

UNIVERSIDADE CATÓLICA DOM BOSCO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Andressa Marques Andreu

Análise de dados e *Machine Learning* na
identificação de possíveis biomarcadores em câncer
de mama

Campo Grande
Mato Grosso do Sul
2023

UNIVERSIDADE CATÓLICA DOM BOSCO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Andressa Marques Andreu

Análise de dados e *Machine Learning* na identificação de possíveis biomarcadores em câncer de mama

“Dissertação apresentada como parte das exigências para obtenção do título de MESTRE EM BIOTECNOLOGIA, no Programa de Pós-Graduação em Biotecnologia da Universidade Católica Dom Bosco - Área de concentração: Biotecnologia”

Autora: Andressa Marques Andreu
Orientador: Prof. Dr. Cristiano Marcelo Espinola Carvalho

Campo Grande
Mato Grosso do Sul
2023

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade Católica Dom Bosco
Bibliotecária Mourâmise de Moura Viana - CRB-1 3360

A555a Andreu, Andressa Marques

Análise de dados e Machine Learning na identificação de possíveis biomarcadores em câncer de mama / Andressa Marques Andreu sob orientação do Prof. Dr. Cristiano Marcelo Espinola Carvalho...-- Campo Grande, MS : 2023.
75 p.: il.

Dissertação (Mestrado em Biotecnologia) - Universidade Católica Dom Bosco, Campo Grande- MS, 2023
Bibliografia: p. 58- 61

1. Biomarcadores. 2. Câncer de mama. 3. Biotecnologia.
4. Machine learning. 5. Medicina de precisão I.Carvalho,
Cristiano Marcelo Espinola. II. Título.

CDD: 660.6

"ANÁLISE DE DADOS E MACHINE LEARNING NA IDENTIFICAÇÃO DE POSSÍVEIS BIOMARCADORES EM CÂNCER DE MAMA"

Autora: ANDRESSA MARQUES ANDREU

Orientador: Prof. Dr. CRISTIANO MARCELO ESPINOLA CARVALHO

TITULAÇÃO: Mestre em Biotecnologia

Área de concentração: Biotecnologia.

APROVADO EM 10 DE MARÇO DE 2023.

Prof. Dr. Cristiano Marcelo Espinola Carvalho (Orientador) – UCDB

Tiago Tognoli de Almeida

Prof. Dr. Tiago Tognoli de Almeida (FACSUL)

Carina Elisei de Oliveira

Profa. Dra. Carina Elisei de Oliveira (UCDB)

Danieli Fernanda Buccini

Profa. Dra. Danieli Fernanda Buccini (UCDB)

Aos pacientes de tumor mamário; que os esforços hoje empreendidos se traduzam em qualidade de vida e que os produtos biotecnológicos sejam acessíveis.

À minha avó, Lenita Silva de Arruda, por vencer, todos os dias, esta enfermidade.

AGRADECIMENTOS

À CAPES, à FUNDECT e à Universidade Católica Dom Bosco pelo custeio do meu curso de mestrado. Também por todo acolhimento, estrutura e pela formação integral fornecida.

À Pró-reitoria de Pesquisa e Pós-Graduação dirigida pelo meu orientador, professor Dr. Cristiano Carvalho, pelos muitos ensinamentos, paciência, compreensão e sabedoria em gerir pessoas. Por ser exemplar em ética profissional e pelo cuidado com a saúde mental de seus orientandos; muito obrigada por marcar positivamente a minha carreira acadêmica.

À Engenheira de computação que tornou possível todas as análises deste trabalho, Thayane Batista, por ter adentrado ao mundo da biologia e me possibilitado conhecer de perto o universo TI. Muito obrigada!

Aos meus alunos de TCC e co-orientados de Iniciação Científica, Thailainy Furtado, Lívia Mota de Lima, Thainá Rejala, Juarez Culau, Gabriela Moraes. Agradeço pela oportunidade de crescemos juntos e de transmitir conhecimentos.

À minha companheira de mestrado, Mirianne dos Reis, obrigada pela motivação nas madrugadas afora, por aguentar os meus surtos e por sempre me dizer: “você nasceu para ministrar aula”. Obrigada por todas as vivências durante a graduação e as pós-graduações. Ter com quem dividir as alegrias, desesperos e conquistas faz, no mínimo, toda a diferença!

Aos demais colaboradores da equipe, programadores e estatísticos; grata por alavancar as análises feitas até aqui.

Às instituições Santa Casa de Campo Grande e Cassems-MS e seus contribuintes, cirurgiões oncologistas, médicos residentes e enfermeiros, obrigada pela parceria até aqui.

Aos meus irmãos, Mayane, Welyson e Daniela e meu cunhado Josemar, pelo incrível desafio de amar pessoas tão distintas e peculiares.

Às sobrinhas mais lindas do mundo, Ana Júlia e Alice, por existirem.

Ao meu “padrasto”, Claudiomir Procksch, pelo amor que transcende laços consanguíneos.

À minha avó, Izabel Andreu Maldonado de Brito (*in memorian*), por me proporcionar uma herança genética tão resiliente.

Ao meu pai, Wilson Andreu, grata pela história de superação que escreveu, por

seus recomeços.

À minha heroína mãe, Marcia Marques, por ser o aplauso mais gratificante:
obrigada pela honra em ser sua filha.

Ao meu esposo, Dr. Luiz Fernando Rodrigues, pela companhia de vida.
Obrigada pelo zelo com a nossa tão recente família.

À Deus, por sempre me mostrar que se há algo de bom, de sábio e de
proveitoso em mim, é d'Ele que provém.

“A Ciência é a arte de ver os detalhes que Deus criou.”
G. Antognolli

BIOGRAFIA DO AUTOR

Andressa Marques Andreu nasceu em vinte e um de maio de 1999 no município de Campo Grande, Mato Grosso do Sul, Brasil. Filha de Márcia Terezinha Ferraz Marques e Wilson Andreu Lima. Musicista de essência, ao final de seu ensino médio encantou-se pelo universo científico. Cursou Biomedicina na Universidade Católica Dom Bosco com Bolsa PROUNI entre os anos de 2017 a 2021. Durante esses anos, foi bolsista do CNPq no Programa Institucional de Bolsas de Iniciação Científica durante três ciclos. Recebeu premiação como Melhor Trabalho de Iniciação Científica no ano de 2020, por meio do qual foi contemplada com bolsa de mestrado pela própria instituição e também bolsa CAPES. Em alguns de seus estudos, avaliou as variações hormonais e comportamentais de ratos frente aos diferentes estímulos musicais, participou de pesquisas para identificação de compostos com potenciais antimicrobianos e antitumorais advindos de amostras de veneno de animais. Conheceu também o universo da Tecnologia da Informação em busca de respostas biológicas promissoras relacionadas ao câncer. Realizou seu mestrado entre 2021 e 2022 e seguiu para a docência.

SUMÁRIO

	Página
LISTA DE FIGURAS	10
LISTA DE TABELAS	13
RESUMO	14
ABSTRACT	15
1 INTRODUÇÃO	16
2. REVISÃO DE LITERATURA	18
OBJETIVOS	26
3.1. OBJETIVO GERAL	29
3.2. OBJETIVOS ESPECÍFICOS	29
CAPÍTULO 1 - Identificação de miRNAs potenciais biomarcadores no câncer de mama	30
Abstract	31
Introduction	32
Materials and Methods	34
Results	36
Discussion/Conclusion	37
References [Numerical]	40
Figure Legends	41
CAPÍTULO 2 - <i>ALDH2</i> , <i>PCK1</i> , famílias <i>ADH</i> e <i>COX</i> : genes relacionados às alterações metabólicas no câncer de mama	43
Resumo	43
Introdução	44
Metodologia	45
Resultados	48
Discussão	54
Conclusão	57
Referências	58
CAPÍTULO 3 BREAST CANCER IA APLICATTION : Programa de Computador com aplicação de <i>machine learning</i> em dados genéticos para identificar possíveis biomarcadores no câncer de mama	62

RESUMO	62
Certificado de Registro de Programa de Computador.....	63
CONSIDERAÇÕES FINAIS	64
Anexo 1 – Pedido de Registro de Programa de Computador	65
Anexo 2 – Parecer Consustanciado pelo CEP	72

LISTA DE FIGURAS

	Página
INTRODUÇÃO/REVISÃO DE LITERATURA	
Figura 1. Modelo de progressão tumoral linear. Este modelo esquematiza a progressão da estrutura ducto-mamária normal para um carcinoma <i>in situ</i> , passando por invasivo e culminando num carcinoma metastático. Os ductos mamários são formados pela membrana basal e uma bicamada de células luminais (camada mais interna) e mioepiteliais (camada mais externa). No carcinoma <i>in situ</i> as células mioepiteliais estão em menor número, devido à degradação da membrana basal, e o número de células luminais é maior. A perda de células mioepiteliais e da membrana basal resulta no carcinoma invasivo, no qual as células tumorais invadem os tecidos adjacentes e podem migrar para órgãos distantes, determinando à formação de metástases. FONTE: Polyak, 2007 (modificado).....	17
Figura 2. Biogênese dos miRNAs. O gene correspondente ao miRNA é transcrito pela enzima RNA polimerase II, formando o pri-miRNA. Tal molécula apresenta uma estrutura <i>hairpin</i> , fundamental para o reconhecimento e processamento pela enzima Drosha, formando um pré-miRNA precursor de ± 70 nucleotídeos. A proteína exportina 5 carrega este miRNA precursor ao citoplasma, para então ser processado pela Dicer, gerando um duplex de miRNA de ± 22 nucleotídeos. Uma das fitas é degradada e a outra permanece no complexo RISC. Fonte: SIQUEIRA e PIRES, 2017.....	20
CAPÍTULO 1	
Figure. 1. Flowchart of the methodology developed in this work. Figura do autor.....	37
Figura 2. MiRNAs differentially expressed in breast carcinomas. In the lines, there are the predicted miRNAs in decreasing order of number of target genes with which they were predicted to interact. In the columns, the target genes studied in this work are arranged. In the center, the spaces filled with color represent that interaction was predicted. Up-regulated and down-regulated miRNAs are, respectively, in blue and red. Color strength represents the strength of the Fold Change. On the right, the exact	

HR values of the miRNAs in the samples available from the TCGA are displayed.
Figurado autor.....37

Fig. 3. Linear correlations between target genes and their miRNAs. Negative linear correlations are in red and positive ones in blue. The greater the intensity of the color, the greater the intensity of the correlation. Figura do autor.....38

CAPÍTULO 2

Figura 1 - Fluxograma da metodologia realizada neste trabalho. A pesquisa consistiu na escolha dos bancos de dados, seguida da análise e manipulação dos dados que foram submetidos a testes estatísticos possibilitando a seleção de genes potenciais marcadores moleculares do metabolismo energético no câncer de mama.
Figurado autor.....44

FIGURA 2. Gráfico de expressão diferencial da via da Glicólise. O gráfico, obtido pelo pacote VolcanoPlot, traça valores de FC (eixo X) e valores de significância (eixo Y). Os genes nas extremidades direita e esquerda são os que apresentaram diferença em perfil de expressão entre o tecido normal e o tumoral. Destaque para as extremidades em ambos os eixos; valor de corte Fold Change: 1,5. Figura do autor..45

FIGURA 2 - Gráfico de expressão diferencial da via fosforilação oxidativa. O gráfico, obtido pelo pacote VolcanoPlot, traça valores de FC (eixo X) e valores de significância (eixo Y). Os genes nas extremidades direita e esquerda são os que apresentaram diferença em perfil de expressão entre o tecido normal e o tumoral. À esquerda, COX7A1 apresentou-se down-regulated, enquanto à direita, COX6A2 e COX6 apresentaram-se Up-regulated. Figura do autor.....46

FIGURA 4. Gráfico de correlação linear da via Glicólise. Foi obtido por intermédio do pacote GGCorPlot sendo baseado na correlação linear do perfil de expressão de genes relacionados à via da glicólise. As correlações lineares positivas estão em vermelho, as negativas, em azul e os genes não correlacionados, em branco. Quanto mais intensa a correlação positiva e/ou negativa, maior a intensidade das cores.
Figurado autor.....47

FIGURA 5 - Gráfico de correlação linear da via fosforilação oxidativa. Foi obtido por intermédio do pacote GGCorPlot sendo baseado na correlação linear do perfil de

expressão de genético da via fosforilação oxidativa. As correlações lineares positivas estão em vermelho, as negativas, em azul e os genes não correlacionados, em branco. Quanto mais intensa a correlação positiva e/ou negativa, maior a intensidade das cores. Figura do autor.....49

LISTA DE TABELAS

	Página
CAPÍTULO 1	
Table 1. Number of MiRNAs predicted for each target gene by the MirWalk platform.....	38
CAPÍTULO 2	
Tabela 1. As doze correlações lineares positivas e negativas mais intensas entre os genes referentes à via da glicólise e seus respectivos valores de correlação.....	48
Tabela 2. As doze correlações lineares positivas e negativas mais intensas entre os genes referentes à via fosforilação oxidativa e seus respectivos valores de correlação.....	49
Tabela 3. Correlação lineare entre os genes selecionados, separados por via.....	53

RESUMO

O tipo de câncer mais comum no mundo é o de mama. Sabe-se que 20 a 30% dos pacientes diagnosticados com tumor primário apresentam recidiva devido à baixa acurácia dos indicadores tradicionais. Oncogenes e genes supressores tumorais têm sua expressão alterada por mecanismos regulatórios como os miRNAs, os quais atuam como biomarcadores em diversos tipos de cânceres. Diante disso, o presente trabalho teve como objetivo identificar genes e miRNAs com potencial a biomarcadores de diagnóstico, prognóstico e estadiamento no câncer de mama por meio de análises de dados e *machine learning*. Foram utilizados dados provenientes do MirWalk, KEGG Pathway e TCGA. Nos capítulos 1 e 2, a expressão diferencial foi calculada pelo algoritmo DESeq2. O cruzamento dos dados foi realizado por meio da IDE RStudio, bem como a aplicação de testes estatísticos e produção de gráficos com os pacotes de repositório Bioconductor: VolcanoPlot e GGCorrPlot. Os miRNAs que apresentaram expressão diferencial com *Fold change* acima de 1,5, predição de interação com mRNAs do gene alvo e correlação linear negativa com pelo menos 3 dos genes envolvidos foram miR-665, miR-378c, miR-184, miR-133b e miR-3161 (capítulo 1). Quanto aos genes, *ADH1A*, *ADH1B*, *ALDH2*, *PCK1*, *COX7A1* e *COX6A2* apresentaram destaque como possíveis participantes da modulação do metabolismo energético do câncer de mama (capítulo 2). No terceiro capítulo, encontra-se um modelo de *machine learning* e seu Registro de Patente, o qual identificou os genes *MMP11* e *FIGF* e os alvos hsa-mir-139 e hsa-mir-145 como possíveis biomarcadores do câncer de mama. Esses transcritos se apresentam como candidatos a marcadores de diagnóstico, prognóstico e estadiamento bem como possíveis contribuições para a terapia do câncer de mama. Novos estudos são sugeridos para validar os achados, bem como a elucidação de mecanismos regulatórios para que esses alvos possam originar ferramentas biotecnológicas importantes para a medicina de precisão.

Palavras-chave: Biomarcadores; Câncer de Mama; Biotecnologia; *machine learning*, Medicina de Precisão.

ABSTRACT

The most common type of cancer in the world is breast cancer. It is known that 20 to 30% of patients diagnosed with a primary tumor present recurrence due to the low accuracy of traditional indicators. Oncogenes and tumor suppressor genes have their expression altered by regulatory mechanisms such as miRNAs, which act as biomarkers in several types of cancer. Therefore, the present work aimed to identify genes and miRNAs with potential biomarkers for diagnosis, prognosis and staging in breast cancer through data analysis and machine learning. Data from MirWalk, KEGG Pathway and TCGA were used. In chapters 1 and 2, the differential expression was calculated by the *DESeq2* algorithm. Data crossing was carried out using the RStudio IDE, as well as the application of statistical tests and graph production with the Bioconductor repository packages: VolcanoPlot and GGCorrPlot. The miRNAs that showed differential expression with Fold change above 1.5, predicted interaction with target gene mRNAs and negative linear correlation with at least 3 of the genes involved were miR-665, miR-378c, miR-184, miR-133b and miR-3161 (chapter 1). As for the genes, *ADH1A*, *ADH1B*, *ALDH2*, *PCK1*, *COX7A1* and *COX6A2* were highlighted as possible participants in the modulation of energy metabolism in breast cancer (chapter 2). In the third chapter, there is a machine learning model and its Patent Registration, which identified the *MMP11* and *FIGF* genes and the hsa-mir-139 and hsa-mir-145 targets as possible biomarkers of breast cancer. These transcripts are presented as candidates for diagnostic, prognostic and staging markers, as well as possible contributions to breast cancer therapy. New studies are suggested to validate the findings, as well as the elucidation of regulatory mechanisms so that these targets can originate important biotechnological tools for precision medicine.

Keywords: Biomarkers; Breast cancer; Biotechnology; *Machine Learning*, Precision Medicine.

1 INTRODUÇÃO

As mudanças nos hábitos de vida e no perfil epidemiológico da população têm transformado o perfil morbi-mortalidade mundial, variando entre doenças infecto-parasitárias a crônico-degenerativas, como o câncer (LANETZKI *et al.*, 2012). O câncer, caracterizado pelo crescimento celular desordenado resultante de alterações na expressão gênica, é um dos principais problemas de saúde pública no mundo, com mais de 10 milhões de mortes por ano (OMS, 2023). Dentre as suas origens, relaciona-se às alterações genéticas, físicas, químicas e biológicas que se perpetuam ao longo da vida (INUMARU; SILVEIRA; NAVES, 2011). No Brasil, são estimados mais de 704 mil para cada ano até 2025 (INCA, 2023).

Segundo o Instituto Nacional do Câncer (INCA, 2023), a neoplasia mais incidente em mulheres e com taxas crescentes a cada ano é o câncer de mama. Com cerca de 73.610 novos casos estimados para cada ano do triênio 2023-25, deixou de ser considerada uma doença única com diferenças histopatológicas e variável prognóstico. Devido ao advento das novas tecnologias, atualmente o câncer de mama é definido como um conjunto de doenças heterogêneas que afetam o mesmo órgão, tendo a mesma origem estrutural e diferentes características histopatológicas, fatores de risco e respostas às terapias (FUKUSHIMA, 2015).

Além disso, dados revelam que o câncer de mama é o tipo de neoplasia mais frequente no mundo (INCA, 2023). Sua heterogeneidade é compreendida a partir de seus aspectos morfológicos, imunohistoquímicos, clínicos e genéticos específicos. Porém, mesmo com os importantes avanços para sua detecção, diagnóstico e tratamento em estágios iniciais, esse câncer ainda é considerado uma das principais doenças potencialmente letais devido ao seu diagnóstico tardio com incidência de 41,89 casos por 100.000 mulheres (INCA, 2022).

MEHTA, GAURAV e KHANNA (2019) enfatizam a importância que estudos moleculares têm na elucidação de mecanismos da oncogênese mamária e na identificação de genes que permitem consideráveis avanços diagnósticos e terapêuticos. O entendimento a nível molecular permite empregar biomarcadores como ferramenta diagnóstica na rotina clínica. Diante disso, esta área de pesquisa se mostra promissora, especialmente em doenças altamente heterogêneas como o câncer de mama. Embora muitos estudos sendo realizados, não há marcadores moleculares disponíveis capazes de detectar uma lesão pré-maligna ou predizer, com eficácia e precisão, o estadiamento do câncer.

Diante disso, o presente trabalho visou selecionar possíveis genes e miRNAs associados ao câncer de mama, por meio de análise de dados e *machine learning* com a utilização de dados públicos mundiais. Pretende-se, em pesquisas posteriores, realizar a validação *in vitro* da expressão diferencial dos alvos moleculares filtrados por estas análises, nas amostras de tecido mamário normal e tumor primário de pacientes sul-matogrossenses.

REVISÃO DE LITERATURA

2.1 Etiologia do câncer

Em um tecido sadio, as células se adaptam às mudanças do meio interno e proliferam-se a fim de suprir alguma necessidade do meio externo, como a de substituir células perdidas ou reparar lesões teciduais, ativadas por sinais de proliferação-repressão. As lesões celulares podem ser reversíveis ou irreversíveis, levando a diferenciação celular ou a apoptose - a qual compreende um importante mecanismo de morte celular programada para remoção de células danificadas do organismo. A desregulação desse mecanismo propicia a carcinogênese, marcada pela mutação de genes e pela proliferação celular descontrolada, características fundamentais ao desenvolvimento do câncer (DOGAN et.al., 2007).

Nesse contexto, o modelo de progressão tumoral linear sugere que o câncer de mama é iniciado por uma hiperplasia epitelial atípica, que se torna um *carcinoma in situ*, progride a um carcinoma invasivo e resulta em um câncer de mama metastático (Figura 1; POLYAK, 2007). No entanto, de acordo com Balmain, Gray e Ponder (2003), nem todos os tumores seguem esta ordem e podem até mesmo não apresentar todos estes estágios.

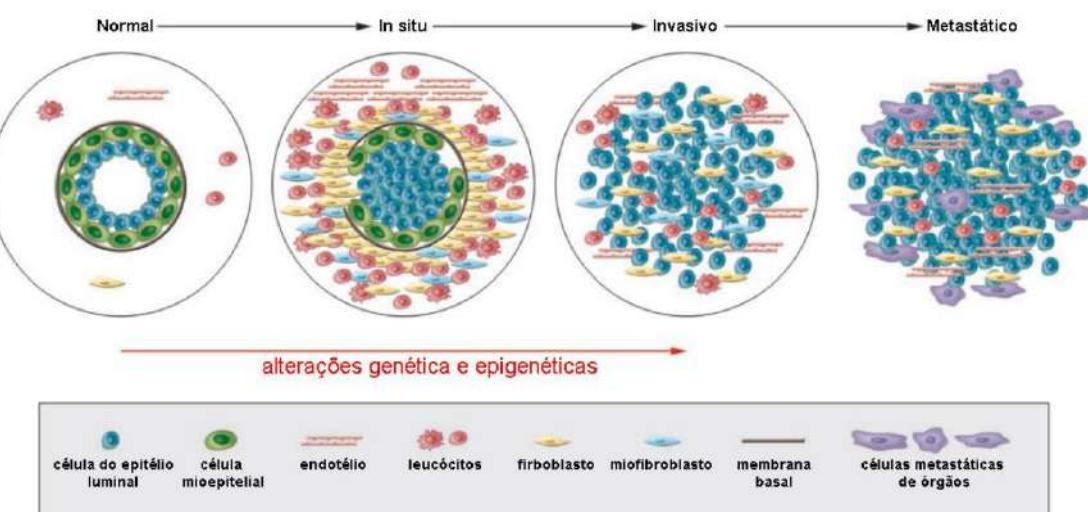


Figura1. Modelo de progressão tumoral linear. Este modelo esquematiza a progressão da estrutura ducto-mamária normal para um carcinoma *in situ*, passando por invasivo e culminando num carcinoma metastático. Os ductos mamários são formados pela membrana basal e uma bicamada de células luminárias (camada mais interna) e mioepiteliais (camada mais externa). No carcinoma *in situ* as células mioepiteliais estão em menor número, devido à degradação da membrana basal, e o número de células luminárias é maior. A perda de células mioepiteliais e da membrana basal resulta no carcinoma invasivo, no qual as células tumorais invadem os tecidos adjacentes e podem migrar para órgãos distantes, determinando à formação de metástases.

FONTE: Polyak, 2007 (modificado).

As causas do câncer de mama ainda não são totalmente conhecidas. Estudos indicam alguns fatores de risco endógenos e exógenos, como a extensão e duração da exposição a hormônios sexuais do ciclo menstrual e derivados de contraceptivos e reposição hormonal (DUMITRESCU & COTARLA, 2005). Além disso, segundo Oldenburg e colaboradores (2007), fatores como: menarca precoce - antes dos 11 anos de idade, menopausa tardia, nuliparidade e primeira gestação após 30 anos também apresentam risco elevado. Outros fatores relacionados são: exposição à radiação ionizante antes dos 35 anos, tabagismo, alcoolismo, obesidade, alta densidade óssea após a menopausa, doença proliferativa benigna da mama e histórico familiar de câncer (ZHENG *et al.*, 1999; GERGER *et al.*, 2007).

Sabe-se que, 20 a 30% das pacientes com diagnóstico de tumor primário de mama apresentarão recorrência da doença em sítios distantes, devido à presença de micrometástases (VELASCO-VELÁZQUEZ *et al.*, 2012). A progressão tumoral, com a invasão adjacente e o desenvolvimento de metástases, é a principal responsável pela alta taxa de mortalidade observada nas pacientes (GUNASINGHE *et al.*, 2012), evidenciando a necessidade de marcadores moleculares capazes de predizer precocemente o estadiamento do câncer.

2.2 Genes *drivers*

Segundo Storchova e Pellman (2004), a etiologia dessa patologia é genética, caracterizada por diversas alterações genômicas, como rearranjos, deleções e amplificações, resultando na modificação da expressão de oncogenes e genes supressores de tumor. De acordo com Nusbaum e Isaacs (2007), em muitas famílias a predisposição ao câncer de mama pode ser atribuída a mutações em dois principais genes, *BRCA1* e *BRCA2*. Estes genes *drivers* são supressores de tumor e apresentam expressão reduzida em amostras tumorais (OLDENBURG *et al.*, 2007), favorecendo a carcinogênese. O *BRCA1*, associado ao reparo do DNA por recombinação homóloga (RH), está localizado no cromossomo 17 (17q21.31), transcreve 31 mRNA, dos quais 24 são codificadores de proteína e possui 24 éxons. Esses transcritos codificam uma fosfoproteína nuclear que, dentre as suas principais funções, garante a integridade genômica, suprimindo o tumor (GHADIRIAN *et al.*, 2009). O *BRCA2* expressa uma proteína que atua em conjunto ao *BRCA1*. Localizado no cromossomo 13 (13q13.1), transcreve 7 mRNA, dos quais 2 são codificadores de proteínas (NCBI, 2020).

Além destes, outros genes supressores de tumor têm se mostrado relacionados ao câncer de mama na literatura. Dentre eles, *CDKN1A* e *Rad51*. O *CDKN1A*, mais conhecido como *p21*, localiza-se no cromossomo 6 (6p21.2) e transcreve 8 mRNA, dos quais 5 são codificadores de proteína. Tal gene codifica um potente inibidor da quinase dependente de ciclina (CDK), funcionando como regulador da progressão do ciclo celular. A expressão desse gene é controlada pela proteína p53, e desempenha papel fundamental na fase S da replicação do DNA e no reparo a danos no DNA. Já o *Rad51*, conhecido por estar envolvido na RH, situa-se no cromossomo 15 (15q15.1). Transcreve 11 mRNA, dos quais 8 são codificadores de uma família de proteínas que inibe a patogênese tumoral (NCBI, 2020).

Outro grupo gênico de importância na etiologia e desenvolvimento dos cânceres são os oncogenes, como *CCND1* e *CDK1*. O *CCND1* possui cinco éxons, localiza-se no cromossomo 11 (11q13.3) e atua favorecendo a proliferação celular exacerbada no câncer de mama (MORENO-BUENO *et al.*, 2003). O *CDK1*, localizado no cromossomo 10 (10q21.2), possui 9 éxons e pertence à família CDK. Codifica uma proteína atuante na transição da fase G2 para a mitose. Ambos estes genes apresentam-se superexpressos em tumores, como o carcinoma mamário (PREVO *et al.*, 2018).

2.3 MiRNAs

Vários mecanismos podem estar associados à redução ou aumento da expressão desses genes. Dentre eles, a regulação pós-transcricional por miRNAs tem se destacado em decorrência da sua intensa associação a processos que favorecem o câncer, como proliferação celular, diferenciação, desenvolvimento e inibição da apoptose. Além disso, estão envolvidos na cascata de invasão-metástase. Tais RNAs não-codificantes (ncRNAs), apresentam um clássico mecanismo de regulação negativa, embora evidências tenham apontado para uma possível regulação positiva, que influencia na regulação da expressão gênica por mecanismos epigenéticos (VAN SCHOONEVELD *et al.*, 2012).

Esses microRNAs (miRNAs) são moléculas de RNA fita simples não-codificadoras de proteínas, compostas de 19 a 25 nucleotídeos. Foram descritos como potentes reguladores pós-transcpcionais da expressão gênica em plantas e animais (LIANG *et al.*, 2016). Segundo Ricarte e Kimura (2006), a biogênese dessas estruturas ocorre no núcleo e posteriormente no citoplasma. Inicia-se com a transcrição de seu

gene, pela RNA polimerase II que gera uma longa cadeia de pri-miRNA. Sua estrutura característica possui um *hairpin*, fundamental para o reconhecimento da RNase Drosha e seu cofator DGCR8 (*DiGeorge syndrome critical region gene 8*), dando origem a um pré-miRNA, de 70 a 100 nucleotídeos. Este é transportado ao citoplasma pela Exportina-5, onde é processado pela RNase III Dicer e origina um duplex de miRNA, com aproximadamente 22 nucleotídeos. O duplex é então incorporado ao complexo RISC (*RNA-induced silence complex*) onde estão as proteínas Argonautas. Apenas uma das fitas do duplex miRNA-miRNA permanece no RISC para regulação pós-transcricional da expressão de seus genes-alvo (Figura 2).

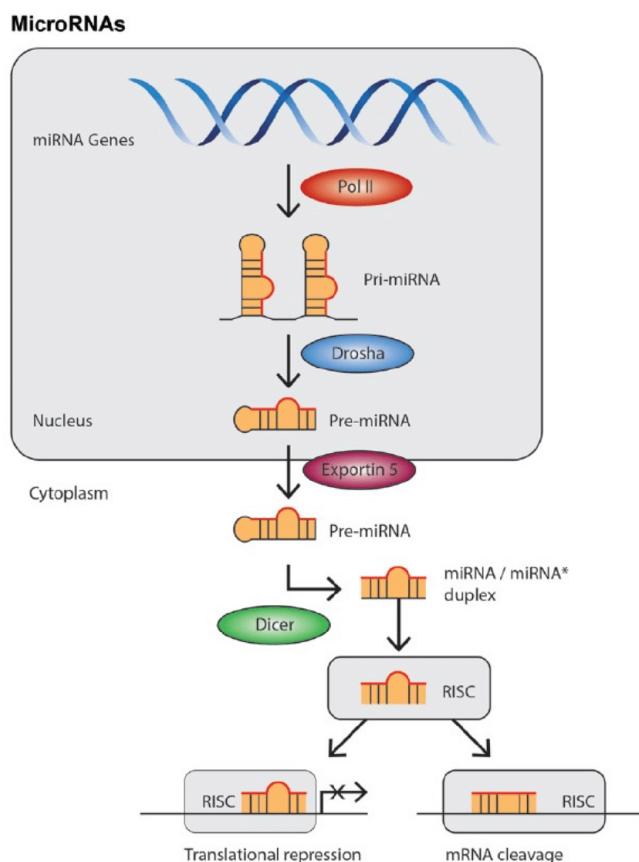


Figura 2. Biogênese dos miRNAs. O gene correspondente ao miRNA é transcrito pela enzima RNA polimerase II, formando o pri-miRNA. Tal molécula apresenta uma estrutura *hairpin*, fundamental para o reconhecimento e processamento pela enzima Drosha, formando um pré-miRNA precursor de \pm 70 nucleotídeos. A proteína exportina 5 carrega este miRNA precursor ao citoplasma, para então ser processado pela Dicer, gerando um duplex de miRNA de \pm 22 nucleotídeos. Uma das fitas é degradada e a outra permanece no complexo RISC.

Fonte: SIQUEIRA e PIRES, 2017.

Para regular a expressão, os miRNAs ligam-se à região 3' UTR do mRNA-alvo por pareamento completo ou parcial (CHAMORRO; ALBUQUERQUE; ROSA, 2015). O grau de complementariedade desses miRNAs com seus mRNA-alvo é o que permite sua ação, podendo ser por inibição da tradução – quando parcial - ou por

degradação do mRNA-alvo – quando o pareamento é completo. Nos mamíferos, o mecanismo principal de atuação é pelo pareamento parcial com o mRNA, promovendo a inibe da tradução.

Assim como os genes codificadores de proteínas, muitos miRNAs têm apresentado mudanças em seu perfil de expressão entre tecidos mamário normal e tumoral. As investigações a respeito dessas alterações no câncer de mama e tecidos adjacentes apontam seu potencial para identificar características clínico-patológicas (VOLINIA *et al.*, 2006). Dessa forma, esses transcritos apresentam relevante potencial biotecnológico para diagnóstico precoce, indicativo de prognóstico e de progressão da doença e pode auxiliar na terapêutica, uma vez que contribui com a melhora na sobrevida das pacientes (YAN *et al.*, 2008).

Os dados de expressão de genes e miRNAs são disponibilizados à comunidade científica em bancos de dados públicos, que possibilitam análises bioinformáticas combinadas por meio de ferramentas de análises de dados, como a *IDE RStudio*. Sendo uma plataforma de código aberto que usa a linguagem de programação R, a *IDE RStudio* é um Ambiente de Desenvolvimento Integrado (IDE) projetado para facilitar o manuseio e fornecer ferramentas de alta performance para pesquisas avançadas. Tem como diferencial ampla representação gráfica e bases estatísticas que garantem a confiabilidade aos resultados (GANDRUD; 2018).

Atualmente, o prognóstico da doença tem sido baseado em parâmetros histológicos e clínicos (tipo/tamanho do tumor, presença/ausência de metástases em linfonodos axilares, presença/ausência dos receptores hormonais de estrogênio (ER) e progesterona (PR) e amplificação do fator de crescimento epidermal 2 (*HER2* ou *ERBB2*). Embora estes fatores sejam de importância prognóstica, a sua capacidade informativa é limitada.

2.4 Análise bioinformática e banco de dados

A importância da execução de análises *in silico*, previamente a análises *in vitro*, se dá mediante a varredura empregada sobre os dados angariados até então, com alta precisão e assertividade. Visto que, ferramentas bioinformáticas confiáveis garantem precisão na mineração de dados, refinam a pesquisa de acordo com o que apresenta maior evidência e confiabilidade estatística e possibilitam identificar padrões moleculares passíveis de serem aplicados na rotina clínica.

A bioinformática apresenta-se como um instrumento capaz de auxiliar na

descoberta de redes de interações que envolvem os genes de interesse, de modo a obter o contexto biológico de todos os genes que são afetados pela instabilidade genômica, contribuindo com a escolha de tratamentos e terapias ideais (CAVA et al., 2015). De acordo com conhecimentos atuais dos sistemas biológicos, tornou-se possível pesquisar a partir de dados experimentais publicados na literatura, em termos de interação molecular e redes de reação, que podem ser encontrados em bancos de dados, como o KEGG PATHWAY (KANEHISA et al., 2021).

O objetivo principal do KEGG tem sido permitir a reconstrução computacional de sistemas biológicos, incluindo a célula, o organismo e o ecossistema, a partir de informações do genoma (KANEHISA et al., 2021). Com a diminuição contínua do custo para sequenciamento de próxima geração e outros métodos de caracterização molecular de alto rendimento, uma grande quantidade de conjuntos de dados é gerada e disponibilizada para acesso público, no campo da pesquisa do câncer.

Neste contexto do câncer, o banco mais utilizado é o *The Cancer Genome Atlas* (TCGA) (DENG et al., 2016). Este banco foi desenvolvido e criado para atender a essas necessidades e seus grandes conjuntos de dados oferecem algumas oportunidades que não precisam de precedentes para a análises sistemáticas e integradas (KANDOTH et al., 2013). O TCGA incentivou a utilização de muitas ferramentas bioinformáticas como o RStudio, no qual é utilizado a linguagem R, orientados para a descoberta, caracterização e priorização de processos celulares que conduzem ao câncer com base em vias, genes ou RNAs (BAILEY et al., 2018). O R é um ambiente para modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento e gráficos, e permite análises estáticas e dinâmicas de programas. Sua interface apresenta o RStudio como um ambiente de desenvolvimento integrado (IDE) de código aberto para a linguagem de programação R, um software que combina vários componentes e pacotes do R, os quais facilitam a sua utilização por pesquisadores de outras áreas, como da saúde (TOPARSLAN et al., 2020).

2.5 Aprendizado de Máquina (machine learning)

O aprendizado de máquina (*machine learning*) é um subconjunto da inteligência artificial, baseada na construção de um modelo a partir de dados de amostra com o intuito de obter previsões ou decisões sem que seja necessário programá-lo explicitamente para realizar tal tarefa utilizando cálculos definidos em diferentes tipos

de métodos (ZHANG et al., 2020).

Esta ferramenta tecnológica tem sido utilizada para diagnosticar e também para classificar o câncer por há alguns anos, entretanto, somente alguns estudos relatam sua importância para o prognóstico no câncer (KOUROU et al., 2015).

A melhoria contínua da tecnologia possibilitou grandes quantidade de dados sobre o câncer no campo da medicina. Porém, embora haja diversas informações armazenadas, a previsão assertiva do diagnóstico de uma neoplasia, continua sendo uma das tarefas que mais desafiam as possibilidades que a tecnologia de aprendizado de máquina trouxe (DOHERTY et al., 2016).

O aprendizado de máquina é diferenciado em dois tipos principais conhecidos como aprendizado não-supervisionado e aprendizado supervisionado, de modo que no primeiro método citado não há dados definidos a serem utilizados como input (entrada) e também não é possível saber como será seu output (saída) durante o processo de aprendizagem. Desta maneira, cabe ao modelo de aprendizagem escolhido encontrar padrões ou descobrir os grupos de dados de entrada (TAN e GILBERT, 2003).

Já o segundo caso, consiste em definir um conjunto de dados a serem utilizados como treinamento, para que seja possível estimar ou mapear os dados de *input* para o *output* esperado. Neste modelo de aprendizagem, o processo para encontrar padrões pode ser desenvolvido como um problema de classificação, de maneira que esta tarefa de classificação se trata de um processo de aprendizagem capaz de categorizar os dados em um conjunto de classes determinadas finitas (KOUROU et al., 2015).

No *machine learning*, algumas tarefas são comuns em ambos os tipos de aprendizados citados. Para casos em que se vê a necessidade de encontrar categorias ou agrupamentos para que assim seja possível distinguir as diferenças entre os dados é utilizada a técnica denominada como *Clustering*, que consiste em segregar as amostras de acordo com suas características atribuindo aos agrupamentos adequados (AAMIR e ZAIDI, 2021).

Outra técnica comumente utilizada é chamada de regressão. Esta técnica consiste em definir uma função de aprendizado que faz o mapeamento dos dados em uma variável, para que assim seja possível a cada nova amostra neste processo estimar a predição do valor de uma variável (CHAKRAVARTY et al., 2020). No terceiro capítulo deste trabalho, utilizou-se as técnicas: Distância Euclidiana, Árvore de

Decisão, SVM (*Support Vector Machine*), MLP (*Multi-Layer Perceptron Classifier*), e Regressão Logística.

O capítulo 1, a seguir, foi escrito dentro das normas da revista *Oncology - Karger Publishers*. A pesquisa do capítulo 3 gerou um Registro de Código de Computador de número BR512022001525-0, cujo certificado e pedido de registro estão contidos no corpo do capítulo e em Anexo 1, respectivamente. O Parecer Consustanciado pelo CEP para aprovação da validação destes resultados - número 3.246.746 - encontra-se em Anexo 2.

Referências

- AAMIR, Muhammad; ZAIDI, Syed Mustafa Ali. Clustering based semi-supervised machine learning for DDoS attack classification. *Journal of King Saud University-Computer and Information Sciences*, v. 33, n. 4, p. 436-446, 2021.
- BALMAIN, Allan; GRAY, Joe; PONDER, Bruce. A genética e genômica do câncer. *Nature genetics*, v. 33, n. 3, p. 238-244, 2003.
- CHAKRAVARTY, Srinivas; DEMIRHAN, Haydar; BASER, Furkan. Fuzzy regression functions with a noise cluster and the impact of outliers on mainstream machine learning methods in the regression setting. *Applied Soft Computing*, v. 96, p. 106535, 2020.
- CHAMORRO, Amanda Coelho; ALBUQUERQUE, Eliane Papa Ambrosio; DELLA ROSA, Valter Augusto. Micro-RNAs e câncer: abordagens e perspectivas/ Micro-RNAs and cancer: approaches and perspectives. *Revista de Pesquisa em Saúde*, v. 16, n. 2, 2016.
- DENG, Yanhong et al. Modified FOLFOX6 with or without radiation versus fluorouracil and leucovorin with radiation in neoadjuvant treatment of locally advanced rectal cancer: initial results of the Chinese FOWARC multicenter, open-label, randomized three-arm phase III trial. *Journal of clinical oncology*, v. 34, n. 27, p. 3300-3307, 2016.
- DING, Li et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. *Cell*, v. 173, n. 2, p. 305-320. e10, 2018.
- DOGAN, Soner et al. Effects of high-fat diet and/or body weight on mammary tumor leptin and apoptosis signaling pathways in MMTV-TGF- α mice. *Breast Cancer Research*, v. 9, n. 6, p. R91, 2007.
- DOHERTY, M. et al. Precision medicine and oncology: an overview of the opportunities presented by next-generation sequencing and big data and the challenges posed to conventional drug development and regulatory approval pathways. *Annals of Oncology*, v. 27, n. 8, p. 1644-1646, 2016.
- DUMITRESCU, R. G.; COTARLA, I. Understanding breast cancer risk-where do we stand in 2005. *Journal of cellular and molecular medicine*, v. 9, n. 1, p. 208-221, 2005.
- FUKUSHIMA, Kassandra Ferreira Pessoa et al. Frequência e fatores associados à síndrome da web axilar em mulheres submetidas à cirurgia de câncer de mama: estudo transversal e retrospectivo. *Springerplus*, v. 4, n. 1, p. 112, 2015.
- GANDRUD, Christopher. Reproducible research with R and R studio. CRC Press, 2013
<http://dx.doi.org/10.1201/9781315382548>.
- GUNASINGHE, NPA Devika et al. Mesenchymal–epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. *Cancer and Metastasis Reviews*, v.31, n. 3-4, p. 469-478, 2012.
- INSTITUTO NACIONAL DO CÂNCER (Brasil).Câncer de mama. [Brasília, DF]: Instituto Nacional do Câncer, 2022. Disponível em: <http://www2.inca.gov.br/>
- INSTITUTO NACIONAL DO CÂNCER (Brasil).Câncer de mama. [Brasília, DF]: Instituto Nacional do Câncer, 2023. Disponível em: <http://www2.inca.gov.br/>
- INUMARU, Lívia Emi; SILVEIRA, Érika Aparecida da; NAVES, Maria Margareth Veloso.

Fatores de risco e de proteção para câncer de mama: uma revisão sistemática. **Cadernos de Saúde Pública**, v. 27, n. 7, p. 1259-1270, 2011.

KANDOTH, Cyriac et al. Mutational landscape and significance across 12 major cancer types. **Nature**, v. 502, n. 7471, p. 333-339, 2013.

KANEHISA, Minoru et al. KEGG: integrando vírus e organismos celulares. Pesquisa de ácidos nucléicos, v. 49, n. D1, pág. D545-D551, 2021.

KOUROU, Konstantina et al. Machine learning applications in cancer prognosis and prediction. **Computational and structural biotechnology journal**, v. 13, p. 8-17, 2015.

LANETZKI, Camila Sanches et al. The epidemiological profile of pediatric intensive care center at hospital israelita albert einstein. **Einstein** (São Paulo), v. 10, n. 1, p. 16-21, 2012.

LIANG, Zhenzhen et al. Diagnostic value of microRNAs as biomarkers for cholangiocarcinoma. **Digestive and Liver Disease**, v. 48, n. 10, p. 1227-1232, 2016.

HORN, M. et al. O uso da microscopia confocal de varredura a laser em microcirurgia para carcinoma de células escamosas invasivo. **British Journal of Dermatology**, v. 156, n. 1, pág. 81-84, 2007.

MEHTA, Gaurav A.; KHANNA, Pooja; GATZA, Michael L. Emerging role of SOX proteins in breast Cancer development and maintenance. **Journal of mammary gland biology and neoplasia**, v. 24, p. 213-230, 2019.

MORENO-BUENO, Gema et al. Cyclin D1 gene (CCND1) mutations in endometrial cancer. **Oncogene**, v. 22, n. 38, p. 6115-6118, 2003.

NCBI National Center for Biotechnology Information. Disponível em: <https://www.ncbi.nlm.nih.gov/>. Acesso em: SET. 2020.

NUSSBAUM, R. L.; MCINNES, R. R.; WILLARD, H. F. Genetic counseling and risk assessment. Thompson and Thompson. **Genetics in Medicine**. 6th ed. Philadelphia: WB Saunders Company, p. 375-89, 2001.

OLDENBURG, R. A. et al. Genetic susceptibility for breast cancer: how many more genes to be found? **Critical reviews in oncology/hematology**, v. 63, n. 2, p. 125-149, 2007.

OMS. Organização Mundial da Saúde. (2022, 18 de setembro). Um relatório sobre saúde. Retirado de https://www.paho.org/bra/index.php?option=com_content&view=article&id=393:cancer<c.

POLYAK, Kornelia; WEINBERG, Robert A. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. **Nature Review Cancer**, v. 9, n. 4, p. 265-273, 2007.

PREVO, Remko et al. A inibição de CDK1 sensibiliza células normais a danos no DNA de uma maneira dependente do ciclo celular. **Cell Cycle**, v. 17, n. 12, pág. 1513-1523, 2018.

RICARTE FILHO, Júlio; KIMURA, Edna Teruko. MicroRNAs: nova classe de reguladores gênicos envolvidos na função endócrina e câncer. **Arquivos Brasileiros de Endocrinologia & Metabologia**, v. 50, p. 1102-1107, 2006.

STORCHOVA, Zuzana; PELLMAN, David. From polyploidy to aneuploidy, genome instability and cancer. **Nature reviews molecular cell biology**, v. 5, n. 1, p. 45-54, 2004.

TAN, Aik Choon; GILBERT, David. Ensemble machine learning on gene expression data for cancer classification. 2003.

TOPARSLAN, Emine; KARABAG, Kemal; BILGE, Ugur. A workflow with R: Phylogenetic analyses and visualizations using mitochondrial cytochrome b gene sequences. **PLoS One**, v. 15, n. 12, p. e0243927, 2020.

VAN SCHOONEVELD, Eleni et al. Expression profiling of cancerous and normal breast tissues identifies microRNAs that are differentially expressed in serum from patients with (metastatic) breast cancer and healthy volunteers. **Breast cancer research**, v. 14, n. 1, p. R34, 2012.

VELASCO-VELÁZQUEZ, Marco et al. CCR5 antagonist blocks metastasis of basal breast cancer cells. **Cancer research**, v. 72, n. 15, p. 3839-3850, 2012.

VOLINIA, Stefano et al. A microRNA expression signature of human solid tumors defines cancer gene targets. **Proceedings of the National Academy of Sciences**, v. 103, n. 7, p. 2257-2261, 2006.

YAN, Li-Xu et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. **Rna**, v. 14, n. 11, p. 2348-2360, 2008.

ZHENG, Wei et al. N-acetyltransferase 1 genetic polymorphism, cigarette smoking, well- done meat intake, and breast cancer risk. **Cancer Epidemiology and Prevention Biomarkers**, v. 8, n. 3, p. 233-239, 1999.

OBJETIVOS

3.1. OBJETIVO GERAL

Selecionar genes e miRNAs candidatos a biomarcadores no câncer de mama por meio de análise de dados e *machine learning*.

3.2. OBJETIVOS ESPECÍFICOS

- 3.2.1. Selecionar possíveis miRNAs reguladores dos genes *BRCA1*, *BRCA2*, *CCND1*, *CDK1*, *CDKN1A* e *Rad51* por meio de análise de dados com os bancos *TCGA* e *MirWalk*;
- 3.2.2. Gerar um *screening* molecular de genes associados ao metabolismo energético no câncer de mama, a partir de análises de dados com os bancos *TCGA* e *KEGG Pathway*;
- 3.2.3. Encontrar possíveis biomarcadores (genes e miRNAs) por meio da aplicação de aprendizado de máquina (*machine learning*) em dados de amostras de câncer.

CAPÍTULO 1 - Identificação de miRNAs potenciais biomarcadores no câncer de mama

Research Article

Identification of potential biomarker miRNAs in breast cancer

Andressa Marques Andreu^a, Thayane Gonçalves da Silva Batista^b, Cristiano Marcelo Espínola Carvalho^c

^aPrograma de Pós-Graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, MS, Brazil

^bDepartamento de Engenharia da Computação, Universidade Católica Dom Bosco, Campo Grande, MS, Brazil

^cPró-reitoria de Pesquisa e Pós-Graduação, Universidade Católica Dom Bosco, Campo Grande, MS, Brazil

Identification of potential biomarker miRNAs in breast cancer

Corresponding Author:

Andressa Marques Andreu

Programa de Pós-graduação em Biotecnologia

Universidade Católica Dom Bosco

Av. Tamandaré, 6000

Campo Grande, MS, 79117-000, Brazil

Tel: (67) 3312-3300

E-mail: dessamarques21@gmail.com

Number of Tables: 1.

Number of Figures: 3.

Word count: 3017 words.

Keywords: MiRNAs, Biomarkers, Breast Cancer.

1 **Abstract**

2 The second most common type of cancer in the world is breast cancer. It is known that 20 to
3 30% of patients diagnosed with a primary tumor have recurrence due to the low accuracy of
4 traditional indicators. Oncogenes and tumor suppressor genes have their expression altered
5 by regulatory mechanisms such as miRNAs, which are post-transcriptional regulators gene
6 expression and act as biomarkers in several cancers. transcriptome data are available in banks
7 that enable bioinformatics analysis, which generate evidence that enhances the assertiveness
8 of further *in vitro* analyses. Given this, the presente work aimed to select possible regulatory
9 miRNAs of *BRCA1*, *BRCA2*, *CCND1* genes, *CDK1*, *CDKN1A* and *Rad51* through *in silico*
10 differential expression analysis in order to identify possible biomarkers in breast carcinomas.
11 The predictions of the modules-regulatory miRNAs-mRNA were performed by MirWalk. A
12 differential expression was Calculated by the DESeq2 algorithm with data from *The Cancer*
13 *Genome Atlas National Cancer Institute Program - TCGA*, generated by *qPCR*, *RNAseq*,
14 *microarray* and other *in vitro* analyses. The crossing of data was performed through the IDE
15 RStudio, as well such as applying statistical tests and producing graphics with repository
16 packages Bioconductor: *VolcanoPlot* and *GGCorrPlot*. The miRNAs that express expression
17 differential with Fold change above 1.5, prediction of interaction with target gene mRNAs and
18 negative linear correlation with at least 3 of the genes involved were miR-665, miR-378c, miR-
19 184, miR-133b and miR-3161. These transcripts present themselves as candidates for
20 diagnosis, prognosis and staging biomarkers and possible contributions to breast cancer
21 therapy. Further studies are suggested to validate the findings, as well as the elucidation of
22 regulatory mechanisms so that these miRNAs can originate biotechnological tools of
23 importance to precision medicine.

24

25 **Introduction**

26 Changes in the population's lifestyle and epidemiological profile have transformed the
27 global morbidity and mortality profile, ranging from infectious-parasitic to chronic-degenerative
28 diseases, such as cancer [1]. Cancer is one of the main public health problems in the world,
29 with 9.6 million deaths per year. Among its origins, it is related to genetic, physical, chemical
30 and biological changes that are perpetuated throughout life [2]. In Brazil, around 600,000 new
31 cases were registered in the 2018-19 period [3].

32 According to the National Cancer Institute, the most common cancer in women and
33 with increasing rates each year is breast cancer. This pathology, with about 66,280 new cases
34 estimated for each year of the 2020-22 triennium [3], is no longer considered a single disease
35 with histopathological differences and a variable prognosis. Due to the advent of new
36 technologies, breast cancer is currently defined as a set of different heterogeneous diseases
37 that affect the same organ, originating from the same structure and with different
38 histopathological characteristics, risk factors and responses to systemic therapies [4].

39 Bertucci, Birnbaum and Gonçalves (2006) [5] emphasize the importance that molecular
40 studies have had in elucidating the mechanisms of breast oncogenesis and in identifying genes
41 that have allowed considerable advances in diagnosis and therapy. Understanding at the
42 molecular level allows the use of molecular markers as a diagnostic tool in clinical routine.
43 Therefore, this area of research shows promise, especially in highly heterogeneous diseases
44 such as breast cancer. And although many studies are underway, there are still no available
45 molecular markers capable of detecting a premalignant lesion or effectively predicting cancer
46 staging.

47 It is known that 20 to 30% of patients diagnosed with a primary breast tumor will present
48 disease recurrence in distant sites, due to the presence of micrometastases [6]. Tumor
49 progression, with adjacent invasion and the development of metastases, is the main
50 responsible for the high mortality rate observed in patients[7], highlighting the need for
51 molecular markers capable of early prediction of cancer staging.

52 According to Nusbaum and Isaacs (2007) [8], predisposition to breast cancer in many
53 families can be attributed to mutations in two main genes, *BRCA1* and *BRCA2*. These driver
54 genes are tumor suppressors and show reduced expression in tumor samples [9], favoring
55 carcinogenesis. *BRCA1*, associated with DNA repair by homologous recombination (RH), is
56 located on chromosome 17 (17q21.31), transcribes 31 mRNA, of which 24 are protein-coding
57 and has 24 exons. These transcripts encode a nuclear phosphoprotein that, among its main
58 functions, ensures genomic integrity, suppressing the tumor [10]. *BRCA2* expresses a protein
59 that acts together with *BRCA1*. Located on chromosome 13 (13q13.1), it transcribes 7 mRNA,

60 of which 2 are protein coding [11].

61 In addition to these, other tumor suppressor genes have been shown to be related to
62 breast cancer in the literature. Among them, *CDKN1A* and *Rad51*. *CDKN1A*, better known as
63 p21, is located on chromosome 6 (6p21.2) and transcribes 8 mRNA, of which 5 are protein-
64 coding. This gene encodes a potent cyclin-dependent kinase (CDK) inhibitor, functioning as a
65 regulator of cell cycle progression. The expression of this gene is controlled by the p53 protein,
66 and it plays a fundamental role in the S phase of DNA replication and in repairing DNA damage.
67 *Rad51*, known to be involved in HR, is located on chromosome 15 (15q15.1). It transcribes 11
68 mRNA, 8 of which encode a family of proteins that inhibit tumor pathogenesis [11].

69 Another gene group of importance in the etiology and development of cancers are the
70 oncogenes, such as *CCND1* and *CDK1*. *CCND1* has five exons, is located on chromosome
71 11 (11q13.3) and acts favoring cell proliferation exacerbated in breast cancer [12]. *CDK1*,
72 located on chromosome 10 (10q21.2), has 9 exons and belongs to the CDK family. It encodes
73 a protein involved in the transition from the G2 phase to mitosis. Both of these genes are
74 overexpressed in tumors, such as breast carcinoma [13].

75 Several mechanisms can be associated with the reduction or increase in the expression
76 of these genes. Among them, post-transcriptional regulation by miRNAs has stood out due to
77 its intense association with processes that favor cancer, such as cell proliferation,
78 differentiation, development and inhibition of apoptosis. Furthermore, they are involved in the
79 invasion-metastasis cascade. Such non-coding RNAs (ncRNAs) present a classic down-
80 regulation mechanism, although evidence has pointed to a possible up-regulation, which
81 influences the regulation of gene expression by epigenetic mechanisms [14].

82 To regulate expression, miRNAs bind to the 3' UTR region of the target mRNA by
83 complete or partial pairing [15]. The degree of complementarity of these miRNAs with their
84 target mRNA is what allows their action, which may be by inhibition of translation – when partial
85 – or by degradation of the target mRNA – when the pairing is complete. In mammals, the main
86 mechanism of action is the partial pairing with mRNA, promoting the inhibition of translation.

87 As well as protein-coding genes, many miRNAs have shown changes in their
88 expression profile between normal and tumoral breast tissue. Investigations about these
89 changes in breast cancer and adjacent tissues point to their potential to identify
90 clinicopathological characteristics [16]. Thus, these transcripts have relevant biotechnological
91 potential for early diagnosis, indicative of prognosis and disease progression, and may help in
92 therapy, as it contributes to an improvement in patient survival [17].

93 Gene and miRNA expression data are made available to the scientific community in
94 public databases, which enable combined bioinformatics analysis through data analysis tools

such as IDE RStudio. As an open-source platform that uses the R programming language, the RStudio IDE is an Integrated Development Environment (IDE) designed to be easy to use and provide high-performance tools for advanced research. Its differential is wide graphical representation and statistical bases that guarantee the reliability of the results [18].

The importance of performing *in silico* analyses, prior to *in vitro* analyses, is given by scanning the data collected so far, with high precision and assertiveness. Since, reliable bioinformatics tools ensure accuracy in data mining, refine research according to what presents greater evidence and statistical reliability, and make it possible to identify molecular patterns that can be applied in clinical routine.

Therefore, the present work aimed to perform the selection of possible regulatory miRNAs of the *BRCA1*, *BRCA2*, *Rad51*, *CDK1*, *CCND1* and *CDKN1A* genes, associated with breast cancer, through an *in silico* analysis, using the IDE R Studio tool, using global public data. It is intended, then, to carry out the *in vitro* validation of the differential expression of miRNAs filtered by data analysis, in samples of normal breast tissue and primary tumor of patients from Mato Grosso do Sul.

110

111 Materials and Methods

112 Selection of target genes

113 For the analysis, four tumor suppressor genes were selected; *BRCA1*, *BRCA2*, *Rad51* 114 and *CDKN1A*, and two oncogenes; *CCND1* and *CDK1*.

115

116 Predictive analysis of miRNA-mRNA regulatory modules

117 The identification of *in silico* predicted miRNA-mRNA regulatory modules was 118 performed using the MirWalk platform (<http://mirwalk.umm.uni-heidelberg.de/>). In it, sequence 119 databases were used - GenBank (<https://www.ncbi.nlm.nih.gov/genbank/> - Accessed in 120 November 2020) and miRBase (<https://www.mirbase.org/> - Accessed in November 2020) to 121 generate the predicted modules. The prediction was performed using the TarPmiR algorithm, 122 developed in Python. This open source platform predicts the interaction between miRNAs and 123 their target genes according to the following main parameters: sequence complementarity, 124 binding site conservation, accessibility to target sites, among others [19]. The output of this 125 analysis constituted a list of miRNAs predicted by gene, which was entered into the IDE R 126 Studio where the other subsequent steps were performed.

127

128 Differential expression analysis of miRNAs

129 Data for the present study were downloaded from the Cancer Genome Atlas (TCGA)

130 of the TCGA-BRCA project (<https://portal.gdc.cancer.gov/>). Differential expression of the
131 predicted miRNAs for each target genes was evaluated between primary tumor (PT) vs non-
132 tumor (NT) breast tissue samples, using the DESeq2 algorithm (available at
133 <http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>) following the
134 suggestions [20]. Data from 1063 TP and 104 NT of breast were used. Among several objects
135 generated as output by DESeq2, there is a matrix with normalized absolute quantification
136 values and a list with Fold Change (FC) values, which represent relative quantification values
137 between the two analyzed conditions. As a calibrator, the expression value in samples of
138 normal breast tissue was used. A cutoff value of +/- 1.5 was assigned to identify up-regulated
139 and down-regulated miRNAs.

140

141 Analysis of differential expression of miRNAs and mRNAs

142 Data for the present study were *download* from The Cancer Genome Atlas (TCGA) of
143 the TCGA-BRCA project (<https://portal.gdc.cancer.gov/>). The differential expression of target
144 gene miRNAs and mRNAs was evaluated between primary tumor (PT) vs non-tumor (NT)
145 breast tissue samples, using the DESeq2 algorithm (available at <http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>) following the suggestions [20]. Data from 1063 TP
146 and 104 NT of breast were used. Among several objects generated as output by DESeq2,
147 there is a matrix with normalized absolute quantification values and a list with Fold Change
148 (FC) values, which represent relative quantification values between the two analyzed
149 conditions. As a calibrator, the expression value in samples of normal breast tissue was used.
150 A cutoff value of +/- 1.5 was assigned to identify up-regulated and down-regulated miRNAs.

151

153 Cross-prediction analysis of regulatory modules and differential expression

154 The list of miRNAs predicted for each target gene by MirWalk was crossed with the
155 miRNAs that showed differential expression according to TCGA data using the IDE R Studio
156 platform. The results of this combined analysis were displayed in scatter plots using the
157 VolcanoPlot package
158 (<https://www.bioconductor.org/help/search/index.html?q=VONCANOPLLOT/>).

159

160 Linear correlation analysis

161 According to the mechanism of downregulation of gene expression by miRNAs
162 described in the literature, there is an inversely proportional interdependence between the
163 expression profile of miRNAs and that of their target genes [21]. To assess this, the linear
164 correlation test was applied to the mRNA expression data of the genes under study and the
165 predicted and differentially expressed miRNAs. Focus was given to the smallest negative linear
166 correlations. To illustrate the results obtained in this analysis, correlation graphs were

167 generated using the GGCorrPlot package, available at
168 <https://www.bioconductor.org/help/search/index.html?q=GGcorPlot/>.

169

170 **Selection of candidate miRNAs for biomarkers**

171 The selection of miRNAs with potential to biomarkers followed the following criteria:
172 being differentially expressed with HR above 1.5, presenting a prediction of interaction with at
173 least three of the six genes under study and being among the highest intensity negative linear
174 correlations with at least three of the target genes.

175 The methodology is represented in the flowchart in Figure 1.

176

177 **Results**

178 **Predicted miRNA-mRNA regulatory modules**

179 For each target gene, a series of miRNAs was predicted, whose quantities are
180 expressed in the table 1:

181

182 *MiRNAs predict interaction with breast cancer driver genes and differential
183 expression between normal and tumoral breast tissue*

184 It was observed that miR-665, miR-449a and miR-1254 were predicted to interact with
185 the six driver genes studied in this work. Furthermore, miR-378c, miR-184, miR-760 and miR-
186 3174 were predicted to interact with five of the six target genes. In the same sense, miR-133b,
187 miR-3610, miR-639 and miR-3161 were predicted for four of the six target genes. All these
188 miRNAs had differential expression above 1.5 with statistical significance. As for the HR
189 values, miR-665, miR-378c and miR-133b were down-regulated, and the others, up-regulated
190 (Shown in Fig. 2).

191

192 *Differentially expressed MiRNAs have a prevalence of negative linear correlations
193 with their target genes*

194 The miRNAs that showed prediction of interaction with driver genes, differential
195 expression above 1.5 and below -1.5 and statistical significance were imputed in the linear
196 correlation analysis. Linear correlations with their respective numerical values are shown in
197 Figure 3.

198 By analyzing the linear correlations, it was observed that miR-665 demarcates the most
199 intense negative correlation of five of the six genes under study. Furthermore, miR-378c
200 negatively correlated with two-thirds of the genes. Nevertheless, miR-133b, miR-184 and miR-
201 3161 negatively correlated with half of the target genes. These five miRNAs concomitantly
202 showed differential expression, interaction prediction and negative linear correlation with at

203 least 50% of the genes under study.

204

205 **Discussion/Conclusion**

206 The MiRNAs that showed interaction prediction, differential expression and negative
207 linear correlation with at least three of the six driver genes studied were miR-665, miR-378c,
208 miR-133b, miR-184 and miR-3161.

209 A study involving 511 patients with squamous cell carcinoma of the hypopharynx
210 showed nine candidate miRNAs as molecular markers of this pathology. Among them, miR-
211 3161 with the same expression profile observed in the present analysis [22]. The same
212 transcript has been correlated by other authors to uncontrolled cell proliferation in prostate
213 cancer. Research indicates that this RNA fragment causes the absence of cell cycle arrest
214 between the G1 and S phases [23].

215 Nevertheless, miR-133b, physiologically expressed in muscle cells, is significantly
216 suppressed in breast cancer [24]. Furthermore, it was found that the higher the tumor grade,
217 the greater the intensity of its expression suppression. Its ability to inhibit cell cloning and
218 metastases in vitro was found, as well as its potential to inhibit carcinogenesis and metastases
219 in lung tissue - one of the main sites of breast cancer metastasis [25].

220 Furthermore, the low expression of miR-133b concurrently with the high expression of
221 miR-184 was correlated with poor survival in colorectal cancer, which is also related to the
222 BRCA1 and BRCA2 genes [26]. Another study pointed out a set of miRNAs involved in the
223 development of breast cancer and promising biomarkers for its detection, one of them being
224 miR-184 [27].

225 Furthermore, miR-378c mediates the metabolic shift in breast cancer cells to alter the
226 lipid metabolism characteristic of the breast tissue, through the PGC-1 β /ERR γ transcription
227 pathway [28]. In therapy, this miRNA-mediated suppression of Runx1 has been shown to
228 alleviate the aggressive phenotype of triple-negative breast cancer cells [29]. Its reduced
229 expression has been associated with increased cell proliferation [30] and lung cancer cell
230 invasion and migration [21]. In addition, the relationship of this expression profile with poor
231 patient survival and recurrence of ovarian cancer was described [31].

232 In the same sense, miR-665 levels were significantly higher in healthy individuals and
233 reduced in individuals with tumors that were larger in size and who had metastases [32].
234 Knowing that the main sites of breast metastases are brain, lung, bone and liver, this transcript
235 is presented in the literature as a diagnostic and prognostic biomarker in hepatocellular
236 carcinoma [32], suppressor of invasion and metastasis in osteosarcomas by inhibition of
237 RAB23 [33] and inhibitor of neuroblastoma cells [34]. In studies with paraffined breast tumors,
238 it has been associated with heritability of mutations in the BRCA1/2 genes [35]. Other studies
239 have shown its regulation in the mesenchymal epithelial transition process, which is related to

240 inflammation in the tissue adjacent to the stroma and favors the processes of invasion and
241 metastasis in breast cancer [36].

242 This shows agreement with the results obtained in this research, demonstrates the
243 relevance of these miRNAs in tumor pathogenesis and development, and highlights their
244 biotechnological potential. Furthermore, knowing the mechanism of regulation of gene
245 expression by miRNA, which advocates inversely proportional expression between miRNA and
246 its target gene, it is thus suggested that up-regulated miRNAs - miR-3161 and miR-184 - are
247 possible regulators of tumor suppressor genes and that the down-regulated miRNAs – miR-
248 665, miR-378c and miR-133b, are possible oncogene regulators.

249 However, given the compatibility of the sequences of the mRNAs of tumor suppressor
250 genes with the main down-regulated miRNAs and of the mRNAs of oncogenes with the up-
251 regulated miRNAs, the possible existence of mechanisms of expression regulation by miRNAs
252 not yet elucidated. Furthermore, the existence of a positive linear correlation between down-
253 regulated miRNAs, such as miR-665, miR-378c and miR-133b - and the tumor suppressor
254 gene CDKN1A, reinforces the possible existence of mechanisms of regulation by miRNAs in
255 a positive way. As it is an imminent apparatus for regulating gene expression, further research
256 is needed for its understanding.

257 This *in silico* analysis and *in vitro* and *in vivo* research used in the theoretical foundation
258 of this study show the potential of miRNAs miR-665, miR-378c, miR-184, miR-133b and miR-
259 3161 as candidates for diagnostic, prognostic and cancer staging, as well as possible
260 contributors to therapy. In this sense, further studies are suggested to validate these findings,
261 as well as to prove the regulation mechanism by miRNA, so that they can become
262 biotechnological tools acting in precision medicine.

263

264 **Statements**

265 Statement of Ethics

266 This study has been reviewed and approved by CONEP, approval number 3,246,746.

267 **Conflict of Interest Statement**

268 The authors have no conflicts of interest to declare.

269 **Funding Sources**

270 There were no sources of funding.

271 **Author Contributions**

272 ANDREU, AM and BATISTA, TGS developed an analysis. ANDREU, AM analyzed the data.

273 ANDREU, AM and CARVALHO, CME wrote the manuscript.

274 **Data Availability Statement**

275 The datasets generated or analyzed in the present study are available from the

276 corresponding author on reasonable request.

References [Numerical]

- 1 LANETZKI, Camila Sanches et al. The epidemiological profile of pediatric intensive care center at hospital israelita albert einstein. **Einstein (São Paulo)**, v. 10, n. 1, p. 16-21, 2012.
- 2 INUMARU, Lívia Emi; SILVEIRA, Érika Aparecida da; NAVES, Maria Margareth Veloso. Fatores de risco e de proteção para câncer de mama: uma revisão sistemática. **Cadernos de Saúde Pública**, v. 27, p. 1259-1270, 2011.
- 3 INCA. Instituto Nacional de Câncer. Ministério da Saúde. Secretaria de Atenção à Saúde. Coordenação de Prevenção e Vigilância. **Estimativas 2019**. Rio de Janeiro. Disponível em: <<http://www.inca.gov.br>>, 2020. Acessado em fevereiro de 2020.
- 4 FUKUSHIMA, Kassandra Ferreira Pessoa et al. Frequência e fatores associados à síndrome da web axilar em mulheres submetidas à cirurgia de câncer de mama: estudo transversal e retrospectivo. **Springerplus**, v. 4, n. 1, p. 112, 2015.
- [5] BERTUCCI, François; BIRNBAUM, Daniel; GONCALVES, Anthony. Proteomics of breast cancer: principles and potential clinical applications. **Molecular & Cellular Proteomics**, v. 5, n. 10, p. 1772-1786, 2006.
- 6 VELASCO-VELÁZQUEZ, Marco et al. CCR5 antagonist blocks metastasis of basal breast cancer cells. **Cancer research**, v. 72, n. 15, p. 3839-3850, 2012
- 7 GUNASINGHE, NPA Devika et al. Mesenchymal–epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. **Cancer and Metastasis Reviews**, v. 31, n. 3-4, p. 469-478, 2012.
- 8 NUSBAUM, Rachel; ISAACS, Claudine. Management updates for women with a BRCA1 or BRCA2 mutation. **Molecular diagnosis & therapy**, v. 11, n. 3, p. 133-144, 2007.
- 9 OLDENBURG, R. A. et al. Genetic susceptibility for breast cancer: how many more genes to be found? **Critical reviews in oncology/hematology**, v. 63, n. 2, p. 125-149, 2007.
- 10 GHADIRIAN, P. et al. The contribution of founder mutations to early-onset breast cancer in French-Canadian women. **Clinical genetics**, v. 76, n. 5, p. 421-426, 2009.
- 11 NCBI National Center for Biotechnology Information. Disponível em: <https://www.ncbi.nlm.nih.gov/>. Acesso em: SET. 2020.
- 12 MORENO-BUENO, Gema et al. Cyclin D1 gene (CCND1) mutations in endometrial cancer. **Oncogene**, v. 22, n. 38, p. 6115-6118, 2003.
- 13 PREVO, Remko et al. A inibição de CDK1 sensibiliza células normais a danos no DNA de uma maneira dependente do ciclo celular. **Cell Cycle**, v. 17, n. 12, pág. 1513-1523, 2018.
- 14 VAN SCHOONEVELD, Eleni et al. Expression profiling of cancerous and normal breast tissues identifies microRNAs that are differentially expressed in serum from patients with (metastatic) breast cancer and healthy volunteers. **Breast cancer research**, v. 14, n. 1, p. R34, 2012.
- 15 CHAMORRO, Amanda Coelho; ALBUQUERQUE, Eliane Papa Ambrosio; DELLA ROSA, Valter Augusto. Micro-RNAs e câncer: abordagens e perspectivas/ Micro-RNAs and cancer: approaches and perspectives. **Revista de Pesquisa em Saúde**, v. 16, n. 2, 2016.
- 16 VOLINIA, Stefano et al. A microRNA expression signature of human solid tumors defines cancer gene targets. **Proceedings of the National Academy of Sciences**, v. 103, n. 7, p. 2257-2261, 2006.
- 17 YAN, Li-Xu et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. **Rna**, v. 14, n. 11, p. 2348-2360, 2008
- 18 GANDRUD, Christopher. **Reproducible research with R and R studio**. CRC Press, 2013 <http://dx.doi.org/10.1201/9781315382548>
- 19 STICHT, Carsten et al. miRWALK: An online resource for prediction of microRNA binding sites. **PLoS one**, v. 13, n. 10, 2018.
- 20 LOVE, Michael I.; HUBER, Wolfgang; ANDERS, Simon. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome biology**, v. 15, n. 12, p. 550, 2014.
- 21 XIA, Hongwei et al. The lncRNA PVT1 promotes invasive growth of lung adenocarcinoma cells by targeting miR-378c to regulate SLC2A1 expression. **Human Cell**, p. 1-10, 2020.
- 22 XU, Xinbo et al. A 3-miRNA signature predicts survival of patients with hypopharyngeal squamous cell carcinoma after post-operative radiotherapy. **Journal of Cellular and Molecular Medicine**, v. 23, n. 12, p. 8280-8291, 2019.
- 23 HERNÁNDEZ PÉREZ, Angélica Paola. Computational analysis of genomic variants affecting predicted microRNA: target interactions in prostate cancer. 2018.
- 24 LI, Xinpeng et al. LncRNA NEAT1 silenced miR-133b promotes migration and invasion of breast cancer cells. **International journal of molecular sciences**, v. 20, n. 15, p. 3616, 2019.
- 25 WANG, Qiu-Yu et al. MiR-133b targets Sox9 to control pathogenesis and metastasis of breast cancer. **Cell death & disease**, v. 9, n. 7, p. 1-14, 2018.
- 26 AKCAKAYA, Pinar et al. miR-185 and miR-133b deregulation is associated with overall survival and metastasis in colorectal cancer. **International journal of oncology**, v. 39, n. 2, p. 311-318, 2011.
- 29 BROWNE, G.; DRAGON, J.A.; HONG, D.; MESSIER, T.L.; GORDON, J.A.; FARINA, N.H.; BOYD, J.R.; VANOUDENHOVE, J.; PEREZ, A.W.; ZAIDI, S.K.; STEIN, G.S.:
- 27 FU, Lun et al. Níveis de expressão sérica de microRNA-382-3p, -598-3p, -1246 e -184 em pacientes com câncer de mama. **Cartas de oncologia**, v. 12, n. 1, pág. 269-274, 2016.
- 28 EICHNER, Lillian J. et al. miR-378* mediates metabolic shift in breast cancer cells via the PGC-1β/ERRγ transcriptional pathway. **Cell metabolism**, v. 12, n. 4, p. 352-361, 2010.
- 30 SHIVAPURKAR, Narayan et al. Circulating microRNAs in patients with hormone receptor-positive, metastatic breast cancer treated with dovitinib. **Clinical and translational medicine**, v. 6, n. 1, p. 1-10, 2017.
- 31 EITAN, Ram et al. Tumor microRNA expression patterns associated with resistance to platinum based

- chemotherapy and survival in ovarian cancer patients. **Gynecologic oncology**, v. 114, n. 2, p. 253-259, 2009.
- 32 QU, Zhen et al. Exosomal miR-665 as a novel minimally invasive biomarker for hepatocellular carcinoma diagnosis and prognosis. **Oncotarget**, v. 8, n. 46, p. 80666, 201733 DONG, Chenhui et al. MicroRNA-665 suppressed the invasion and metastasis of osteosarcoma by directly inhibiting RAB23. **American journal of translational research**, v. 8, n. 11, p. 4975, 2016.
- 34 PRASHAD, Nagindra. O miR-665 tem como alvo c-MYC e HDAC8 para inibir o crescimento celular de neuroblastoma murino. **Objetivo inicial**, v. 9, n. 69, p. 33186, 2018.
- 35 TANIC, Miljana et al. MicroRNA expression signatures for the prediction of BRCA1/2 mutation-associated hereditary breast cancer in paraffin-embedded formalin-fixed breast tumors. **International journal of cancer**, v. 136, n. 3, p. 593-602, 2015.
- 36 HU, Yuanchang et al. miR-665 promotes hepatocellular carcinoma cell migration, invasion, and proliferation by decreasing Hippo signaling through targeting PTPRB. **Cell death & disease**, v. 9, n. 10, p. 1-13, 2018.

Figure Legends

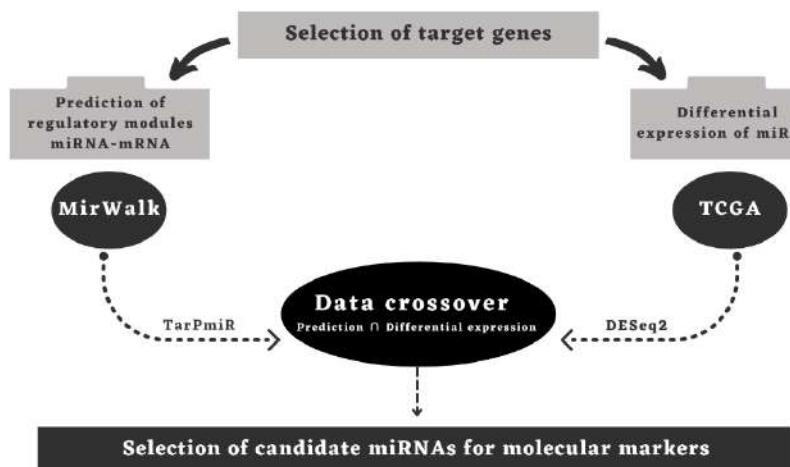


Fig. 1. Flowchart of the methodology developed in this work.

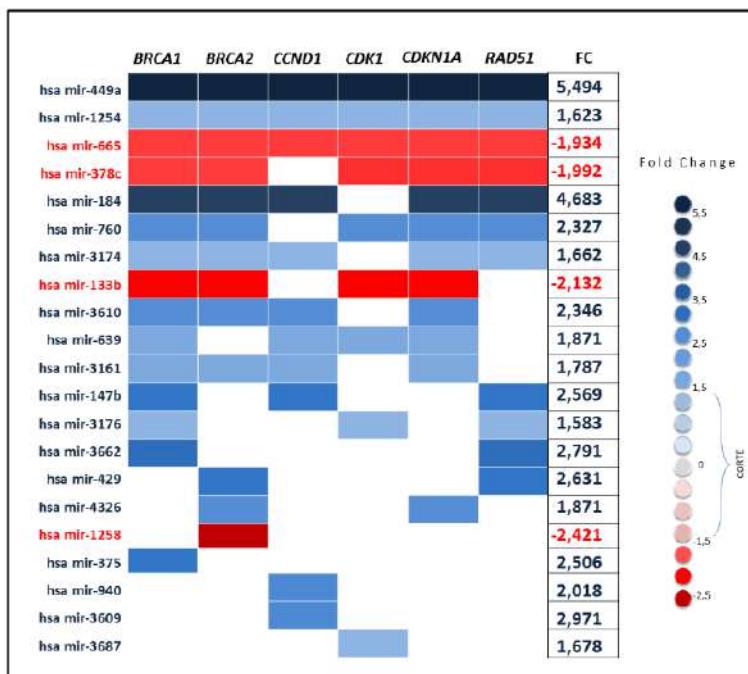


Fig. 2. MiRNAs differentially expressed in breast carcinomas. In the lines, there are the predicted miRNAs in decreasing order of number of target genes with which they were predicted to interact. In the columns, the target genes studied in this work are arranged. In the center, the spaces filled with color represent that interaction was predicted. Up-regulated and down-regulated miRNAs are, respectively, in blue and red. Color strength represents the strength of the Fold Change. On the right, the exact HR values of the miRNAs in the samples available from the TCGA are displayed.

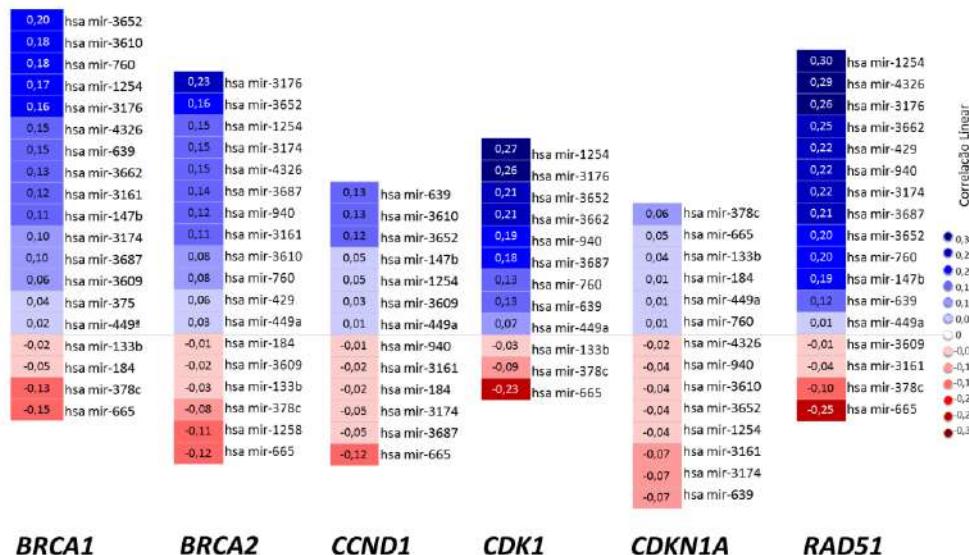


Fig. 3. Linear correlations between target genes and their miRNAs. Negative linear correlations are in red and positive ones in blue. The greater the intensity of the color, the greater the intensity of the correlation.

Table 1. Number of MiRNAs predicted for each target gene by the MirWalk platform.

Genes	Predicted miRNAs (quantity)
BRCA1	19612
BRCA2	3896
RAD51	11257
CDKN1A	9715
CCND1	2773
CDK1	5859

CAPÍTULO 2 - *ALDH2, PCK1, famílias ADH e COX: genes relacionados às alterações metabólicas no câncer de mama*

Andressa Marques Andreu

Thayane Gonçalves Da Silva Batista

Cristiano Marcelo Espinola Carvalho (Orientador)

dessamarques21@gmail.com-(67)99121-1710 / thayane.batista.ec@gmail.com –

(67) 99120-0761 / cristiano@ucdb.br –(67)99982-1997

Universidade Católica Dom Bosco. Programa de Pós-Graduação em Biotecnologia. Avenida Tamandaré, 6000 – Jardim Seminário – CEP: 79117-900. Campo Grande - MS, Brasil. Fone: 55 67 3312-3300.

Resumo

O câncer de mama desenvolve-se a partir da proliferação desregulada de células mamárias e progride a fim de adquirir capacidades específicas do fenótipo maligno que orquestram o desenvolvimento e progressão tumoral. A alta demanda energética estimula a realocação de vias metabólicas com alteração na expressão de suas enzimas a fim de originar e sustentar características do câncer. A variação no perfil de expressão desses genes pode servir como biomarcadores de diagnóstico e prognóstico na rotina médica. Assim, o presente estudo objetivou gerar um *screening* molecular de genes associados ao metabolismo energético no câncer de mama para identificar possíveis marcadores moleculares. As vias metabólicas foram obtidas pela plataforma *KEGG PATHWAYS - Kyoto Encyclopedia of Genes and Genomes Pathways* e os dados de expressão, pelo banco *The Cancer Genome Atlas Program do National Cancer Institute – TCGA*. A análise dos dados foi realizada por linha de comando na *IDE R Studio*. As vias selecionadas foram glicólise e fosforilação oxidativa. Foram obtidos os genes referentes a cada via metabólica e avaliado a expressão diferencial com os dados disponíveis no TCGA. As análises de expressão diferencial foram realizadas por meio do algoritmo *DESeq2*. Realizou-se teste estatístico de correlação linear entre os genes de cada via. Os gráficos foram gerados por meio de pacotes disponíveis no repositório *Bioconductor: VolcanoPlot* e *GGCorPlot*. Os genes *ADH1A, ADH1B, ALDH2, PCK1* referentes à via da glicólise – e *COX7A1* e *COX6A2* – relacionados à fosforilação oxidativa, apresentaram valores de correlação linear positiva acima de 80% entre si. Esta interdependência e a variação no perfil de expressão observada entre o tecido mamário normal e canceroso sugerem a relevância na patogênese tumoral e podem ser utilizados como biomarcadores de diagnóstico, prognóstico e estadiamento no câncer de mama. Desse modo, corroboraram com o surgimento de novas ferramentas biotecnológicas e com a medicina de precisão.

Introdução

A neoplasia mais incidente em mulheres e com taxas crescentes a cada ano é o câncer de mama. Com cerca de 73.610 novos casos estimados para cada ano do triênio 2023-25 (INCA, 2023), deixou de ser considerada uma doença única com diferenças histopatológicas e variável prognóstico. Devido ao advento das novas tecnologias, atualmente o câncer de mama é definido como um conjunto de doenças heterogêneas que afetam o mesmo órgão, tendo a mesma origem estrutural e diferentes características histopatológicas, fatores de risco e respostas às terapias (FUKUSHIMA, 2015).

Além disso, dados revelam que o câncer de mama é o tipo de neoplasia mais frequente no mundo (INCA, 2023). Sua heterogeneidade é compreendida a partir de seus aspectos morfológicos, imunohistoquímicos, clínicos e genéticos específicos. Porém, mesmo com os importantes avanços para sua detecção, diagnóstico e tratamento em estágios iniciais, esse câncer ainda é considerado uma das principais doenças potencialmente letais devido ao seu diagnóstico tardio com incidência de 41,89 casos por 100.000 mulheres (INCA, 2022).

No contexto do câncer, diversos mecanismos fisiológicos são alterados para o seu surgimento. Um dos principais destes é a desregulação do metabolismo energético. A teoria canônica para explicar o metabolismo é embasada nos estudos de Otto Warburg. Pioneiro no estudo da respiração, Warburg constatou o aumento excessivo da absorção de glicose pelas células tumorais, o que é associado ao metabolismo elevado dessas células em comparação com as normais. Para atender às demandas obrigatórias do câncer, as células cancerosas aumentam a absorção e metabolização desse carboidrato, o que possibilita avaliações radiográficas por exames de monitoramento como a tomografia por emissão de pósitrons pela fluorodeoxiglucose (HSU; SABATINI, 2008).

Porém, o que não se esperava da teoria é que a glicose não seria metabolizada convencionalmente. Warburg descobriu que as células preferem metabolizar a glicose pela glicólise não seguindo à fosforilação oxidativa. Vale ressaltar que, no que se refere a saldo energético, o processo de glicólise proporciona a obtenção de duas moléculas de ATP, ao passo que a fosforilação oxidativa, 26 moléculas de ATP. Tal comportamento parece paradoxal ao metabolismo do câncer. Alguns autores explicaram que, devido ao microambiente tumoral ser muito denso, há redução do suprimento de vasos sanguíneos no local e os distúrbios metabólicos nas células

cancerosas surgem da seleção natural de células adaptadas ao microambiente. No entanto, essa evidência foi verificada mesmo na presença de oxigênio anos mais tarde (LIBERTI; LOCASALE, 2020).

Diante disso, o presente estudo objetivou gerar um *screening* molecular de genes associados ao metabolismo energético no câncer de mama, a partir de análises bioinformáticas, a fim de identificar possíveis marcadores moleculares para a detecção precoce e estadiamento de carcinomas mamários.

Metodologia

Os novos alvos moleculares foram selecionados por meio de análises bioinformáticas combinadas. Os dados de expressão do transcriptoma de amostras de carcinomas mamários foram analisados, a fim de gerar um *screening* de alvos moleculares diferencialmente expressos. Esses dados foram provenientes de bancos de dados públicos específicos (TCGA e KEGG Pathway) e a análise foi realizada por intermédio de algoritmos escritos na linguagem R, executados por meio da IDE R Studio.

2.1 Escolha dos bancos de dados

Os bancos de dados foram escolhidos de acordo com a quantidade e qualidade dos dados, a frequência com que os bancos foram citados na literatura, a frequência da atualização dos dados disponibilizados e a facilidade para acesso e *download* desses dados. Os bancos selecionados possuem *links* diretos para acesso e *download* a partir do ambiente do *RStudio*, por meio de *APIs* em pacotes específicos, disponíveis no repositório *Bioconductor*.

2.1.1 Banco de dados de transcriptomas das amostras de carcinomas mamários: TCGA – The Cancer Genome Atlas Program do National Cancer Institute

O TCGA foi escolhido mediante a quantidade de amostras por tipo tumoral disponibilizado; a qualidade dos dados, visto que, é um banco curado e está constantemente sendo atualizado; e devido aos pacotes disponibilizados pelo *Bioconductor*, os quais permitem acessar e baixar os dados de expressão requeridos diretamente pelo *RStudio*. Sendo um banco de dados do *National Cancer Institute*, o TCGA é alimentado por vários grupos de pesquisa de diversos países (TOMCZAK; CZERWINSKA; WIZNEROWICZ, 2015).

**2.1.2 Bancos de dados de vias metabólicas para enriquecimento funcional:
KEGG PATHWAYS - Kyoto Encyclopedia of Genes and Genomes Pathways**

O mapa de vias KEGG PATHWAYS é um diagrama de rede de interação/reAÇÃO molecular utilizado em experimentos de organismos específicos com informações genéticas que podem ser generalizados a outras espécies. Todos os mapas são desenhados manualmente com o software interno chamado KegSketch, que gera o arquivo KGML +, disponibilizado para download. Os mapas podem ser acessados na caixa de pesquisa na parte superior da página KEGG PATHWAYS ou pela ferramenta KEGG Mapper, ou ainda, por linha de comando na *IDE RStudio*, com foi utilizado neste trabalho (KANEHISA et al; 2016).

2.2 Seleção das vias metabólicas

As vias metabólicas foram escolhidas de acordo com as principais formas de obtenção de energia das células e que aparecem alteradas no contexto do câncer de mama. Foram selecionadas as duas principais vias do metabolismo energético para serem exploradas neste trabalho: Glicólise e Fosforilação Oxidativa. Os genes de ambas as vias disponibilizados no banco KEGG PATHWAYS foram acessados por linha de comando para a execução das filtragens nas análises posteriores.

2.3 Filtragem/triagem dos dados

Para gerar uma lista de genes de cada via foram utilizadas funcionalidades básicas de programação na linguagem R, por meio da *IDE RStudio*. A partir das bibliotecas *TCGAbiolinks* e *KEGGREST* foi possível acessar os dados disponíveis nos bancos de dados. A filtragem dos dados de expressão foi realizada a partir da biblioteca *KEGGREST*, que cruzou os dados do TCGA com o KEGG PATHWAYS filtrando apenas os dados referentes aos genes de cada via escolhida.

2.4 Análise dos dados

A análise dos dados foi realizada por meio da linguagem de programação R, a qual permitiu analisar os bancos de dados primários e secundários e interagir com pacotes pré-desenvolvidos por cientistas de vários países. O uso da linguagem R no Ambiente de Desenvolvimento Integrado – *IDE RStudio* possibilitou a utilização dos códigos e banco de dados públicos por linha de comando e permitiu a construção de análises estatisticamente comprovadas e tecnologicamente relevantes.

2.4.1 Análise da Expressão Diferencial

A expressão diferencial dos genes de cada via foi avaliada em grupos de

amostras de tumor primário (TP) vs tecido mamário não-tumoral (NT) disponíveis no TCGA, por meio do algoritmo *DESeq2* (LOVE; HUBER; ANDERS, 2014). Dentre vários objetos que foram gerados como *output* pelo *DESeq*, um *data frame* com valores de quantificação absoluta normalizados e valores de *Fold Change* (FC) para cada gene do transcriptoma foram gerados. FC corresponde a objetos numéricos que representam o valor da quantificação relativa de cada gene sobre um calibrador. Foi usado como calibrador o valor de expressão dos genes em amostras de tecido mamário normal. O *DESeq2* é um algoritmo amplamente utilizado para estimação de FC e dispersão de dados de *RNAseq*. Os valores de quantificação relativa foram obtidos em ensaios de expressão gênica por meio das metodologias de qPCR e *microarray* ou por meio de análises *in silico*, com a utilização de algoritmos específicos como o *DESeq*. O uso deste algoritmo permitiu realizar diversas análises comparativas de expressão diferencial, uma vez que, a partir dos dados de *RNAseq*, cada amostra possui seu valor de quantificação absoluta para cada gene do transcriptoma. Atribui-se um valor de corte 1,2 para a expressão diferencial, buscando incluir na análise genes que tivessem expressão diferencial discrepante entre tecido mamário normal e tumoral. A análise de expressão diferencial dos genes no *DESeq2* foi rodada uma única vez para todas as vias, enquanto que a filtragem dos genes por via escolhida foi realizada posteriormente ao *DESeq2* ter sido rodado. Os resultados da análise de expressão diferencial filtrada pelos genes das vias metabólicas escolhidas, foram exibidos em gráficos de dispersão com o uso do pacote *VolcanoPlot*.

2.4.2 Análise de Correlação Linear

A fim de enriquecer a análise e buscar identificar interdependência no perfil de expressão entre os genes de cada uma das vias, os dados de expressão normalizados pelo *DESeq2* foram submetidos à análise de correlação linear. A análise foi rodada separadamente por via e expostos em gráficos gerados por meio da utilização do pacote *GGCorPlot*.

2.5 Seleção dos genes com potencial a marcadores moleculares

Foram selecionados quatro genes referentes à via glicolítica e dois referentes a fosforilação oxidativa, perfazendo um total de seis genes. Foram escolhidos genes que apresentaram variação no perfil de expressão acima de 120% entre tecido normal e tumoral de mama e que apresentaram correlação linear positiva acima de 80% entre

si.

A metodologia desenvolvida neste trabalho está esquematizada na figura 1.

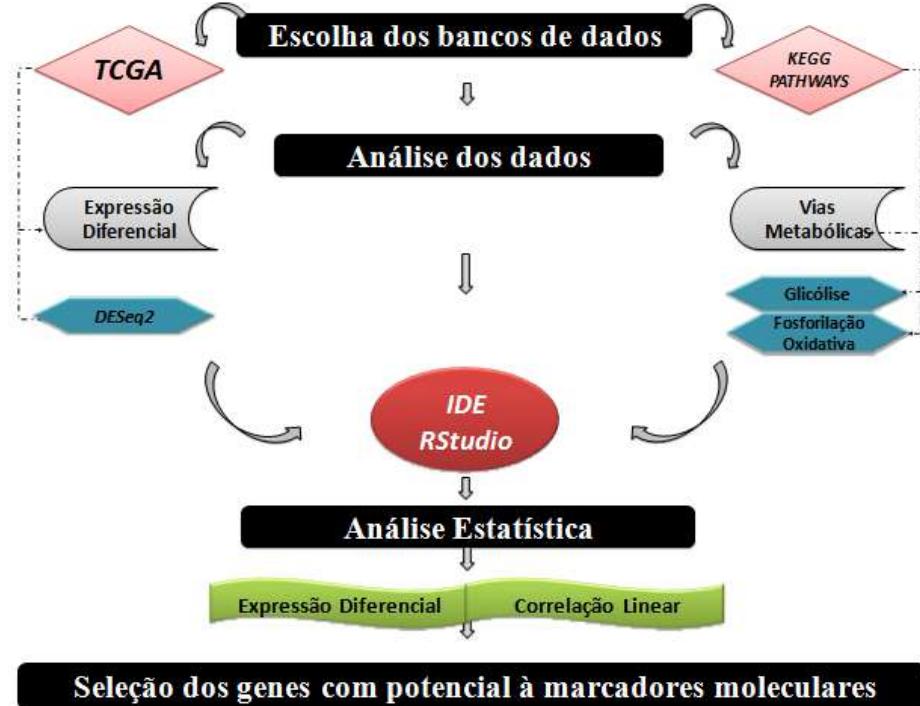


Figura 1 - Fluxograma da metodologia realizada neste trabalho. A pesquisa consistiu na escolha dos bancos de dados, seguida da análise e manipulação dos dados que foram submetidos a testes estatísticos possibilitando a seleção de genes potenciais marcadores moleculares do metabolismo energético no câncer de mama.

Resultados

3.1 Seleção de Vias do Metabolismo Energético: Glicólise e Fosforilação Oxidativa

No presente trabalho, foram selecionadas 2 vias relacionadas ao metabolismo energético: Glicólise e Fosforilação Oxidativa.

Estas vias foram selecionadas em decorrência aos estudos de Otto Warburg, no qual foi identificado que o câncer aumenta a absorção de glicose para as células tumorais para atender as necessidades de multiplicação exacerbada, angiogênese, fuga da detecção pelo sistema imunológico e diversas outras características. Contudo, esse carboidrato não é metabolizado da forma convencional. Segundo Warburg, as células passam a metabolizar glicose pela glicólise não seguindo para a fosforilação oxidativa. Entretanto, no processo da fosforilação oxidativa há a obtenção de 26 moléculas de ATP, enquanto no processo da glicólise é adquirido apenas duas. Isto pode ser devido à uma redução do suprimento de oxigênio nos vasos sanguíneos no tecido tumoral, criando uma seleção natural onde se originam distúrbios metabólicos como esse devido a adaptação das células a esse microambiente ou mesmo como

consequência da sinalização aberrante devido à ativação oncogênica. Este distúrbio metabólico foi identificado na presença de oxigênio anos mais tarde (LIBERTI; LOCASALE, 2020).

3.2 Expressão Diferencial de Genes das Vias de Metabolismo Energético

3.2.1 Glicólise

No gráfico de dispersão *VolcanoPlot* da via da glicólise (figura 2), encontrou-se variação no perfil de expressão de 68 genes. Como critério de análise, foi estabelecido valor de corte para variações de expressão acima de 120%, para mais ou para menos. Dessa forma, os genes *ENO2*, *PKLR* e *PGK2* apresentaram perfil de expressão *up-regulated*, ou seja, superexpressos, com valores de FC entre 1,2 e 3,5. Já os genes *ADH1A*, *ADH1B*, *ADH4*, *ADH1C*, *ALDH2*, *PCK1*, *ALDOC* e *ACSS2* mostraram-se *down-regulated*, ou seja, com expressão reduzida. Seus valores de FC ficaram entre -1,2 a aproximadamente -5.

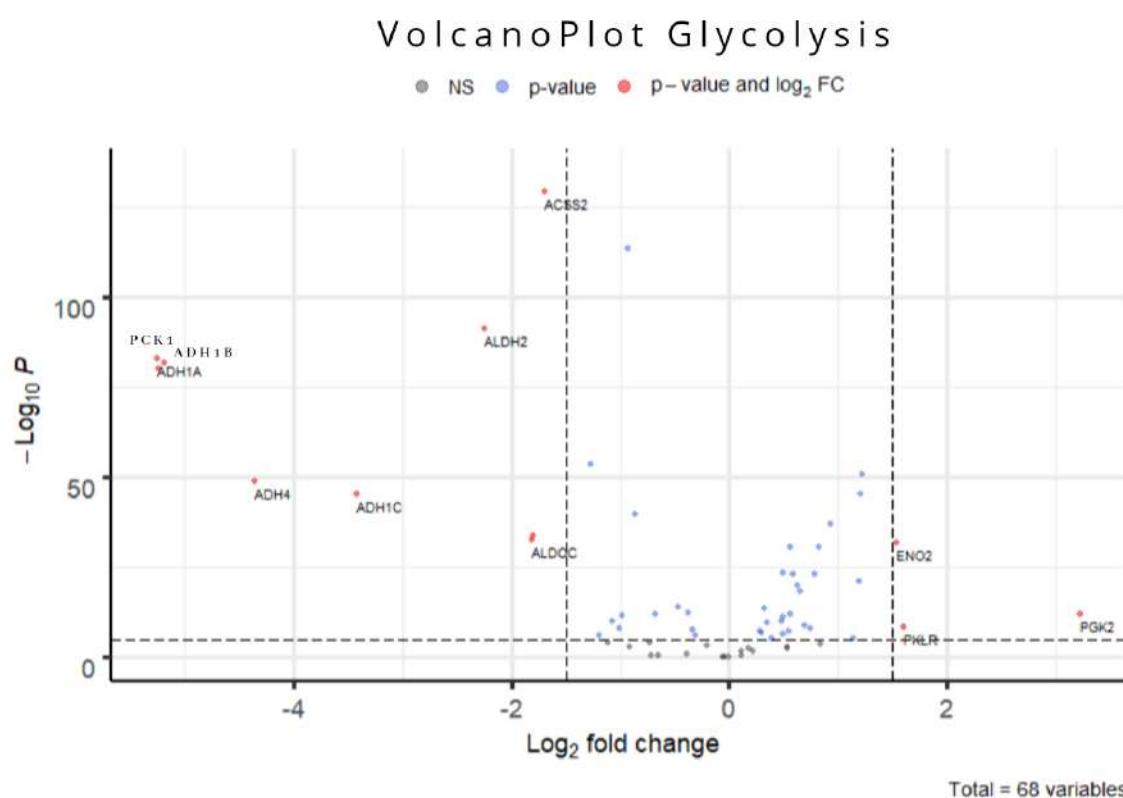


FIGURA 2. Gráfico de expressão diferencial da via da Glicólise. O gráfico, obtido pelo pacote VolcanoPlot, traça valores de FC (eixo X) e valores de significância (eixo Y). Os genes nas extremidades direita e esquerda são os que apresentaram diferença em perfil de expressão entre o tecido normal e o tumoral. Destaque para as extremidades em ambos os eixos; valor de corte Fold Change: 1,5.

3.2.2 Fosforilação oxidativa

No gráfico *VolcanoPlot* da via da fosforilação oxidativa (figura 3), encontrou-se variação no perfil de expressão de 133 genes. Como critério de análise, foi estabelecido valor de corte para variações de expressão acima de 120%, para mais ou para menos. Dessa forma, os genes COX6A2 e COX6 apresentaram perfil de expressão *up-regulated*, ou seja, superexpressos, com valores de FC acima de 2,5. Já o gene COX7A1 mostrou-se *down-regulated*, ou seja, com expressão reduzida. Seu valor de FC foi aproximadamente -2,5.

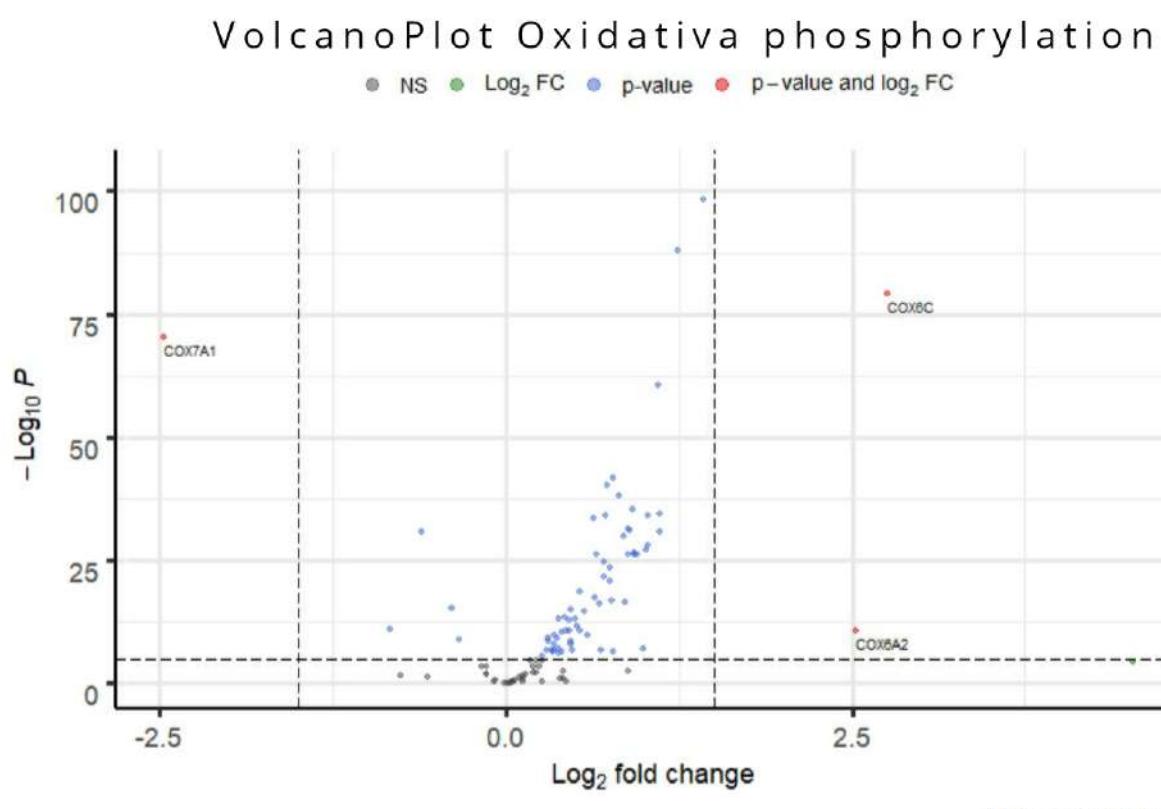


FIGURA 3 - Gráfico de expressão diferencial da via fosforilação oxidativa. O gráfico, obtido pelo pacote VolcanoPlot, traça valores de FC (eixo X) e valores de significância (eixo Y). Os genes nas extremidades direita e esquerda são os que apresentaram diferença em perfil de expressão entre o tecido normal e o tumoral. À esquerda, COX7A1 apresentou-se down-regulated, enquanto que à direita, COX6A2 e COX6 apresentaram-se Up-regulated.

3.3 Avaliação da existência de Correlação Linear entre os dados de expressão

Com o intuito de avançar no compreendimento dos fatores biológicos responsáveis pela regulação de genes com expressão alterada no câncer, analisamos as correlações lineares existentes entre os genes selecionados na via da Glicólise e Fosforilação Oxidativa.

Com o algoritmo *GGCorrPlot* foi possível montar gráficos baseados na busca de uma correlação linear entre os dados de expressão gênica dos genes-alvos relacionados às vias metabólicas selecionadas. Os dados imputados no algoritmo foram obtidos como resultado do *DESeq2*, sendo uma matriz com valores de quantificação absoluta normalizada para os genes-alvos com expressão diferencial em todas as amostras (TP e NT).

O valor de r está sempre entre 1 e -1. Se r está próximo de 1, há uma forte correlação positiva. Se r está próximo a -1, há uma forte correlação negativa. Caso o r esteja próximo de 0, não há correlação linear entre os valores analisados.

3.3.1 Glicólise

O gráfico *GGCorrPlot* de correlação da expressão de genes relacionada à via da glicólise (figura 4) apresentou correlação linear entre 66 genes, perfazendo um total de 2178 correlações lineares. Destas, 1230 são correlações lineares positivas (56,47%), 886, correlações lineares negativas (40,68) e 62 não apresentaram correlação (2,85%), com valor igual à zero.

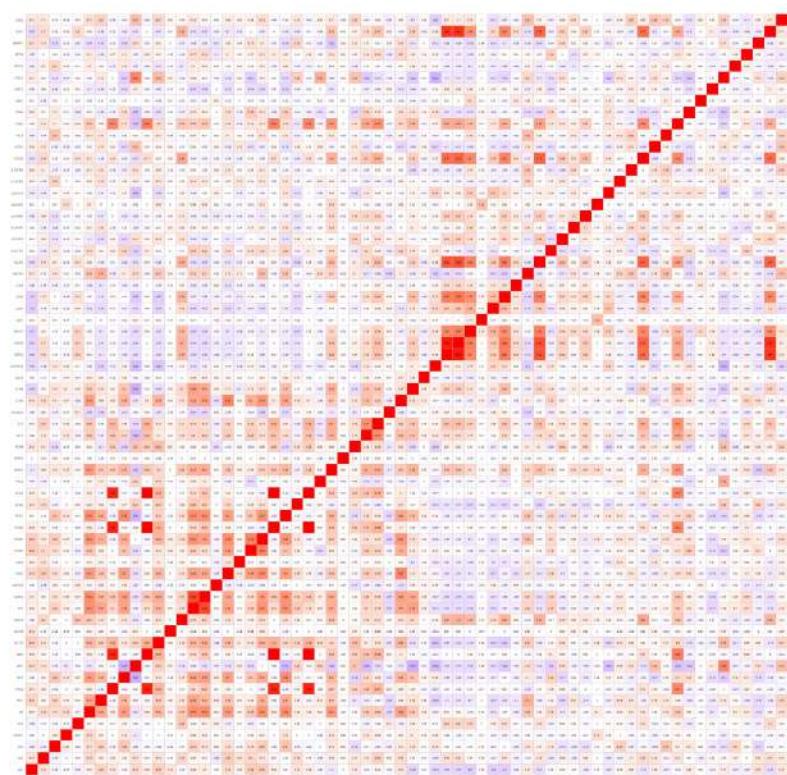


FIGURA 4. Gráfico de correlação linear da via Glicólise. Foi obtido por intermédio do pacote *GGCorrPlot* sendo baseado na correlação linear do perfil de expressão de genes relacionados à via da glicólise. As correlações lineares positivas estão em vermelho, as negativas, em azul e os genes não correlacionados, em branco. Quanto mais intensa a correlação positiva e/ou negativa, maior a intensidade das cores.

Na tabela a seguir (tabela 1), estão apresentadas as correlações lineares

positivas e negativas mais intensas e seus respectivos valores numéricos de correlação.

Tabela 1. As doze correlações lineares positivas e negativas mais intensas entre os genes referentes à via da glicólise e seus respectivos valores de correlação.

CORRELAÇÕES LINEARES			
POSITIVAS		NEGATIVAS	
Genes Correlacionados	Valores de correlação	Genes Correlacionados	Valores de correlação
<i>PGAM2/FBP2</i>	1,0	<i>FBP1/PFKP</i>	-0,34
<i>ENO3/PGAM2</i>	1,0	<i>FBP1/LDHB</i>	-0,34
<i>ENO3/FBP2</i>	1,0	<i>G6PC3/LDHAL6B</i>	-0,33
<i>ADH1A/ADH1B</i>	0,98	<i>ENO1/FBP1</i>	-0,31
<i>PFKM/ENO3</i>	0,96	<i>FBP1/DLAT</i>	-0,28
<i>PFKM/PGAM2</i>	0,96	<i>FBP1/LDHAL6B</i>	-0,28
<i>PFKM/FBP2</i>	0,96	<i>FBP1/ALDH1A3</i>	-0,26
<i>ADH1B/PCK1</i>	0,93	<i>FBP1/PGM1</i>	-0,25
<i>ADH1A/PCK1</i>	0,92	<i>LDHAL6B/ALDOA</i>	-0,24
<i>TPI1/GAPDH</i>	0,86	<i>G6PC3/PGM2</i>	-0,24
<i>ADH1B/ALDH2</i>	0,84	<i>FBP1/PGM2</i>	-0,23
<i>ADH1A/ALDH2</i>	0,82	<i>MINPP1/PDHA1</i>	-0,23

3.3.2 Fosforilação oxidativa

O gráfico *GGCorPlot* de correlação da expressão de genes relacionados à via da fosforilação oxidativa (figura 5) apresentou correlação linear com 104 genes, perfazendo um total de 5408 correlações lineares. Destas, 4534 são correlações lineares positivas (83,84 %), 809, correlações lineares negativas (14,96 %) e 65 não apresentaram correlação (1,2 %), estando no gráfico com valor igual à zero.

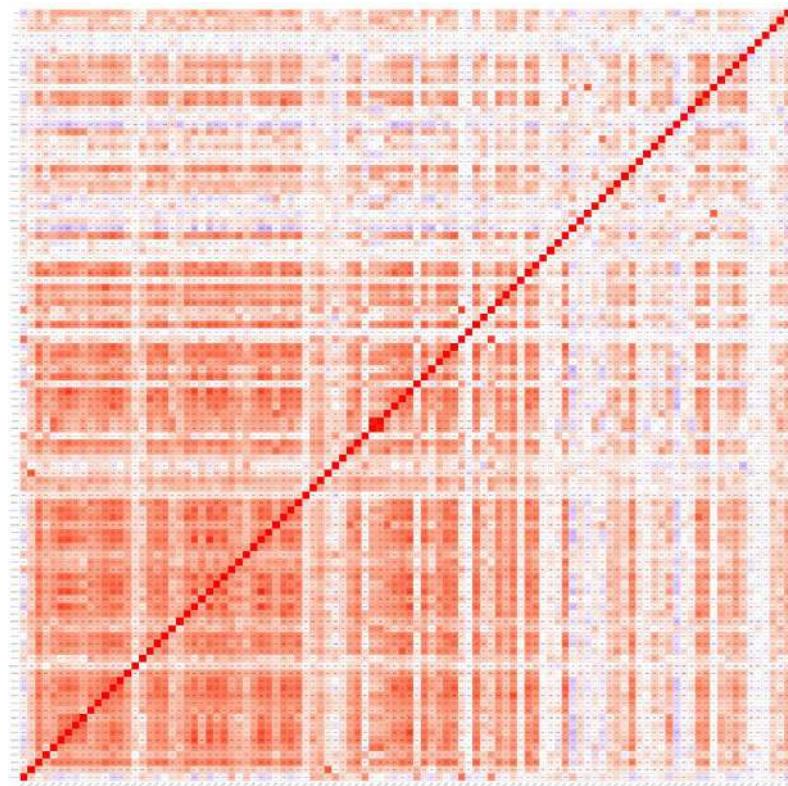


FIGURA 5 - Gráfico de correlação linear da via fosforilação oxidativa. Foi obtido por intermédio do pacote GGCorPlot sendo baseado na correlação linear do perfil de expressão de genético da via fosforilação oxidativa. As correlações lineares positivas estão em vermelho, as negativas, em azul e os genes não correlacionados, em branco. Quanto mais intensa a correlação positiva e/ou negativa, maior a intensidade das cores.

Na tabela a seguir (tabela 2), estão apresentadas as correlações lineares positivas e negativas mais intensas e seus respectivos valores numéricos de correlação:

Tabela 2. As doze correlações lineares positivas e negativas mais intensas entre os genes referentes à via fosforilação oxidativa e seus respectivos valores de correlação.

CORRELAÇÕES LINEARES			
POSITIVAS		NEGATIVAS	
Genes Correlacionados	Valores de correlação	Genes Correlacionados	Valores de correlação
<i>UQCRH/UQCRHL</i>	0,97	<i>ATP6AP1/SDHD</i>	-0,31
<i>NDUFB7/NDUFA13</i>	0,93	<i>ATP6V1A/NDUFS7</i>	-0,31
<i>COX7A1/COX6A2</i>	0,88	<i>ATP6V1A/COX4/2</i>	-0,29
<i>UQCR11/NDUFS7</i>	0,88	<i>ATP6V1A/NDUFB7</i>	-0,29
<i>NDUFB7/NDUFA11</i>	0,88	<i>ATP6V1A/NDUFB8</i>	-0,29
<i>NDUFA13/NDUFA11</i>	0,88	<i>ATP6V1A/NDUFA11</i>	-0,28
<i>UQCR11/NDUFA13</i>	0,85	<i>ATP6V0A2/UQCRQ</i>	-0,28
<i>NDUFA11/NDUFS6</i>	0,85	<i>ATP6V0A2/NDUFS7</i>	-0,26
<i>UQCR11/NDUFA11</i>	0,84	<i>ATP6V1A/NDUFA13</i>	-0,26
<i>UQCR11/NDUFB7</i>	0,83	<i>ATP6V1A/UQCR11</i>	-0,24
<i>NDUFS7/NDUFB7</i>	0,83	<i>ATP6V1A/NDUFV1</i>	-0,24
<i>NDUFS7/NDUFA13</i>	0,83	<i>ATP6V0A2/NDUFA10</i>	-0,24

Discussão

Dos genes referentes à via da glicólise, *ENO2*, *PKLR* e *PGK2* foram os que apresentaram aumento de expressão no tecido mamário tumoral em relação ao tecido mamário normal, com significância estatística. Localizados nos cromossomos 12 (p13.31), 1 (q22) e 6 (p12.3), respectivamente, codificam proteínas com funções catalíticas responsáveis por diferentes etapas dessa via metabólica as quais são alteradas no metabolismo energético do câncer (NCBI; 2021).

Envolvida na quarta etapa da sub-via que sintetiza piruvato a partir de D-gliceraldeído 3-fosfato na via glicolítica, a enolase 2 (*ENO2*) se relaciona com a transição epitélio mesenquimal e angiogênese no câncer de mama (KIM et al, 2019). Além disso, apresenta valor prognóstico no câncer colorretal (PAN et al, 2020). Já a piruvato quinase (*PKLR*), que participa da quinta etapa da mesma sub-via de síntese de piruvato, quando desregulada, mostrou correlação com a manutenção do estado transformado do câncer de mama, impedindo o processo de apoptose (JAGUS; JOSHI; BARBER, 1999). Em estudos posteriores, o quimioterápico 5-fluorouracil promoveu apoptose mediada por esta enzima de maneira independente da p53 em células de câncer de cólon e mama (GARCIA et al, 2011). Não obstante, a fosfoglicerato quinase 2 (*PGK2*), relacionada com o processo de espermatogênese e responsável pela segunda etapa da glicólise, apresenta regulação positiva em cânceres de pulmão, cólon e fígado (SCATENA, 2008).

Em relação aos genes da via glicolítica regulados negativamente, encontram-se *ADH1A*, *ADH1B*, *ADH1C*, *ADH4*, *ALDH2*, *PCK1*, *ALDOC* e *ACSS2*. Localizam-se nos cromossomos 4 (q23 – os quatro primeiros pertencentes à família ADH), 12 (q24.12 - *ALDH2*), 17 (q11.2 - *ALDOC*) e 20 (q13.31 - *PCK1* / q11.22 - *ACSS2*) (NCBI,2020).

Os genes ADH codificam enzimas glicolíticas conhecidas como álcool desidrogenase e são sintetizadas primordialmente no fígado. Desenvolvem atividade catalítica na transformação de uma molécula de álcool em aldeído, com a adição de um H⁺ ao carreador de elétrons NAD, gerando NADH⁺⁺H⁺ (DE ANDRADE, 2017). A família gênica, principal responsável pela oxidação de álcool etílico em ácido acético, é constituída por 8 genes (MARTINS, 2013) dos quais 4, apareceram regulados negativamente no presente trabalho.

Mutações na família ADH são fortemente correlacionadas ao risco de

desenvolvimento de câncer de mama, embora o mecanismo carcinogênico nos possíveis subtipos tumorais não seja muito bem esclarecido (DENNIS et al., 2010; MARTINS, 2013). Mecanismos sugeridos referem-se à mutagênese ocasionada pelo acúmulo de acetaldeído, por perturbação do metabolismo em resposta ao hormônio estrogênio - causada também pelo acetaldeído, pela indução de estresse oxidativo e/ou afetando as vias do folato e do metabolismo do carbono, unidade constituinte fundamental da glicose (DUMITRESCU, 2005; DE ANDRADE, 2017).

Um estudo que avaliou o perfil de expressão de cânceres de mama em mulheres indianas revelou a baixa expressão dos genes *ADH1A* e *ADH1B* e constatou sua correlação com a desregulação dos processos de ciclo celular, adesão, metástase, replicação do DNA, sinalização PPAR e metabolismo do lipídio (MALVIA, 2019). Outro estudo envolvendo 1969 pacientes e 2244 controles com 7 populações diferentes sugeriu que o alelo mutante de *ADH1C* pode influenciar o efeito do álcool na carcinogênese e no desenvolvimento do câncer mamário (MAO, 2015). Além disso, a baixa expressão do gene *ADH4*, potencial biomarcador em carcinoma hepatocelular (WEI, 2012), foi correlacionada com metástases em linfonodos axilares e com graus mais avançados do carcinoma mamário, os quais apresentam a ocorrência de metástases (ZHANG et al, 2011).

Ademais, a aldeído desidrogenase 2 (*ALDH2*) atua na segunda etapa da via secundária que sintetiza o acetato a partir do etanol. Mutações e polimorfismos desta enzima têm sido associados ao desenvolvimento de diversos cânceres, como o de mama (KAWASE, 2009), colorretal, esôfago e fígado (LI, 2016). Neste mesmo sentido, a fosfoenolpiruvato carboxicinase citosólica (*PCK1*) que catalisa a descarboxilação e fosforilação reversíveis de oxaloacetato, atua como a enzima limitadora na gliconeogênese. Expressões alteradas do gene *PCK1* são encontradas nos cânceres de vários órgãos, incluindo cólon, pulmão e pele, e estão ligadas ao aumento do metabolismo anabólico e da proliferação celular. Baixa expressão deste gene é sugerida como marcador molecular de diagnóstico em câncer de fígado além de ser correlacionada com o mau prognóstico dos pacientes (LIU, 2018).

Com relação aos genes referentes à via da fosforilação oxidativa, três pertencentes à família citocromo oxidase apresentaram-se diferencialmente expressos: *COX7A1*, com expressão reduzida e, *COX6A2* e *COX6*, com expressão elevada. São localizados respectivamente nos cromossomos 19 (q13.12), 16 (p11.2) e 8. As proteínas desta família formam um complexo multimérico de subunidades que

acopla a transferência de elétrons do citocromo c para o oxigênio molecular e favorece o gradiente eletroquímico de prótons por meio da membrana mitocondrial interna para a síntese de ATP (NCBI,2020). Este complexo enzimático desempenha papel fundamental na regulação da produção de energia mitocondrial. Mostra-se indispensável no desenvolvimento de células-tronco embrionárias, na formação de prostaglandinas em estados patogênicos e em processos inflamatórios. Apresenta correlação com a tumorigênese de câncer colorretal (DIXON, 2004).

O polipeptídeo 7A1 do citocromo c oxidase (COX7A1) apresenta-se superexpresso fisiologicamente em tecidos com alta demanda energética, como cérebro, coração, fígado e, patologicamente, em formas agressivas de células de câncer de mama. Estudos sugerem que a variação de ganho e perda de sua função evidencia que esta subunidade é fundamental para que as células do carcinoma mamário mantenham taxas mais elevadas de proliferação, formação de clones e invasão (ZHANG, 2016). Por outro lado, a superexpressão do gene COX6A2 foi correlacionada com o metabolismo aberrante de lipídio por células do câncer mamário (VISWESWARAN, 2020), o que pode estar relacionado com o emagrecimento das pacientes. Já a subunidade COX6, teve sua superexpressão verificada em carcinoma espinocelular ex-adenoma pleomórfico (MARIANO, 2018).

Diante do exposto, este estudo evidencia que as famílias gênicas ADH e COX, responsáveis pela expressão dos grupos enzimáticos álcool desidrogenase e citocromo c oxidase são fundamentais para as alterações no metabolismo energético que originam e sustentam o câncer de mama. Analisando-se os dados obtidos na correlação linear (tabela 3), evidencia-se a interdependência entre os genes *ADH1A* e *ADH1B* com 98% de linearidade positiva e *COX7A1* e *COX6A2*, com 88%. Além disso, *ALDH2* apresentou correlação linear positiva de 82% com *ADH1A* e 84% com *ADH1B*. Não obstante, *PCK1* também correlacionou-se positivamente com esses dois genes, apresentando 92 e 93% de linearidade com *ADH1A* e *ADH1B*, respectivamente.

Tabela 3. Correlação linear entre os genes selecionados, separados por via.

	Glicólise				Fosforilação Oxidativa		
	Genes	<i>ADH1A</i>	<i>ADH1B</i>	<i>ALDH2</i>	<i>PCK1</i>	<i>COX7A1</i>	<i>COX6A2</i>
Glicólise	<i>ADH1A</i>	1	98%	82%	92%	-	-
	<i>ADH1B</i>	98%	1	84%	93%	-	-
	<i>ALDH2</i>	82%	84%	1	76%	-	-
	<i>PCK1</i>	92%	93%	76%	1	-	-
Fosforilação Oxidativa	<i>COX7A1</i>	-	-	-	-	1	88%
	<i>COX6A2</i>	-	-	-	-	88%	1

Esta interdependência com significância biológica e estatística, bem como a intensa variação no perfil de expressão destes genes entre o tecido mamário normal e canceroso sugere que estes genes apresentam potencial como biomarcadores para diagnóstico, prognóstico e estadiamento no câncer de mama.

Conclusão

As análises bioinformáticas combinadas empregadas no presente estudo vêm de encontro com os resultados obtidos em validações *in vitro* e estudos clínicos encontradas na literatura, que apresentam mesmo grau de confiabilidade e significância estatística, que o algoritmo *DESeq* para a obtenção de valores de quantificação relativa. Diante do exposto, pode-se inferir que, neste estudo, os genes *ADH1A*, *ADH1B*, *ALDH2*, *PCK1*, *COX7A1* e *COX6A2* apresentaram potencial como biomarcadores de diagnóstico, prognóstico e estadiamento no câncer de mama e podem corroborar com o surgimento de novas ferramentas biotecnológicas e com a medicina de precisão. Validações *in vitro* para esses genes selecionados serão realizadas em trabalhos posteriores nas amostras provenientes das pacientes sul-mato-grossenses a fim de avaliar as evidências aqui identificadas a nível celular.

Referências

- ABDUL-RASOOL, S. et al. An evaluation of molecular markers for improved detection of breast cancer metastases in sentinel nodes. **Journal of clinical pathology**, v. 59, n. 3, p. 289-297, 2006.
- BALMAIN, Allan; Gray, Joe; PONDER, Bruce. The genetics and genomics of cancer. **Nature genetics** , v. 33, n. 3, p. 238-244, 2003.
- BERTUCCI, François; BIRNBAUM, Daniel; GONCALVES, Anthony. Proteomics of breast cancer: principles and potential clinical applications. **Molecular & Cellular Proteomics**, v. 5, n. 10, p. 1772-1786, 2006.
- DE ANDRADE, Ikaro Alves et al. BRCA1, BRCA2, ALDH and ADH Family: Genes related to alcoholism and female breast cancer. **Revista de Patologia do Tocantins**, v. 4, n. 1, pág. 39-64, 2017.
- DENKO, Nicholas C. Hypoxia, HIF1 and glucose metabolism in the solid tumour. **Nature Reviews Cancer**, [s.l.], v. 8, n. 9, p.705-713, 14 ago. 2008. **Springer Science and Business Media LLC**. <http://dx.doi.org/10.1038/nrc2468>.
- DENNIS, Jessica, et al. "Alcohol consumption and the risk of breast cancer among BRCA1 and BRCA2 mutation carriers." **The Breast** 19.6 (2010): 479-483.
- DIXON, Dan A. Deregulated post-transcriptional control of COX-2 gene expression in cancer. **Projeto farmacêutico atual** , v. 10, n. 6, pág. 635-646, 2004.
- FUKUSHIMA, Kassandra Ferreira Pessoa et al. Frequency and factors associated with axillary web syndrome in women undergoing breast cancer surgery: a cross-sectional and retrospective study. **Springerplus** , v. 4, n. 1, p. 112, 2015.
- GANDRUD, Christopher. Reproducible research with R and R studio. **CRC Press**, 2013 <http://dx.doi.org/10.1201/9781315382548>
- GARCÍA, María Angel et al. The chemotherapeutic drug 5-fluorouracil promotes PKR-mediated apoptosis in a p53-independent manner in colon and breast cancer cells. **PLoS One**, v. 6, n. 8, p. e23887, 2011.
- GUNASINGHE, NPA Devika et al. Mesenchymal–epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. **Cancer and Metastasis Reviews**, v. 31, n. 3-4, p. 469-478, 2012.
- HANAHAN, D.; WEINBERG, R. A. Hallmarks of Cancer: The Next Generation. **Cell**, v. 144, p. 646-674, 2011.
- HSU, Peggy P.; SABATINI, David M.. Cancer Cell Metabolism: Warburg and Beyond. **Cell**, [s.l.], v. 134, n. 5, p.703-707, set. 2008. **Elsevier BV**.

- [http://dx.doi.org/10.1016/j.cell.2008.08.021.](http://dx.doi.org/10.1016/j.cell.2008.08.021)
- INCA. Instituto Nacional de Câncer. Ministério da Saúde. Secretaria de Atenção à Saúde. Coordenação de Prevenção e Vigilância. **Estimativas 2019**. Rio de Janeiro. Disponível em: <http://www.inca.gov.br>, 2020. Acessado em fevereiro de 2020.
- INUMARU, Lívia Emi; SILVEIRA, Érika Aparecida da; NAVES, Maria Margareth Veloso. Risk and protective factors for breast cancer: a systematic review. **Cadernos de Saúde Pública**, v. 27, p. 1259-1270, 2011.
- JAGUS, Rosemary; JOSHI, Bhavesh; BARBER, Glen N. PKR, apoptosis and cancer. **The international journal of biochemistry & cell biology**, v. 31, n. 1, p. 123-138, 1999.
- KANEHISA, Minoru; FURUMICHI, Miho; TANABE, Mao; SATO, Yoko; MORISHIMA, Kanae. KEGG: new perspectives on genomes, pathways, diseases and drugs. **Nucleic Acids Research**, [s.l.], v. 45, n. 1, p.353-361, 28 nov. 2016. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gkw1092>.
- KAWASE, Takakazu et al. Interaction of the effects of alcohol drinking and polymorphisms in alcohol-metabolizing enzymes on the risk of female breast cancer in Japan. **Journal of epidemiology**, v. 19, n. 5, p. 244-250, 2009.
- KIM, Baek Gil et al. The compression-induced expression of the glycolysis genes in CAFs correlates with EMT and the expression of the angiogenesis gene in breast cancer. **Biologia das comunicações**, v. 2, n. 1, pág. 1-15, 2019.
- LANETZKI, Camila Sanches et al. The epidemiological profile of pediatric intensive care center at hospital israelita albert einstein. **Einstein (São Paulo)**, v. 10, n. 1, p. 16-21, 2012.
- LI, Rui et al. ALDH2 gene polymorphism in different types of cancers and its clinical significance. **Life sciences**, v. 147, p. 59-66, 2016.
- LIBERTI, Maria V.; LOCASALE, Jason W.. The Warburg Effect: How Does it Benefit Cancer Cells?. **Trends In Biochemical Sciences**, [s.l.], v. 41, n. 3, p.211-218, mar. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.tibs.2015.12.001>
- LIEMBURG-APERS, Dania C.; WILLEMS, Peter H. G. M.; KOOPMAN, Werner J. H.; GREFTE, Sander. Interactions between mitochondrial reactive oxygen species and cellular glucose metabolism. **Archives Of Toxicology**, [s.l.], v. 89, n. 8, p.1209-1226, 6 jun. 2015. **Springer Science and Business Media LLC.** <http://dx.doi.org/10.1007/s00204-015-1520-y>.
- LIU, Meng-Xi et al. Metabolic reprogramming by PCK1 promotes TCA cataplerosis,

- oxidative stress and apoptosis in liver cancer cells and suppresses hepatocellular carcinoma. *Oncogene*, v. 37, n. 12, p. 1637-1653, 2018.
- MALVIA, Shreshtha et al. Study of breast cancer gene expression profiles in Indian women. *Relatórios científicos*, v. 9, n. 1, pág. 1-15, 2019.
- MAO, Qunxia et al. The alcohol dehydrogenase 1C genotype (rs698) and breast cancer: a meta-analysis. *Asia Pacific Journal of Public Health*, v. 27, n. 2, pág. NP36-NP46, 2015.
- MARIANO, Fernanda Viviane et al. Genomic profile of a squamous cell carcinoma ex-pleomorphic adenoma compared to a squamous cell carcinoma of the head and neck. *Revista Brasileira de Otorrinolaringologia*, v. 84, n. 3, pág. 393-397, 2018.
- MARTINS, Otávio Augusto. "EFFECT OF THE CONSUMPTION OF ALCOHOLIC BEVERAGES ON THE ORGANISM - A REVIEW." *Revista Eletrônica de Educação e Ciência* 3.2 (2013): 07-10.
- NCBI National Center for Biotechnology Information. Disponível em: <https://www.ncbi.nlm.nih.gov/>. Acesso em: SET. 2020.
- PAN, Xiaohang et al. Valor prognosis of the enolase gene family in colon cancer. Medical Science Monitor: *International Medical Journal of Experimental and Clinical Research*, v. 26, p. e922980-1, 2020.
- POLYAK, Kornelia; WEINBERG, Robert A. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nature Reviews Cancer*, v. 9, n. 4, p. 265-273, 2009.
- SCATENA, Roberto et al. Glycolytic enzyme inhibitors in cancer treatment. *Opinião de especialista em drogas experimentais*, v. 17, n. 10, pág. 1533-1545, 2008.
- STORCHOVA, Zuzana; PELLMAN, David. From polyploidy to aneuploidy, genome instability and cancer. *Nature reviews Molecular cell biology*, v. 5, n. 1, p. 45-54, 2004.
- TOMCZAK, Katarzyna; CZERWINKA, Patrycja; WIZNEROWICZ, Maciej. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (pozn)*, Poland, p.68-77, 2015.
- VELASCO-VELÁZQUEZ, Marco et al. CCR5 antagonist blocks metastasis of basal breast cancer cells. *Cancer research*, v. 72, n. 15, p. 3839-3850, 2012.
- VISWESWARAN, Malini et al. Aberrant lipid metabolism as an emerging therapeutic strategy to target cancer stem cells. *Stem Cells*, v. 38, n. 1, pág. 6-14, 2020.
- WEI, Rong-Rong et al. Identification of ADH4 as a new and potential prognostic marker

- in hepatocellular carcinoma. **Oncología médica**, v. 29, n. 4, pág. 2737-2743, 2012.
- ZHANG, Kezhong et al. COX7AR is a stress-inducible mitochondrial COX subunit that promotes breast cancer malignancy. **Relatórios científicos**, v. 6, n. 1, pág. 12/01/2016.
- ZHANG, Na et al. Detection and clinical significance of Notch 1 methylation in breast cancer and proliferative intraductal lesions of the breast. **Zhonghua bing li xue za zhi** = jornal chinês de patologia, v. 40, n. 5, pág. 324, 2011.

CAPÍTULO 3 BREAST CANCER IA APLICATTION : Programa de Computador com aplicação de *machine learning* em dados genéticos para identificar possíveis biomarcadores no câncer de mama

AUTORES: JOÃO GABRIEL DE LIMA CASTRO, JOÃO MATEUS CABRITA RODRIGUES, THAYANE GONÇALVES DA SILVA BATISTA, ANDRESSA MARQUES ANDREU, LOHRENN BRYANDA MAIA, MARCOS ALVES, CRISTIANO MARCELO ESPINOLA CARVALHO.

RESUMO

O câncer de mama é a neoplasia mais incidente entre mulheres e a segunda mais frequente no mundo. Devido ao aumento estimado de 625 mil casos novos de câncer por ano previstos pelo INCA (Instituto Nacional de Câncer) na projeção de 2020-2022 com cerca de 66.280 novos casos, este se tornou uma preocupação mundial. O diagnóstico precoce do câncer de mama influência diretamente nos possíveis tratamentos, no prognóstico da doença e a taxa de sobrevivência. Este trabalho teve como objetivo encontrar possíveis biomarcadores por meio da aplicação de aprendizado de máquina (*machine learning*) em amostras de câncer. A pesquisa seguiu 3 caminhos paralelos que utilizam das mesmas técnicas e métodos. Os segmentos são separados em genes que atuam nos diferentes estágios, em miRNA relacionados ao câncer de mama e por fim genes e miRNAs em um único dataset (conjunto de dados). Foi utilizado o algoritmo DESeq2 para normalizar os dados genéticos obtidos do banco TCGA (The Cancer Genome Atlas) utilizando linguagem R. A partir dos resultados obtidos foram aplicados dois algoritmos de filtragem, Distancia Euclidiana e Árvore de decisão, em linguagem python, assim obtendo valores de classificação, para que houvesse maior precisão na aplicação das técnicas de *machine learning*. Foram utilizadas quatro técnicas para classificar a qualidade dos alvos escolhidos, são elas: o SVM (Support vector machine), o MLP Classifier (Multi-layer Perceptron classifier), a Árvore de decisão (Decision Tree) e a Regressão logística (Logistic Regression). Dentre os resultados, os alvos que se destacaram foram os genes MMP11 e FIGF, os quais se mostraram relevantes nos estágios iniciais do câncer de mama e também para classificar os dados conjuntos nas análises. Já entre os miRNAs estudados, os alvos hsa-mir-139 e hsa-mir-145 demonstraram resultados promissores. Os resultados vão de acordo com a literatura validando a metodologia aplicada, o que indica a possibilidade de definir possíveis biomarcadores capazes de influenciar no diagnóstico precoce e a taxa de sobrevida no câncer de mama. Sugere-se validação com estudos *in vitro* e *in vivo* para confirmar os resultados aqui obtidos.

Palavras-chave: Biomarcadores; Câncer de Mama; Computação; Aprendizado de Máquina.

Certificado de Registro de Programa de Computador



INPI
Instituto Nacional da Propriedade Industrial
Assinado Digitalmente

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA ECONOMIA
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512022001525-0**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 01/12/2020, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: Breast Cancer IA Application

Data de criação: 01/12/2020

Titular(es): MISSÃO SALESIANA DE MATO GROSSO - UNIVERSIDADE CATÓLICA DOM BOSCO - UCDB (UCDB)

Autor(es): MARCOS ALVES; CRISTIANO MARCELO ESPINOLA CARVALHO; ANDRÉ LUIZ SANTANA TREU AFONSO; JOÃO LUIZ AGUIAR TAKAYAMA; THAYANE GONÇALVES DA SILVA BATISTA; ANDRESSA MARQUES ANDREU; LORHENN BRYANDA LEMES MAIA; MICHEL ÂNGELO CONSTANTINO DE OLIVEIRA

Linguagem: PYTHON; R

Campo de aplicação: BL-02; BL-07; IF-07; SD-07

Tipo de programa: IA-01; IT-03; TC-02

Algoritmo hash: SHA-512

Resumo digital hash:

fd080eaea37ac5e6ba26c2172dcd8c8818a61e2e2551e5b753779d245e06b6cbfbb9784aa157b0c104dd3259a336f6241fa52874eee9d0ba4e36779385c5d106

Expedido em: 28/06/2022

Aprovado por:

Joelson Gomes Pequeno

Chefe Substituto da DIPTO - PORTARIA/INPI/DIRPA Nº 02, DE 10 DE FEVEREIRO DE 2021

CONSIDERAÇÕES FINAIS

O presente estudo teve como intuito identificar alvos moleculares capazes de serem utilizados como marcadores moleculares no câncer de mama. No primeiro capítulo, foram identificados, por análise de dados, miRNAs possíveis reguladores de oncogenes e genes supressores de tumores mamários (*BRCA1*, *BRCA2*, *CCND1*, *CDK1*, *CDKN1A* e *Rad51*), sendo: miR-665, miR-378c, miR-184, miR-133b e miR-3161. Esses transcritos apresentaram perfil de expressão diferencial em amostras de câncer de mama e tecido mamário normal. De mesmo modo analítico, no segundo capítulo, foram identificados genes que orquestram o metabolismo energético do câncer de mama com expressão diferencial entre as condições estudadas. Ademais, o programa de computador baseado em aprendizagem de máquina (*machine learning*) também identificou genes e miRNAs promissores: *MMP11*, *FIGF* e; mir-139 e mir-145. Essas informações demonstram que os objetivos deste trabalho foram alcançados com sucesso. Essas biomoléculas podem auxiliar na rotina clínica como ferramentas diagnósticas mais precisas e eficazes que as convencionais, marcadores de estadiamento e de estágios tumorais e, assim, proporcionar melhorias na terapêutica dos pacientes. Sugere-se validação com estudos *in vitro* e *in vivo* para confirmar os resultados aqui obtidos.

Anexo 1 – Pedido de Registro de Programa de Computador



22/06/2022 870220054758

18.08



29409191951329784

Pedido de Registro de Programa de Computador - RPC

Número do Processo: 512022001525-0

Dados do Titular

Titular 1 de 1

Nome ou Razão Social: MISSÃO SALESIANA DE MATO GROSSO - UNIVERSIDADE CATÓLICA DOM BOSCO - UCDB (UCDB)

Tipo de Pessoa: Pessoa Jurídica

CPF/CNPJ: 03226149001587

Nacionalidade: Brasileira

Qualificação Jurídica: Associação com intuito não econômico

Endereço: Avenida Tamandaré, nº 6000. Jardim Seminário

Cidade: Campo Grande

Estado: MS

CEP: 79002-970

País: Brasil

Telefone: (67) 3312-3580

Fax: -

Email: s-inova@ucdb.br

Dados do Programa

Data de Criação: 01/12/2020

- § 2º do art. 2º da Lei 9.609/98: "Fica assegurada a tutela dos direitos relativos a programa de computador pelo prazo de cinquenta anos contados a partir de 1º de janeiro do ano subsequente ao da sua publicação ou, na ausência desta, da sua criação"

Título: Breast Cancer IA Application

Algorítimo hash: SHA-512 - Secure Hash Algorithm

Resumo digital hash: fd080eaea37ac5e6ba26c2172cd8c8818a61e2e2551e5b753779d24
5e06b6cbfb9784aa157b0c104dd3259a336f6241fa52874eee9d0ba4
e36779385c5d106

§1º e Incisos VI e VII do §2º do Art. 2º da Instrução Normativa: O titular é o responsável único pela

PETICIONAMENTO ELETRÔNICO

Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 22/06/2022 às 18:08, Petição 870220054758

transformação, em resumo digital hash, dos trechos do programa de computador e demais dados considerados suficientes para identificação e caracterização, que serão motivo do registro. O titular terá a inteira responsabilidade pela guarda da informação sigilosa definida no inciso III, § 1º, art. 3º da Lei 9.609 de 19 de fevereiro de 1998.

Linguagem: PYTHON

R

Campo de Aplicação: IF07-CIÊNC INFO (SISTEMA DE INFORMAÇÃO, REDE DE INFORMAÇÃO, TEORIA DA INFORMAÇÃO, FLUXO DE INFORMAÇÃO)
BL02-GENÉTICA (CITOGENÉTICA, ENGENHARIA GENÉTICA, GENOTIPO, HEREDITARIEDADE, MELHORAMENTO GENÉTICO, GEN, GENÉTICA DAS POPULAÇÕES)
BL07-BIOQUÍMICA (AMINOÁCIDO, PROTEÍNA, HORMÔNIO, FENÔMENO BIOQUÍMICOS: BLOSSÍNTSE - FERMENTAÇÃO - OSMOSE, ETC)
SD07-MEDICINA (ALOPÁTICA, HEMEOPÁTICA, PREVENTIVA, TROPICAL, NUCLEAR, MEDICINA DO TRABALHO, LEGAL, DE URGÊNCIA)

Tipo de Programa: IA01-INTELIGÊNCIA ARTIFICIAL

IT03 - INSTRUMENTAÇÃO BIOMÉDICA

TC02 - PESQUISA OPERACIONAL

**PETICIONAMENTO
ELETRÔNICO**

Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 22/06/2022 às 18:08, Petição 870220054758

Dados do Autor**Autor 1 de 8****Nome:** CRISTIANO MARCELO ESPINOLA CARVALHO**CPF:** 61504319168**Nacionalidade:** Brasileira**Qualificação Física:** Professor do ensino superior**Endereço:** Rua Figueira da Foz, 37, bairro Portal Panamá**Cidade:** campo grande**Estado:** MS**CEP:** 79113-115**País:** BRASIL**Telefone:** (67) 999 821997**Fax:****Email:** cristiano@ucdb.br**Autor 2 de 8****Nome:** ANDRÉ LUIZ SANTANA TREU AFONSO**CPF:** 05694599152**Nacionalidade:** Brasileira**Qualificação Física:** Estudante de Graduação**Endereço:** Rua Severino Marques, 169, São Francisco,**Cidade:** campo grande**Estado:** MS**CEP:** 79009-730**País:** BRASIL**Telefone:** (67) 994 65367**Fax:****Email:** andr.luiz.santana@gmail.com**Autor 3 de 8****PETICIONAMENTO
ELETRÔNICO**

Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 22/06/2022 às 18:08, Petição 870220054758

Nome: JOÃO LUIZ AGUIAR TAKAYAMA
CPF: 05726746198
Nacionalidade: Brasileira
Qualificação Física: Estudante de Graduação
Endereço: Rua José Ramão Canteiro, 110, vila Manoel da costa lima
Cidade: campo grande
Estado: MS
CEP: 79040-180
País: BRASIL
Telefone: (67) 996 416225
Fax:
Email: takayama585@gmail.com

Autor 4 de 8

Nome: THAYANE GONÇALVES DA SILVA BATISTA
CPF: 03588054146
Nacionalidade: Brasileira
Qualificação Física: Estudante de Graduação
Endereço: Rua Atlântica, 135, Loteamento Porto Belo
Cidade: campo grande
Estado: MS
CEP: 79042-185
País: BRASIL
Telefone: (67) 991 200761
Fax:
Email: ta_thayane03@hotmail.com

Autor 5 de 8**PETICIONAMENTO
ELETRÔNICO**

Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 22/06/2022 às 18:08, Petição 870220054758

Nome: ANDRESSA MARQUES ANDREU
CPF: 04136251173
Nacionalidade: Brasileira
Qualificação Física: Mestrando
Endereço: Rua dos Mariscos, 292, Coophavila II
Cidade: campo grande
Estado: MS
CEP: 79097-180
País: BRASIL
Telefone: (67) 991 211710
Fax:
Email: dessamarques21@gmail.com

Autor 6 de 8

Nome: LORHENN BRYANDA LEMES MAIA
CPF: 74245848120
Nacionalidade: Brasileira
Qualificação Física: Doutorando
Endereço: Rua São Felipe, 176 - Morada dos deuses
Cidade: campo grande
Estado: MS
CEP: 79117-438
País: BRASIL
Telefone: (67) 999 292752
Fax:
Email: lorhenn.maia@gmail.com

Autor 7 de 8

**PETICIONAMENTO
ELETRÔNICO**

Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 22/06/2022 às 18:08, Petição 870220054758

Nome: MICHEL ÂNGELO CONSTANTINO DE OLIVEIRA
CPF: 76700291100
Nacionalidade: Brasileira
Qualificação Física: Professor do ensino superior
Endereço: Rua Monte Negro, 110, Vila Ieda
Cidade: campo grande
Estado: MS
CEP: 79050-580
País: BRASIL
Telefone: (67) 992 494575
Fax:
Email: michel@ucdb.br

Autor 8 de 8

Nome: MARCOS ALVES
CPF: 55483089120
Nacionalidade: Brasileira
Qualificação Física: Professor do ensino superior
Endereço: Estrada NS 4, 31, Chácara Recando do Pedrinho, Chácara dos Poderes
Cidade: campo grande
Estado: MS
CEP: 79037-813
País: BRASIL
Telefone: (67) 999 571467
Fax:
Email: marcos@ucdb.br

Declaração de Veracidade - DV

Nome: DECLARAÇÃO_DE_VERACIDADE Cert.DIGITAL.pdf

PETICIONAMENTO ELETRÔNICO Esta solicitação foi enviada pelo sistema Peticionamento Eletrônico em 22/06/2022 às 18:08, Petição 870220054758

DECLARAÇÃO DE VERACIDADE - CLIENTE

Em atendimento à Instrução Normativa em vigor eu, MISSÃO SALESIANA DE MATO GROSSO - UNIVERSIDADE CATÓLICA DOM BOSCO - UCDB (UCDB), CNPJ: 03.226.149/0015-87, declaro, para fins de direito, sob as penas da Lei e em atendimento ao art. 2º do Decreto nº 2.556², de 20 de abril de 1998, que as informações feitas no formulário eletrônico de programa de computador – e-Software, são verdadeiras e autênticas.

Fico ciente através desse documento que a falsidade dessa declaração configura crime previsto no Código Penal Brasileiro e passível de apuração na forma da Lei.

Ciente das responsabilidades pela declaração apresentada, firmo a presente.

-----(assinado digitalmente)-----

MISSAO
SALESIANA DE
MATO
GROSSO:0322614
9001587

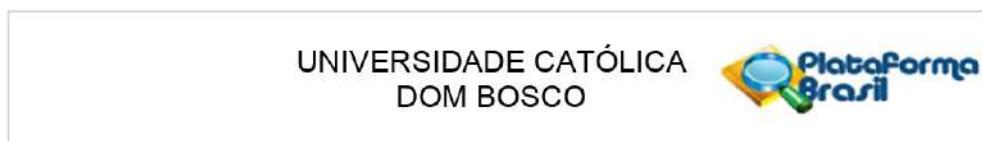
Assinado de forma
digital por MISSAO
SALESIANA DE MATO
GROSSO:032261490015
87
Dados: 2022.06.22
16:59:34 -04'00'

DECRETO N° 2.556, DE 20 DE ABRIL DE 1998
Art. 1º Os programas de computador poderão, a critério do titular dos respectivos direitos, ser registrados no Instituto Nacional da Propriedade Industrial - INPI.

Art. 2º A veracidade das informações de que trata o artigo anterior são de inteira responsabilidade do requerente, não prejudicando eventuais direitos de terceiros nem acarretando qualquer responsabilidade do Governo.

29409191951329784

Anexo 2 – Parecer Consustanciado pelo CEP



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Avaliação de possíveis marcadores moleculares em carcinomas mamários

Pesquisador: Cristiano Marcelo Espinola Carvalho

Área Temática:

Versão: 3

CAAE: 83349318.5.0000.5162

Instituição Proponente: Universidade Católica Dom Bosco

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.246.746

Apresentação do Projeto:

O projeto se propõe realizar uma caracterização molecular das amostras tumorais do tipo mais incidente entre as mulheres sul-mato-grossenses – o câncer de mama e, avaliar possíveis marcadores moleculares de progressão tumoral em carcinomas mamários, avaliar os genes caretakers BRCA1, BRCA2, RAD51, p21, CCND1 e CDK1 - integrantes da maquinaria de reparo acionada para quebras dupla-fita (DBSs) do DNA.

Objetivo da Pesquisa:

Pretende-se com este estudo realizar a caracterização molecular das amostras provenientes de pacientes do Mato Grosso do Sul e avaliar genes caretakers -supressores de tumor, como candidatos à marcadores moleculares para predição de malignidade, agressividade e progressão tumoral de carcinomas mamários.

Avaliação dos Riscos e Benefícios:

Riscos:

Os resultados finais podem demonstrar que os genes analisados não servem como marcadores moleculares.

Benefícios:

Sendo o câncer de mama o tipo mais incidente nas mulheres sul-mato-grossenses, com 820 novos casos esperados para os anos de 2016/2017 no estado de Mato Grosso do Sul, com 460

Endereço: Av. Tamandaré, 6000

Bairro: Jardim Seminário

CEP: 79.117-900

UF: MS **Município:** CAMPO GRANDE

Telefone: (67)3312-3723

E-mail: cep@ucdb.br

**UNIVERSIDADE CATÓLICA
DOM BOSCO**



Continuação do Parecer: 3.246.746

ocorrendo na capital Campo Grande (INCA, 2016), pesquisas que busquem contribuir para a diminuição da mortalidade dessas pacientes afetam positivamente o estado como um todo, além contribuir para o desenvolvimento da pesquisa aplicada à saúde, fortalecendo o vínculo entre a bancada e a clínica médica, através da associação de médicos e pesquisadores do estado de Mato Grosso do Sul, que unem seus conhecimentos em favor da inovação.

Comentários e Considerações sobre a Pesquisa:

A pesquisa é pertinente e fundamental para o avanço dos conhecimentos na área da saúde. O projeto foi reapresentado para apreciação ética e atendeu a todas as recomendações prévias do parecer emitido pela CONEP e confirmado por este CEP.

Considerações sobre os Termos de apresentação obrigatória:

Adequados.

Recomendações:

sem recomendações.

Conclusões ou Pendências e Lista de Inadequações:

Aprovado.

Considerações Finais a critério do CEP:

O colegiado acompanha o parecer do relator.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJECTO_1058350.pdf	22/03/2019 10:47:52		Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE.pdf	22/03/2019 10:38:19	Cristiano Marcelo Espínola Carvalho	Aceito
Declaração de Instituição e Infraestrutura	santa.pdf	07/06/2018 16:12:07	Márcio Luís Costa	Aceito
Declaração de Instituição e Infraestrutura	Autorizacao_CASSEMS_UCDB.pdf	13/02/2018 19:32:29	Cristiano Marcelo Espínola Carvalho	Aceito
Folha de Rosto	folhaDeRosto_carcinoma_mamario.pdf	13/02/2018 19:30:18	Cristiano Marcelo Espínola Carvalho	Aceito
Projeto Detalhado / Brochura	Projeto_carcinoma_mamario.pdf	13/02/2018 19:28:34	Cristiano Marcelo Espínola Carvalho	Aceito

Endereço: Av. Tamandaré, 6000

Bairro: Jardim Seminário

CEP: 79.117-900

UF: MS

Município: CAMPO GRANDE

Telefone: (67)3312-3723

E-mail: cep@ucdb.br

UNIVERSIDADE CATÓLICA
DOM BOSCO



Continuação do Parecer: 3.246.746

Investigador	Projeto_carcinoma_mamario.pdf	13/02/2018 19:28:34	Cristiano Marcelo Espínola Carvalho	Aceito
--------------	-------------------------------	------------------------	--	--------

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

CAMPO GRANDE, 05 de Abril de 2019

Assinado por:

Karla de Toledo Cândido Muller
(Coordenador(a))

Endereço: Av. Tamandaré, 6000	CEP: 79.117-900
Bairro: Jardim Seminário	
UF: MS	Município: CAMPO GRANDE
Telefone: (67)3312-3723	E-mail: cep@ucdb.br