

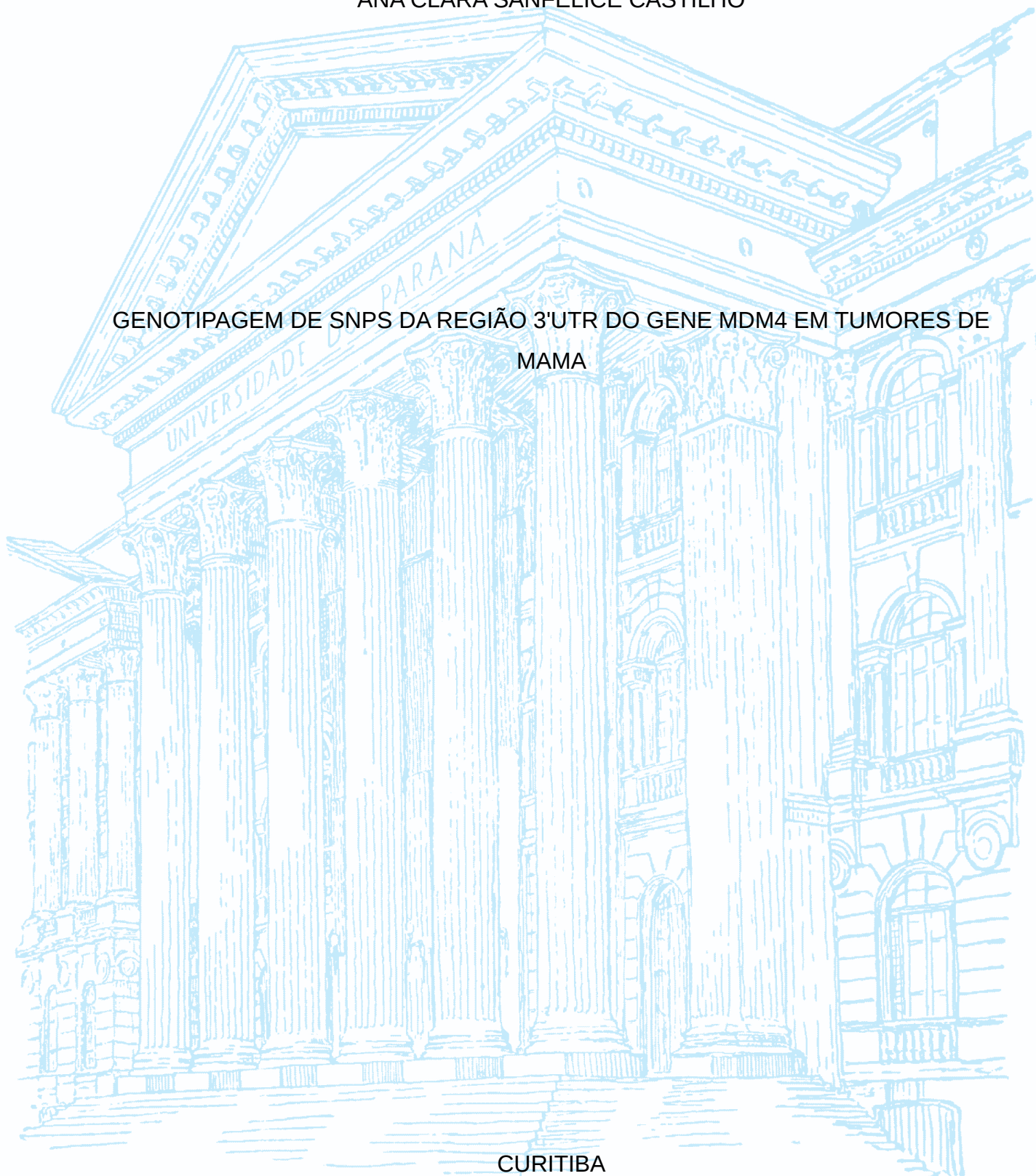
UNIVERSIDADE FEDERAL DO PARANÁ

ANA CLARA SANFELICE CASTILHO

GENOTIPAGEM DE SNPS DA REGIÃO 3'UTR DO GENE MDM4 EM TUMORES DE
MAMA

CURITIBA

2021



ANA CLARA SANFELICE CASTILHO

GENOTIPAGEM DE SNPS DA REGIÃO 3'UTR DO GENE MDM4 EM TUMORES DE
MAMA

Monografia apresentada ao Curso de Ciências Biológicas, Setor de Ciências Biológicas, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Ciências Biológicas

Orientador(a): Prof(a). Dr(a). Karin Braun Prado

CURITIBA

2021

TERMO DE APROVAÇÃO

ANA CLARA SANFELICE CASTILHO

GENOTIPAGEM DO GENE *MD4M* EM TUMORES DE MAMA

Monografia apresentada ao Curso de Ciências Biológicas, Setor de Ciências Biológicas, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Ciências Biológicas.

Prof(a). Dr(a)./Msc. _____

Orientador(a) – Departamento _____, INSTITUIÇÃO

Prof(a). Dr(a)./Msc. _____

Departamento _____, INSTITUIÇÃO

Prof(a). Dr(a)./Msc. _____

Departamento _____, INSTITUIÇÃO

Cidade, ____ de _____ de 201_.

AGRADECIMENTOS

À minha família, principalmente aos meus pais, por sempre terem me proporcionado a melhor educação possível e terem me apoiado em todas as minhas escolhas, inclusive a de ser bióloga.

Aos meus amigos e companheiros da biologia, Laura, Milena, Natália, Natalie e Gabriel. Um agradecimento especial à Natalie, que também é minha companheira de laboratório e sempre me auxiliou muito, sem ela esse trabalho não teria sido possível. Sou muito grata por ter feito amizades tão maravilhosas nesse curso.

Ao Junior, pelo interminável apoio em todos os momentos durante a realização desse trabalho e por ser meu parceiro em tantas coisas.

À minha orientadora Karin, pelo auxílio durante o desenvolvimento desse projeto e pelos quase 3 anos de orientação.

Aos colegas do LGMH, pelas ajudas no laboratório e convivência.

À professora Jaqueline Carvalho de Oliveira, por ter cedido as amostras que foram utilizadas na realização deste trabalho.

Ao Valter Antonio de Baura pelo auxílio no sequenciamento de todas as amostras utilizadas neste trabalho.

Agradeço também ao financiamento do CNPq para que esse projeto pudesse ser realizado.

Por último, mas não menos importante, agradeço ao meu cachorrinho, Banzé, por toda a companhia que me fez nesse ano tão difícil e por ser um companheiro tão amado. Ele nem sabe, mas me ajudou muito!

“As coisas que não saem conforme o planejado em sua vida são em geral mais importantes, mais formativas a longo prazo, do que as que saem. É preciso esperar o inesperado, aceitá-lo”

Maggie O’Farrell

RESUMO

O câncer de mama é uma doença heterogênea desencadeada por uma diversidade de fatores intrínsecos e extrínsecos. Dentre os fatores intrínsecos, as causas genéticas são o principal alvo de pesquisas, pois frequentemente são encontrados genes envolvidos direta ou indiretamente na neoplasia mamária. Um desses genes é o *MDM4*, que apresenta uma variedade de polimorfismos de nucleotídeo único (SNPs) e regula negativamente a proteína p53, supressora de tumor. Esses polimorfismos são relacionados a susceptibilidade a alguns tipos de câncer, como o *rs4245739* (A/C), o qual, na presença do alelo C, é associado ao menor risco de câncer de mama. Além disso, o alelo C do SNP *rs10900596* (C/T) foi predito como alvo do mir-1257 em análises *in silico*. A ligação desse microRNA no mRNA do *MDM4* interfere na sua tradução e, assim, na regulação negativa da p53. A menor expressão do *MDM4* possivelmente diminui o risco de neoplasia mamária. Nesse contexto, reconstruímos o haplótipo do *MDM4* genotipando os SNPs *rs56200417* (C/T), *rs10900596* (T/C) e *rs10900597* (C/T), localizados na região 3'UTR do gene. Essas variantes estão em desequilíbrio de ligação entre si e foram genotipadas por sequenciamento Sanger. Desse modo, avaliamos a combinação dos haplótipos dessas três variantes em 156 amostras de tumor de mama e 183 amostras controle de pessoas saudáveis. Encontramos o haplótipo TCT em maior frequência no grupo controle, o que corrobora a nossa hipótese de que o alelo C na variantes *rs10900596* pode ser um fator de proteção ao câncer de mama. Isto pode ser decorrente da afinidade de microRNAs a esses alelos, interferindo na tradução do oncogene *MDM4*.

Palavras-chave: câncer de mama. *rs4245739*. *rs10900596*. Haplótipo.

ABSTRACT

Breast cancer is a heterogeneous disease that can be triggered by a diversity of intrinsic and extrinsic factors. Within the intrinsic ones, the genetic causes are the researcher's main target, as genes are frequently found to be directly or indirectly involved in breast neoplasm. One of these genes is the MDM4, which has a variety of Single Nucleotide Polymorphisms (SNPs) and negatively regulates the p53 protein, a tumor suppressor. These polymorphisms are related with some types of cancer susceptibility, like rs4245739 (A/C), which is associated with lower risk of breast cancer in the presence of the C allele. Besides that, the rs10900596 (C/T) C allele was predicted to be mir-1257's target in in silico analyses. The binding of this microRNA on MDM4's mRNA interfere its translation and, therefore, the negative regulation of p53. MDM4's lower expression possibly decreases the risk of breast neoplasm. In this context, we reconstructed MDM4's haplotype by genotyping the SNPs rs56200417 (C/T), rs10900596 (T/C), rs10900597 (C/T), localized in the gene's 3'UTR region. These variants are in linkage disequilibrium with each other and were genotyped by Sanger sequencing. Thus, we evaluated the haplotypes combinations of these three variants in 156 breast tumor samples and 183 healthy control samples. We found the TCT haplotype in higher frequency in the control group, which corroborates our hypothesis that the C allele rs10900596 may be a breast cancer protection factor. This is supported by the microRNA's affinity for this allele, preventing the MDM4 oncogene's translation.

Key words: breast cancer. *rs4245739*. *rs10900596*. Haplotype.

LISTA DE FIGURAS

| | |
|--|----|
| FIGURA 1 – <i>MDM4</i> E A REGULAÇÃO NEGATIVA DA P53..... | 16 |
| FIGURA 2 – ESTRUTURA DO <i>MDM4</i> | 17 |
| FIGURA 3 – SEQUÊNCIA AMPLIFICADA DO <i>MDM4</i> | 21 |
| FIGURA 4 – ELETROFORESE DE VERIFICAÇÃO DA AMPLIFICAÇÃO DO FRAGMENTO..... | 22 |
| FIGURA 5 – ELETROFORESE COM AMOSTRAS DE TUMOR AMPLIFICADAS NA PCR..... | 25 |
| FIGURA 5 – LEITURA DO SEQUENCIAMENTO – SNP <i>rs56200417</i> | 26 |
| FIGURA 6 - LEITURA DO SEQUENCIAMENTO – SNP <i>rs10900596</i> | 26 |
| FIGURA 7 - LEITURA DO SEQUENCIAMENTO – SNP <i>rs10900597</i> | 26 |

LISTA DE TABELAS

| | |
|--|----|
| TABELA 1 – SEQUÊNCIAS DOS <i>PRIMERS</i> UTILIZADOS PARA AMPLIFICAÇÃO DA SEQUÊNCIA ALVO..... | 20 |
| TABELA 2 – FREQUÊNCIAS ALÉLICAS DAS VARIANTES <i>rs56200417</i> , <i>rs10900596</i> e <i>rs10900597</i> NO GRUPO CONTROLE E NAS PACIENTES..... | 27 |
| TABELA 3 – FREQUÊNCIAS GENOTÍPICAS DAS VARIANTES <i>rs56200417</i> , <i>rs10900596</i> e <i>rs10900597</i> NO GRUPO CONTROLE E NAS PACIENTES..... | 27 |
| TABELA 4 - FREQUÊNCIAS HAPLOTÍPICAS NO GRUPO CONTROLE E NAS PACIENTES - ORDEM DAS VARIANTES: <i>rs56200414</i> , <i>rs10900596</i> , <i>rs10900597</i> | 29 |
| TABELA 5 - ANÁLISE DO ALELO C DO SNP <i>rs10900596</i> EM LUMINAL A E B.... | 30 |
| TABELA 6 - FREQUÊNCIAS HAPLOTÍPICAS DO GRUPO CONTROLE E DAS PACIENTES PARA AS VARIANTES <i>rs4245739</i> , <i>rs56200417</i> , <i>rs10900596</i> , <i>rs10900597</i> | 30 |
| TABELA 7 - VALORES DE LD ENTRE A VARIANTE <i>rs4245739</i> E AS VARIANTES <i>rs56200417</i> , <i>rs10900596</i> E <i>rs10900597</i> | 32 |

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO..... | 11 |
| 2 | REVISÃO BIBLIOGRÁFICA..... | 14 |
| 2.1 | CÂNCER DE MAMA..... | 14 |
| 2.2 | FAMÍLIA <i>MDM</i> | 15 |
| 2.2.1 | <i>MDM4</i> | 16 |
| 2.3 | HAPLÓTIPOS..... | 18 |
| 2.4 | POLIMORFISMOS DO <i>MDM4</i> | 19 |
| 3 | MATERIAL E MÉTODOS..... | 20 |
| 3.1 | SNPs..... | 20 |
| 3.2 | AMOSTRAS..... | 20 |
| 3.3 | PCR E REAÇÃO DE SEQUENCIAMENTO..... | 21 |
| 3.4 | ANÁLISE DAS SEQUÊNCIAS..... | 23 |
| 3.5 | ANÁLISE DOS DADOS..... | 24 |
| 3.6 | ANÁLISE DO HAPLÓTIPO COM A VARIANTE <i>rs4245739</i> | 24 |
| 4 | RESULTADOS E DISCUSSÃO..... | 25 |
| 4.1 | AMOSTRAS SELECIONADAS PARA AS ANÁLISES ESTATÍSTICAS..... | 25 |
| 4.2 | ESTATÍSTICAS DAS AMOSTRAS DAS PACIENTES..... | 26 |
| 4.3 | FREQUÊNCIAS ALÉLICAS, GENOTÍPICAS E HAPLOTÍPICAS..... | 26 |
| 4.3.1 | Grupo controle e pacientes..... | 26 |
| 4.3.2 | Subtipos tumorais das pacientes..... | 29 |
| 4.4 | ANÁLISES ESTATÍSTICAS COM A VARIANTE <i>rs4245739</i> PARA MONTAGEM DE HAPLÓTIPO..... | 30 |
| 4.4.1 | Frequências haplotípicas do grupo controle e das pacientes para as variantes <i>rs4245739</i> , <i>rs56200417</i> , <i>rs10900596</i> , <i>rs10900597</i> | 30 |
| 5 | CONSIDERAÇÕES FINAIS..... | 33 |
| | REFERÊNCIAS..... | 35 |

1 INTRODUÇÃO

O câncer de mama é um dos tipos de câncer que mais acomete mulheres no mundo todo. Segundo o Instituto Nacional do Câncer, em 2020, o câncer que mais acometeu mulheres no Brasil foi o de mama, representando cerca de 29% dos novos casos da doença (INCA, 2020).

Sabe-se que o câncer de mama é uma doença heterogênea, unindo fatores intrínsecos (causas genéticas) e fatores extrínsecos (como idade, hábitos alimentares e de atividade física, tabagismo, alcoolismo, contato com substância tóxicas por longos períodos, etc) para o seu desenvolvimento (INCA, 2020).

As neoplasias mamárias são comumente classificadas quanto ao seu tipo histológico, subtipo molecular, grau tumoral, presença/ausência de receptores hormonais de estrógeno e progesterona e presença/ausência de linfonodos. Dentro de cada uma dessas características pode-se avaliar o prognóstico do paciente. Nesse trabalho, as características utilizadas para as análises foram: subtipo molecular, presença ou ausência de linfonodos e presença ou ausência de receptores hormonais

Conforme a classificação molecular dos tumores de mama, podemos diferenciá-los em 5 subtipos: luminal A, luminal B, tumores com superexpressão de *HER2*, *basal-like* e *normal-like*. Essas classificações foram estabelecidas a partir da análise de microarranjos de DNA e de um conjunto de genes mínimos (PAM50) (HENG, 2017).

De forma geral, os tumores do tipo luminal possuem o melhor prognóstico, todos expressam os receptores hormonais (RE e/ou RP), mas não têm superexpressão de *HER2* – com exceção de alguns tumores luminal B (EROLES et al., 2012); os tumores com superexpressão do *HER2* não apresentam os receptores hormonais, mas têm superexpressão do oncogene *HER2* (Abreu; Koifman, 2002); os *basal-like* também não são positivos para os receptores hormonais, assim como não têm superexpressão do *HER2* e representam o pior prognóstico entre todos os subtipos (Sørli et al., 2003).

Muitos são os genes que podem estar envolvidos no câncer de mama, seja direta ou indiretamente. Neste trabalho, o gene em questão é o *MDM4*, o qual está envolvido na via de regulação da proteína supressora de tumor p53 (Haupt et al. ,

1997). O gene *MDM4* é homólogo ao *MDM2*, o qual já tem suas funções bastante compreendidas, sendo a principal delas regular negativamente a p53. O gene *MDM2* e a proteína p53 têm um feedback de autorregulação, isto é, em condições normais, existe um mecanismo no qual a p53 ativa o *MDM2* para que, depois de exercer suas funções, possa ser bloqueada (Picksley; Lane, 1993). No entanto, se essa regulação não for finamente controlada, a p53 não será devidamente regulada e pode ocorrer uma desregulação na via de replicação celular, resultando em um crescimento descontrolado de células e possivelmente originando um tumor.

A proteína *MDM4* é capaz de formar um heterodímero de forte ligação com a proteína do *MDM2*, o qual tem grande capacidade de degradar a p53 (Karni-Schmidt et al., 2016). Além disso, estudos recentes encontraram altos níveis de expressão do *MDM4* em tumores de mama e indicaram uma possível relação entre seus SNPs (Polimorfismos de Nucleotídeo Único) e a susceptibilidade ao câncer (Atwal et al., 2009).

Uma análise do nosso grupo de pesquisa por meio de ferramentas de bioinformática indicou uma possível interação entre um microRNA (o mir-1257) e um SNP do *MDM4*, o *rs10900596* (T/C), localizado na região 3'-UTR do gene. Outros miRNAs, como miR-887 e miR-191, já foram alvos de pesquisas por conta de sua interação com o SNP *rs4245739* (A/C), também localizado na região 3'-UTR do *MDM4*. Nessas pesquisas, foi identificada maior afinidade do miR-191 e do miR-887 pelo alelo C do SNP *rs4245739*. Isso aponta o alelo C como fator de menor risco para alguns tipo de câncer (como de ovário e de mama), já que, na presença desses microRNAs, a expressão do *MDM4* é diminuída, o que evitaria a degradação do p53 e permitiria que ele exercesse sua função de fator de transcrição (Liu et al., 2013; Wynendaele et al., 2010).

A partir dessa análise, selecionamos no presente estudo os polimorfismos *rs10900596* (T/C) e outros dois polimorfismos – *rs56200417* (C/T) e *rs10900597* (C/T), além do SNP *rs4245739*, que foi previamente avaliado em outro estudo, para a formação do haplótipo do *MDM4*. As três variantes genotipadas no presente trabalho estão localizadas na região 3'UTR do *MDM4* e estão em total desequilíbrio de ligação entre si, formando um bloco de haplótipo. Uma região de 704 pb que engloba as referidas variantes foi sequenciada. Para cada SNP foram determinadas as frequências alélicas, genotípicas e as combinações haplotípicas em amostras tumorais de mama, que foram comparadas com aquelas obtidas em amostras

controle de pessoas saudáveis. Além disso, foi possível também analisar as frequências genóticas e haplotípicas mais comuns entre os subtipos tumorais presentes na amostra analisada, bem como se há alguma relação entre os genótipos dos polimorfismos e as características tumorais descritas para cada amostra.

2 REVISÃO BIBLIOGRÁFICA

2.1 CÂNCER DE MAMA

O câncer de mama, segundo o Instituto Nacional de Câncer (INCA), é o segundo tipo de câncer mais comum entre as mulheres no Brasil e no mundo, ficando atrás apenas do câncer de pele não melanoma. Ainda segundo o INCA, no Brasil, anualmente, 29% dos novos casos de câncer são de neoplasia de mama.

Este tipo de câncer é bastante heterogêneo e pode ocorrer e se apresentar em diferentes formas (Sørli et al., 2001). A forma clínica mais comum de classificá-los é por tipos histológicos, divididos em dois grupos: carcinoma in situ e carcinoma invasivo. Dentro dos carcinomas in situ, há dois subtipos, o carcinoma ductal in situ e o carcinoma lobular in situ. Já dentro dos carcinomas invasivos, há 6 subtipos: carcinoma ductal invasivo (o qual corresponde a maior parte dos tumores de mama malignos), lobular invasivo, medular, colóide, tubular e outros tipos. Os tumores especializados, ou seja, que são facilmente diferenciados histologicamente, representam os melhores prognósticos, diferentemente dos tumores ductais, que são pouco diferenciados (ABBAS, 2016).

Além da classificação histológica, há a classificação molecular, proposta por Perou et al., 2000, e por Sorlie et al., 2001. Esses pesquisadores utilizaram microarranjos de DNA para classificar os tumores em 5 subtipos: luminal A, luminal B, *basal-like*, superexpressão de HER2+ e *normal-like* (Perou et al., 2000). Posteriormente, mais um subtipo foi definido, o Claudin-low (Herschkowitz et al., 2007).

Dentre esses subtipos, o luminal A representa 60% dos tumores mamários e também o melhor prognóstico, geralmente com características imunoistoquímicas positivas para RE e/ou RP, sem superexpressão do *HER2* e com baixas taxas de proliferação celular (EROLES et al., 2012). O luminal B é semelhante ao A, pois também é positivo para RE e/ou RP, embora em menor nível, e pode apresentar expressão do *HER2*, seu prognóstico é pior que do luminal A (EROLES et al., 2012).

O subtipo de tumor com superexpressão de *HER2* é negativo para ambos os receptores hormonais, mas apresenta alta expressão do *HER2* e altos níveis de proliferação celular, possuindo o segundo pior prognóstico entre os outros

subtipos (Abreu; Koifman, 2002). O grupo basal-like explicita a ausência dos receptores hormonais e não expressa o *HER2*, além de ser positivo para CK5 (Citoqueratinas de padrão basal) e EGFR (Receptor de fator de crescimento epidérmico); esse grupo apresenta o pior prognóstico dentre todos os subtipos (EROLES et al., 2012). Por último, o *normal-like*, também é negativo para os três marcadores imunoistoquímicos (ER-/PR-/HER2-), mas, como também é negativo para CK5 e EGFR, não é classificado como *basal-like*. Os tumores *normal-like* apresentam um prognóstico intermediário quando comparados com os demais (EROLES et al., 2012).

Apenas 5-10% do total de casos de câncer de mama têm causa hereditária (INCA, 2021), mas sabe-se que nos casos hereditários os principais genes envolvidos são o *BRCA1*, *BRCA2* (genes 1 e 2 do Câncer de Mama) e *tp53*, todos considerados genes supressores de tumor, uma vez que mutações em qualquer um desses genes aumentam consideravelmente o risco de desenvolver câncer de mama (Leistner-Segal et al., 2001). A maioria dos casos da doença, no entanto, são decorrentes de uma combinação de vários fatores extrínsecos (como alimentares, hormonais, hábitos de vida, tabagismo, alcoolismo entre outros) e fatores intrínsecos (fatores genéticos como o efeito combinado de vários genes).

A proteína p53, conhecida como “guardiã do genoma”, é codificada pelo gene *tp53*. A inativação desse gene ou inibição da proteína certamente aumentará consideravelmente o risco de uma pessoa desenvolver câncer (Hollstein et al., 1991). O primeiro gene a ser descrito como regulador da p53 foi o *MDM2*, encontrado em altos níveis de expressão em tecidos tumorais (Momand et al., 1992).

2.2 FAMÍLIA *MDM*

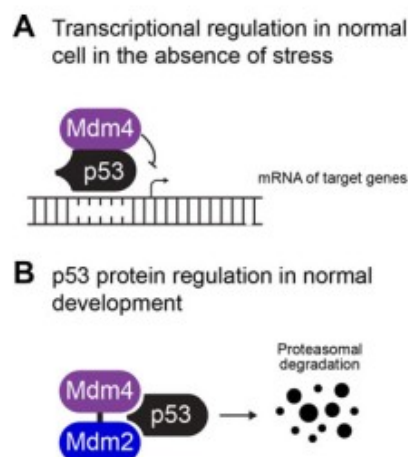
Desde a descoberta da relação entre o *MDM2* e a p53, a família de genes *MDM* (*Mouse Double Minute*) vem sendo alvo de estudos no câncer. Essa família compreende os genes *MDM2* e *MDM4* e regula as respostas ao estresse celular em condições normais (Karni-Schmidt et al., 2016), mas é principalmente conhecida por regular negativamente a proteína p53 (Haupt et al., 1997). O gene *MDM2* é o principal regulador negativo da p53 por meio de sua capacidade de ubiquitinizá-la e bloqueá-la de suas funções podendo se ligar a ela de diferentes maneiras: seu

transcrito apresenta duas isoformas, uma com 1,2kB - responsável por inibir p53 - e outra versão menor, que é capaz de formar um dímero com a versão maior e sequestrá-la de sua função, aumentando os níveis do supressor de tumor (Evans et al., 2001).

2.2.1 MDM4

O gene *MDM4* é homólogo ao *MDM2* e ainda não tem suas funções totalmente compreendidas, mas sabe-se que seu produto é capaz de formar um heterodímero de forte ligação com o produto do *MDM2*, o qual degrada a p53. Além disso, o *MDM4* também pode regular negativamente e de forma direta a p53 pelo seu transcrito de função supressora (Francoz et al., 2006), no entanto, não apresenta a capacidade de ubiquitiná-la como faz o *MDM2* (Shafdan, Lopez-Parares, Yuan, 2012) (FIGURA 1). Ainda, altos níveis de expressão do *MDM4* são verificados em tumores de mama e estudos recentes têm mostrado que o *MDM4* e seus polimorfismos de nucleotídeo único (SNPs) estão relacionados à susceptibilidade ao câncer de mama (Atwal et al., 2009; Haupt et al., 2017).

FIGURA 1 - *MDM4* E REGULAÇÃO NEGATIVA DA p53

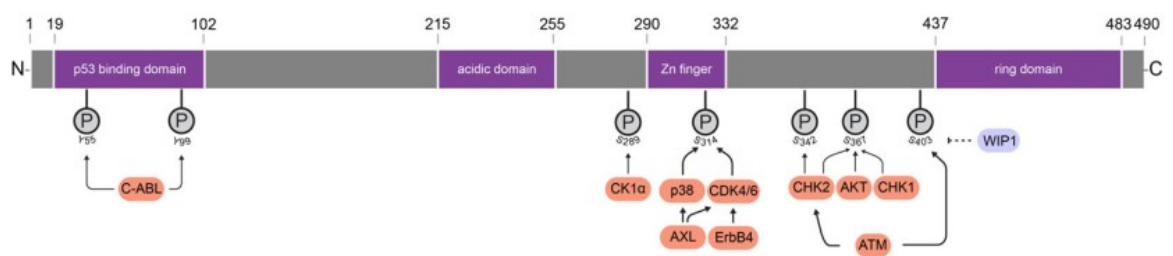


FONTE: adaptado de Haupt et al. (2019).

LEGENDA: em A, MDM4 inibe a p53 se ligando diretamente a ela. Em B, MDM4 se liga a MDM2 levando a degradação da p53.

O gene *MDM4* está localizado em 1q32 e tem aproximadamente 41 kB e 11 éxons. As isoformas de seus transcritos podem ser geradas por *splicing* alternativo ou por modificações pós-traducionais como fosforilação e ubiquitinação (Colacurcio et al., 2013). Entre as isoformas, no entanto, 2 são mais comuns: *MDM4-FL* (*full-length*) e *MDM4-S* (*short-length*), sendo a primeira mais estável do que a segunda. A isoforma *MDM4-S* não possui o *ring-domain*, responsável pela interação com o *MDM2*, porém é uma forte inibidora da p53 (Rallapalli et al. , 2003). Apesar de ter baixa expressão na maioria dos tecidos normais do corpo, o *MDM4* é comumente expresso em tecidos mamários (Haupt et al., 2015). Em sua estrutura, há quatro domínios principais e altamente conservados que permitem sua interação com a p53 e com a proteína *MDM2*, como citado no parágrafo anterior (FIGURA 2).

FIGURA 2 – ESTRUTURA DO *MDM4*



Fonte: Haupt et al. (2019).

LEGENDA: da esquerda para direita: região N-terminal (responsável pela interação com a p53); *acid domain* (estabiliza ligação com a p53); *zinc finger* (estabiliza ligação com o transcrito do *MDM2*); *ring finger* (responsável pela interação com o transcrito do *MDM2*).

Estudos anteriores do gene *MDM2* mostraram que mesmo uma pequena mudança de pares de base na fita de DNA afeta a via de sinalização da p53, além de um SNP específico ter sido detectado (rs2279744, SNP 309 – T/G) e relacionado ao desenvolvimento de tumores, dentre eles, o de mama (Poyurovsky & Prives, 2006). A partir dessa descoberta, a estrutura do haplótipo do *MDM2* foi determinada em diferentes populações e constatou-se que o alelo G do SNP rs2279744 tinha maior frequência no haplótipo de células tumorais (Atwal et al., 2007). Nesse estudo, também foi notado que essa maior frequência desvia do padrão dos modelos de seleção neutra, o que possivelmente indica que a via da p53 pode ser acionada por uma pressão de seleção evolutiva.

Como o *MDM4* é um homólogo do *MDM2* e também está envolvido na via de regulação da p53, é possível que contenha em seu haplótipo SNPs funcionais que demonstrem variações alélicas em células de tumor. Isso foi demonstrado em estudo de Atwal e colaboradores, em 2009, no qual esses pesquisadores verificaram que em populações caucasianas os alelos de menor frequência estavam associados ao maior risco de câncer de mama ou de câncer precoce no ovário (Atwal et al., 2009). Dessa forma, foi possível demonstrar que o desequilíbrio de ligação entre as variantes desse gene pode ser usado para selecionar Tag SNPs e blocos de haplótipos envolvidos em doenças (Prat et al., 2010).

2.3 HAPLÓTIPOS

Um bloco de haplótipo representa um conjunto de alelos que estão muito próximos e, por isso, sofrem pouca ou nenhuma recombinação, levando ao desequilíbrio de ligação (LD); existem níveis de LD, sendo um bloco de haplótipo os alelos em desequilíbrio total de ligação, ou seja, não há nenhuma recombinação entre eles. Como diversas variantes são herdadas juntas nos blocos, é possível que, entre elas, haja alguma que esteja relacionada a alguma doença (Liang, 2013). A análise feita a partir de um bloco de haplótipo é mais precisa porque não é feita com base em apenas um único SNP, já que envolve diversas variantes próximas, e dessa forma garante que nenhuma - ou pouca - informação seja perdida (Schaffner et al., 2003). Assim que os blocos são definidos, calcula-se a frequência deles na população, encontrando os haplótipos mais comuns em diferentes grupos/populações e, se o resultado for positivo para uma associação de doença, pode-se procurar um SNP específico como causador de determinado fenótipo, ou mesmo identificar um haplótipo recorrente em determinado fenótipo (Schaffner et al., 2003).

2.4 POLIMORFISMOS DO *MDM4*

Kulkarni et al. (2009) selecionaram 15 SNPs do *MDM4* - dentre eles o rs10900596 e o rs10900597 - em alto desequilíbrio de ligação para verificar se havia alguma correlação entre o haplótipo do *MDM4* e o aparecimento precoce de câncer

de mama. Esses pesquisadores verificaram que o alelo C de rs1563828 (T/C) era mais frequente em tumores de mama sem receptor de estrogênio em amostras de populações caucasianas. Além disso, notaram que havia dois haplótipos principais nessa população, sendo que 70% dos haplótipos encontrados apresentavam o alelo C e 25% o alelo T (Kulkarni et al. , 2009). Isso indica que na presença de um desses dois alelos, os demais alelos do haplótipo são previsíveis, devido ao desequilíbrio de ligação. Isso facilita uma possível associação entre o genótipo de um único SNP com um fenótipo, sabendo que é possível prever o genótipo de outras variantes em desequilíbrio de ligação com esse SNP (Kulkarni et al. , 2009).

Há diversos SNPs (polimorfismos de nucleotídeos únicos) ao longo do *MDM4*, e alguns deles já têm seu papel reconhecido em alguns tipos de câncer. Um exemplo é o SNP rs4245739 (A/C), localizado na região 3' UTR do gene, o qual possui um sítio de ligação para o microRNA (miR-191) que, ligado ao transcrito do gene, bloqueia sua tradução (Liu et al., 2013). MicroRNAs são RNAs não codificantes de 18-25 pb que podem atuar como fatores epigenéticos, já que muitos são capazes de controlar a expressão gênica sem alterar a sequência de DNA. A grande maioria dos miRNA interfere na expressão dos genes ligando-se na região 3'UTR do transcrito (Kim et al., 2009). No estudo de Wynendaele e colaboradores (2010) o SNP rs4245739 (A/C), foi genotipado em amostras de pacientes com tumor de ovário e constatou-se que o miR-191 apresentava maior afinidade pelo alelo C do que pelo A. Assim, o alelo A foi associado ao aumento da expressão do *MDM4* e consequentemente, ao maior risco de câncer de ovário (Wynendaele et al., 2010). Posteriormente, em 2013, LIU e colaboradores identificaram que o mesmo acontece no câncer de mama, ou seja, o alelo C foi associado ao menor risco de desenvolver câncer de mama e o alelo A ao maior risco (Liu et al. , 2013).

3 MATERIAL E MÉTODOS

Os experimentos práticos foram realizados no Laboratório de Genética Molecular Humana (LGMH), localizado no Setor de Ciências Biológicas da UFPR.

3.1 SNPS

Os SNPs alvo desse trabalho são: *rs4245739(A/C)*, *rs56200417(C/T)*, *rs10900596(T/C)*, *rs10900597 (C/T)*. Os três últimos foram sequenciados no presente trabalho. Todos estão localizados na região 3'-UTR do *MDM4*.

O alelo A do SNP *rs4245739* é o ancestral, e o alelo C é o de menor frequência (ENSEMBL, 2021). Quanto ao o SNP *rs56200417*, o alelo ancestral é o T, enquanto o C é o alelo de menor frequência (ENSEMBL, 2021). Já para *rs10900596*, o alelo ancestral é o C e o de menor frequência é o T (ENSEMBL, 2021). Por último, o alelo C é ancestral do *rs10900597*, e o alelo T é o de menor frequência (ENSEMBL, 2021).

3.2 AMOSTRAS

As amostras de tumor são provenientes de pacientes com câncer de mama do Hospital Nossa Senhora das Graças, em Curitiba-PR. A coleta das amostras foi realizada com aprovação do comitê de ética em Pesquisa em Seres Humanos do Setor de Ciências da Saúde, UFPR, sob o número CAAE: 67029617.4.0000.0102 e todas as pacientes assinaram o termo de consentimento livre e esclarecido.

Há 362 amostras de tumor de mama foram gentilmente cedidas pelo Laboratório de Citogenética Humana e Oncogenética (LabCHO) da UFPR. A classificação dos subtipos moleculares não foi encontrada em 80 amostras. As demais amostras foram classificadas como luminal A (n=81), luminal B (n=133), luminal (n = 19), HER2+ (n= 18) e triplo negativo (n=31). Além disso, as amostras foram classificadas quanto à histologia, presença/ausência de linfonodo, presença/ausência dos receptores hormonais de estrogênio e progesterona, porcentagem de Ki-67 e grau do tumor.

Já o grupo controle envolve 384 amostras provenientes do banco de amostras do Laboratório de Imunogenética e Histocompatibilidade (LIGH), localizado no Setor de Ciências Biológicas da UFPR. São amostras de pessoas do sexo feminino, acima de 40 anos e auto declaradas brancas.

3.3 PCR e REAÇÃO DE SEQUENCIAMENTO

Os primers foram desenhados com auxílio das ferramentas primer blast, olygoanalyzer e MFE primer e estão descritos na TABELA 1

TABELA 1 - SEQUÊNCIAS DOS *PRIMERS* UTILIZADOS PARA AMPLIFICAÇÃO DA SEQUÊNCIA ALVO

| | |
|-----------------------|--------------------------------|
| <i>Primer forward</i> | 5'-CCGGGGTTTTTCACTGCTTC -3' |
| <i>Primer reverso</i> | 5'-CATCCAAATGAAGAGGTTTGAAAG-3' |

Os SNPs sequenciados estão marcados na FIGURA 3, bem como os locais de anelamento dos primers forward e reverso.

FIGURA 3 – SEQUENCIA AMPLIFICADA DO *MDM4*

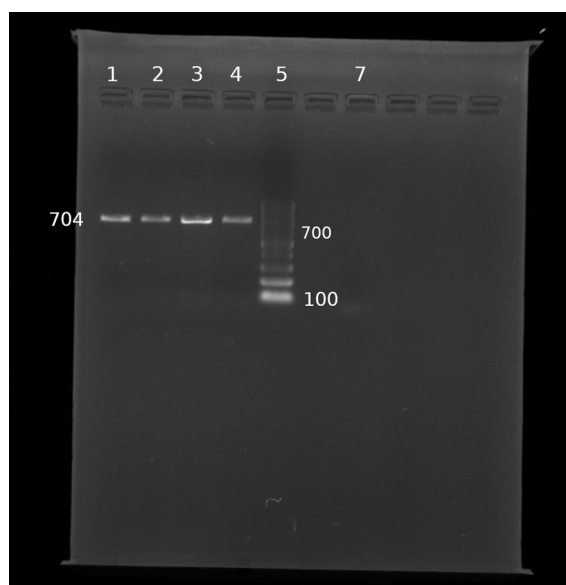
5'-
CCGGGGTTTTTCACTGCTTCTGTTAGCACTAAGTACTTAGACGATCCTAAGATATGTGCTTGAGCC
 GAATTCATCTTTACTTGTAGGAACTTTAAACTATTTCTTTTCTTTTCTTTTTTTTTTTTTTTTA**CTTG**
 AGATGGAGTTTTGCTCTTGTGCGCCAGGCTGGAGTGCAGTGGAGTGATCTCGGCTCACTGCAAC
 CTCTGCCTCCCGGGTTCAAATGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGT
 GCACCACCATGTCTGGCTAATTTTGTATTTTAGTAGAGATGGTTTCCCATGTTGGTCAGGCTGGTC
 TCGAACTCCTGACCTCAGGTCATCCACCCACCTCAGCCTCGCAAAGTGCTGAGATTACAGGCATG
 AGCCACAGCGCCAGCTTAAACTATTTTCTTGGTCTGTTTTTGATTTTCTTTTTTCTTTGCCACTGC
 GGTACAGATTTTTTTACTCACTGCCACTAACTAAAGCAAGGCATAGTTTATATGTGAAGTGTTCAG
 AGTTTACTGCTATAAGGAACTTCAAATACTGACATT**ACCTTTTAGCTGTAGTTATTGGGACCATG**
 TG**CTCTGGTTTTCTGGAGACTGCCAAATTGCTCCCATTTTTCTGCATCCCACCTGGTTTTCTTTCTG**
 CATGTCCCCTTCA**CTTTCAAACCTCTTCATTGGATG** -3'

LEGENDA sequência FASTA da fita antisense e indicação do local de anelamento dos primers forward (em verde) e reverso (em amarelo), além da indicação da localização dos SNPs, respectivamente rs56200417 em marrom, rs10900596 em rosa, rs10900597 em azul.

As amostras foram amplificadas pelo método da PCR convencional utilizando 5ul de DNA a 20pmol/ul, tampão 1x, dNTP 0.2mM, MgCl₂ 3mM, os primers forward e reverso 0.2uM e a Taq 0.1U/ul, em um volume final de 20ul. As condições de ciclagem no termociclador foram: 95° por 5 minutos, seguido por um ciclo repetido 30 vezes de 95° por 30 segundos, 59.8° por 30 segundos e 72° por 30 segundos e para finalizar 72° durante 5 minutos e 10° durante 10 minutos.

As ampliações foram verificadas por meio de corridas de eletroforese em gel de agarose 0.8% corado com *Unisafe Dye* (fluorescente sob luz ultravioleta), na condição de 80V, 70mA, 4W, durante 1 hora (FIGURA 4).

FIGURA 4 - ELETROFORESE DE VERIFICAÇÃO DE AMPLIFICAÇÃO DO FRAGMENTO



FONTE: o autor (2021)

LEGENDA: em 1, 2,3 e 4 estão as amostras utilizadas para padronização da reação de PCR, respectivamente: linhagem MCF-7, 35-LU5, 62-MOJ e 63-PSS, sendo as três últimas provenientes de pessoas saudáveis do painel de controle do LGMH. Em 5 o ladder de 100 pb e em 7 o controle negativo.

As amostras que amplificaram foram selecionadas para o sequenciamento, que possui as seguintes etapas: etapa de purificação, na qual foram utilizados 5ul do produto de PCR das amostras. Para cada amostra, foram adicionados 2ul do mix de purificação, que contém 0.4ul da exonuclease I (20U/ul), 1ul de FastAP (1U/ul) e 0.6 ul de água ultra pura. Depois disso, elas foram levadas para o termociclador durante 60 minutos a 20°C seguidos por 20 minutos a 80°C. Etapa da reação de sequenciamento: preparo do mix do sequenciamento que contém 5.6ul de água ultrapura, 0.2ul do *primer*, 1.9ul de tampão e 0.3ul de *big dye*, em um volume final de 8ul. Devido à presença de uma região poli-T próxima ao SNP *rs56200417*, apenas um mix com o *primer reverso* foi preparado (dessa forma, a região poli-T localizadas no final da fita 3'-5' não interferiu na análise dos 3 SNPs). Por último, 8ul desse mix de sequenciamento foram pipetados em cada poço de uma placa **MicroAmp** de 96 poços e em seguida foram adicionados 2ul dos produtos de PCR purificados. Após isso, a placa foi para o termociclador durante 2 horas e meia.

Depois da reação de sequenciamento, as amostras passaram por uma nova reação de purificação: foram adicionados 2,5 µL de EDTA 125 mM gelado e 30 µL de etanol absoluto a cada amostra. Em seguida, foi realizada incubação por 15 min à temperatura ambiente e centrifugação à 18°C, 1300 rpm por 45 min. O sobrenadante foi descartado e a placa centrifugada invertida por 6 segundos até 400 rpm. Foram adicionados 30 µL de etanol 70% e a placa foi centrifugada a 25°C, 1300 rpm por 5 min. O sobrenadante foi descartado e a placa novamente centrifugada invertida por 6 segundos a 400 rpm. A placa, então, foi mantida a temperatura ambiente por 20 min para a evaporação do etanol. As amostras ficaram armazenadas a -20°C até serem sequenciadas. Antes de serem injetadas no sequenciador 3500 *Series Genetic Analyzers (The Applied Biosystems™)*, as amostras foram ressuspensas em 10 µL de Hi-Di Formamida (*Life Technologies*).

3.4 ANÁLISE DAS SEQUÊNCIAS

As sequências obtidas foram alinhadas à sequência consenso de referência obtida no NCBI (National Center for Biotechnology Information). Assim, foi possível definir os alelos presentes em cada variante para cada amostra. Os

alinhamentos foram feitos por meio do programa *Mutation Surveyour*® versão 3.30 (*SoftGenetics*).

3.5 ANÁLISE DOS DADOS

As frequências alélicas, genótípicas foram calculadas pelo pacote “*genetics*” implementado na linguagem de programação R, bem como o equilíbrio de *Hardy-Weinberg*, os cálculos de qui-quadrado e o teste exato de Fisher para as tabelas de contingência (WARNES, 2009). Algumas análises de correlação foram feitas por meio do software SPSS 23 (IBM, 2020). Já as frequências dos haplótipos foram calculadas pelo programa PHASE 2.1.1 (STEPHENS et al., 2001; STEPHENS & SCHEET, 2005). O valor de p igual ou menor que 0,05 foi adotado como limiar de significância para todos os testes estatísticos realizados.

3.6 ANÁLISE DO HAPLÓTIPO COM A VARIANTE *rs4245739*

Para nossas análises, utilizamos também as genotipagens do SNP *rs4245739* (A/C), realizadas em um subconjunto das amostras aqui estudadas e que foram feitas pelo Laboratório de Citogenética Humana e Oncogenética (LABCho). Entre as amostras previamente genotipadas e coincidentes com o presente estudo, 17 amostras tumorais e 43 amostras do grupo controle foram empregadas na construção dos haplótipos.

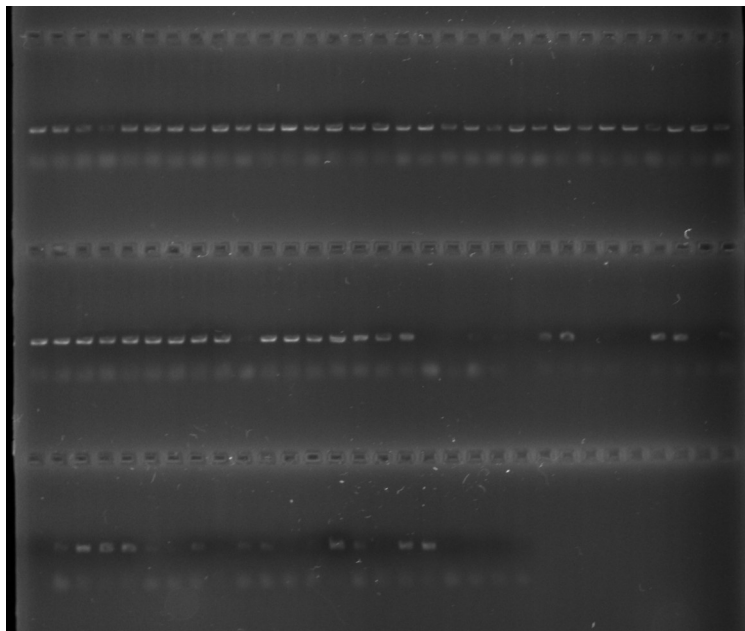
4 RESULTADOS E DISCUSSÃO

4.1 AMOSTRAS SELECIONADAS PARA AS ANÁLISES ESTATÍSTICAS

Entre as 362 amostras de tumor, 146 não amplificaram na PCR e foram excluídas deste projeto. Outras 60 não tiveram sequências de qualidade no sequenciamento e também foram excluídas. Dessa forma, restaram 156 amostras de tumor de mama, que tiveram seus alelos anotados para cada variante selecionada. IMAGEM 5 exemplifica amostras de tumor amplificadas na PCR.

Entre as 384 amostras controle, 25 também não amplificaram na PCR ou não tiveram boas sequências no sequenciamento e foram excluídas do trabalho. Assim, 183 amostras tiveram seus alelos anotados para cada variante.

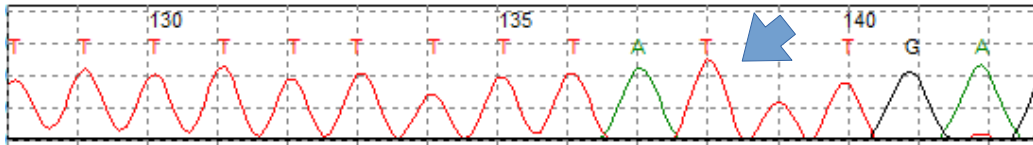
FIGURA 5 – ELETROFORESE COM AMOSTRAS DE TUMOR AMPLIFICADAS NA PCR



FONTE: a autora (2021).

As figuras 5, 6 e 7 são exemplos da leitura de sequenciamento dos SNPs realizados nas amostras controle e de tumores.

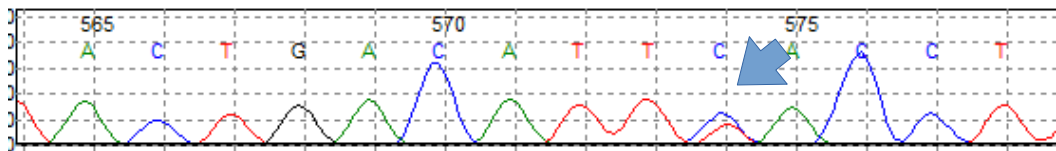
FIGURA 6 - LEITURA DO SEQUENCIAMENTO - SNP *rs56200417*



Fonte: a autora (2021).

LEGENDA: indivíduo com genótipo TT para a variante *rs56200417*. A numeração indica a posição da variante na sequência lida – neste caso, posição 138.

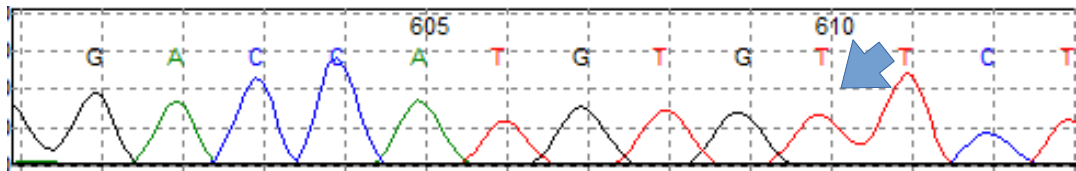
FIGURA 7 – LEITURA DO SEQUENCIAMENTO - SNP *rs10900596*



Fonte: a autora (2021).

LEGENDA: indivíduo com genótipo TC para a variante *rs10900596*. A numeração indica a posição da variante na sequência lida – neste caso, posição 574.

FIGURA 8 - LEITURA DO SEQUENCIAMENTO – SNP *rs10900597*



Fonte: a autora (2021).

LEGENDA: indivíduo com genótipo TT para a variante *rs10900597*. A numeração indica a posição da variante na sequência lida – neste caso, posição 610.

4.2 ESTATÍSTICA DAS AMOSTRAS DAS PACIENTES

A presença de acometimento dos linfonodos em câncer de mama aumenta as chances da ocorrência de metástase, sendo um fator negativo para o prognóstico (Fisher; Slack & Bross, 1969). Entre as 156 amostras de tumor de mama sequenciadas, cerca de 44% (N=69) são negativas para linfonodos, 28% (N=44) são

positivas e 26% (N=42) não estão classificadas quanto à presença ou não de linfonodos.

A presença/ausência de receptores hormonais também é uma característica molecular importante para o prognóstico da paciente, principalmente porque a presença de receptores de estrógeno e progesterona permite tratamentos hormonais (Mason et al., 1983). Entre as 156 amostras tumorais, 57% (N= 89) eram positivas para ambos os receptores (estrógeno e progesterona), 6.4%(N= 10) eram positivas para estrógeno e negativas para progesterona, 15% (N=23) eram negativas para ambos e apenas uma era negativa para estrógeno e positiva para progesterona. O restante não estava classificado quanto a presença de receptores hormonais.

Outro fator analisado foi a porcentagem de Ki-67, um marcador de proliferação celular. Geralmente porcentagens mais altas (>30%) de Ki-67 são associadas a um prognóstico ruim, já que indicam que o tumor está com taxas altas de proliferação, enquanto porcentagens baixas (<10%) são associadas a um prognóstico melhor (Nishimura et al., 2010). Cerca de 32% (N=50) das nossas amostras possuíam Ki-67>30%, 20% (N=32) possuíam Ki-67<10% e 18% (N= 28) apresentavam níveis intermediários (10%<Ki-67<30%). O restante das amostras não estava classificado quanto a porcentagem deste marcador.

As amostras foram também classificadas quanto ao grau do tumor, avaliado em 3 categorias: I, II e III - sendo que quanto maior o grau tumoral, pior o prognóstico. Entre as amostras tumorais, 40% (N= 63 amostras) foram classificadas em grau I e II e 20% (n = 31) das amostras eram grau III. As demais amostras, 40%, não foram classificadas para essa variável.

4.3 FREQUÊNCIAS ALÉLICAS, GENOTÍPICAS E HAPLOTÍPICAS

4.3.1 Grupo controle e pacientes

As frequências alélicas (TABELA 2) e genotípicas (TABELA 3) de cada polimorfismo estudado foram determinadas para as amostras do grupo controle e para as amostras tumorais. A distribuição das frequências genotípicas de todos os SNPs, em ambas as amostras, encontram-se em equilíbrio de Hardy-Weinberg (TABELA 3).

Não foram observadas diferenças significativas na distribuição das frequências alélicas e genotípicas entre o grupo controle e as amostras de tumor de mama.

A ausência de significância deve-se à heterogeneidade genética do câncer (McCullough, 2013). É raro encontrar diferenças genéticas significativas quando se analisa o genótipo sem correlacioná-lo com características histológicas e moleculares dos tumores.

TABELA 2 – FREQUÊNCIAS ALÉLICAS DAS VARIANTES *rs56200417*, *rs10900596* e *rs10900597* NO GRUPO CONTROLE E NAS PACIENTES

| SNP | Alelo | CONTROLES | | PACIENTES | | |
|-------------------|-------|------------------------------|------------|------------------------------|------------|------------|
| | | Número de indivíduos (N=183) | Frequência | Número de indivíduos (N=156) | Frequência | Valor de p |
| <i>rs56200417</i> | C | 128 | 0.35 | 128 | 0.41 | p=0.17 |
| | T | 238 | 0.65 | 184 | 0.59 | p=0.11 |
| <i>rs10900596</i> | C | 238 | 0.65 | 191 | 0.61 | p=0.27 |
| | T | 128 | 0.35 | 121 | 0.39 | p=0.34 |
| <i>rs10900597</i> | T | 234 | 0.64 | 184 | 0.59 | p=0.21 |
| | C | 132 | 0.36 | 128 | 0.41 | p=0.21 |

FONTE: a autora (2021)

TABELA 3 – FREQUÊNCIAS GENOTÍPICAS DAS VARIANTES *rs56200417*, *rs10900596* e *rs10900597* NO GRUPO CONTROLE E NAS PACIENTES

| SNP | Genótipo | CONTROLES | | | PACIENTES | | |
|-------------------|----------|------------------------------|------------|---|------------------------------|------------|---|
| | | Número de indivíduos (N=183) | Frequência | Equilíbrio de Hardy-Weinberg (valor de p) | Número de indivíduos (N=156) | Frequência | Equilíbrio de Hardy-Weinberg (valor de p) |
| <i>rs56200417</i> | TT | 81 | 0.44 | p=0.26 | 58 | 0.37 | p=0.31 |
| | TC | 77 | 0.42 | | 69 | 0.44 | |

| | | | | | | | |
|------------|----|----|------|--------|----|------|--------|
| | CC | 25 | 0.14 | | 29 | 0.19 | |
| rs10900596 | TT | 25 | 0.14 | p=0.32 | 25 | 0.16 | p=0.60 |
| | TC | 77 | 0.42 | | 71 | 0.46 | |
| | CC | 81 | 0.44 | | 60 | 0.38 | |
| rs10900597 | TT | 79 | 0.43 | p=0.1 | 59 | 0.38 | p=0.31 |
| | TC | 75 | 0.41 | | 66 | 0.42 | |
| | CC | 29 | 0.16 | | 31 | 0.20 | |

FONTE: a autora (2021).

Os haplótipos foram representados conforme a ordem em que ocorrem no cromossomo: *rs56200417*, *rs10900596*, *rs10900597*. Foram observados 6 combinações haplotípicas (TABELA 4), sendo que os haplótipos TCT e CTC foram os mais frequentes nos dois grupos amostrais. No grupo controle essas combinações correspondem a 97% dos haplótipos possíveis na população brasileira. Sabendo que essas variantes estão em desequilíbrio de ligação e, portanto, são herdadas em blocos de haplótipos, é possível, a partir de um único alelo, inferir o restante do haplótipo sem a necessidade de sequenciar toda a região.

Obseamos uma diferença significativa ($p=0,049$) entre a frequência do haplótipo TCT em pessoas saudáveis e em tumores. Essa combinação confere maior proteção ao desenvolvimento tumoral (OR = 0,73) (TABELA 4). Interessante notar que esse haplótipo contém o alelo C do SNP *rs10900596*, como analisado previamente pelo nosso grupo de pesquisa em um estudo *in silico*, o qual previu que esse alelo é alvo do microRNA-1257 (Schmidt, 2016). A possível ligação do miRNA-1257 ao alelo C presente no mRNA do gene MDM4 tem o potencial de interferir na expressão desse gene. A ausência da proteína MDM4 é importante para que não haja inibição da p53 e ela possa exercer suas funções supressoras tumorais (Shvarts et al., 1996). Esse resultado corrobora a nossa hipótese de que o alelo C do SNP *rs10900596* está relacionado ao menor risco de desenvolvimento do câncer de mama, caracterizando a combinação haplotípica TCT como possível marcadora de prognóstico no câncer de mama. As outras combinações haplotípicas contendo

os 3 SNPs não se apresentaram significativamente diferentes entre os controles e as amostras tumorais.

TABELA 4 - FREQUÊNCIAS HAPLOTÍPICAS NO GRUPO CONTROLE E NAS PACIENTES - ORDEM DAS VARIANTES: *rs56200414*, *rs10900596*, *rs10900597*

| Haplótipos | CONTROLES | | PACIENTES | | Valor de p |
|------------|------------------------------|------------|------------------------------|------------|--|
| | Número de indivíduos (N=183) | Frequência | Número de indivíduos (N=156) | Frequência | |
| TCT | 231 | 0.63 | 174 | 0.558 | P=0.049 OR=0.73 (CI = 95% : 0.425 – 1.04 |
| CTC | 124 | 0.34 | 121 | 0.388 | p=0.21 |
| CCC | 1 | 0.002 | 2 | 0.006 | p=0.89 |
| TCC | 5 | 0.02 | 8 | 0.026 | p=0.40 |
| CCT | 2 | 0.003 | 5 | 0.016 | p=0.32 |
| CTT | 3 | 0.005 | 2 | 0.006 | p=1 |

FONTE: a autora (2021).

LEGENDA CI: intervalo de confiança.

4.3.2 Subtipos tumorais das pacientes

As amostras tumorais analisadas pertenciam na sua maioria ao subtipo luminal (66%). Sendo assim, verificamos se o alelo C do *rs10900596* do gene *MDM4* relacionava-se com o os subtipos Luminal A e Luminal B. Constatamos que há uma diferença significativa na distribuição do alelo C entre os tumores luminais A e B ($p=0.048$). 9% da amostras tumorais classificadas como Luminal A não portavam o alelo C e 22% dos tumores Luminal B também não apresentavam esse alelo (TABELA 5) . Os tumores não portadores do alelo C tem maior chance de serem do tipo Luminal B (OR = 2,826). Esses resultados indicam que o alelo C está associado a um tipo tumoral de melhor prognóstico clínico.

TABELA 5 - ANÁLISE DO ALELO C DO SNP *rs10900596* EM LUMINAL A E B

| Subtipo tumoral <i>rs10900596</i> | Luminal A (N=44) | Luminal B (N=59) |
|--------------------------------------|---------------------|---------------------|
| Presença do alelo C | 0.91 | 0.78 |
| Ausência do alelo C | 0.09 | 0.22 |

Fonte: a autora (2021).

LEGENDA: IC- intervalo de confiança.

4.4 ANÁLISES ESTATÍSTICAS COM A VARIANTE *rs4245739* PARA MONTAGEM DE HAPLÓTIPO

Para montagem do haplótipo completo do *MDM4* incluindo a variante já estudada (*rs4245739*) e as variantes desse estudo (*rs56200417*, *rs10900596*, *rs10900597*) utilizamos uma sub-amostra de tumores e de controles que foram anteriormente genotipadas pelo Laboratório de Citogenética Humana e Oncogenética.

4.4.1 Frequências haplotípicas do grupo controle e das pacientes para as variantes *rs4245739*, *rs56200417*, *rs10900596*, *rs10900597*

Dentre estes quatro polimorfismos, dois deles são alvos principais das nossas análises por suas interações com miRNAs, *rs4245739* com mir-191 e miR887 (Kohler, 2020) e *rs10900596* alvo de interação com mir-1257 (Maximiano, 2019). Ambos microRNAs possuem maior afinidade pelo alelo C desses polimorfismos. As frequências haplotípicas foram determinadas nas duas amostras de populações deste trabalho (controle saudável e tumores de mama) (TABELA 6).

TABELA 6 - FREQUÊNCIAS HAPLOTÍPICAS DO GRUPO CONTROLE E DAS PACIENTES PARA AS VARIANTES *rs4245739*, *rs56200417*, *rs10900596*, *rs10900597*

| | CONTROLES | | PACIENTES | | |
|------------|--------------------------------|------------|--------------------------------|------------|------------|
| Haplótipos | Número de indivíduos (N=43) | Frequência | Número de indivíduos (N=17) | Frequência | Valor de p |

| | | | | | |
|-------------|-----------|-------------|-----------|-------------|---------------|
| ATCT | 40 | 0.45 | 20 | 0.59 | p=0.31 |
| ACTC | 17 | 0.21 | 0 | 0.00 | p=0.01 |
| CTCT | 14 | 0.17 | 1 | 0.03 | p=0.04 |
| CCTC | 12 | 0.13 | 12 | 0.35 | p=0.04 |
| CTCC | 3 | 0.03 | 0 | 0.00 | p=0.67 |
| ATCC | 0 | 0.00 | 1 | 0.03 | p=0.85 |

FONTE: a autora (2021).

Entre os três haplótipos encontrados em frequências significativamente diferentes entre o grupo controle e o grupo de pacientes, dois estão em maior frequência no grupo controle: ACTC e CTCT. Já o terceiro, CCTC, está em maior frequência nas pacientes.

Vale ressaltar que o polimorfismo *rs4245739*, diferentemente dos outros três, não está em desequilíbrio total de ligação com as outras variantes, o que aumenta a variação nas combinações de haplótipo. Os valores de LD da referida variante com as outras estão na TABELA 7: há dois parâmetros distintos, r^2 e D' . O r^2 basicamente nos informa a ocorrência ou não de recombinação no haplótipo (HILL & ROBERTSON, 1968). Já o D' indica a razão da diferença entre as frequências haplotípicas observadas e esperadas pelas frequências alélicas observadas, sendo também informativo quanto a eventos de recombinação (LEWONTIN, 1964). Em ambos os parâmetros, quanto mais próximos de 1, menos recombinação está ocorrendo, e portanto, mais alto é o desequilíbrio de ligação. Diz-se que desequilíbrio total de ligação acontece quando os dois parâmetros são iguais a 1.

O haplótipo ACTC não apareceu em nenhuma paciente, diferente do que era esperado, já que, teoricamente, os alelos A e T, respectivamente das variantes *rs4245739* e *rs10900596*, eram considerados como possíveis fatores de risco ao câncer de mama. No caso do SNP *rs4245739*, trabalhos prévios já haviam relacionado o alelo C ao menor risco de câncer de mama em populações europeias (Gansmo et al., 2015) e chinesas (Liu et al., 2013). No entanto, esses trabalhos não encontraram correlação entre o alelo A e o maior risco de desenvolvimento da doença ou piora no prognóstico. Nesse caso, o fato do haplótipo ACTC ter sido encontrado apenas em pessoas saudáveis aponta que os alelos A e T dos SNPs

avaliados não foram associados ao risco de desenvolver câncer de mama entre as amostras estudadas.

O haplótipo CTCT está significativamente em maior frequência no grupo controle do que nas pacientes. Como citado no parágrafo anterior, pesquisas recentes apontaram o alelo C da variante *rs4245739* associado ao menor risco de desenvolvimento de câncer de mama (Gansmo et al., 2015). Além disso, o alelo C da variante *rs10900596* foi predito como alvo do mir-1257 pelo nosso grupo de pesquisa. Esse resultado corrobora ainda mais a hipótese de que os alelos C dessas variantes estão associados ao menor risco de desenvolvimento de câncer de mama. E, ainda, a simultaneidade desses dois alelos pode ser um fator de proteção ao câncer de mama.

Por último, o haplótipo CCTC aparece em maior frequência nas pacientes do que no grupo controle. Isso nos permite inferir que a combinação dos alelos C dos SNPs *rs4245739* e *rs10900596* confere maior proteção do que a presença de apenas um alelo C desses dois SNPs no haplótipo. Ainda assim, é importante ressaltar o baixo número amostral (n=17) utilizado nessas análises, sendo necessários experimentos com um maior número amostral para confirmar este resultado.

TABELA 7 - VALORES DE LD ENTRE A VARIANTE *rs4245739* E AS VARIANTES *rs56200417*, *rs10900596* E *rs10900597*

| Variante 1 | Variante 2 | valor de r^2 | valor de D' |
|------------------|-------------------|----------------|-------------|
| <i>rs4245739</i> | <i>rs56200417</i> | 0.89 | 1.00 |
| <i>rs4245739</i> | <i>rs10900596</i> | 0.86 | 0.97 |
| <i>rs4245739</i> | <i>rs10900597</i> | 0.85 | 1.00 |

FONTE: a autora (2021).

5 CONSIDERAÇÕES FINAIS

É importante que os SNPs do *MDM4* continuem sendo genotipados e avaliados quanto as suas frequências alélicas e genotípicas, correlacionando-as ao risco de desenvolvimento de doenças, como o câncer de mama. Possivelmente, há mais de um tipo de combinação haplotípica que represente risco de câncer de mama, ao mesmo tempo em que há mais de uma combinação que representa menor risco. Neste trabalho, por exemplo, concluímos que o haplótipo CTCT do *MDM4* pode ser utilizado como um possível marcador de menor risco de desenvolvimento de câncer de mama.

Também é válido que mais amostras sejam genotipadas para o SNP *rs4245739* para confirmação das correlações entre os haplótipos de pessoas com câncer de mama e pessoas saudáveis.

REFERÊNCIAS

ABREU, E. DE; KOIFMAN, S. Fatores prognósticos no câncer da mama feminina. **Revista Brasileira de Cancerologia**, v. 48, n. 1, p. 113–131, 2002.

ABUL K. ABBAS, A. H. L. **tahir99-VRG & vip.persianss.ir**. 2016.

ATWAL, G. S.; BOND, G. L.; METSUYANIM, S.; et al. Haplotype structure and selection of the MDM2 oncogene in humans. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 11, p. 4524–4529, 2007.

ATWAL, G. S.; KIRCHHOFF, T.; BOND, E. E.; et al. Altered tumor formation and evolutionary selection of genetic variants in the human MDM4 oncogene (Proceedings of the National Academy of Sciences of the United States of America (2009) 106, 25 (10236-10241) DOI: 10.1073/pnas. 0901298106). **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 29, p. 12208, 2009.

COLACURCIO, D. J.; YEAGER, A.; KOLSON, D. L.; JORDAN-SCIUTTO, K. L.; AKAY, C. Calpain-mediated degradation of MDMx/MDM4 contributes to HIV-induced neuronal damage. **Molecular and Cellular Neuroscience**, v. 57, p. 54–62, 2013. Elsevier Inc. Disponível em: <<http://dx.doi.org/10.1016/j.mcn.2013.10.003>>. .

EVANS, S. C.; VISWANATHAN, M.; GRIER, J. D.; et al. An alternatively spliced HDM2 product increases p53 activity by inhibiting HDM2. **Oncogene**, v. 20, n. 30, p. 4041–4049, 2001.

ENSEMBL. Curitba, 2021 em: <
https://www.ensembl.org/Homo_sapiens/Variation/Explore?db=core;r=1:204552861-204553861;v=rs10900597;vdb=variation;vf=4693292>. Acesso em: 18 ago. 2021

EROLE, P.; BOSCH, A.; ALEJANDRO PÉREZ-FIDALGO, J.; LLUCH, A. Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. **Cancer Treatment Reviews**, v. 38, n. 6, p. 698–707, 2012. Elsevier Ltd.

FRANCOZ, S.; FROMENT, P.; BOGAERTS, S.; et al. Mdm4 and Mdm2 cooperate to inhibit p53 activity in proliferating and quiescent cells in vivo. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 9, p. 3232–3237, 2006.

GANSMO, L. B.; ROMUNDSTAD, P.; BIRKELAND, E.; et al. MDM4 SNP34091 (rs4245739) and its effect on breast-, colon-, lung-, and prostate cancer risk. **Cancer Medicine**, v. 4, n. 12, p. 1901–1907, 2015.

HAUPT, S.; BUCKLEY, D.; PANG, J. M. B.; et al. Targeting Mdmx to treat breast cancers with wild-type p53. **Cell Death and Disease**, v. 6, n. 7, 2015. Nature Publishing Group.

HAUPT, S.; MEJÍA-HERNÁNDEZ, J. O.; VIJAYAKUMARAN, R.; KEAM, S. P.; HAUPT, Y. The long and the short of it: The MDM4 tail so far. **Journal of Molecular Cell Biology**, v. 11, n. 3, p. 231–244, 2019.

HAUPT, S.; VIJAYAKUMARAN, R.; MIRANDA, P. J.; et al. The role of MDM2 and MDM4 in breast cancer development and prevention. **Journal of molecular cell biology**, v. 9, n. 1, p. 53–61, 2017.

HAUPT, Y.; MAYAT, R.; KAZAZT, A.; ORENT, M. letters to nature Mdm2 promotes the rapid degradation of p53. **Nature**, v. 387, n. May, p. 296–299, 1997.

HENG, Y. J.; LESTER, S. C.; TSE, G. M. K.; et al. The molecular basis of breast cancer pathological phenotypes. *Journal of Pathology*, v. 241, n. 3, p. 375–391, 2017.

HERSCHKOWITZ, J. I.; SIMIN, K.; WEIGMAN, V. J.; et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. **Genome Biology**, v. 8, n. 5, p. 1–17, 2007.

IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp. Oct 20, 2020.

KARNI-SCHMIDT, O.; LOKSHIN, M.; PRIVES, C. The Roles of MDM2 and MDMX in Cancer. **Annual Review of Pathology: Mechanisms of Disease**, v. 11, n. March, p. 617–644, 2016.

KIM, V. N.; HAN, J.; SIOMI, M. C. Biogenesis of small RNAs in animals. **Nature Reviews Molecular Cell Biology**, v. 10, n. 2, p. 126–139, 2009.

KULKARNI, D. A.; VAZQUEZ, A.; HAFFTY, B. G.; et al. A polymorphic variant in human MDM4 associates with accelerated age of onset of estrogen receptor negative breast cancer. **Carcinogenesis**, v. 30, n. 11, p. 1910–1915, 2009.

LEISTNER-SEGAL, S.; BITTELBRUNN, A.; BIAZÚS, J.; et al. Genética e câncer de mama. , v. 21, n. 2, p. 191–197, 2001.

LIU, J.; TANG, X.; LI, M.; et al. Functional MDM4 rs4245739 genetic variant, alone and in combination with P53 Arg72Pro polymorphism, contributes to breast cancer susceptibility. **Breast Cancer Research and Treatment**, v. 140, n. 1, p. 151–157, 2013.

MASON, B. H.; YEE, L. H.; HOLDAWAY, I. M.; MULLINS, P. R.; KAY, R. G. Progesterone and Estrogen Receptors as Prognostic Variables in Breast Cancer. **Cancer Research**, v. 43, n. 6, p. 2985–2990, 1983.

MCCULLOUGH, A. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. **Yearbook of Pathology and Laboratory Medicine**, v. 2013, p. 300–301, 2013. Mosby, Inc. Disponível em: <<http://dx.doi.org/10.1016/j.ypat.2012.10.022>>. .

MOMAND, J.; ZAMBETTI, G. P.; OLSON, D. C.; GEORGE, D.; LEVINE, A. J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. **Cell**, v. 69, n. 7, p. 1237–1245, 1992.

NISHIMURA, R.; OSAKO, T.; OKUMURA, Y.; et al. Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer. **Experimental and Therapeutic Medicine**, v. 1, n. 5, p. 747–754, 2010.

PICKSLEY, S. M.; LANE, D. P. What the papers say: The p53-mdm2 autoregulatory feedback loop: A paradigm for the regulation of growth control by p53? **BioEssays**, v. 15, n. 10, p. 689–690, 1993.

POYUROVSKY, M. V.; PRIVES, C. Unleashing the power of p53: Lessons from mice and men. **Genes and Development**, v. 20, n. 2, p. 125–131, 2006.

PRAT, A.; GUO, H.; GANESAN, P.; et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. **Breast Cancer Research**, v. 12, n. 5, p. 68–86, 2010. Disponível em: <<http://breast-cancer-research.com/content/12/5/R68%0ARESEARCH>>. .

RALLAPALLI, R.; STRACHAN, G.; TUAN, R. S.; HALL, D. J. Identification of a domain within MDMX-S that is responsible for its high affinity interaction with p53 and high-level expression in mammalian cells. **Journal of Cellular Biochemistry**, v. 89, n. 3, p. 563–575, 2003.

SCHAFFNER, S. F.; GABRIEL, S. B.; PATTERSON, N.; et al. Haplotype structure in the human genome. , v. 29, n. october, p. 217, 2003.

SCHMIDT, D.F. **Padronização de ensaios em linhagens tumorais de mama para a análise de MIRNAs**. Trabalho de graduação (Bacharelado em Biomedicina) – Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, 2016.

SHVARTS, A.; STEEGENGA, W. T.; RITECO, N.; et al. MDMX: A novel p53-binding protein with some functional properties of MDM2. **EMBO Journal**, v. 15, n. 19, p. 5349–5357, 1996.

SØRLIE, T.; PEROU, C. M.; TIBSHIRANI, R.; et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 19, p. 10869–10874, 2001.

SØRLIE, T.; TIBSHIRANI, R.; PARKER, J.; et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 14, p. 8418–8423, 2003.

WYNENDAELE, J.; BÖHNKE, A.; LEUCCI, E.; et al. An illegitimate microRNA target site within the 3' UTR of MDM4 affects ovarian cancer progression and chemosensitivity. **Cancer Research**, v. 70, n. 23, p. 9641–9649, 2010.