 2013 com o registro da Query DataSets for GSE46355. Essa base de dados foi utilizada para identificar microRNAs com expressão alterada em câncer de mama tipo Luminal A, com potencial para serem utilizados como biomarcadores pra identificação em biópsia líquida.

A base GSE46355, embora apresente um baixo volume de dados foi escolhida devido ao alto padrão de qualidade dos experimentos realizados para a obtenção dos miRNAs circulantes. O perfil elaborado propõe-se a identificar miRNAs circulantes que são diferencialmente expressos em mulheres com câncer de mama tipo Luminal A (ER positivo, PR positivo, HER2 / *neu* negativo) em comparação com pacientes saudáveis para investigar sua associação com parâmetros clínico patológico comuns, e estudar a sua eficácia como biomarcador diagnóstico sérico ou plasmático no ambiente clínico.

Usando um algoritmo de mineração de dados e o ambiente de prototipagem do programa MATLAB versão R2016a, realizamos a decomposição por valores singulares (para a análise de sistemas multivariáveis) capaz de mensurar o cosseno da distância entre dois pontos para comparação e classificação, e projetá-los graficamente em um espaço reduzido. para visualizar dos perfis dos pacientes. Como metodologia foi utilizado o modelo logístico, também conhecido como Regressão Logística ou Regressão Logit, utilizada para previsão de riscos, efeito classificatório e determinação de características.

3.2. Modelo 1 (SVD)

A técnica matemática da álgebra linear SVD (*Singular Value Decomposition*) pode ser aplicada a um termo documento matriz para encontrar documentos relevantes a partir de palavras de consulta usando máquina de busca no contexto de recuperação informacional, permitindo a análise da semântica latente (ou seja, oculta) em um documento contendo

palavras (Deerwester, et al., 1990). Em um vetor modelo espacial o SVD não mostrou a capacidade de estabelecer relações não notórias e agrupar itens relacionados (vetores), sendo assim podemos considerar um bom método determinístico.

Formalmente, o nexos do SVD nada mais é do que uma matriz A representada por um conjunto derivado de matrizes, da mesma forma que um número pode ser fatorado por exemplo, para um conjunto de números primos. Outra representação plausível é pensar em SVD como um conjunto de matrizes que fornecem representações numericamente diferentes de dados sem a perda do significado semântico, como a representação em diferentes números. Para entender o conceito matemático de SVD, suponha que A seja um conjunto de matrizes de números reais ou números complexos compostos de m , linhas por n , colunas. Uma nova matriz com uma decomposição de valor singular da matriz A pode ser realizada:

$$A=U\Sigma V^T \quad (9)$$

Sendo U uma matriz de m linhas por m colunas ortogonais e Σ é uma matriz diagonal $m \times n$, com números reais e não negativos. A matriz V^T é conhecida como transposição conjugada. Desde os valores diagonais de Σ são classificados em ordem decrescente Σ é uma função direta da matriz A e caracteriza os valores singulares dessa matriz, ordenando-os dos valores mais significativos para os menos significativos. Considerando um subconjunto de valores singulares de tamanho $k < n$, podemos obter A_k uma matriz aproximada de a matriz A :

$$A_k = U_k \Sigma_k V_k^T \quad (10)$$

Assim, a aproximação dos dados depende de quantos valores singulares são utilizados (10) Neste caso, o número de valores singulares k é também o rank da matriz A_k , indicando quantas linhas e quantas colunas na matriz A_k são linearmente independentes. A possibilidade de extrair informações com base em um menor número de parâmetros é parte da razão para o sucesso desta técnica, já que permite a compressão/descompressão de dados, com um tempo de execução que não aumenta exponencialmente com o aumento do tamanho da matriz, tornando a análise viável. Um conjunto de dados representado por um número menor de parâmetros do que o conjunto original tem uma tendência para agrupar itens que não seriam agrupados se usássemos o conjunto de dados original (9). Isso poderia explicar

por que clusters derivado de SVD podem expor relações não triviais existentes no conjunto de dados originais itens (11).

$$\begin{aligned}
 A_k &= U_k \Sigma K V K^T \\
 A_k &= U_k (\Sigma K V K^T) \\
 A_k &= U_k D K \qquad (11)
 \end{aligned}$$

A justificativa para usar somente o D_k é que ele tem k linhas ao invés de m linhas de A_k , logo D_k é composto de combinações lineares de colunas U_k , que por sua vez dá a relação $A \approx A_k$ que é representado por D_k .

3.3. Modelo de Regressão Logística Modificada

No modelo proposto não há a necessidade de uma poda variável, agregamos o termo quadrático à solução de um sistema de equações para determinar o valor dos parâmetros associados a α_i^* . Esses parâmetros estão relacionados com a expressão do miRNA. As variáveis que estão associadas à expressão gênica e não têm um papel discriminatório em nenhum dos modelos de classificação são indiretamente removidas do modelo, pois seus parâmetros associados α_i^* são zero ou próximos de zero. O símbolo α_i^* refere-se a valores específicos obtidos após a aplicação do novo modelo de regressão baseado em logística proposto e o símbolo α_i é usado antes da aplicação do modelo. O primordial para o desenvolvimento deste modelo é um termo estabilizador que permite a atribuição de valores aos parâmetros α_i^* , possibilitando que o sistema tenha uma solução única.

Após obtenção dos dados por experimentos de microarray, estes serão representados por uma matriz $\mathbf{A} = \{x_{i,j}\}$ com m linhas e n colunas, onde as linhas tem por finalidade representar os pacientes e as colunas representando os genes de miRNAs. O valor de cada posição $x_{i,j}$ representa os níveis de expressão de um determinado miRNA j para um paciente i . Iremos omitir a indicação da linha i nos elementos do vetor x . Isso é $x = \{x_1, x_2, \dots, x_n\}$ toda vez que a linha i à qual x se refere é clara no contexto. Associado a cada linha i é $P_i(x) = 0/1$ que informa a origem do perfil do gene (nenhum membro ou membro de um subtipo de câncer de mama).

A função logit expressa para cada paciente é dada por:

$$P_i(x) = g_i(x) / (1 + g_i(x)), \quad (1)$$

Onde

$$g_i(x) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n),$$

Para $i = 1, 2, \dots, m$ e \exp será uma função exponencial ($\exp(x) = e^x$). A regressão logística consiste em encontrar um vetor $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_n)^T$ para ajustar o conjunto de equações (1).

Observamos que quando $g_i(x)$ tende a zero, $P_i(x)$ também será zero. Se $g_i(x)$ tende ao infinito, $P_i(x)$ irá aproximar-se da unidade.

Sendo $P_i(x)$ a probabilidade, e $C_i(x)$ a chance de o evento ocorrer temos:

$$C_i(x) = P_i(x) / (1 - P_i(x)). \quad (2)$$

Substituindo a equação (2) em (1), teremos:

$$C_i(x) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n). \quad (3)$$

Para implementação do método, usaremos $\hat{C}_i(x) \approx C_i(x) = (0.99 / (1 - 0.99))$ apesar de $C_i(x)$ quando a chance está relacionada a $P_i(x) = 1$. Caso $P_i(x) = 0$, vamos considerar $\hat{C}_i(x) \approx C_i(x) = (0.01 / (1 - 0.01))$.

Considerando $b_i = \log(\hat{C}_i(x))$ e atribuindo logaritmo em ambos os lados (3), um modelo algébrico linear é criado para determinar α :

$$b_i = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n, \quad (4)$$

Para $i = 1, 2, \dots, m$. Seja $\bar{e} = (1, \dots, 1)^T$ um vetor de m e $b = [b_1, b_2, \dots, b_m]^T$. O sistema de equações lineares (4) será representado por:

$$\mathbf{B}\alpha = \mathbf{b}, \quad (5)$$

Com $\mathbf{B} = [\bar{e} \mathbf{A}]$.

O sistema (5) possui um número infinito de soluções, desde $n + 1 \gg m$. É comum contornar essa dificuldade, podendo o modelo e mantendo apenas um pequeno subconjunto dos n genes. Esse procedimento se assemelha à seleção de recursos na mineração de dados -

uma área de pesquisa aberta. Propomos a utilização de um termo estabilizador no modelo de regressão logística encontrado nos trabalhos de LINNIKL (1961) e GOLUB (1965) e posteriormente em ABREU SIQUEIRA CAIFFA (2009) e MENARD (2010) que permite a atribuição de valores a Parâmetros α minimizando a soma quadrada dos resíduos $(\mathbf{B}\alpha - \mathbf{b})$ somada aos quadrados de α , permitindo que o sistema tenha uma solução única. Então, para atribuir uma solução a (5), estamos considerando que:

$$\text{Minimizando } f(\alpha) = \alpha^T \alpha + (\mathbf{B}\alpha - \mathbf{b})^T (\mathbf{B}\alpha - \mathbf{b}). \quad (6)$$

Como $f(\alpha)$ é uma função convexa, o argumento α^* que minimiza (6) é dado pela derivada de $f(\alpha)$ onde α será igual a zero. Resultando no seguinte sistema de equações lineares:

$$(\mathbf{I} + \mathbf{B}^T \mathbf{B})\alpha = \mathbf{B}^T \mathbf{b}, \quad (7)$$

Sendo \mathbf{I} uma matriz de identidade de dimensão n . Observe que a matriz de identidade não permite que o sistema (7) tenha posto deficiente. A solução ótima α^* em (6) obtida através da solução de (7) é única. É atingido em α_i^* próximo a zero toda vez que a expressão do gene i for irrelevante para o cálculo da função *logit*. Logo, dada uma consulta query $q = [q_1, q_2, \dots, q_n]$ com os níveis de expressão de n genes, a probabilidade de q , estar relacionada a um câncer é dada por:

$$P(q) = g(q)/(1 + g(q)) \quad (8)$$

Onde

$$g(q) = \exp([1 \ q^T]\alpha).$$

4. RESULTADOS E DISCUSSÃO

4.1. Modelo 1 (SVD)

Como mencionado anteriormente, um termo documento típico matriz é muito grande e muitas vezes muito disperso, SVD atua como um método para reduzir a dimensionalidade deste espaço original e construir um subespaço sem grande perda do detalhamento. Com SVD, os recursos sobrevivendo com menor frequência ocorrem em um dado documento são excluídos do subespaço e, como tal, o “ruído” da matriz original é reduzido. Essa perspectiva nos levou a aplicar a SVD na classificação do câncer de mama na amostra do conjunto de dados GSE46355.

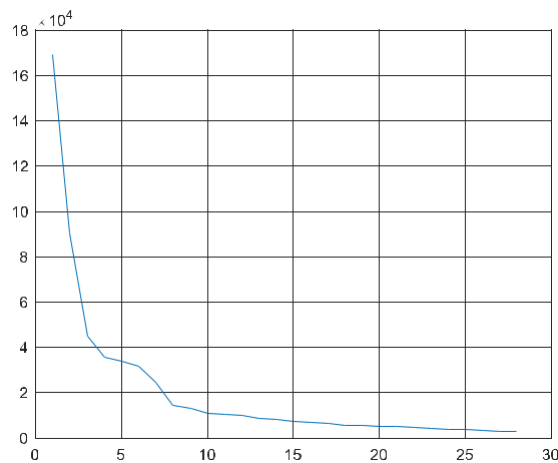


Gráfico 1: Valores Singulares não normalizados oriundos da aplicação da SVD à matriz de dados. Presença de cauda longa demonstra que não houve separação clara entre grupos.

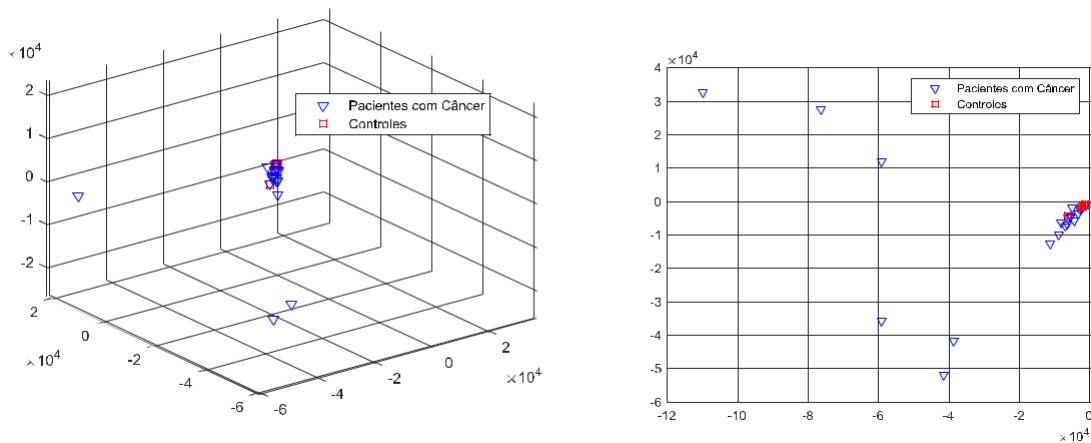


Figura 2: Visualização da amostra de Câncer de Mama do conjunto de dados GSE46355, após aplicação da técnica de *SVD* para redução de dimensionalidade. Vetores foram projetados no espaço D3 em (A) e no espaço D2 em (B) utilizando o método descrito por Marcolino et.al. (2010).

Para a amostra de mama não houve separação clara entre o subtipo Luminal A de câncer de mama e os pacientes saudáveis. A aplicação desta metodologia para a classificação do câncer de mama não produziu os resultados esperado como a separação em clusters ou subconjuntos. Presumimos que o fraco desempenho de classificação se deva à enorme semelhança entre as sondas. Quando o cosseno do ângulo formado entre cada vetor representando um indivíduo e qualquer outro vetor no conjunto de dados é computado, apontando para vetores muito semelhantes posicionados quase ao lado. (MARCOLINO et al., 2010; XU et al., 2011 e WU et al., 2015).

4.2. Modelo de Regressão Logística Modificada

Propomos um modelo baseado em regressão logística que desenvolvemos para classificar a amostra de câncer de mama com base em dados de expressão de microarray com todos os recursos incluídos e sem necessidade de redução da matriz de dados de microarray. Este modelo usa a função Logit como classificador (COSTA, F. M. R., 2017).

Para a aplicação do método de regressão logística, duas classes codificadas como 1 e 0 são atribuídas a cada amostra de paciente e uma computam as probabilidades que deram algumas variáveis explicativas que um paciente pertence à classe codificada. Aplicamos a metodologia de classificação proposta ao sistema criado a partir do conjunto de dados de

câncer de mama mencionado, com o objetivo de realizar uma classificação binária entre o subtipo Luminal A de câncer de mama e entre amostra dos pacientes saudáveis.

A metodologia apresentou excelente desempenho na tarefa de classificação. O gráfico 1 ilustra esse resultado para sistema criado a partir do conjunto de dados GS46355. Continuamente, quando um modelo é apresentado, o princípio de que menos é sempre mais é seguido e, portanto, a possibilidade de redução variável é frequentemente explorada (BAZZOLI e LAMBERT-LACROIX, 2018; LI et al., 2018; LIEN et al., 2018; et al., 2019).

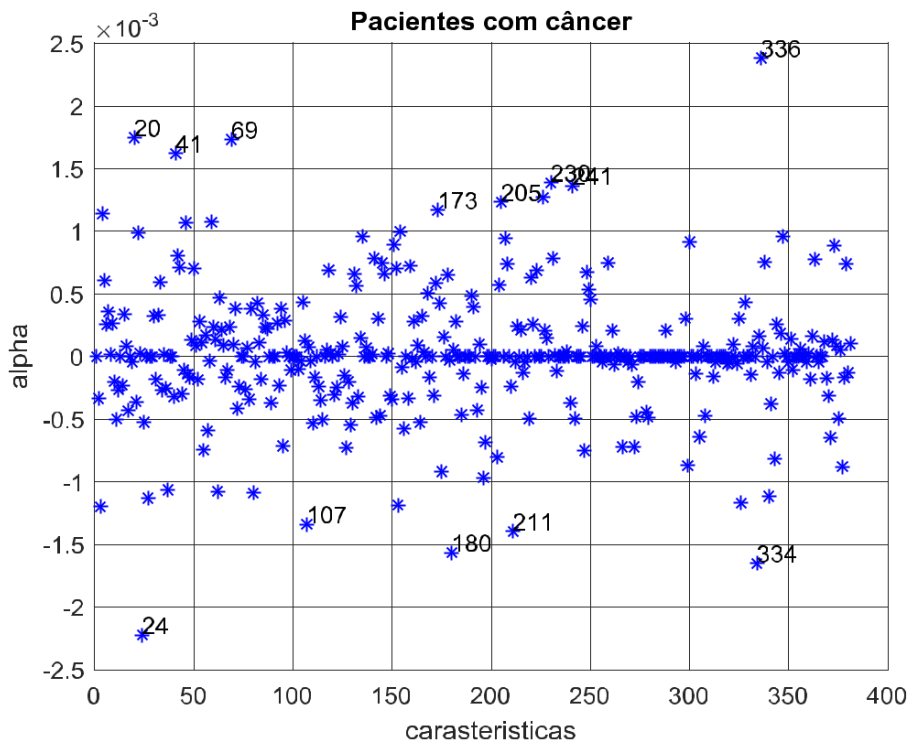


Gráfico 2: Valores dos pesos associados aos miRNAs. Observar que os pesos (α_i) positivos contribui para a ocorrência da doença. Já os negativos, são fatores de proteção (possivelmente). O ponto de corte foi realizado $\pm 1,2 \times 10^{-3}$.

Abaixo apresentamos os miRNAs correspondentes os pontos afixados no gráfico 2 para os 8 α positivos e os 5 α negativos após realizar o corte em $+1,2 \times 10^{-3}$ e $-1,2 \times 10^{-3}$ respectivamente.

Positivos	miRNAs	Posição no gráfico
	hsa-miR-124-	20
	hsa-miR-139-3p	41
	hsa-miR-182	69
	hsa-miR-411	205
	hsa-miR-483-5p	226
	hsa-miR-486-3p	230
	hsa-miR-494	241
	hsa-miR-628-5p	336
Negativos	miRNAs	Posição no gráfico
	hsa-miR-126	24
	hsa-miR-212	107
	hsa-miR-362-5p	180
	hsa-miR-429	211
	hsa-miR-625	334

Tabela 1: miRNAs (atributos) com os *ai* mais positivos e mais negativos do sistema criado a partir do conjunto de dados GS46355 do gráfico 2.

Todas as combinações possíveis de dados foram exploradas e o novo modelo proposto foi capaz de classificar todas as amostras em todos os conjuntos de dados, considerando todas as possíveis combinações de dados, com bom desempenho. Para avaliar o poder discriminatório do método proposto, também realizamos análise de sensibilidade e especificidade. Para todas as combinações possíveis de dados explorados, a faixa de valores de sensibilidade e especificidade é de 0,9.

	Média Harmônica	Sensibilidade Média	Especificidade Média	Área ROC Média das 100 áreas
Pacientes com Câncer	0,92	0,99	0.89	0.89

Tabela 2: Resultados da probabilidade de pacientes virem a desenvolverem câncer de mama Luminal A em estágio inicial para amostra após classificação no modelo proposto (valores médios após 100 folders).

A probabilidade de erro de classificação é a propriedade mais importante de um classificador porque quantifica a capacidade preditiva do classificador (BAZZOLI e LAMBERT-LACROIX, 2018; LI et al., 2018; DING et al., 2018; GÁLVEZ et al., 2018). ; PEREDA et al., 2018) A distribuição do rótulo do recurso é conhecida pelo conjunto de dados usado e, portanto, o erro verdadeiro pode ser encontrado com exatidão. Para uma rodada de taxa de classificação incorreta de validação cruzada, pegamos um subconjunto aleatório da amostra e aplicamos o novo modelo de regressão logística e o termo de estabilização para os conjuntos de dados encurtados e para os subconjuntos criados, para avaliar o desempenho da classificação.

Para reduzir a variabilidade, cem rodadas (folders) de validação cruzada foram realizadas. Os valores de corte foram definidos e, para estes, o modelo de regressão logística modificado classificou corretamente para todas as rodadas de pacientes extraídas. do conjunto de dado GSE46355. Isso também revela que a remoção de subconjuntos não afetou a estrutura organizacional da matriz.

4.3. Avaliação dos miRNAs na literatura.

Abaixo apresentamos as referências e suas correlações onde os miRNAs definidos por nosso sistema de classificação são citados.

4.3.1. Alfas Negativos

De acordo com os dados obtidos por TURGUT COSAN D et al., (2016), o miR-126 pode ser mais eficaz na inibição da metástase, pois segundo seus estudos houve uma diminuição na expressão de miR-126 quando as células MCF-7 e MDA-MB-231 foram transfectadas com o inibidor de miR-126. Além disso, a presença e ausência de miR-126 modulou as expressões gênicas da sinalização de VEGF / PI3K / AKT e MAPK em MCF-7 e MDA-MB-231

Já ZHU N et al., (2011) mostrou que o miR-126 pode ter como alvo tanto o VEGFA quanto o PIK3R2, e sua expressão foi diminuída (regulado negativamente) no câncer de mama humano, implicando que o miR-126 pode desempenhar um papel na gênese e crescimento do tumor e regulando a via de sinalização VEGF / PI3K / AKT.

Em seus estudos WANG CZ et al., (2015) revelou que o miR-126 desempenha um papel de supressão na invasão de células de câncer de mama humano através da repressão direta de uma desintegrina e metaloprotease 9 (ADAM9), via inibição da invasão celular pela diminuição da regulação de ADAM9 no desenvolvimento do câncer de mama.

Sobre o miR-212 DAMAVANDI Z et al., (2016) constatou que a deleção e perda de heterozigose no locus 17p13.3 foram relatadas em 49% dos cânceres de mama sendo consistente com a deleção frequente do locus 17p13.3 em células de tumor de mama, onde os dados de expressão gênica demonstraram uma significativa desregulação de miR-212 e miR-132 em tecidos de câncer de mama relacionado a tal fator.

Segundo estudos apresentados por XIE M et al., (2018) miR-132 / -212 eram comumente superexpressos em tumores e células de câncer de mama resistentes a DOX. A expressão de miR-132 / -212 silenciada induziu acumulação de DOX em células MCF-7 / ADR, enquanto a sobre-expressão de miR-132 / -212 levou a efluxo de DOX baseado em proteína de resistência ao cancro da mama (BCRP) em células MCF-7, seus achados

demonstraram um novo aspecto da via miR-132 / -212-PTEN-AKT / NF-κB-BCRP na geração de resistência ao câncer de mama e fornece um método potencial para reverter a resistência aos medicamentos.

Em seus estudos atuais HANIEH H et al., (2015) sugerem um novo mecanismo baseado em miRNA elucidando as propriedades anti-metastáticas dos agonistas de Ahr, sugerindo a possibilidade de usar o miR-212/132 para controlar metástases em pacientes com câncer de mama, através da supressão de SOX4.

Em estudo recente NI F et al., (2016) evidenciou que inibição do miR-362-5p demonstrou inibir significativamente a proliferação celular, migração e invasão de células MCF7 de câncer de mama humano; sugerindo que o miR-362-5p possui potencial para atuar como um novo alvo terapêutico no tratamento do câncer de mama e relata o gene supressor de tumor CYLD é um alvo direto do miR-362-5p, que a expressão ectópica do miR-362-5p reprime a expressão de CYLD, enquanto o tratamento com inibidor de miR-362-5p induz a expressão da proteína CYLD e diminui a expressão de NF-κB na via de sinalização a jusante.

Já WANG C et al., (2015) relatou que o miR-429 é supra regulado em duas linhagens de células TNBC (MDA-MB-231 e MDA-MB-468) e inibição do miR-429 pode resgatar parcialmente a apoptose induzida pelo δ-tocotrienol (Vitamina E) nas células MDA-MB-231. Identificou como um dos genes alvo do miR-429 como inibidor da proteína de apoptose ligado ao X (XIAP). Sugerindo que a ativação do miR-429 pelo δ-tocotrienol pode ser uma abordagem eficaz para a prevenção e tratamento do câncer de mama triplo negativo.

Em estudos realizados por ZHOU WB et al., (2016), seus dados demonstram que a inibição da expressão do miR-625 suprime a proliferação e migração celular, o HMGA1 foi confirmado como alvo direto do miR-625, isso sugere que o miR-625 pode ser um promissor biomarcador prognóstico e um potencial alvo terapêutico para o câncer de mama. A diminuição do miR-625 esteve intimamente associada ao receptor de estrogênio, ao receptor do fator de crescimento epidérmico humano 2 e ao estágio clínico.

4.3.2. Alfas Positivos

Em seus estudos, FENG T et al., (2015) demonstrou que miR-124 funciona como um miRNA supressor do crescimento e desempenha um papel importante na inibição da tumorigênese, visando o CDK4. Foi identificado e confirmado que a quinase 4 dependentes

de ciclina (CDK4) era um alvo direto do miR-124. A super expressão de miR-124 suprimiu a expressão da proteína CDK4 e atenuou a viabilidade celular, a proliferação e a progressão do ciclo celular em células de câncer de mama MCF-7 e MDA-MB-435S in vitro. A superexpressão de CDK4 resgatou parcialmente o efeito inibitório do miR-124 nas células do câncer de mama. Além disso, descobrimos que a super expressão de miR-124 efetivamente reprimiu o crescimento tumoral em experimentos com animais xenoenxertados.

Segundo CAI WL et al., (2018), regulação negativa do miR-124 foi associada a características clínicas agressivas e menor sobrevida livre de metástases ósseas e sobrevida global ratificando o estudo realizado por FENG T et al., (2015). A restauração de miR-124 suprimiu, enquanto a inibição de miR-124 promoveu a metástase óssea de células de câncer de mama in vivo. Assim, a identificação de um eixo desregulado miR-124/ IL-11 ajuda a elucidar os mecanismos das metástases do câncer de mama para o osso, revela novos marcadores prognósticos e facilita o desenvolvimento de novos alvos terapêuticos para tratar e até prevenir metástases ósseas do câncer de mama.

Em seus resultados YU J, SHEN W et al., (2017) indicaram que o miR-182 pode promover proliferação e migração celular em câncer de mama triplo negativo (TNBC), via regulação negativa do FOXF2, identificado como um alvo direto do miR-182. O miR-182 pode servir como um alvo potencial no tratamento com TNBC. A expressão de miR-182 foi significativamente maior nos tecidos TNBC do que nos tecidos normais combinados, e foi significativamente maior nas células MDA-MB-231 do que nas células MCF7. O *knockdown* miR-182 inibiu a proliferação e migração de células MDA-MB-231, enquanto a super expressão de miR-182 promoveu marcadamente a proliferação e migração de células MCF7.

Em seus achados ZHAN Y, LI X et al., (2017) descreveram o papel do miR-182 na colonização e macrometástases no câncer de mama pela primeira vez, e prometendo o diagnóstico ou a terapia da metástase do câncer de mama. Eles mostraram que o miR-182 exerceu sua função através do direcionamento de seu supressor SNAI1. Além disso, maior nível de expressão de miR-182 foi detectado em linfonodos metastáticos, comparado com tecidos primários pares de tumores. Além disso, o nível de expressão de miR-182 foi negativamente correlacionado com o nível de SNAI1 nesses espécimes clínicos.

ZHANG Y et al., (2016) relata que miR-411-5p (anteriormente chamado de miR-411) lançou luz sobre o mecanismo relacionado ao miR-411-5p na progressão do câncer de mama e, o eixo miR-411-5p / GRB2 / Ras é potencial para ser alvo molecular para a terapia do câncer de mama. Pois a expressão de miR-411-5p foi regulada negativamente em tecidos de câncer de mama em comparação com seus tecidos adjacentes não-neoplásicos adjacentes,

uma vez que foi investigado o alvo e o mecanismo do miR-411-5p no câncer de mama usando mimética e inibidor, e demonstramos o envolvimento da ativação de GRB2 e Ras. Expressão ectópica de miR-411-5p suprimiu a proliferação, migração e invasão de células de câncer de mama, enquanto a baixa expressão de miR-411-5p exibiu o efeito oposto. Além disso, demonstrou-se que a GRB2 estava significativamente super expressa em tecidos de câncer de mama em comparação com tecidos normais, e a baixa expressão de GRB2 teve uma sobrevida global mais longa em comparação com a alta expressão de GRB2 no câncer de mama.

Em seus estudos ZHANG M et al., (2016) para identificação de biomarcadores em destaque no câncer de mama com microarray de microRNAs, relatou que miR-483-5p miR-142-3p, e miR-483-3p são potenciais fatores-chave para o entendimento do mecanismo molecular do câncer de mama, afetando a função fisiológica normal da célula.

Na identificação de expressão de miRNA em linfonodos sentinela de pacientes com câncer de mama via sequenciamento de RNA para acurácia diagnóstica SOL D et al., (2019) em seus achados sugerem que há uma associação de miRNAs com metástase do linfonodo sentinela e também que os miRNAs funcionam como biomarcadores no que diz respeito à escolha da terapia e prognóstico da doença. Em seus resultados de RNA-seq revelaram que miR-1-3p e miR-486-3p foram reprimidos em pacientes com câncer de mama com metástase SLN.

Em sua conclusão, MACEDO T et al., (2017) apresentou resultados que suportam a hipótese de que miR-183 e miR-494 desempenham um papel fundamental na metástase do câncer de mama, e o miR-183 pode atuar como um oncogene, visando a proteína RB1 nas células MDA-MB-231. No presente estudo, a expressão de ambos os miRNAs foi avaliada nas linhagens celulares MDA-MB-231 e MDA-MB-468 BC. Ensaios funcionais revelaram que a super expressão de miR-183 ou miR-494 modulou a proliferação e migração de células *in vitro* em ambas as linhagens.

5. CONCLUSÕES

Embora a base GSE46355 referente ao artigo *Identification and Validation of Oncologic miRNA Biomarkers for Luminal A-like Breast Cancer* (McDermott AM, Miller N, Wall D, Martyn LM, Ball G, et al. 2014), apresente um baixo volume de dados, seu perfil e associação com parâmetros clínico patológico comuns, permite a obtenção de resultados mais sólidos quanto a sua eficácia como biomarcadores diagnósticos circulantes no ambiente clínico.

Como metodologia McDERMOTT (2014) e colaboradores obtiveram seus dados de 20 pacientes, sendo 10 com o diagnóstico de câncer de mama tipo Luminal A e 10 participantes de controle saudáveis. O perfil de expressão de miRNAs circulantes foi realizado para as 20 amostras, realizaram PCR quantitativo em tempo real para 667 miRNAs, utilizando cartões microfluídicos A e B, cada um contendo iniciadores e sondas para 381 miRNAs específicos mais 3 controles e termociclados em um instrumento Applied Biosystems 7900 HT.

A análise estatística foi realizada usando o Minitab versão 16.0 (Minitab Ltd, Coventry UK). O teste de normalidade de Kolmogorov-Smirnov foi realizado. Os dados foram transformados em log (log 10) para análise quando a distribuição não normal foi identificada. A significância e as associações dos níveis circulantes de miRNA foram determinadas usando o teste U de Mann-Whitney, teste t, ANOVA, correlação de Spearman's Rho ou Pearson, conforme apropriado. Resultados com valor de p menor que 0,05 foram considerados significativos. Utilizou-se análise de regressão logística binária e geraram-se curvas ROC de característica de operação do receptor para avaliar a capacidade dos miRNAs escolhidos para distinguir entre casos de câncer e controles. Isso foi realizado individualmente e para combinações de miRNAs.

Os autores, apresentaram 76 miRNAs com expressão diferencial na circulação de mulheres com câncer de mama tipo Luminal A em comparação com aquelas que não têm câncer de mama. Sendo que o perfil de miRNA de três biomarcadores de miRNA associados a tumores circulantes (*miR-29a*, *miR-181a* e *miR-652*) para o câncer de mama, os quais, em combinação, fornecem um perfil de sensibilidade e especificidade de 0,80.

Já o Modelo de Regressão Logística Modificada utilizado, apresentou excelente desempenho na tarefa de classificação. Todas as combinações possíveis de dados foram exploradas e o novo modelo proposto foi capaz de classificar todas as amostras em todos os

conjuntos de dados, considerando todas as possíveis combinações de dados, com bom desempenho. Para avaliar o poder discriminatório do método proposto, também realizamos análise de sensibilidade e especificidade. Para todas as combinações possíveis de dados explorados, a faixa de valores de sensibilidade e especificidade é de 0,9, superior ao apresentado pelo autor oriundo os dados.

Após análise na literatura referente aos miRNAs com valores de alfas negativos, constatamos que todos estão relacionados a capacidade de inibição da proliferação da doença, alguns por regulação negativa (diminuição da expressão). Observamos também que todos desempenham tal papel pois estão envolvidos na inibição de algum fator essencial para a via de sinalização, regulação dos fatores de crescimento, atividade ou expressão de seus constituintes proteicos.

Já no que se refere aos miRNAs com valores de alfas positivos, o único comportamento análogo entre eles que foi identificado foi a sua super expressão em todos os estudos realizados e assim como os de valores negativos todos desempenham tal papel pois estão envolvidos na inibição de algum fator essencial para alguma via seja de sinalização, regulação dos fatores de crescimento, atividade ou expressão de seus constituintes proteicos. Os miR-139-3p e miR-628-5p não foram encontrados na literatura relacionando-os a câncer de mama.

Os achados suportam a hipótese de que os perfis de expressão de miRNA circulante não podem atuar como uma janela direta sobre a atividade tumoral e nos leva a questionar o mecanismo pelo qual eles entram na corrente sanguínea, além de seu papel funcional, se algum, na circulação periférica. Esses processos permanecem pouco entendidos. MiRNAs podem entrar na circulação periférica após secreção seletiva de células tumorais ou microvesículas circulantes (MITCHELL et al., 2008). Outras células no microambiente tumoral também podem secretar miRNAs. Enquanto isso, outra escola de pensamento sugere que os miRNAs possam ser detectáveis na circulação como consequência do vazamento passivo de células apoptóticas e necróticas (OGAWA et al., 2010). Na realidade, é provável que ambas as teorias sejam verdadeiras, com evidências acumuladas para apoiar ambas as propostas.

Não obstante o trabalho de Alysson dos Santos não tenha tido como objetivo explícitos estudar aspectos particulares para biomarcadores, ele foi capaz de diferenciar tecidos saudáveis de tecidos tumorais obtendo seu melhor resultado quando utilizando o modelo ALR com alta sensibilidade e especificidade. Ele também utilizou um algoritmo ingênuo para selecionar um conjunto de 20 miRNAs dentre os quais 10 miRNAs com o

maior e os 10 miRNAs com menores valores de α (SANTOS, 2015). Após essa seleção, foi realizada uma busca na literatura sobre a relação desses miRNAs com o câncer de mama. A busca não trouxe relação significativa dos miRNAs em questão com o câncer de mama. Um dos aspectos que deve ser levado em consideração é o perfil da base utilizada.

Posto que nosso objetivo fosse não descortinar o processo do câncer. Seria significativo em uma nova etapa agregar os resultados obtidos anteriormente juntamente com os nossos em análises futuras de bases com o intuito de obter um kit diagnóstico para detecção precoce da doença (fase assintomática). Importante conjugar os diferentes dados de miRNAs de diferentes técnicas e os resultados mais recentes.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Abreu, M.N.S., Siqueira, A.L., Caiaffa, W.T. (2009) **Regressão logística ordinal em estudos epidemiológicos**. Rev Saúde Pública 2009;43(1): 183-94.
- Barros ACSD, Leite KRM. **Classificação molecular dos carcinomas de mama. Uma visão contemporânea**. Rev Bras Mastologia. 2015;25(4):146-55.
- Bazzoli,C. and Lambert-Lacroix,S. (2018) **Classification based on extensions of LS-PLS using logistic regression: application to clinical and multiple genomic data**. BMC Bioinformatics 19:314.
- Cai WL, Huang WD, Li B, Chen TR, Li ZX, Zhao CL, Li HY, Wu YM, Yan WJ, Xiao JR. **microRNA-124 inhibits bone metastasis of breast cancer by repressing Interleukin-11**. Mol Cancer. 2018 Jan 17;17(1):9. doi: 10.1186/s12943-017-0746-0.
- COSTA, F. M. R. **Microarray---based breast cancer classification using logistic regression and behond**. Tese (Doutorado Bioinformática) – Universidade Federal de Minas Gerais. Belo Horizonte, 2017.
- Damavandi Z , Torkashvand S , Vasei M , Soltani BM , Tavallaei M , Mowla SJ **Aberrant Expression of Breast Development-Related MicroRNAs, miR-22, miR-132, and miR-212, in Breast Tumor Tissues**. J Breast Cancer. 2016 Jun;19(2):148-55. doi:10.4048/jbc.2016.19.2.148. Epub 2016 Jun 24.
- Deerwester, S.C., Dumais, S.T., Landauer, T.K., et al. (1990) **Indexing by Latent Semantic Analysis**. *Journal of the American Society for Information Science*, 41, 391-407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASII>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9)
- Ding,Y. *et al.* (2018) **Identification of a gene-expression predictor for diagnosis and personalized stratification of lupus patients**. PLoS ONE,13(7): e0198325.
- Estimativa 2018: incidência de câncer no Brasil / Instituto Nacional de Câncer José Alencar Gomes da Silva**. Coordenação de Prevenção e Vigilância. – Rio de Janeiro: INCA, 2017.
- Feng T, Shao F, Wu Q, Zhang X, Xu D, Qian K, Xie Y, Wang S, Xu N, Wang Y, Qi C. **miR-124 downregulation leads to breast cancer progression via LncRNA-MALAT1 regulation and CDK4/E2F1 signal activation**. Oncotarget. 2016 Mar 29;7(13):16205-16. doi: 10.18632/oncotarget.7578.

Fotini M. Kouri, Lisa A. Hurley, Weston L. Daniel, et al. **miR-182 integrates apoptosis, growth, and differentiation programs in glioblastoma.** *Genes Dev.* 2015, 29. Access the most recent version at doi:10.1101/gad.257394.114

Gálvez, J.M. *et al.* (2018) **Multiclass classification for skin cancer profiling based on the integration of heterogeneous gene expression series.** *PLoS ONE* 13(5): e0196836.

Golub, G. (1965). **Numerical methods for solving linear least squares problems.** *Numerische Mathematik.* June 1965, Volume 7, Issue 3, pp 206-216

Hanieh H. **Aryl hydrocarbon receptor-microRNA-212/132 axis in human breast cancer suppresses metastasis by targeting SOX4.** *Mol Cancer.* 2015 Sep 17;14:172. doi: 10.1186/s12943-015-0443-9.

Li, P. *et al.* (2014) **Gene regulatory network inference and validation using relative change ratio analysis and time-delayed dynamic Bayesian network.** *Journal on Bioinformatics and Systems Biology*, 2014:12.

Li, Z. *et al.* (2018) **Efficient feature selection and classification for microarray data.** *PLoS ONE*, 13(8): e0202167.

Lien, T.G. *et al.* (2018) **Integrated analysis of DNAmethylation and gene expression using high-dimensional penalized regression: a cohort study on bone mineral density in postmenopausal women.** *BMC Medical Genomics*, 11:24.

Linnik, Y. **Method of Least Squares and Principles of the Theory of Observations.** Translated from Russian by R. C. ELANDT. New York: Pergamon Press 1961.

Macedo T, Silva-Oliveira RJ, Silva VAO, Vidal DO, Evangelista AF, Marques MMC. **Overexpression of mir-183 and mir-494 promotes proliferation and migration in human breast cancer cell lines.** *Oncol Lett.* 2017 Jul;14(1):1054-1060. doi: 10.3892/ol.2017.6265. Epub 2017 May 26.

Marcolino, L.S., Couto, B.R.G.M. and Santos, M.A.D. (2010) **Genome Visualization in Space,** *Proceedings of IWPACBB*, 225-232.

Marôco, J. (2014). *Análise Estatística com o SPSS Statistics*, 6ª ed.

McDermott AM, Miller N, Wall D, Martyn LM, Ball G, Sweeney KJ, et al. (2014) **Identification and Validation of Oncologic miRNA Biomarkers for Luminal A-like Breast Cancer**. PLoS ONE 9(1): e87032.

Menard, S. (2010). **Logistic Regression: From Introductory to Advanced Concepts and Applications**. Thousand Oaks, CA: Sage.

Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, et al. (2008) **Circulating microRNAs as stable blood-based markers for cancer detection**. Proc Natl Acad Sci U S A 105 (30): 10513-10518. Epub 12008 Jul 10528.

Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, et al.. (2008) **Circulating microRNAs as stable blood-based markers for cancer detection**. Proc Natl Acad Sci U S A 105 (30): 10513–10518. Epub 12008 Jul 10528.

Ni F, Gui Z, Guo Q, Hu Z, Wang X, Chen D, Wang S . **Downregulation of miR-362-5p inhibits proliferation, migration and invasion of human breast cancer MCF7 cells**. Oncol Lett. 2016 Feb;11(2):1155-1160. Epub 2015 Dec 3.

Ogawa R, Tanaka C, Sato M, Nagasaki H, Sugimura K, et al. (2010) **Adipocyte-derived microvesicles contain RNA that is transported into macrophages and might be secreted into blood circulation**. Biochem Biophys Res Commun 398 (4): 723-729. Epub 2010 jul 2017.

Pereda, E. *et al.* (2018) **The blessing of Dimensionality: Feature Selection outperforms functional connectivity-based feature transformation to classify ADHD subjects from EEG patterns of phase synchronization**. PLoS ONE, 13(8): e0201660.

Ping Huang, Jie Xi, ShikaiLiu. **MiR-139-3p induces cell apoptosis and inhibits metastasis of cervical cancer by targeting NOB1**. Biomedicine & Pharmacotherapy Volume 83, October 2016, Pages 850-856.

Santos, Alysson. **Early Breast Cancer Detection Using Logistic Regression Models**. Dissertação (Mestrado Bioinformática) – Universidade Federal de Minas Gerais. Belo Horizonte, 2017 xxii, 57 f. : il.

Srivastava, A., Goldberger, H., Dimtchev, A. et al. **Circulatory miR-628-5p is downregulated in prostate cancer patients**. Tumor Biol. (2014) 35: 4867. <https://doi.org/10.1007/s13277-014-1638-1>

Sun D, Zhong J, Wei W, Chen X, Liu J, Hu Z. **Identification of microRNA expression in sentinel lymph nodes from patients with breast cancer via RNA sequencing for diagnostic accuracy.** *J Gene Med.* 2019 Apr;21(4):e3075. doi: 10.1002/jgm.3075. Epub 2019 Mar 3.

Turgut Cosan D, Oner C, Mutlu Sahin F. **Micro RNA-126 coordinates cell behavior and signaling cascades according to characteristics of breast cancer cells.** *Bratisl Lek Listy.* 2016;117(11):639-647. doi: 10.4149/BLL_2016_124.

VARELLA, Dráuzio. **Câncer de mama - Identificar a doença precocemente teve impacto insignificante no Brasil.** Folha de São Paulo. São Paulo, 9 de jun de 2019. Disponível em: <https://www1.folha.uol.com.br/colunas/drauziovarella/2019/06/cancer-de-mama.shtml> Acesso em: 22 de junho 2019.

Wang C, Ju H, Shen C, Tong Z. **miR-429 mediates δ -tocotrienol-induced apoptosis in triple-negative breast cancer cells by targeting XIAP.** *Int J Clin Exp Med.* 2015 Sep 15;8(9):15648-56. eCollection 2015.

Wang CZ, Yuan P, Li Y. **MiR-126 regulated breast cancer cell invasion by targeting ADAM9.** *Int J Clin Exp Pathol.* 2015 Jun 1;8(6):6547-53. eCollection 2015.

Wu, M., et al. (2015) **Prioritization Of Nonsynonymous Single Nucleotide Variants For Exome Sequencing Studies Via Integrative Learning On Multiple Genomic Data.** *Nature Scientific Reports* | 5:14955 | DOI: 10.1038/srep14955.

Xie M, Fu Z, Cao J Liu Y, Wu J, Li Q, Chen Y. **MicroRNA-132 and microRNA-212 mediate doxorubicin resistance by down-regulating the PTEN-AKT/NF- κ B signaling pathway in breast cancer.** *Biomed Pharmacother.* 2018 Jun;102:286-294. doi: 10.1016/j.biopha.2018.03.088. Epub 2018 Mar 22.

Xu, L., et al. (2011) **Functional Cohesion of Gene Sets Determined by Latent Semantic Indexing of PubMed Abstracts.** *PlosOne*, 6(4), e18851.

Yu J, Shen W, Gao B, Zhao H, Xu J, Gong B. **MicroRNA-182 targets FOXF2 to promote the development of triple-negative breast cancer.** *Neoplasma.* 2017;64(2):209-215. doi:10.4149/neo_2017_206.

Zhan Y, Li X, Liang X, Li L, Cao B, Wang B, Ma J, Ding F, Wang X, Pang D, Liu Z. **MicroRNA-182 drives colonization and macroscopic metastasis via targeting its suppressor SNAI1 in breast cancer.** *Oncotarget.* 2017 Jan 17;8(3):4629-4641. doi: 10.18632/oncotarget.13542.

Zhang K, Wang YW, Wang YY, Song Y, Zhu J, Si PC, Ma R. **Identification of microRNA biomarkers in the blood of breast cancer patients based on microRNA profiling.** *Gene*. 2017 Jul 1;619:10-20. doi: 10.1016/j.gene.2017.03.038. Epub 2017 Mar 27.

Zhang M, Liu D, Li W, Wu X, Gao C, Li X. **Identification of featured biomarkers in breast cancer with microRNA microarray.** *Arch Gynecol Obstet*. 2016 Nov;294(5):1047-1053. Epub 2016 Jun 30.

Zhang W, Qian P, Zhang X, Zhang M, Wang H, Wu M, Kong X, Tan S, Ding K, Perry JK, Wu Z, Cao Y, Lobie PE, Zhu T. **Autocrine/Paracrine Human Growth Hormone-stimulated MicroRNA 96-182-183 Cluster Promotes Epithelial-Mesenchymal Transition and Invasion in Breast Cancer.** *J Biol Chem*. 2015 May 29;290(22):13812-29. doi: 10.1074/jbc.M115.653261. Epub 2015 Apr 14.

Zhang Y, Xu G, Liu G, Ye Y, Zhang C, Fan C, Wang H, Cai H, Xiao R, Huang Z, Luo Q. **miR-411-5p inhibits proliferation and metastasis of breast cancer cell via targeting GRB2.** *Biochem Biophys Res Commun*. 2016 Aug 5;476(4):607-613. doi: 10.1016/j.bbrc.2016.06.006. Epub 2016 Jun 3.

Zhao H, Shen J, Medico L, Wang D, Ambrosone CB, Liu S (2010) **A Pilot Study of Circulating miRNAs as Potential Biomarkers of Early Stage Breast Cancer.** *PLoS ONE* 5(10): e13735.

Zhou WB, Zhong CN, Luo XP, Zhang YY, Zhang GY, Zhou DX, Liu LP. **miR-625 suppresses cell proliferation and migration by targeting HMGA1 in breast cancer.** *Biochem Biophys Res Commun*. 2016 Feb 19;470(4):838-44. doi: 10.1016/j.bbrc.2016.01.122. Epub 2016 Jan 22.

Zhu N, Zhang D, Xie H, Zhou Z, Chen H, Hu T, Bai Y, Shen Y, Yuan W, Jing Q, Qin Y. **Endothelial-specific intron-derived miR-126 is down-regulated in human breast cancer and targets both VEGFA and PIK3R2.** *Mol Cell Biochem*. 2011 May;351(1-2):157-64. doi: 10.1007/s11010-011-0723-7. Epub 2011 Jan 20.