

Daniela Marques Saccaro

**Verificação de mutações somáticas em câncer de mama subtipo
luminal de pacientes jovens**

São Paulo

2020

Daniela Marques Saccaro

**Verificação de mutações somáticas em câncer de mama subtipo
luminal de pacientes jovens**

Tese apresentada à Faculdade de Medicina da Universidade de
São Paulo para a obtenção do título de Doutor em Ciências

Área de Concentração: Oncologia

Orientadora: Profa. Dra. Maria Aparecida Azevedo Koike
Folgueira

São Paulo

2020

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Saccaro, Daniela Marques
Verificação de mutações somáticas em câncer de mama
subtipo luminal de pacientes jovens / Daniela
Marques Saccaro. -- São Paulo, 2019.
Tese (doutorado)--Faculdade de Medicina da
Universidade de São Paulo.
Programa de Oncologia.
Orientadora: Maria Aparecida Azevedo Koike
Folqueira.

Descritores: 1.Pacientes 2.Neoplasias da mama
3.DNA 4.Genes 5.Mutação 6.Sequenciamento de
nucleotídeos em larga escala

USP/FM/DBD-460/19

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Agradecimentos

Diversas pessoas foram importantes para o desenvolvimento deste trabalho. Dessa maneira gostaria de agradecer previamente a todos.

À professora Dra. Maria Aparecida Azevedo Koike Folgueira pela dedicação e suporte durante todo o período de doutorado.

À Dra. Maria Mitzi Brentani pelos os conselhos e ensinamentos.

À equipe do Biobanco pela animada companhia no dia a dia do laboratório, tanto nos bons como nos maus momentos.

Aos profissionais da área de patologia do hospital pelo suporte.

À equipe administrativa da pós-graduação pelas informações e rápido retorno.

À todos os membros do meu grupo de pesquisa, pelo auxílio e apoio durante as diferentes etapas do projeto, especialmente para as queridas pesquisadoras Lúcia, Simone, Rose e Fátima, pelo companheirismo e atenção durante todo o período do doutorado.

Para finalizar, um agradecimento especial a minha família e amigos por toda a paciência e carinho.

Resumo

Saccaro DM. Verificação de mutações somáticas em câncer de mama subtipo luminal de pacientes jovens [Tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2020.

A maioria dos cânceres de mama são decorrentes de mutações somáticas que ocorrem ao longo da vida. As mulheres acima dos 50 anos têm maior risco de desenvolver a doença, porém cerca de 4% dos casos ocorre em mulheres muito jovens, com idade entre 20 e 34 anos. Nesta faixa etária, o câncer de mama pode apresentar pior prognóstico, associado a uma maior frequência de tumor de alto grau histológico, elevada taxa de proliferação celular, presença de invasão linfovascular e maior chance de doença triplo negativa. Apesar de pouco frequente em pacientes jovens, o câncer de mama é o câncer mais frequente em mulheres jovens e o subtipo luminal é o que corresponde ao maior número de casos nesta faixa etária. Entretanto, até o momento, poucos estudos foram dedicados à caracterização molecular do câncer de mama nestas pacientes. Logo, o objetivo deste trabalho é realizar a detecção de variantes somáticas em um painel de genes em câncer de mama subtipo luminal de pacientes jovens (≤ 35 anos). O painel incluiu potenciais genes *driver* identificados previamente em câncer de mama luminal de pacientes jovens, além de outros candidatos provenientes da literatura e bancos de dados de mutações somáticas. O DNA foi obtido de amostra tumoral emblocada em parafina e sangue periférico para o sequenciamento. Dentre as 19 amostras analisadas, 9 apresentaram variantes patogênicas em 13 genes diferentes. Duas pacientes destacaram-se por apresentar tanto variantes patogênicas somáticas no tumor como germinativas no sangue. Destas, uma paciente apresentou quatro variantes, sendo três alterações somáticas, uma alteração *nonsense* em *PRKARIA* e duas *missenses*, uma em *SEMA6D* e outra em *PTEN*, assim como uma alteração germinativa *missense* em *BRCA2*. Em uma única paciente foram identificadas duas alterações, ambas *frameshift*, sendo uma variante somática em *TP53* e uma germinativa em *BRCA1*. Dentre as variantes exclusivamente somáticas, três participantes apresentaram variantes em *GATA3*, todas *frameshift*, sendo este o gene mais frequentemente alterado nesta coorte. Uma destas participantes apresentou, em concomitância, uma variante *frameshift* em *GRHL2*. Outra paciente apresentou uma variante *frameshift* em *MEN1*. Uma única paciente apresentou 6 alterações somáticas, uma *nonsense* em *NF1* e 5 *missenses* em *CACNA1E*, *FAT2*, *NF1* e *AHNAK*. Outras duas pacientes apresentaram alterações *missenses* em *PIK3CA* e *AKT1*, cada um em uma paciente. Os genes relatados, exceto *FAT2* e *NF1*, foram previamente descritos em câncer luminal de mulheres jovens, resultado que comprova a sua ação na causalidade do câncer. Novas pesquisas devem ser realizadas para comprovar essas informações.

Palavras chave: pacientes, neoplasia da mama, DNA, genes, mutação e sequenciamento de nucleotídeos em larga escala.

Abstract

Saccaro DM. Somatic mutations in a gene panel in luminal breast cancer from young patients [Thesis]. São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2020.

Most breast cancers are due to somatic mutations that occur along the person lifetime. Women over 50 years have a higher risk of developing breast cancer, but about 4% of the cases occur in very young women, aged 20 to 34 years. In this age group, breast cancer may have a worse prognosis associated with a higher frequency of tumors with high histological grade, high cell proliferation rate, occurrence of lymphovascular invasion and higher chance of triple negative disease. Although uncommon in young patients, breast cancer is the most common cancer in young women and the luminal subtype accounts for the largest number of breast cancer cases in this age group. However, few studies have reported the molecular characterization of luminal cancer in young patients. Therefore, the objective of this study is to detect somatic variants in a panel of genes in luminal breast cancer from young patients (≤ 35 years). The panel included potential driver genes previously identified in luminal breast cancer from young patients, as well as other candidates reported in the literature and somatic mutation databases. DNA from paraffin-embedded tumor samples and peripheral blood mononuclear cells were obtained for next generation sequencing. Among the 19 samples analyzed, 9 presented pathogenic variants identified in 13 different genes. Two patients present both somatic pathogenic variants on tumor and germline on blood. Of these, one patient presented four variants, three somatic alterations, one nonsense in *PRKARIA* and two missenses in *SEMA6D* and *PTEN*, as well as a missense germinative alteration in *BRCA2*. Two variants were identified in a single patient, both frameshift, one somatic variant in *TP53* and one germline in *BRCAl*. Among the exclusively somatic pathogenic alterations, three participants presented *GATA3* variants, all frameshift. *GATA3* was the most frequently altered gene in this cohort. One of these participants presented, simultaneously, a somatic frameshift variant in *GRHL2*. Another patient presented a somatic frameshift variant in *MEN1*. A single patient presented 6 somatic variants, one nonsense in *NF1* and 5 missenses in *CACNA1E*, *FAT2*, *NF1* and *AHNAK*. Two other patients had missense alterations in *PIK3CA* and *AKT1*, each in one patient. The reported genes, except *FAT2* and *NF1*, were previously described in luminal cancer of young women. These result indicated that these genes have important functions on breast cancer of young patients. New research need to be perform to prove these information.

Keywords: patients, breast cancer, DNA, genes, large-scale nucleotide mutation and sequencing.

SUMÁRIO

| | |
|-----------------------------------------------------------------------------------|----|
| 1. INTRODUÇÃO..... | 1 |
| 1.1. Mutações Somáticas..... | 2 |
| 2. OBJETIVO..... | 9 |
| 3. CASUÍSTICA E MÉTODOS..... | 10 |
| 3.1 Pacientes jovens com câncer de mama..... | 10 |
| 3.2 Processamento do sangue para separação de células mononucleares..... | 10 |
| 3.3 Extração do DNA de células mononucleares do sangue..... | 11 |
| 3.4 Extração do DNA de tecido tumoral emblocado em parafina..... | 12 |
| 3.5 Sequenciamento de DNA de tecido tumoral e células mononucleares do sangue.... | 14 |
| 3.6 Seleção de Genes para o Painel de Mutações Somáticas..... | 20 |
| 3.7 Análise das variantes detectadas no sequenciamento do painel de genes..... | 24 |
| 3.7.1 Ferramentas de predição de variantes <i>missenses</i> patogênicas..... | 24 |
| 4.0 RESULTADOS..... | 28 |
| 4.1 Variantes gênicas somáticas..... | 30 |
| 4.1.1 Variantes gênicas somáticas patogênicas..... | 35 |
| 4.1.2 Variantes gênicas germinativas patogênicas..... | 37 |
| 5.0 DISCUSSÃO..... | 38 |
| 6.0 CONCLUSÃO..... | 42 |
| REFERÊNCIAS BIBLIOGRÁFICAS..... | 43 |
| ANEXO..... | 46 |

1. Introdução

O câncer de mama é o carcinoma mais frequentemente diagnosticado e a principal causa de morte entre mulheres na população mundial. De acordo com o *Global Cancer Observatory* foram estimados para a América Latina em torno de 572.000 novos casos de câncer mama para o ano 2020 (GLOBOCAN, 2018). No Brasil o câncer de mama é o câncer mais frequente entre mulheres, com exceção do câncer de pele não melanoma. Em pacientes brasileiras, a incidência de câncer de mama é elevada em diversas regiões do país, especialmente no Sudeste e Sul. Entretanto a região Norte é uma exceção pois a taxa de incidência deste câncer é semelhante ao de câncer de colo do útero (INCA, 2019).

Diversos fatores podem influenciar o desenvolvimento e progressão do câncer de mama como a idade (incidência crescente de acordo com o aumento da idade), histórico familiar de câncer de mama e ovário, histórico hormonal (menarca precoce e menopausa tardia) e histórico reprodutivo (1^ogestação tardia e ausência de gestação). Em relação à idade, as mulheres acima dos 50 anos são as mais propensas a ser diagnosticadas com câncer de mama. Entretanto 4,0% dos casos ocorrem em mulheres jovens, ou seja, com menos de 40 anos (FOSP 2015). De acordo com diversos estudos, as pacientes jovens com câncer de mama apresentam um pior prognóstico com maior índice de mortalidade e menor sobrevida global em comparação às pacientes pós-menopausadas (Han et al., 2004; Bollet et al., 2007).

As pacientes jovens tendem a apresentar maiores taxas de recidiva local, uma vez que podem ser diagnosticadas com mais frequência com câncer de mama de alto grau histológico, elevada taxa de proliferação e presença de invasão linfovascular, em relação às registradas em pacientes idosas (Ribnikar et al., 2015). Outra característica histológica predominante em pacientes jovens é o diagnóstico de carcinoma ductal invasivo, que, em torno de 50% dos casos, já apresentam comprometimento de linfonodos (Copson, et al., 2013).

Neste contexto podem-se citar alguns estudos clínico-patológicos realizados em tumores de mama provenientes de mulheres de diferentes faixas etárias. As pacientes jovens (≤ 40 anos) foram diagnosticadas com tumores de maior tamanho e com comprometimento de linfonodos (Anders et al., 2008). Outra pesquisa comparou o

tamanho dos tumores de mulheres jovens e pós-menopausadas sem metástase, e constatou que 60% dos tumores das pacientes jovens apresentaram-se em estádios mais avançados da doença (II e III) (Gajdos et al., 2000).

Em relação ao diagnóstico de câncer de mama considera-se o tipo histológico, grau histológico, expressão de receptores hormonais e fator de crescimento epitelial HER2 (WHO, 2018). Além disso, identifica-se o subtipo molecular do tumor; Luminal A, Luminal B, HER2 e Basal (Sorlie et al., 2001; Perou et al., 2000). O subtipo Luminal A apresenta alta expressão dos receptores de estrogênio e progesterona (ER+/PR+), baixo índice de proliferação celular, HER2 negativo e expressão de citoqueratinas 7/8 e 18/19. O subtipo Luminal B apresenta padrão de expressão semelhante ao Luminal A, porém com maiores taxas de proliferação celular e/ou expressão de HER2. Os outros dois subtipos são basal e HER2. Eles não expressam os receptores hormonais (ER-/PR-), e apresentam alto grau histológico e nuclear. Entretanto, há uma grande diferença entre esses dois subtipos. O HER2 é caracterizado pela expressão elevada deste oncogene (ER-/PR-/HER+). O subtipo basal é referido por muitos autores como triplo negativo (ER-/PR-/HER-), no entanto o paralelismo é parcial. Os tumores basais em geral expressam citoqueratinas 5, 6, 17 e/ou fator de crescimento epidérmico (EGF). Assim a partir do uso desses marcadores imuno-histoquímicos é possível estimar fenotipicamente o perfil do subtipo basalóide (Paik et al., 2004; Van't Veer et al., 2002).

1.1 Mutações Somáticas

O câncer de mama é caracterizado pela heterogeneidade tumoral, uma vez que as alterações genéticas do tecido epitelial mamário são responsáveis pelo desenvolvimento dos diferentes subtipos moleculares (Hanahan et al., 2011). Essas alterações genéticas podem ser mutações pontuais, deleções, amplificações e translocações, que ocorrem em células germinativas e/ou somáticas. As mutações germinativas são detectadas nos gametas (óvulos e espermatozóides), sendo transmitida após a fecundação para as gerações celulares subsequentes. Em contrapartida as mutações somáticas são provocadas pelo modo de vida de cada indivíduo ou por agentes carcinogênicos. No caso das mulheres com câncer de mama, a exposição endógena ao hormônio estrogênio é considerado um importante fator promotor de tumor. Outros fatores exógenos podem promover alterações hormonais como a

realização de terapia de reposição hormonal. Esses fatores carcinogênicos favorecem a promoção de modificações nos processos de crescimento, proliferação e sinalização celular (Balmain et al., 2003).

A maioria dos cânceres de mama (90%) é decorrente de mutações somáticas. Existem dois tipos de mutações somáticas: *driver* e *passenger*. As mutações *driver* proporcionam uma vantagem seletiva, portanto estão diretamente envolvidas na gênese do câncer e progressão tumoral. Elas são promovidas pela ativação de proto-oncogenes em oncogenes ou pela perda de função de genes supressores de tumor. As mutações *passenger* podem contribuir para a ocorrência da carcinogênese, mas apenas quando associadas às mutações *driver* (Greenman et al., 2007).

O grande desafio atual é identificar quais são as alterações *driver*, uma vez que as mutações *passenger* constituem a maioria das mutações somáticas (Garraway, L. & Lander, E. 2013). Novos algoritmos estão sendo desenvolvidos para a detecção de genes *driver*, sendo a maioria baseado na frequência ou padrão de mutações em tumores, assim como na patogenicidade prevista (Tokheim et al., 2016). A identificação precisa de genes *driver* é fundamental para o avanço das pesquisas e personalização do tratamento dos pacientes (Althubaiti et al., 2019).

Através de consulta no Catálogo de Mutações Somáticas (COSMIC), um total de 55.940 amostras de câncer de mama foram registradas, destas 12.341 apresentaram mutações somáticas, sendo principalmente nos genes *TP53* (26%), *PIK3CA* (19%), *ESR1* (8%), *CDH1* (11%), *GATA3* (10%) e *MAP3K1* (8%) (<https://cancer.sanger.ac.uk/cosmic>, pesquisa realizada em 19/02/2020).

O biólogo Theodor Boveri foi o primeiro a propor, que as mutações somáticas causam o câncer (Knudson, 2000; 2002). A primeira constatação deste conceito foi a demonstração de que a leucemia mielóide crônica por ocorrer através de uma translocação cromossômica recorrente, formando o cromossomo Filadélfia, conforme demonstrado por Nowell e Hungerford em 1960. Essa descoberta foi realizada através da técnica de citogenética, provavelmente o primeiro método utilizado para a exploração do genoma do câncer. Desde então, diversos oncogenes ativados por translocações cromossômicas foram identificados (Mitelman, 2000).

Além da técnica de citogenética outras abordagens sistemáticas do genoma possibilitaram a descoberta de genes importantes, como o mapeamento de genes de pacientes com histórico familiar, a análise da perda de alelos por deleções homozigotas e heterozigotas, e a transfecção de oncogenes (Weir et al., 2004).

Recentes descobertas foram realizadas como o sequenciamento de nova geração (*Next Generation Sequencing*, NGS). Ele possibilita a decodificação de todo o exoma ou genoma, através de uma caracterização molecular abrangente. O avanço da tecnologia NGS permitiu que os grupos de pesquisa sobre câncer utilizassem as técnicas de sequenciamento para identificar mutações em um maior número de amostras, assim como a descoberta de genes alterados com baixa frequência, de alterações em sítios de splicing e mutações em reguladores epigenéticos (Watson et al., 2013).

Além do sequenciamento do genoma e exoma foram elaboradas novas técnicas moleculares como os painéis de genes, através do qual é possível sequenciar amostras em um painel composto por um número determinado de genes relacionados a cânceres específicos. Uma variedade de painéis gênicos tem sido proposta para auxiliar no diagnóstico das pacientes com câncer de mama, do qual é possível identificar mutações em genes *driver* e indicar um tratamento personalizado. Esta técnica molecular apresenta algumas vantagens em relação às demais, o menor custo e o menor tempo necessário para o sequenciamento das amostras, fatores que favorecem a sua aplicação na rotina laboratorial (Gray et al., 2015).

O sequenciamento do painel de genes é realizado através da técnica *target sequencing* com a amplificação de múltiplas regiões genômicas simultaneamente (multiplex PCR). Essa técnica possibilita a análise de diversos genes, assim como o sequenciamento de amostras com determinadas limitações, como pouca quantidade de DNA e/ou DNA fragmentado, como o proveniente de tecido emblocado em parafina (Hoogstraat et al., 2015).

A partir da ampla aplicação do sequenciamento genômico em larga escala foram desenvolvidos modelos matemáticos direcionados para avaliar a evolução dos tumores, assim como auxiliar na identificação de genes mutados em diversos cânceres, como os que estão correlacionados a diferentes subtipos, como os subtipos moleculares de câncer de mama (Stephens et al., 2012).

A análise de mutações associada à utilização de novos algoritmos promoveram a definição de um padrão específico de frequência de mutações somáticas denominada assinatura mutacional. O perfil das assinaturas mutacionais é composto por diferentes combinações de nucleotídeos, que contém seis tipos de substituição (C> A, C> G, C> T, T> A, T> C e T> G) e 96 possíveis combinações de trinucleotídeos, validadas através do genoma de referência humano Hg19 (Nik-Zainal et al., 2016).

Um total de 12 assinaturas foi documentado em amostras de tumores de mama. As assinaturas 1 e 5 estão associadas à idade ao diagnóstico. As assinaturas 2 e 13 estão relacionadas à alterações epigenéticas promovidas pelas enzimas citidina desaminases da família APOBEC. As assinaturas 3 e 8 referem-se à deficiência dos genes supressores tumorais *BRCA1* e *BRCA2*. As assinaturas 6, 20 e 26 estão relacionadas à modificações no processo de reparo de DNA. As assinaturas 17, 18 e 30 são de etiologia desconhecida (Nick-Zaidal e Morganello, 2017).

Atualmente as inovações tecnológicas, sequenciamento do genoma e exoma completo, possibilitaram a detecção de alterações genéticas em grande escala, favorecendo a identificação de um maior número de assinaturas mutacionais (Helleday, Eshtad & Nik Zainal, 2014). A aplicação dessas técnicas moleculares promoveram a geração de um abrangente catálogo genômico, denominado Consórcio do Genoma do Câncer Internacional (ICGC). Ele foi gerado após o lançamento do *Cancer Genome Project* (CGP) no Reino Unido em 2000 e *The Cancer Genome Atlas* (TCGA) nos Estados Unidos em 2006. Os grupos de trabalho do TCGA e CGP disponibilizaram análises genômicas completas de diversos tipos de câncer, assim como efetuaram a caracterização molecular de cânceres comuns de diferentes populações localizadas em variadas regiões geográficas (Cancer Genomic Atlas, 2013).

De acordo com dados do TCGA (<https://portal.gdc.cancer.gov/>), até o dia 19/02/2020 um total de 8813 casos de câncer de mama foram depositados neste banco de dados, referente a mulheres na faixa etária de 20 - 80 anos, foram descritos como genes mais frequentemente alterados; *TP53* (36,89%), *PIK3CA* (34,43%), *CDH1* (15,27%), *GATA3* (14,14%) e *MAP3K1* (9,22%).

Em comparação aos estudos de pacientes pós-menopausadas, são poucas as pesquisas relacionadas à caracterização molecular do tumor de pacientes jovens. Para

fornecer mais informações sobre essas pacientes, algumas publicações relevantes estão descritas a seguir.

A partir de um grande número de pacientes com câncer de mama, 2500 mulheres, foi realizado um estudo comparativo entre mulheres jovens (≤ 45 anos) e idosas (≥ 55 anos), a partir de dados provenientes de estudos do TCGA e METABRIC. Como técnicas moleculares foram realizados transcriptomas (RNAseq) análises de metilação e proteínas (*Reverse Phase Protein Array*). Pela análise de transcriptoma foram identificados 5 genes com taxas de mutação significativamente diferentes entre os tumores de pacientes pré-menopausadas e pós-menopausadas, respectivamente; *CDH1* (14,5% vs 2,9%), *GATA3* (9,4% vs 18,6%), *MLL3* (13,3% vs 4,9%), *GPS2* (0,8% vs 3,9%) e *PI3KCA* (48,7% vs 37,3%). Destacaram-se em amostras de pacientes jovens a hipermetilação do gene *ESR1*, processo que promoveu modificações nas vias de sinalização desse gene. Pela análise integrativa dessas técnicas, os tumores das pacientes jovens apresentaram uma hiperexpressão das vias de sinalização de *EGFR* (fator de crescimento epitelial), *TGF β* (fator de necrose tumoral) e integrina/laminina (adesão celular) (Liao et al., 2015).

Um estudo de expressão gênica por microarray avaliou uma coorte de mulheres jovens com câncer de mama (≤ 45 anos) e idosas (≥ 65 anos). As amostras de pacientes jovens apresentaram uma desregulação das vias de transdução de sinal dos genes *PI3K*, *MYC* e *B-catenina*, associada a um pior prognóstico, resultado oposto ao obtido para as mulheres idosas (Anders et al., 2008).

Em um estudo direcionado para pacientes com câncer de mama subtipo luminal foram analisados tumores de 47 mulheres jovens da Coréia do Sul (≤ 35 anos). Essa coorte apresentou um maior índice mutacional (2,4 mutações por Mb), em relação ao referente perfil de pacientes relatadas em estudos do TCGA (1,12 mutações por Mb). Ambas as alterações, somáticas e germinativas, foram avaliadas e destacaram-se por promover alterações especialmente em genes de reparo do DNA; *BRCA1*, *BRCA2*, *TP53* e *MSH1*. Em relação a modificações no número de cópias, foram detectadas ampliações nos genes *CCND1*, *ERBB2*, *RPS6KB1*, *ZNF703*, e deleções nos genes *NCOR1* e *MAP2K3*. As assinaturas mais frequentemente detectadas em ambos os grupos etários foram relacionadas à enzima APOBEC, família proteica com função citidina deaminase, e transições C>T associadas à idade. Além destas considerações,

após a análise do perfil transcricional, as pacientes jovens foram classificadas em três subgrupos; Grupo A; pacientes luminal A (HER2 negativo) com ativação das vias de ER alpha, alteração da via de sinalização de *IGF1R*, e regulação negativa de *PLK1*; Grupo B: pacientes luminal B (HER2 negativo) com mutações no gene *GATA3*, ativação da transição epitélio-mesenquimal, instabilidade cromossômica, inativação do sistema imune e desregulação da via do reparo de DNA dupla fita; Grupo C: pacientes luminal B (HER2 positivo) com a promoção da ativação do sistema imune (*TNF*, *IFN-γ*, células T e proteínas CD28) (Park et al., 2019).

Em outro estudo realizado em pacientes coreanas pré-menopausadas na faixa etária com câncer de mama recidivado, entre 25 e 49 anos, foram analisados tumores de mama através do sequenciamento completo do exoma. Dentre as 63 pacientes jovens avaliadas foram detectadas, pelo menos, uma alteração em vias de sinalização relacionada à regulação do ciclo celular e modificações em p53. As mutações somáticas mais frequentes nesta coorte ocorreram nos genes *TP53*, *PIK3CA* e *GATA3*, seguido por *ERBB2*, *MYC*, *KMT2C*, *PRKARIA*, *CDKN1B*, *CCND1* e *MLH1*. Assim como no estudo anterior, nesta pesquisa foi realizada uma comparação entre pacientes jovens coreanas e pacientes com o mesmo perfil de estudos do TCGA e ICGC, isto é, jovens com recidiva do tumor. Entre os resultados relatados destacou-se a maior prevalência de alterações no gene *TP53* em pacientes da coorte asiática (Yap et al., 2018).

Outra pesquisa também identificou assinaturas mutacionais características de uma coorte asiática de pacientes jovens com câncer de mama. Neste trabalho foi realizado o exoma completo e sequenciamento do RNA de amostras de tumores de 76 pacientes japonesas jovens na faixa etária de 25-39 anos. A principal causa de mutagênese desta coorte foi associada à assinatura molecular APOBEC, seguida pela deficiência funcional dos genes *BRCA1* e *BRCA2*. Além das mutações somáticas, também foram detectadas mutações germinativas prevalentes nos genes *BRCA2*, *TP53* e *CHEK2* (Kanke et al., 2017).

Um estudo recente teve como objetivo realizar a caracterização molecular de pacientes jovens com câncer de mama de países da América Latina, incluindo Chile, Colômbia, Costa Rica e México. Atualmente este estudo está em andamento e os resultados prévios foram publicados. O sequenciamento do exoma de 12 amostras tumorais (2 luminal A e 10 triplo negativo) revelou uma média de 3,9 alterações por

MB (*missenses* e deleções). Entre os genes mais frequentemente alterados destacam-se *TP53*, *RBI*, *PIK3CA* e *AHNAK*, além de genes de grandes tamanhos como *MUC* e *TTN* (Oliver et al., 2019).

O nosso grupo também avaliou o perfil mutacional de tumores de mama subtipo luminal (HER2 negativo) de pacientes jovens (≤ 35 anos). As amostras de tumores foram submetidas ao sequenciamento do exoma completo, que promoveu a identificação de variantes somáticas em potenciais genes associados ao câncer como *GRHL2*, *PIK3API*, *CACNA1E*, *SEMA6D*, *SMURF2*, *RSBN1* e *MTHFD2*. As informações desse exoma foram integradas aos dados de 29 pacientes jovens, com o mesmo perfil (luminal HER2 negativo), provenientes do banco de dados COSMIC. Observou-se, por amostra, uma mediana de seis genes acometidos por variantes somáticas *missenses*, cujos genes mais frequentemente alterados foram *PIK3CA*, *GATA3*, *TP53* e *MAP2K4*. Do total de 37 tumores, 16 (43%) apresentaram variantes somáticas *missenses* em pelo menos um gene de reparo do DNA (*ATR*, *BAP1*, *ERCC6*, *FANCD2*, *FANCL*, *MLH1*, *MUTYH*, *PALB2*, *POLD1*, *POLE*, *RAD9A*, *RAD51* e *TP53*). Além disso, 20 de 37 tumores (54%) apresentaram variantes somáticas patogênicas, isto é, alterações *frameshift* ou *nonsense*, em pelo menos um gene envolvido no processo de transcrição gênica (Encinas et al., 2018).

2.Objetivo

O objetivo deste trabalho é ampliar a detecção de variantes gênicas somáticas em tumores de pacientes jovens brasileiras (≤ 35 anos) diagnosticadas com câncer de mama subtipo luminal, através da utilização de um painel gênico composto por potenciais genes *drivers*, identificados em estudo anterior do nosso grupo, além de outros candidatos provenientes da literatura e bancos de dados de mutações somáticas.

3. Casuística e Métodos

3.1 Pacientes jovens com câncer de mama

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo (CAAE: 887663218.0.0000.0065). Um total de 44 pacientes jovens (≤ 35 anos) foram convidadas, aceitaram participar do estudo e assinaram o Termo de Consentimento Livre e Esclarecido (TCLE).

As pacientes foram selecionadas a partir do levantamento de mulheres jovens com diagnóstico de câncer de mama, atendidas no período de janeiro de 2018 a fevereiro de 2019, no Instituto do Câncer do Estado de São Paulo (ICESP). O processo de seleção foi conduzido através do acesso ao Sistema TASY de Gestão (*Philips Healthcare*), que contém informações clínico-patológicas de pacientes do ICESP. A inclusão das pacientes neste estudo foi realizada de acordo com os seguintes critérios:

| Crítérios de inclusão de pacientes com câncer de mama | |
|--------------------------------------------------------------|-----------------------------------|
| Idade ao diagnóstico | Entre 18 a ≤ 35 anos |
| Tipo histológico do tumor | Carcinoma mamário ductal invasivo |
| Subtipo molecular | Luminal A e B |
| Imuno-histoquímica | ER+ HER2- |
| Tecido emblocado em parafina | material presente no ICESP |
| Celularidade do tecido tumoral | $\geq 40\%$ |

As pacientes selecionadas, conforme com estas características, foram contatadas via telefone, do qual verificamos o interesse e disponibilidade em participar do estudo. No dia da consulta médica, após o atendimento médico, apresentamos o projeto de pesquisa as pacientes e esclarecemos possíveis dúvidas. Após esse procedimento, aplicamos o TCLE e realizamos a coleta de sangue.

3.2 Processamento do sangue para separação de células mononucleares

O volume de 15 ml de sangue foi coletado por punção venosa no antebraço em tubos com ácido etilenodiaminotetra-acético (EDTA) e processado pela equipe do Biobanco do ICESP. Inicialmente foi realizada a separação do plasma do sangue total, através de uma centrifugação, a 3000g durante 10 minutos a temperatura ambiente. As células mononucleares do sangue (*buffy coat*) e as hemácias foram transferidas para

tubos Falcon de 50 ml e solubilizadas em tampão Tris-EDTA (TE 1X) pH 8,0 (10 mM Tris-HCl, 1 mM EDTA) até atingir o volume de 40ml. As amostras foram homogeneizadas em vórtex por 2 minutos e posterior centrifugação a 2300g por 7 minutos. O sobrenadante foi descartado com cuidados para não perder o *buffy coat* e foi adicionado novamente TE 1X até o volume de 20 ml. As amostras foram novamente homogeneizadas em vórtex por 2 minutos e posterior centrifugação a 2300g por 7 minutos. Na última etapa desse processo, o sobrenadante foi descartado, mantendo um volume de aproximadamente 200 µl por pellet, cuidadosamente homogeneizado por ressuspensão. As amostras foram transferidas para microtubos de 1,5ml, centrifugadas por 5 minutos em velocidade máxima (11.0000g) e acondicionadas em freezer a -80°C.

3.3 Extração do DNA de células mononucleares do sangue

Os microtubos contendo as células mononucleares (*buffy coat*) foram retirados do freezer e mantidos em temperatura ambiente. Como o volume inicial ideal era de 200 µl, o reagente TE 1x foi utilizado para completar o volume desejado. Após o preparo das amostras, foi realizada a extração do DNA com a aplicação do *Kit QIAmp DNA Blood Mini* (Quiagen, 51104). A metodologia deste procedimento está descrita abaixo.

- Lise enzimática

Na primeira etapa de extração do DNA foi efetuada a lise celular, pela adição de 20 µl de solução protease (Proteinase K) e 200 µl de tampão de lise (*Buffer AL*) em cada amostra. O material foi homogeneizado em vórtex durante 15 segundos e incubado a 56°C por 40 minutos com agitação a 650 rpm (*Eppendorf Thermomixer*). Após a incubação, foram adicionados 200 µl de etanol 100%, homogeneizados por pulso no vórtex e breve centrifugação.

O volume total do material lisado foi transferido para uma coluna (previamente colocada dentro do tubo de coleta de 2 ml) e centrifugada a 6000 g por 1 minuto, com posterior descarte do filtrado. Uma etapa opcional foi realizada quando o lisado não passava completamente através da coluna, através de uma centrifugação em velocidade máxima durante 1 minuto.

- Lavagem

A etapa de lavagem foi utilizada para a retirada de possíveis impurezas presentes na coluna de extração, tais como sais e proteínas. Na primeira etapa, 500 µl de *Buffer AW1* são adicionados às colunas com centrifugação a 6000 g por 1 minuto e posterior descarte do filtrado. Na sequência foram aplicados 500 µl do *Buffer AW2* com centrifugação em velocidade máxima por 3 minutos e descarte do filtrado. Para a maior eficiência deste procedimento, uma centrifugação adicional foi realizada em velocidade máxima por 1 minuto.

- Eluição

Para a eluição do DNA, as colunas foram transferidas para microtubos de 1,5 mL *DNase-free* e adicionados 100 µL de *UltraPure™ DNase/RNase-Free Distilled Water* (Invitrogen, 10977-023) no centro da coluna, incubado a temperatura ambiente por 5 minutos e centrifugado a 6.000 g por 1 minuto. As colunas foram removidas e adicionadas em outro tubo de 1,5 ml e o processo de eluição foi repetido por mais uma vez, para obter uma segunda eluição.

- Quantificação do DNA

Uma alíquota de cada amostra foi utilizada para quantificar e avaliar a integridade do DNA genômico. A quantificação das amostras foi realizada no aparelho *Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific Inc)*. A partir de uma alíquota de 1 µl é possível obter uma avaliação quantitativa dos seguintes fatores: concentração de DNA (ng/µl), pureza referente à presença de proteínas (260/280nm, medida ideal: 1,8 a 2,0) e contaminação com outros compostos (260/230 nm, medida ideal: 1,8 a 2,0). Após essa análise, os microtubos contendo as amostras de DNA de células mononucleares do sangue foram armazenados a -20°C.

3.4 Extração do DNA de tecido tumoral emblocado em parafina

Para a utilização do material parafinado, inicialmente os blocos e as respectivas lâminas, coradas com hematoxilina-eosina (HE), foram avaliados pelo médico-patologista do Serviço de Patologia do ICESP, do qual foi realizada a identificação das regiões com maior concentração de células tumorais. As regiões com mais de 40% de

células tumorais, foram delimitadas e posteriormente cortadas no micrótomo em cinco cortes de 20 μm por amostra. Esse material foi adicionado em microtubos de 1,5ml e acondicionado a 4°C. O procedimento para a extração do DNA das amostras está descrito a seguir.

- Desparafinização

Para ter acesso às amostras de tecido tumoral, a parafina deve ser retirada. Nesse processo foi aplicado 1 ml de xilol 100% em cada microtubo contendo os cortes de tecido tumoral. As amostras foram homogeneizadas em vórtex durante 10 segundos e incubadas em duas etapas: 65°C por 10 minutos e 67°C por 45 minutos. Após a incubação, as amostras foram centrifugadas em velocidade máxima por 2 minutos e o sobrenadante foi descartado.

Na sequência, o volume de 1ml de etanol 100% foi adicionado nas amostras, homogeneizado em vórtex por 10 segundos e centrifugado em velocidade máxima por 2 minutos, seguido pelo descarte do sobrenadante. Esse processo foi repetido por 2 vezes. Para a remoção total do etanol, os microtubos contendo os pellets das amostras foram invertidos e mantidos com a tampa aberta durante 10 minutos em temperatura ambiente. Como procedimento adicional, esse material foi incubado a 37°C, até a completa secagem do etanol. Apenas após todo esse processo, deve-se prosseguir para a extração de DNA do tecido tumoral. Para esse procedimento foi utilizado o kit *QIAamp® DNA FFPE Tissue* (Qiagen - 56404), conforme descrição abaixo.

- Lise enzimática

As soluções de lise, *Buffer ATL* (180 μl) e *Proteinase K* (20 μl), foram aplicadas no pellet de cada amostra e homogeneizadas em vórtex durante 10 segundos. Duas etapas de incubação foram realizadas: 56°C por 4 horas em agitação a 650 rpm (*Eppendorf Thermomixer*) e 90°C por 30 minutos. Uma breve centrifugação foi efetuada para retirada de resíduos da tampa. Para garantir remoção de RNA, em cada foram adicionados 8 μL de *RNase A* (100 mg/ μL), homogeneizados por vórtex por 15 segundos e incubado por 2 minutos. Na sequência foi realizada uma nova lise celular

através da aplicação do volume de 200 μ l de *Buffer AL* e 200 μ l de etanol 100%. O material contendo esses reagentes foi homogeneizado em vórtex durante 10 segundos, seguido por uma breve centrifugação.

- Lavagem

Para a lavagem do material, as amostras foram cuidadosamente transferidas para as colunas, centrifugadas a 6000 g por 1 minuto com posterior descarte do filtrado. Na sequência foi adicionado nas amostras o volume de 500 μ l de *Buffer AW1*, com centrifugação a 6000 g por 1 minuto e descarte do filtrado. O *Buffer AW2*, contendo o volume de 500 μ l, foi seguidamente adicionado nas amostras, centrifugado a 6000 g por 1 minuto, posterior descarte do filtrado. Uma centrifugação adicional, em velocidade máxima por 3 minutos, foi realizada para garantir a remoção de qualquer resíduo das membranas de cada coluna.

- Eluição

As colunas foram transferidas para novos microtubos de 1,5ml, no qual foram adicionados em cada amostra 30 μ l de *UltraPure™ DNase/RNase-Free Distilled Water* (Invitrogen, 10977-023). Durante a eluição esse material foi mantido por 5 minutos em temperatura ambiente e centrifugado em velocidade máxima por 1 minuto. As colunas foram removidas e adicionadas em outro tubo de 1,5 mL. O processo de eluição foi repetido por mais uma vez para obter uma segunda eluição das amostras. Uma alíquota de 1 μ l da amostra de DNA foi avaliada quantitativamente, através da utilização do equipamento *Nanodrop 1000 Spectrophotometer* (Thermo Fisher Scientific Inc). Os tubos contendo as amostras de DNA do tecido tumoral foram armazenados a -20°C até o momento do uso.

3.5 Sequenciamento de DNA de tecido tumoral e células mononucleares do sangue

Para o preparo das amostras de tumor e sangue para o sequenciamento é necessário realizar a construção da biblioteca de DNA. A primeira etapa deste processo refere-se ao controle de qualidade das amostras, que foram quantificadas no

equipamento *Qubit* com o reagente *Qubit dsDNA BR Assay* (*Thermo Fisher Scientific Inc*), seguida por análise da integridade do DNA no equipamento *TapeStation 2200* com o *Genomic DNA Screen Tape* (*Agilent*). A análise das amostras no *Qubit* disponibiliza a concentração de DNA (ng/ul) e a avaliação no *Tape Station* fornece informações sobre a qualidade dos fragmentos de DNA ou DIN (*DNA Integrity Number*), cujo valor varia de 1 a 10.

- Fragmentação do DNA

O protocolo *SureSelect XT HS Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library* (*Agilent - G9612A*) foi realizado para a construção da biblioteca de DNA. A primeira etapa deste procedimento consiste na fragmentação das amostras de DNA extraídas do tumor e sangue. De acordo com o protocolo são necessárias amostras contendo até 200 ng de DNA diluídos em 50 µl de solução tampão *TE Low IX* (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA). De maneira que, para cada amostra, foram ajustadas as concentrações de DNA conforme essas especificações. Na sequência o material foi transferido para microtubos e colocado, por 120 segundos, no equipamento *E220 Focused-Ultrasonicator* (*COVARIS*). Em seguida, as amostras foram centrifugadas por 10 segundos e homogeneizadas em vórtex em alta velocidade por 5 segundos. Esse procedimento foi repetido por 2 vezes. Ao final do processo, as amostras com 50 µl de DNA fragmentado foram colocadas em tubos *low binding* (*Axygen-MCT-150_L-C*) de 1,5ml e acondicionadas em freezer a -80°C.

- Reparo das extremidades dos fragmentos de DNA

Na próxima etapa, as extremidades do DNA fragmentado foram reparadas para a eficaz ligação dos próximos componentes (*barcode*, *primer* e *index*). Nesta etapa as amostras de DNA foram retiradas do freezer -80, mantidas em temperatura ambiente e transferidas para placas de 96 poços. Em seguida, foi aplicada 20 µL da solução enzimática *End Repair/dA Tailing* (16 ul de Tampão *End Repair-A Tailing* e 4 ul Enzima *End Repair-A Tailing*), que preenche as brechas da cadeia simples de DNA (5 a 10 bases). A solução foi homogeneizada por ressuspensão de 15 a 20 vezes e

incubada de acordo com os seguintes ciclos: 20°C por 15 minutos, 72°C por 15 minutos e 4°C por tempo indeterminado.

- Ligação do Barcode-Adaptador Molecular

Na sequência ocorreu a adição do barcode-adaptador molecular, ou seja, uma pequena sequência de DNA conhecida. Na microplaca contendo as amostras de DNA foi aplicado o volume de 25 µl da solução *Ligation Master Mix* (23 µl de Tampão de Ligação e 2 µl de *T4 DNA ligase*) e 5 µl da solução de *Oligo Mix*, que contém os adaptadores e o barcode molecular. O material foi homogeneizado por ressuspensão de 15 a 20 vezes e incubado a 20°C durante 30 minutos e 4°C por tempo indeterminado.

- Purificação das amostras usando *AMPure beads*

As amostras foram purificadas por *AMPure beads* (*Agencourt AMPure A63881- Beckman Coulter*). Neste processo, os fragmentos de DNA foram capturados por esferas magnéticas e atraídos às respectivas paredes do poço por uma estante magnética. Inicialmente nas microplacas, foi adicionado o volume 80 µl de *beads* para cada amostra, em seguida estas foram homogeneizadas por ressuspensão de 15 a 20 vezes e incubadas por 5 minutos em temperatura ambiente. Na sequência, a microplaca foi colocada em estante magnética até o clareamento da solução e posterior descarte do sobrenadante.

Na próxima etapa 200 µl de etanol 70% foram adicionados em cada amostra e removidos após 1 minuto. O processo de aplicação do etanol foi repetido por duas vezes. Para a secagem deste reagente, as amostras foram mantidas por 30 segundos em estante magnética e incubadas a 37°C por 2 minutos.

Em seguida ocorreu a eluição das amostras com a aplicação de água ultra pura (*UltraPure™ DNase/RNase-Free Distilled Water*), que induz a quebra da ligação magnética entre as *beads* e o DNA, promovendo a obtenção de um material purificado composto apenas por partículas de DNA. Para esse processo o volume de 35 µl de água ultra pura foi aplicado na microplaca, homogeneizado por ressuspensão de 15 a 20 vezes e incubado por 2 minutos em temperatura ambiente. A microplaca foi colocada em suporte magnético até o clareamento da solução, em seguida o volume de 34,5 µl do sobrenadante foi transferido para uma nova microplaca.

- 1º etapa de amplificação do DNA por PCR

Após a purificação, foi aplicado outro componente o *primer index*, que liga-se aos fragmentos de DNA contendo o *barcode* e os adaptadores. Ele é composto por uma pequena sequência de DNA, que tem como função atuar como um identificador de cada amostra e possibilitar a formação de *pools*. Cada amostra é composta por dois fragmentos de DNA (*forward/reverse*, 5'/3'). O *barcode* é composto por sequências exclusivas para cada fragmento de DNA, enquanto que o *index* contém sequências específicas para cada amostra, ou seja, ambos os fragmentos de DNA da amostra apresentam um mesmo *index*.

A adição do *index* nas amostras ocorre seguidamente a aplicação da solução *Mix* para amplificação do material. De maneira que em cada amostra na microplaca foram adicionados os volumes de 13,5 µl da solução *Mix* (10 µl tampão 5x *Herculase II*; 0,5 µl 100 mM de mix de dNTPs; 2 µl *Primer Forward*, 1 µl da *DNA Polimerase Herculase II Fusion*) e 2 µl de apropriado *Primer Index* com homogeneização por ressuspensão de 15 a 20 vezes.

Para a amplificação do material foram realizadas incubações específicas para cada tipo de amostra (sangue/tumor). De maneira que para as amostras de DNA do tumor foram aplicados 11 ciclos; 98 °C por 2 minutos, 98 °C por 30 segundos (11 ciclos), 60 °C por 30 segundos (11 ciclos), 72 °C por 1 minuto (11 ciclos), 72 °C por 5 minutos e 4 °C. Para as amostras de DNA do sangue foram utilizados 8 ciclos; 98 °C por 2 minutos, 98 °C por 30 segundos (8 ciclos), 60 °C por 30 segundos (8 ciclos), 72 °C por 1 minuto (8 ciclos), 72 °C por 5 minutos e 4 °C.

- Purificação das bibliotecas amplificadas usando *AMPure beads*

As bibliotecas amplificadas foram purificadas com a utilização de *beads* magnéticas, conforme o procedimento descrito anteriormente, porém, para esta etapa, foram utilizados os volumes de 50 µl de *beads* no início do procedimento e ao final do processo 15 µl de água ultrapura para a eluição do material. A quantificação das amostras foi realizada no equipamento *Tape Station 2200*. Nesta etapa foi utilizado o *chip D1000* para a quantificação do DNA (pg/ul) e obtenção dos tamanhos dos fragmentos em pares de bases (pb). A microplaca contendo a biblioteca amplificada foi estocada em freezer a -20°C para a continuidade do protocolo no dia seguinte.

- Hibridização das amostras para a captura da biblioteca

A hibridização das amostras corresponde a ligação do fragmento de DNA a sonda (*Capture Library*). A sonda contém as *probes* correspondentes ao design dos 64 genes selecionados para o painel, previamente elaborado pelo nosso grupo de pesquisa. De acordo com especificações do protocolo, cada amostra a ser hibridizada foi ajustada para apresentar, no máximo, uma concentração de 1000 ng de DNA em um volume final 12ul de água.

Nesta etapa a microplaca contendo a biblioteca amplificada foi transferida para microtubos de 0,2 ml, do qual foram aplicados os reagentes compostos por 5 ul da solução *Sure Select XT HS e XT Low Input Blocker Mix*, em seguida homogeneizados por ressuspensão de 15 a 20 vezes e incubado conforme os seguintes ciclos; 95 °C por 5 minutos, 65 °C por 10 minutos e 65 °C por 1 minuto. Neste momento, foi dada uma pausa no aparelho de PCR e com os tubos dentro deste equipamento foram adicionados 13 ul do Mix (6 µl de *Sure Select Fast Hybridization Buffer*, 0,5 µl de 25% *RNase Block solution*, 2 µl de *Capture Library* (<3Mb) e 4,5 µl de água Ultra Pura). O material foi homogeneizado por ressuspensão de 15 a 20 vezes, e submetido a 12 ciclos com variação de temperatura 65°C por 1 minuto, 37 °C por 3 segundos e 65 °C por tempo indeterminado.

- Captura das amostras pelas *beads* de estreptavidina

O protocolo de construção da biblioteca de DNA utiliza beads de estreptavidina para capturar as sondas hibridizadas com as regiões de interesse. Para um processo eficaz, deve-se realizar uma preparação prévia das beads de estreptavidina. Inicialmente em novo microtubo *low binding* foi adicionado o volume de 50 µl de *Dynabeads MyOne Streptavidin T1 (Thermo Fisher Scientific)* e 200 µl de tampão *SureSelect Binding*. A solução contendo as beads de estreptavidina foi homogeneizada por ressuspensão de 15 a 20 vezes e mantida em estante magnética por 5 a 10 minutos para o clareamento com posterior remoção do sobrenadante. O processo de preparação das beads de estreptavidina foi repetido por 4 vezes.

O volume de 30 µl da biblioteca amplificada foi transferido para microtubos *low binding* contendo a solução de 200 µl de *Streptavidin Magnetic Beads* previamente preparada conforme descrito acima. O material foi homogeneizado por ressuspensão por 5 a 8 vezes e incubado a 22°C sob agitação a 1400 rpm durante 30

minutos. Em seguida, colocado em estante magnética por 2 minutos, e em seguida foi realizado o descarte do sobrenadante cuidadosamente e com posterior retirada da placa da estante magnética.

- Lavagem

Nesta etapa ocorreu a lavagem do material com a adição de 200 µl de *SureSelect Wash Buffer 1* nos microtubos contendo as amostras, seguida por homogeneização em vórtex e utilização de suporte magnético por 1 minuto para o clareamento da solução, e posterior descarte do sobrenadante. O volume de 200 µl de *SureSelect Wash Buffer 2*, previamente aquecido a 70°C, foi adicionado às amostras, homogeneizado em vórtex por 8 segundos e incubado a 70° C durante 5 minutos sob agitação a 900 rpm. Em seguida este foi colocado em suporte magnético durante 1 minuto para o clareamento da solução e posterior descarte do sobrenadante. O procedimento de lavagem do material com o reagente *Sure Select Wash Buffer 2* foi repetido por 6 vezes. No final do processo foi realizada a eluição das amostras com o volume de 25 µl de água ultra pura, adicionado no material com homogeneização por ressuspensão por 8 vezes.

- 2º etapa de amplificação da biblioteca capturada

Uma segunda etapa de amplificação da biblioteca foi realizada com a aplicação de 25 µl da solução Mix (10 µL de 5x *Herculase II Reaction Buffer*, 1 µl *Herculase II Fusion DNA Polymerase*, 100 mM dNTPs, 1 µl *SureSelect Post-capture Primer Mix* e 12,5 µl de água Ultra pura) nos microtubos contendo as amostras. O material foi homogeneizado em vórtex e incubado conforme os seguintes ciclos; 98 °C por 2 minutos, 12 ciclos com variação de temperatura: 98 °C por 30 segundos, 60°C por 30 segundos, 72 °C por 1 minuto, ao final da incubação, 72 °C por 5 minutos e 4°C por tempo indeterminado.

- Purificação das bibliotecas capturadas e amplificadas usando AMPure beads

O processo de purificação foi realizado conforme o procedimento anteriormente descrito, porém inicialmente foi aplicado o volume de 50 µl de beads e ao final do processo 25 µl de água ultra pura. Na última etapa do procedimento foi utilizado o aparelho *Tape Station 2200* e chip *D1000 High Sensivity* para a

quantificação do DNA (pg/ul) e obtenção dos fragmentos por pares de bases (pb). O material contendo a biblioteca de DNA, proveniente do tumor e sangue, foi estocado em freezer a -20°C .

- Pool de amostras para o sequenciamento

As amostras foram ajustadas em concentrações de DNA iguais, através da diluição do material conforme a concentração da amostra mais diluída. O pool de amostras foi sequenciados no aparelho *NextSeq500TM* (*Illumina*) do Laboratório SELA (Sequenciamento em Larga Escala) da Faculdade de Medicina da Universidade de São Paulo (USP).

3.6 Seleção de Genes para o Painel de Mutações Somáticas

Para identificar o perfil mutacional das pacientes selecionadas neste estudo, um painel de genes foi elaborado para sequenciamento-alvo de nova geração, composto por genes potencialmente *drivers* em jovens com câncer de mama. É importante destacar que o painel foi confeccionado em conjunto à outro projeto de nosso grupo, que incluía pacientes com câncer de mama triplo-negativo (TN), portanto alguns genes selecionados foram provenientes desse subtipo de câncer de mama.

Os genes integrantes do painel foram detectados através de análise *in silico*, a partir dos dados mutacionais de um total de 106 pacientes jovens (≤ 35 anos) com tumores de mama subtipos luminal (LM) e triplo negativo (TN), provenientes de diferentes fontes. Em relação aos tumores com o subtipo luminal, foram analisadas as seguintes pacientes jovens; 8 provenientes de estudo publicado por nosso grupo (Encinas et. al., 2018), 29 identificadas por meio de busca em literatura (Kan et. al., 2018) e 29 reportadas no banco de dados do COSMIC (Nik-Zainal et. al., 2016, n=12; TCGA, n=17). Entre as amostras de pacientes jovens com tumores triplo-negativos foram avaliadas 18 pacientes identificadas por meio de literatura (Kan et. al., 2018) e 22 provenientes do banco de dados do COSMIC (Nik-Zainal et. al., 2016, n=18; TCGA, n=14).

Cada gene mutado, identificado pela análise *in silico*, foi avaliado para a seleção de genes potencialmente *drivers*, de acordo com os seguintes critérios descritos a seguir. O primeiro filtro baseou-se na seleção de genes com mutações provavelmente

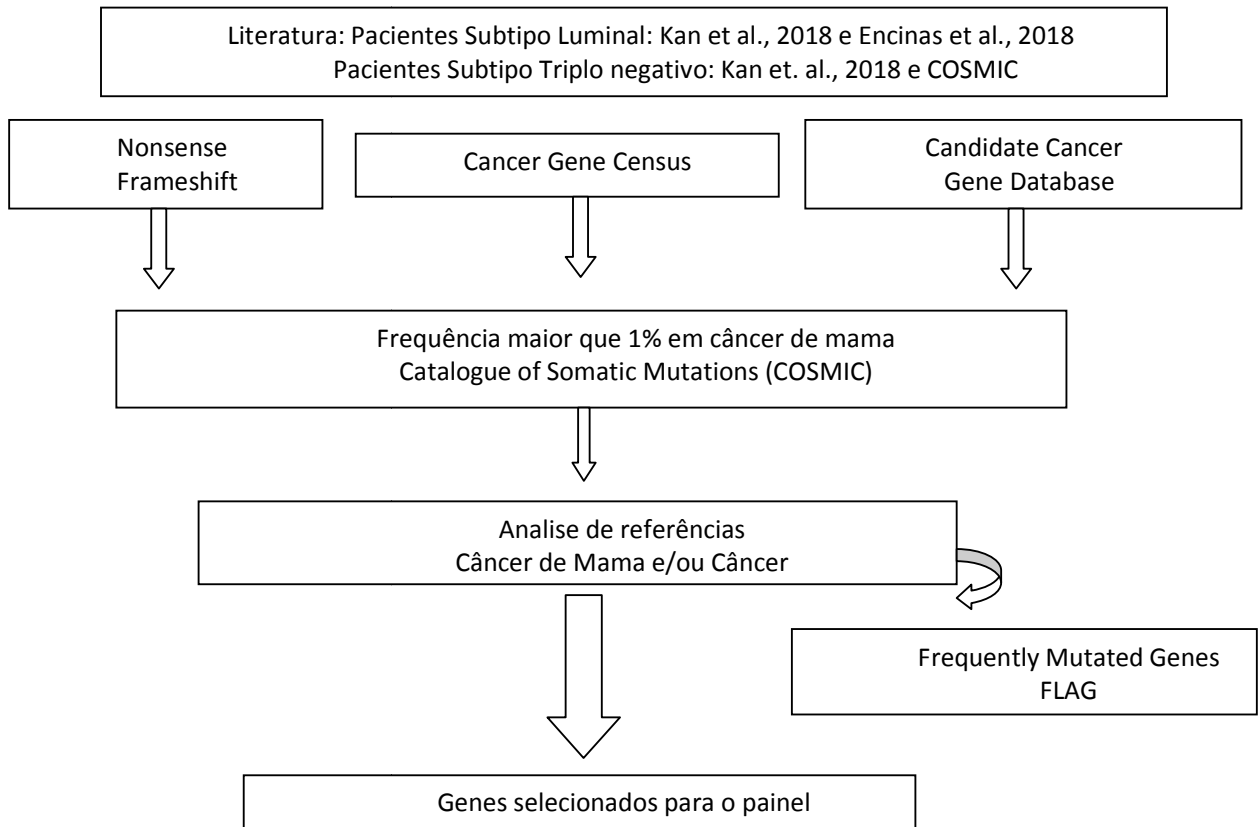
patogênicas, *frameshift* e *nonsense*. Adicionalmente todos os genes identificados neste filtro, foram investigados quanto à sua relevância na literatura, em relação à sua correlação direta com a progressão de câncer. Para esta análise utilizamos o banco de dados *Cancer Gene Census* (CGC; <https://cancer.sanger.ac.uk/census>), que possui uma equipe de curadoria responsável por constantemente atualizar os genes diretamente envolvidos em malignidades, de acordo com a literatura mais recente. Como análise adicional, também investigamos os genes em outro banco de dados, denominado *Candidate Cancer Gene Database* (CCGD; <http://ccgd-starrlab.oit.umn.edu/about.php>), referente a pesquisa de potenciais genes *drivers* correspondente a aplicação da técnica de inserção por transposon.

Após a seleção dos genes conforme os três filtros descritos acima, tipo de mutação, CGC e CCGD, outro filtro foi aplicado para a seleção dos genes com frequência maior ou igual a 1%, correspondente as variantes detectadas em câncer de mama. Os genes selecionados nestes filtros foram analisados individualmente quanto à sua relevância na literatura em câncer de mama e câncer em geral. Os genes que não preenchiam estes critérios foram descartados. Nesse aspecto, a única exceção à essa análise foram os genes catalogados no banco de dados CGC, uma vez que os mesmos são resultado de levantamento em literatura.

O último filtro aplicado baseou-se na exclusão de genes de acordo com o estudo Shyr et. al. 2014, no qual esse grupo de pesquisa desenvolveu uma lista de genes denominados *flags* (*frequently mutated genes*), definidos como genes de grandes tamanhos e alta frequência de mutações, porém sem potencial *driver*. Os genes *flags* identificados neste estudo foram descartados. De acordo com estas considerações, 11 genes foram excluídos; *TTN*, *CSMD1*, *DNAH17*, *DNAH2*, *DST*, *KMT2C*, *MACF1*, *RP1L1*, *SYNE1*, *UTRN*, *VPS13D*, *ZFH3*.

Após a aplicação dos critérios descritos, 64 genes foram selecionados para compor o painel gênico, 55 genes potenciais *drivers* previamente identificados e 9 genes adicionais (*ANLN*, *CAMK1G*, *MTHFD2*, *NCOA3*, *PRKD1*, *RAD51*, *RAD9A*, *SEMA6D* e *PARP4*), cujas funções biológicas são importantes no contexto do câncer e com variantes patogênicas detectadas em pelo menos 3 ferramentas de predição, conforme publicação de Encinas et al., 2018. A figura 1 apresenta as etapas de seleção dos genes para compor o painel gênico.

Figura 1: Processo de seleção dos genes para o painel gênico



O painel de genes elaborado neste estudo é composto por 24 genes supressores tumorais e 6 oncogenes. Outros genes selecionados não apresentam evidências sobre o seu modo de ação como *ANLN*, *ATAD2B*, *ATR*, *ATXN1*, *CACNA1E*, *CAMK1G*, *CSMD3*, *CSPP1*, *DALRD3*, *ERBB2*, *ERBB4*, *ESR1*, *FAT4*, *GRHL2*, *GRM3*, *HERC2*, *HUWE1*, *LYST*, *MTHFD2*, *NCOA3*, *PARP4*, *PCDH10*, *PIK3AP1*, *POLD1*, *PRKD1*, *PTPN13*, *RAD51*, *RAD9A*, *RSBN1*, *SEMA6B*, *SMURF2*, *SPEN*, *TNC* e *UBR5*. Novas pesquisas serão necessárias para definir a classificação destes genes.

Em torno de 40% dos genes foram considerados potencialmente *drivers*; *AKT1*, *APC*, *ARID1A*, *ATM*, *BAP1*, *BRCA1*, *BRCA2*, *CDH1*, *DICER1*, *EP300*, *ERBB2*, *FBXW7*, *GATA3*, *MAP2K4*, *MEN1*, *MET*, *NF1*, *NOTCH1*, *PIK3CA*, *PIK3R1*, *PTEN*, *PRKARIA*, *RBI*, *SETD2*, *SMARCA4* e *TP53*.

Os dados referentes à identificação dos genes como potencialmente *driver*, oncogene e supressor tumoral foram adquiridos a partir de consulta na base de dados *Intogen* (<https://www.intogen.org>, 25/09/2019). A tabela 2 apresenta os genes selecionados conforme a sua classificação como potencialmente *driver*, oncogene e supressor tumoral, assim como a identificação destes em diferentes estudos.

O programa *Topp Gene* (<https://toppgene.cchmc.org/prioritization.jsp>) foi utilizado para agrupar os 64 genes selecionados, de acordo com processos celulares relacionados a promoção do câncer, do qual esses genes participam em diferentes categorias: Função Molecular: hormonal, proteína quinase, transcrição e ubiquitinação; Processo Biológico: metabolismo, ciclo celular, motilidade e cromatina; Transdução de Sinal: reparo do DNA. A tabela 3 apresenta os genes selecionados neste estudo e os respectivos processos celulares.

De acordo com o programa *Topp Gene*, 64 genes foram agrupados em 9 categorias, conforme apresentado a seguir:

- Hormonal: *ARID1A, BRCA1, EP300, ESR1, NCOA3, PIK3R1, RB1 e SMARC4*.
- Proteína Quinase: *AKT1, APC, ATM, ATR, CAMK1G, ERBB2, ERBB4, MAP2K4, MET, MTOR, NF1, PIK3AP1, PIK3CA, PRKAR1A, PRKD1, PTEN, PTPN13 e RB1*
- Transcrição: *BRCA1, EP300, ERBB4, ESR1, FBXW7, GATA3, GRHL2, MEN1, MTOR, NCOA3, NOTCH1, RB1, SMARC4, SPEN e TP53*.
- Ubiquitinação: *AHNAK, APC, BRCA1, ERBB2, FAT4, FBXW7, HERC2, PCDH10, PIK3R1, PRKAR1A, RB1 e TP53*.
- Metabolismo: *AKT1, ATM, ATR, ATXN1, BRCA1, BRCA2, CSMD3, DALRD3, DICER1, EP300, FBXW7, GATA3, GRHL2, GRM2, HERC2, HUWE1, MAP2K4, MEN1, MET, MTHFD2, PARP4, PIK3CA, POLD1, RAD51, RAD9A, SETD2, SMARC4, TP53 e UBR5*.
- Ciclo Celular: *AKT1 ANLN, APC, ATM, ATR, BAP1, BRCA1, BRCA2, CSPPI, DICER1, EP300, ESR1, FBXW7, GATA3, MEN1, MET, MTOR, NOTCH1, PRKAR1A, PTEN, RAD51, RAD9A, RB1, SETD2 e TP53*.
- Motilidade: *AKT1, ANLN, APC, ATM, CACNA1E CDH1, ERBB2, ERBB4, FAT2, FBXW7, GATA3, LYST, MAP3K1, MED23, MET, MTOR, NF1, NOCT1, PIK3CA, PIK3R1, PRKD1, PTEN, SEMA6D, SETD2, SMARCA4, SMURF2, TNC e TP53*.
- Cromatina: *ARID1A, ATAD2B, ATM, BAP1, BRCA1, BRCA2, DICER1, EP300, ESR1, GATA3, HUWE1, MEN1, NCOA3, PRKD1, RB1, RSBN1, SETD2, SMARC4, TP53 e UBR5*.
- Reparo do DNA: *ATM, ATR, BAP1, BRCA1, BRCA2, HERC2, PARP4, PILD1, RAD51, RAD9A e TP53*.

3.7. Análise das variantes detectadas no sequenciamento do painel de genes

O software *SureCall* (*Agilent SureCall* 4.1.1) foi utilizado para realizar o alinhamento das sequências de DNA e efetuar a chamada de variantes. A partir deste programa um arquivo de extensão vcf foi gerado para cada paciente.

Para a detecção das alterações somáticas, os arquivos referentes às variantes presentes nas amostras de tumor, foram anotados individualmente no *Variant Effect Predictor Program* (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP). A partir deste software foram obtidos os seguintes dados: efeitos das variantes nos genes e proteínas, frequência alélica global conforme o banco de dados GnomAD (<https://gnomad.broadinstitute.org/>) e as informações referente a patogenicidade das amostras em ferramentas de predição

Os dados gerados no arquivo VEP foram filtrados utilizando como critérios a frequência alélica global <1% e o efeito das variantes classificados como moderado e alto (variantes *missense*, *inframe*, *frameshift* e *nonsense*). Em seguida, cada variante foi revisada manualmente para obter o número de leituras alteradas em relação a sequência referência. As variantes detectadas no tumor foram comparadas às alterações presentes no sangue e as variantes presentes em ambos foram consideradas germinativas. De maneira que as alterações presentes exclusivamente no tumor foram consideradas somáticas.

3.7.1 Ferramentas de predição de variantes *missenses* patogênicas

A aplicação do software VEP possibilitou a utilização de diferentes ferramentas de predição para a identificação de alterações *missenses* patogênicas: *Sift*, *Polyphen*, *FATHMM*, *Mutation Assessor*, *Mutation Taster* e *REVEL*. As alterações *missenses* classificadas como deletérias, em pelo menos 4 do total de 6 preditores, foram consideradas patogênicas. A descrição dos preditores utilizados para classificação da patogenicidade das variantes *missenses* identificadas neste estudo está apresentada a seguir.

SIFT (*Sorting Intolerant From Tolerant*) é um algoritmo que avalia se a substituição de um aminoácido é prejudicial à função da proteína. Este software realiza a análise da conservação de proteínas, através da utilização de sequências protéicas homólogas (Vaser et al., 2016). Ele apresenta uma pontuação que varia de 0 a 1, do qual

a substituição de aminoácidos é prevista como prejudicial, se a pontuação for menor ou igual a 0,05 (<https://sift.bii.a-star.edu.sg>).

PolyPhen-2 (Phenotyping Polymorphism v2) é um programa que prevê o possível impacto das substituições de aminoácidos na estabilidade de proteínas. Este preditor estima a probabilidade de uma mutação *missense* ser prejudicial com base em uma combinação de diversas propriedades: anotações de sequências de proteínas, perfis de conservação proteica e características estruturais (Adzhubei et al., 2013). A pontuação deste programa varia de 0 a 1, do qual as variantes *missenses* com pontuação maior que 0,45 são consideradas patogênicas (<http://genetics.bwh.harvard.edu/pph2/dokuwiki/start>).

FATHMM (Functional Analysis Through Hidden Markov Models) é um software que avalia variantes *missenses* através da análise de alterações em domínios proteicos conservados. Além disso, este preditor realiza a identificação de mutações *driver* em amostras de tumores (Shihab et al., 2013). A pontuação neste programa varia de 0 a 1, cujas as alterações *missenses* com pontuação maior que 0,85 são consideradas patogênicas (<http://fathmm.biocompute.org.uk>).

Mutation Assessor é uma ferramenta de predição baseada em análises evolutivas entre espécies. Este preditor utiliza informações referentes a conservação evolutiva de aminoácidos e homólogos de proteínas para prever o impacto funcional das substituições de aminoácidos. A pontuação deste programa varia de 0 a 1, do qual as alterações *missenses* com pontuação maior que 0,65 são consideradas patogênicas (<http://mutationassessor.org/r3>).

MutationTaster é um preditor que analisa variantes *missenses* pelo cálculo da probabilidade de uma alteração ser classificada como mutação patogênica ou polimorfismo. (<http://www.mutationtaster.org/info/documentation.html>). A pontuação deste programa varia de 0 a 1, do qual as alterações *missenses* com pontuação maior que 0,50 são consideradas patogênicas.

REVEL (rare exome variant ensemble learner) é um software que realiza a integração de dados de um conjunto composto por 13 preditores: *MutPred*, *FATHMM*, *VEST*, *PolyPhen*, *SIFT*, *PROVEAN*, *MutationAssessor*, *MutationTaster*, *LRT*, *GERP*, *SiPhy*, *phyloP* e *phastCons*. Este software realiza a distinção entre variantes *missenses* patogênicas e neutras (Ioannidis, et al., 2016). A pontuação deste programa varia de 0 a 1, do qual as alterações *missenses* com pontuação maior que 0,55 são consideradas patogênicas (<https://omictools.com/revel-tool>).

Tabela 2: Classificação dos genes selecionados e distribuição nos diferentes estudos

| Gene | Classificação do Gene | | | Câncer de Mama LM | | Câncer de Mama TN | |
|---------|-----------------------|--------|----------|-------------------|----------------------|-------------------|----------------------------|
| | Simbolo | Driver | Oncogene | Supressor | Encinas et al., 2018 | Kan et al., 2018 | Literatura e base de dados |
| AHNAK | | | | | | | |
| AKT1 | | | | | | | |
| ANLN | | | | | | | |
| APC | | | | | | | |
| ARID1A | | | | | | | |
| ATAD2B | | | | | | | |
| ATM | | | | | | | |
| ATR | | | | | | | |
| ATXN1 | | | | | | | |
| BAP1 | | | | | | | |
| BRCA1 | | | | | | | |
| BRCA2 | | | | | | | |
| CACNA1E | | | | | | | |
| CAMK1G | | | | | | | |
| CDH1 | | | | | | | |
| CSMD3 | | | | | | | |
| CSPP1 | | | | | | | |
| DALRD3 | | | | | | | |
| DICER1 | | | | | | | |
| EP300 | | | | | | | |
| ERBB2 | | | | | | | |
| ERBB4 | | | | | | | |
| ESR1 | | | | | | | |
| FAT2 | | | | | | | |
| FAT4 | | | | | | | |
| FBXW7 | | | | | | | |
| GATA3 | | | | | | | |
| GRHL2 | | | | | | | |
| GRM3 | | | | | | | |
| HERC2 | | | | | | | |
| HUWE1 | | | | | | | |
| LYST | | | | | | | |
| MAP2K4 | | | | | | | |
| MAP3K1 | | | | | | | |
| MED23 | | | | | | | |
| MEN1 | | | | | | | |
| MET | | | | | | | |
| MTHFD2 | | | | | | | |
| MTOR | | | | | | | |
| NCOA3 | | | | | | | |
| NF1 | | | | | | | |
| NOTCH1 | | | | | | | |
| PARP4 | | | | | | | |
| PCDH10 | | | | | | | |
| PIK3AP1 | | | | | | | |
| PIK3CA | | | | | | | |
| PIK3R1 | | | | | | | |
| POLD1 | | | | | | | |
| PRKAR1A | | | | | | | |
| PRKD1 | | | | | | | |
| PTEN | | | | | | | |
| PTPN13 | | | | | | | |
| RAD51 | | | | | | | |
| RAD9A | | | | | | | |
| RB1 | | | | | | | |
| RSBN1 | | | | | | | |
| SEMA6D | | | | | | | |
| SETD2 | | | | | | | |
| SMARCA4 | | | | | | | |
| SMURF2 | | | | | | | |
| SPEN | | | | | | | |
| TNC | | | | | | | |
| TP53 | | | | | | | |
| UBR5 | | | | | | | |

LM: luminal e TN: triplo negativo

4. Resultados

Dentre as 44 pacientes jovens com câncer de mama selecionadas para este estudo, 19 apresentaram amostras de tumores com concentração de DNA suficiente para construção da biblioteca e posterior sequenciamento. Destas, todas as pacientes tiveram o diagnóstico de carcinoma ductal invasivo, sendo que a maioria apresentou câncer de mama subtipo luminal B (16 Luminal B e 3 Luminal A), moderadamente diferenciado com grau histológico 2, estadio II ou III, e elevada expressão do receptor de estrogênio (70% - 100%). Em relação à faixa etária, as pacientes foram diagnosticadas com câncer de mama em idades entre 26 e 35 anos, correspondente a uma média de 31 anos. O perfil clínico das pacientes está descrito na tabela 1.

Tabela 1: Características clínico-patológicas das pacientes jovens com câncer de mama

| Paciente | Idade | Luminal | ER | PR | Her2 | Tipo Hist | Estadio | Grau Hist | Ki67 | Cel Tumor | TNM |
|----------|-------|---------|------|------|------|-----------|---------|-----------|------|-----------|------|
| 800 | 26 | B | 100% | 10% | neg | CDI | IIIb | 3 | 60% | 75% | T4N2 |
| 803 | 31 | B | 80% | 70% | neg | CDI | IIIb | 3 | 30% | 80% | T4N3 |
| 805 | 33 | B | 70% | 70% | neg | CDI | Iib | 2 | 30% | 65% | T2N1 |
| 809 | 34 | B | 80% | 20% | neg | CDI | IIIa | 3 | 50% | 60% | T3N1 |
| 811 | 34 | B | 100% | 10% | neg | CDI | Iib | 2 | 50% | 55% | T2N1 |
| 812 | 33 | A | 100% | 100% | neg | CDI | IIIa | 2 | 10% | 65% | T2N3 |
| 814 | 28 | B | 95% | 90% | neg | CDI | IIIb | 2 | 70% | 70% | T4N2 |
| 818 | 33 | B | 95% | 40% | neg | CDI | IIIa | 2 | 20% | 60% | T3N2 |
| 819 | 28 | A | 95% | neg | neg | CDI | Iia | 1 | 10% | 60% | T2N0 |
| 821 | 35 | B | 40% | 30% | neg | CDI | I | 2 | 35% | 70% | T1N0 |
| 822 | 27 | A | 100% | 20% | neg | CDI | IIIb | 2 | 10% | 50% | T4N0 |
| 823 | 29 | B | 100% | 80% | neg | CDI | Iia | 2 | 13% | 75% | T2N0 |
| 828 | 33 | B | 100% | 20% | neg | CDI | Iia | 2 | 30% | 40% | T2N0 |
| 829 | 34 | B | 90% | 30% | neg | CDI | Iib | 2 | 25% | 40% | T2N1 |
| 830 | 33 | B | 100% | 15% | neg | CDI | Iia | 2 | 20% | 40% | T2N0 |
| 837 | 29 | B | 90% | 90% | neg | CDI | IIIb | 3 | 50% | 95% | T4N0 |
| 838 | 32 | B | 100% | 100% | neg | CDI | I | 2 | 15% | 60% | T1N0 |
| 839 | 33 | B | 20% | 60% | neg | CDI | Iib | 3 | 30% | 75% | T2N1 |
| 842 | 34 | B | 90% | neg | neg | CDI | Iia | 2 | 20% | 50% | T1N1 |

ER: receptor de estrogênio, PR: receptor de progesterona, neg: negativo, Her2: fator de crescimento epidérmico receptor 2, Tipo Hist: tipo histológico, CDI: carcinoma ductal invasivo, Grau Hist: grau histológico e Cel Tumor: células tumorais

Após a análise dos dados do sequenciamento das amostras de tumor das pacientes, uma parcela destas foram excluídas (4 pacientes: 803, 814, 830 e 839). Além disso, 3 genes componentes do painel não foram considerados na análise (*HERC2*, *PARP4e* *RBI*). Ambas as amostras e genes foram retirados do estudo por apresentar uma cobertura de baixa qualidade, devido a leitura de nucleotídeos abaixo do limiar estabelecido (menos de 20 nucleotídeos ou *reads* avaliados em grande parte dos *exons* dos genes selecionados). A tabela 1 (suplementar) apresenta a mediana da porcentagem de *exons* com cobertura considerada inadequada. Como cada gene possui múltiplos

exons, o valor percentual dos *exons* de cada gene foi extraído a partir de sua média, de forma individual para cada amostra. Além disso, foram calculadas as porcentagens referentes à mediana de bases de qualidade inadequada dos *exons* por todos os genes de cada amostra individual, assim como a mediana de bases de *exons* de cada gene individualmente considerando todas as amostras.

As amostras em que a mediana de bases de *exons* com sequenciamento de qualidade inadequada para análise ultrapassou 30% foram excluídas (tabela 1). Portanto, a amostra 814, com mediana de 68.2% de bases de *exons* e sequenciamento de qualidade inadequada para análise, foi excluída (tabela 1).

A tabela 1 (suplementar) permite visualizar de forma global a qualidade das bases dos *exons* das amostras. Entretanto, a mesma deve ser aplicada como um mapa de direcionamento e não para a aplicação final. Com exceção de genes que claramente apresentam má qualidade global (*HERC2*, *PARP4* e *RBI*), outros genes contêm variação em seus valores e portanto suas respectivas variantes devem ser analisadas individualmente quanto à aspectos de cobertura e qualidade de *exon* (no caso, o *exon* no qual a variante está localizada).

O gene *PIK3CA* apresenta 3 variantes somáticas neste estudo, identificadas nas pacientes 809, 811 e 822. Estas alterações foram analisadas individualmente, especificamente em relação ao *exon* que contém a alteração. A descrição desta análise está apresentada abaixo.

Paciente 809:

A variante apresentou boa cobertura (*reads* alterados/*reads* totais: 70/266). A variante está localizada no *exon* 10. Ao contrário de alguns *exons* do gene, globalmente (entre todas as amostras), o *exon* 10 apresenta média de porcentagem de bases de má qualidade em frequências pouco elevadas (média: 22%; min: 2%, max: 100%). Olhando especificamente para a amostra 809, percebemos que a frequência de bases com baixa cobertura é nula (0%), portanto, mesmo que globalmente a qualidade de todos os *exons* de *PIK3CA* não seja alta, consideramos que a variante localizada nesse paciente pode ser considerada.

Paciente 811:

A variante apresentou boa cobertura, porém com baixa frequência alélica (*reads* alterados/*reads* totais: 12/77; frequência alélica=15%). No entanto, tal frequência respeita o mínimo estipulado (10%). A variante está localizada no *exon* 19. O *exon* apresenta qualidade ruim em 6 pacientes, tendo as restantes 0% de bases de baixa

qualidade, sendo que a amostra 811 está inclusa nessas (de boa qualidade). Portanto, consideramos que a variante localizada nesse paciente pode ser considerada.

Paciente 822:

A variante apresentou boa cobertura (*reads* alterados/*reads* totais: 45/170). A variante está localizada no exon21. O *exon* possui boa qualidade em todas as amostras, inclusive em amostras de má qualidade, como a 814 (5% de bases com má qualidade). Portanto, consideramos que a variante localizada nesse paciente pode ser considerada.

"Concluindo, as três variantes encontradas em *PIK3CA* no presente estudo podem ser consideradas como fidedignas."

4.1 Variantes gênicas somáticas

Dentre as 19 pacientes deste estudo, 15 foram consideradas para a análise das amostras de tumores. Destas, foram detectadas variantes somáticas em 11 pacientes (63%); 800, 805, 809, 811, 819, 821, 822, 828, 829, 838 e 842 em 21 genes diferentes; *GATA3*, *MEN1*, *BRCA2*, *HUWE1*, *RAD51*, *PIK3CA*, *CAMK1G*, *SPEN*, *CACNA1E*, *AHNAK*, *NF1*, *BAP1*, *FAT2*, *TNC*, *NOTCH1*, *TP53*, *GRHL2*, *AKT1*, *PTEN*, *SEMA6D* e *PRKARIA*.

As variantes somáticas detectadas correspondem a 31 alterações, sendo estas 23 *missenses*, 2 *nonsenses*, 1 deleção e 5 inserções. Elas apresentaram uma frequência alélica média no tumor de 0,25, assim como uma média de 34 bases alteradas, relacionadas ao número mínimo de 10 bases e máximo de 161 bases. Em relação às bases referência foram registradas uma média de 81 bases, do qual obteve-se uma leitura mínima de 11 bases e máxima de 310 bases.

A seguir estão apresentadas as características clínico-patológicas de 11 pacientes jovens com câncer de mama identificadas com alterações somáticas. Na tabela 2 (suplementar) estão descritas as informações sobre essas variantes somáticas.

A primeira amostra de tumor avaliada neste estudo refere-se a paciente 800. Ela é a mulher mais jovem incluída neste estudo (26 anos), com diagnóstico de câncer de mama subtipo luminal B, grau histológico 3 e estadio IIIb. A variante somática detectada foi considerada patogênica, pois corresponde a uma alteração *frameshift* no gene *GATA3*, referente a inserção de um nucleotídeo com frequência

alélica no tumor de 0,419 e uma leitura de 86 bases de referência em relação a 36 bases alteradas. A tabela 4 apresenta a variante somática detectada na paciente 800.

Tabela 4: Variante somática detectada na amostra de tumor da paciente 800

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | MTaster |
|-------|-----------|------|-------|----|--------|--------|----------|-----------|-----------|------|----------|--------|-------|-----------|---------|
| GATA3 | FRA | 6 | 10 | ND | C | CG | 0,419 | 36 | 86 | ND | ND | ND | ND | ND | ND |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, FRA: frameshift e ND: não determinado

O total de quatro variantes somáticas foram detectadas na amostra de tumor da paciente 805. Ela apresenta câncer de mama subtipo luminal B diagnosticado aos 33 anos, estadió IIB e grau histológico 2. Em relação às variantes somáticas, foram detectadas uma alteração *frameshift* no gene *MEN1* considerada patogênica e três alterações *missenses* identificadas como benignas. A variante *frameshift* no gene *MEN1* refere-se a deleção de um nucleotídeo com frequência alélica de 0,133 e com uma leitura de 12 bases alteradas. A tabela 5 apresenta as variantes somáticas detectadas na paciente 805.

Tabela 5: Variantes somáticas detectadas na amostra de tumor da paciente 805

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | MTaster |
|-------|-----------|------|-------|--------|--------|--------|----------|-----------|-----------|---------|----------|---------|---------|-----------|---------|
| MEN1 | FRA | 2 | 11 | ND | TC | T | 0.133 | 12 | 90 | ND | ND | ND | ND | ND | ND |
| BRCA2 | MIS | 27 | 13 | C3404Y | G | A | 0.37 | 10 | 27 | B(0.25) | D(0.72) | B(0.00) | B(0.32) | B(0.45) | B(0.08) |
| HUWE1 | MIS | 65 | X | S3052L | G | A | 0.412 | 21 | 51 | B(0.11) | B(0.00) | B(0.36) | B(0.15) | B(0.59) | B(0.40) |
| RAD51 | MIS | 4 | 15 | T76M | C | T | 0.127 | 14 | 110 | D(0.06) | B(0.00) | B(0.39) | B(0.07) | ND | B(0.08) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, FRA: frameshift, MIS, missense, D: deletério, B: benigno e ND: não determinado

A paciente 809 apresentou diagnóstico de câncer de mama subtipo luminal B aos 34 anos, grau histológico 3 e estadió IIIa. Nesta paciente foi identificada uma variantes somática no gene *PIK3CA*, correspondente a uma alteração *missense* (E542K) considerada patogênica. Esta variante foi classificada como deletéria de acordo com quatro programas de predição, sendo referente a frequência alélica de 0,263 e com uma leitura com 266 bases de referência em relação a 70 bases alteradas. Na tabela 6 está descrita a variante somática detectada na paciente 809.

Tabela 6: Variante somática detectada na amostra de tumor da paciente 809

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | M.Taster |
|--------|-----------|------|-------|-------|--------|--------|----------|-----------|-----------|---------|----------|---------|---------|-----------|----------|
| PIK3CA | MIS | 10 | 3 | E542K | G | A | 0.263 | 70 | 266 | D(0.04) | D(0.90) | B(0.61) | D(0.74) | B(0.44) | D(0.81) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, MIS, missense, D: deletério e B: benigno

Em relação a paciente 811, duas variantes somáticas foram detectadas na amostra de tumor. Ela apresenta câncer de mama subtipo luminal B diagnosticado aos 34 anos, grau histológico 2 e estadió IIB. As variantes somáticas detectadas são alterações *missenses* identificadas nos genes *PIK3CA* (F909L) e *CAMK1G* (R466P), do qual ambas foram consideradas benignas. Na tabela 7 estão descritas as variantes somáticas detectadas na paciente 811.

Tabela 7: Variantes somáticas detectadas na amostra de tumor da paciente 811

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | M.Taster |
|--------|-----------|------|-------|-------|--------|--------|----------|-----------|-----------|---------|----------|---------|---------|-----------|----------|
| CAMK1G | MIS | 12 | 1 | R466P | G | C | 0.109 | 12 | 110 | B(0.12) | B(0.00) | B(0.69) | B(0.53) | B(0.35) | D(0.38) |
| PIK3CA | MIS | 19 | 3 | F909L | C | G | 0.156 | 12 | 77 | B(0.11) | B(0.38) | B(0.75) | D(0.86) | B(0.41) | D(0.58) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, MIS, missense, D: deletério e B: benigno

A paciente 819 apresentou o maior número de variantes somáticas, um total de onze alterações. Ela foi diagnosticada com câncer de mama subtipo luminal A aos 28 anos, grau histológico 1 e estadió IIA. Nesta paciente foram detectadas uma alteração nonsense no gene *NFI* (E1181*) e dez alterações missenses nos genes, duas variantes em *AHNAK* e uma variante em diferentes genes; *SPEN*, *CACNA1E*, *BAP1*, *FAT2*, *ANLN*, *TNC*, *NOTCH1* e *NFI*.

Tabela 10: Variantes somáticas detectadas na amostra de tumor da paciente 819

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | M.Taster |
|---------|-----------|------|-------|--------|--------|--------|----------|-----------|-----------|---------|----------|---------|---------|-----------|----------|
| SPEN | MIS | 11 | 1 | T1756I | C | T | 0.269 | 14 | 52 | B(0.12) | B(0.10) | D(0.08) | B(0.23) | B(0.60) | B(0.08) |
| CACNA1E | MIS | 39 | 1 | L1773M | C | A | 0.152 | 16 | 105 | D(0.01) | D(0.92) | D(0.96) | D(0.85) | B(0.41) | D(0.58) |
| AHNAK | MIS | 5 | 11 | D1342A | T | G | 0.309 | 25 | 81 | D(0.00) | D(0.98) | B(0.03) | B(0.43) | D(0.79) | D(0.81) |
| AHNAK | MIS | 5 | 11 | D1342Y | C | A | 0.305 | 25 | 82 | D(0.00) | D(0.99) | B(0.03) | B(0.46) | D(0.79) | D(0.81) |
| NFI | MIS | 27 | 17 | L1183M | C | A | 0.333 | 11 | 33 | D(0.00) | D(0.97) | B(0.73) | D(0.57) | D(0.82) | D(0.58) |
| NFI | NON | 27 | 17 | E1181* | G | T | 0.314 | 11 | 35 | ND | ND | ND | ND | ND | D(0.81) |
| BAP1 | MIS | 14 | 3 | P598S | G | A | 0.321 | 18 | 56 | B(0.50) | B(0.02) | B(0.49) | B(0.23) | B(0.06) | B(0.31) |
| FAT2 | MIS | 23 | 5 | G4294S | C | T | 0.368 | 14 | 38 | D(0.01) | D(0.75) | D(0.71) | D(0.55) | B(0.39) | D(0.51) |
| ANLN | MIS | 2 | 7 | A40V | C | T | 0.245 | 12 | 49 | B(0.37) | B(0.00) | B(0.43) | B(0.02) | B(0.24) | B(0.08) |
| TNC | MIS | 18 | 9 | R1741Q | C | T | 0.351 | 20 | 57 | B(0.36) | D(0.75) | B(0.57) | D(0.68) | D(0.71) | B(0.34) |
| NOTCH1 | MIS | 25 | 9 | R1523H | C | T | 0.283 | 15 | 53 | B(0.11) | B(0.00) | B(0.64) | B(0.11) | B(0.26) | B(0.08) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, MIS, missense, D: deletério e B: benigno

O total de duas variantes somáticas foram identificadas na paciente 821, diagnosticada com câncer de mama subtipo luminal B aos 35 anos, grau histológico 2 e estadio I. As variantes somáticas detectadas foram uma alteração *missense* no gene *TNC* (R1675M) classificada como benigna e uma alteração *frameshift* no gene *TP53* considerada patogênica. A variante no gene *TP53* corresponde a inserção de um nucleotídeo, referente a frequência alélica no tumor de 0,346 e com uma leitura de 78 bases referência em relação a 27 bases alteradas. Na tabela 11 estão descritas as variantes somáticas detectadas na paciente 821.

Tabela 11: Variantes somáticas detectadas na amostra de tumor da paciente 821

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | M.Taster |
|------|-----------|------|-------|--------|--------|--------|----------|-----------|-----------|----------|----------|----------|----------|-----------|----------|
| TP53 | FRA | 4 | 17 | ND | A | AG | 0.346 | 27 | 78 | ND | ND | ND | ND | ND | ND |
| TNC | MIS | 17 | 9 | R1675M | C | A | 0.256 | 10 | 39 | B (0.22) | B (0.10) | B (0.55) | B (0.40) | B (0.21) | D (0.81) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, FRA; frameshift, MIS, missense, D: deletério, B: benigno e ND: não determinado

A paciente 822 apresentou diagnóstico de câncer de mama subtipo luminal A aos 27 anos, grau histológico 2 e estadio IIb. Um total de 3 variantes somáticas foram detectadas nesta paciente (1 *missense* e 2 *frameshift*). A alteração *missense* foi identificada no gene *PIK3CA* (H1047R) e classificada como benigna. Diferentemente desta, duas alterações *frameshift* foram identificadas nos genes *GATA3* e *GRHL2* e consideradas patogênicas, sendo ambas referentes a inserção de nucleotídeos. A alteração *frameshift* no gene *GATA3* deve-se a inserção de dois nucleotídeos, referente a frequência alélica de 0,252 e uma leitura de 310 bases de referência em relação a 78 bases alteradas. O gene *GRHL2* apresentou uma variante *frameshift* relacionada a inserção de um nucleotídeo. Esta variante apresenta frequência alélica de 0,271 e uma leitura de 107 bases referência em relação a 29 bases alteradas. Na tabela 12 estão descritas as variantes somáticas detectadas na paciente 822.

Tabela 12: Variantes somáticas detectadas na amostra de tumor da paciente 822

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | M.Taster |
|--------|-----------|------|-------|--------|--------|--------|----------|-----------|-----------|----------|----------|----------|----------|-----------|----------|
| GATA3 | FRA | 6 | 10 | ND | A | ACC | 0.252 | 78 | 310 | ND | ND | ND | ND | ND | ND |
| GRHL2 | FRA | 3 | 8 | ND | C | CT | 0.271 | 29 | 107 | ND | ND | ND | ND | ND | ND |
| PIK3CA | MIS | 21 | 3 | H1047R | A | G | 0.265 | 45 | 170 | B (0.11) | B (0.08) | B (0.80) | D (0.75) | B (0.15) | D (0.81) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, FRA; frameshift, MIS, missense, D: deletério, B: benigno e ND: não determinado

A paciente 828 apresentou o diagnóstico de câncer de mama subtipo luminal B aos 33 anos, grau histológico 2 e estadio IIa. Duas variantes somáticas foram detectadas nesta paciente, correspondentes a alterações *missenses* nos genes *RAD51* (V112A) e *AHNAK* (V3972I), ambas classificadas como benignas. Na tabela 14 estão descritas as variantes somáticas detectadas na paciente 828.

Tabela 14: Variantes somáticas detectadas na amostra de tumor da paciente 828

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M.Asses. | M.Taster |
|-------|-----------|------|-------|--------|--------|--------|----------|-----------|-----------|----------|----------|----------|----------|----------|----------|
| AHNAK | MIS | 5 | 11 | V3972I | C | T | 0.137 | 22 | 161 | D (0.02) | B (0.02) | B (0.00) | B (0.09) | B (0.34) | B (0.08) |
| RAD51 | MIS | 17 | 15 | V112A | T | C | 0.228 | 13 | 57 | B (0.07) | B (0,00) | B (0.07) | B (0.09) | ND | B (0.08) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, MIS, missense, B: benigno e ND: não determinado

Assim como no caso anterior, duas variantes somáticas *missenses* foram identificadas na amostra de tumor da paciente 829. Esta paciente apresenta câncer de mama subtipo luminal B diagnosticado aos 34 anos, grau histológico 2 e estadio IIb. Uma alteração missense foi detectada no gene *AHNAK* e classificada como benigna. Outra alteração *missense*, referente ao *AKT1* (P423L), foi considerada deletéria, conforme cinco ferramentas de predição. Esta alteração apresenta frequência alélica de 0,373 e uma leitura de 22 bases alteradas. Na tabela 15 está descrita a variante somática detectada na paciente 829.

Tabela 15: Variantes somáticas detectadas na amostra de tumor da paciente 829

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M.Asses. | M.Taster |
|-------|-----------|------|-------|--------|--------|--------|----------|-----------|-----------|----------|----------|----------|----------|----------|----------|
| AKT1 | MIS | 13 | 14 | P423L | G | A | 0.373 | 22 | 59 | D (0.01) | D (0.98) | B (0.62) | D (0.84) | D (0.99) | D (0.81) |
| AHNAK | MIS | 5 | 11 | N3135K | A | C | 0.333 | 15 | 45 | B (0.20) | B (0.08) | B (0.00) | B (0.09) | B (0.03) | B (0.22) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutation Taster, MIS, missense, D: deletério e B: benigno.

A amostra de tumor da paciente 838 apresentou uma variante somática. A paciente foi diagnosticada com câncer de mama subtipo luminal B aos 32 anos, grau histológico 2 e estadio I. A variante somática corresponde a uma alteração *frameshift* no gene *GATA3*, considerada patogênica. Esta alteração ocorreu devido a inserção de um nucleotídeo, referente a frequência alélica de 0,349 e com uma leitura de 67 bases alteradas. Na tabela 16 está descrita a variante somática detectada na paciente 838.

Tabela 16: Variantes somáticas detectadas na amostra de tumor da paciente 838

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | MTaster |
|-------|-----------|------|-------|----|--------|--------|----------|-----------|-----------|------|----------|--------|-------|-----------|---------|
| GATA3 | FRA | 6 | 10 | ND | T | TG | 0,349 | 67 | 192 | ND | ND | ND | ND | ND | ND |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutantion Taster, FRA; frameshift e ND: não determinado

A paciente 842 apresentou diagnóstico de câncer de mama subtipo luminal B aos 34 anos, grau histológico 2 e estadio IIa. Nesta paciente foram detectadas três variantes somáticas (1 *nonsense* e 2 *missenses*), todas consideradas patogênicas. Uma alteração *nonsense* foi detectada no gene *PRKARIA* (R96*), referente a frequência alélica no tumor de 0,11 e com uma leitura de 91 bases referência em relação a 10 bases alteradas. A variante *missense* foi identificada no gene *SEMA6D* (I406T) e considerada deletéria de acordo com quatro programas de predição. Da mesma maneira, resultado semelhante foi identificado para a variante patogênica no gene *PTEN* (L220V). Esta alteração apresenta a frequência alélica de 0,346 e uma leitura de 78 bases referência em relação a 27 bases alteradas. Na tabela 17 estão descritas as variantes somáticas detectadas na paciente 842.

Tabela 17: Variantes somáticas detectadas na amostra de tumor da paciente 842

| Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref | SIFT | PolyPhen | FATHMM | REVEL | M. Asses. | MTaster |
|---------|-----------|------|-------|-------|--------|--------|----------|-----------|-----------|---------|----------|---------|---------|-----------|---------|
| SEMA6D | MIS | 15 | 15 | I406T | T | C | 0,156 | 10 | 64 | D(0.03) | D(0.95) | B(0.11) | D(0.61) | B(0.46) | D(0.54) |
| PRKAR1A | NON | 4 | 17 | R96* | C | T | 0,11 | 10 | 91 | ND | ND | ND | ND | ND | D(0.81) |
| PTEN | MIS | 7 | 10 | L220V | C | G | 0,346 | 27 | 78 | B(0.50) | B(0.33) | D(0.84) | D(0.75) | D(0.78) | D(0.81) |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, M. Asses: Mutation Assessor, M. Taster: Mutantion Taster, MIS, missense, NON: nonsense, D: deletério, B: benigno e ND: não determinado

4.1.1 Variantes gênicas somáticas patogênicas

Em 9 de 11 pacientes descritas acima, foram detectadas 17 variantes somáticas patogênicas em 13 genes. A figura 2 apresenta o processo de identificação das variantes somáticas patogênicas e na tabela 18 estão descritas as informações sobre essas variantes e os respectivos genes.

Figura 2: Processo de identificação de variantes somáticas patogênicas

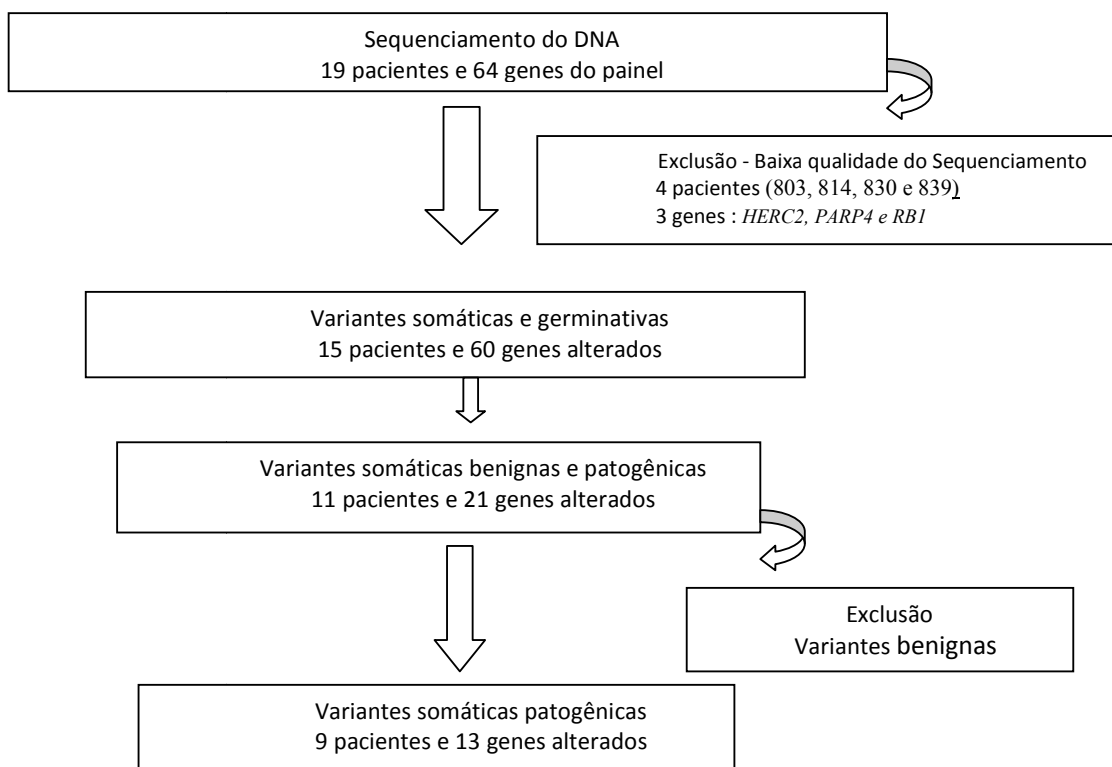


Tabela 18: Variantes somáticas consideradas patogênicas

| Pacientes | Gene | Alteração | Exon | Chrom | AA | Al Ref | Al Alt | Freq. Al | Bases Alt | Bases Ref |
|-----------|---------|-----------|------|-------|----------|--------|--------|----------|-----------|-----------|
| 800 | GATA3 | FRA | 6 | 10 | Inserção | C | CG | 0,419 | 36 | 86 |
| 838 | GATA3 | FRA | 6 | 10 | Inserção | T | TG | 0,349 | 67 | 192 |
| 822 | GATA3 | FRA | 6 | 10 | Inserção | A | ACC | 0,252 | 78 | 310 |
| 822 | GRHL2 | FRA | 3 | 8 | Inserção | C | CT | 0,271 | 29 | 107 |
| 821 | TP53 | FRA | 4 | 17 | Inserção | A | AG | 0,346 | 27 | 78 |
| 805 | MEN1 | FRA | 2 | 11 | Deleção | TC | T | 0,133 | 12 | 90 |
| 819 | CACNA1E | MIS | 39 | 1 | L1773M | C | A | 0,152 | 16 | 105 |
| 819 | FAT2 | MIS | 23 | 5 | G4294S | C | T | 0,368 | 14 | 38 |
| 819 | NF1 | MIS | 27 | 17 | L1183M | C | A | 0,333 | 11 | 33 |
| 819 | NF1 | NON | 27 | 17 | E1181* | G | T | 0,314 | 11 | 35 |
| 819 | AHNAK | MIS | 5 | 11 | D1342A | T | G | 0,309 | 25 | 81 |
| 819 | AHNAK | MIS | 5 | 11 | D1342Y | C | A | 0,305 | 25 | 82 |
| 809 | PIK3CA | MIS | 10 | 3 | E542K | G | A | 0,263 | 70 | 266 |
| 829 | AKT1 | MIS | 13 | 14 | P423L | G | A | 0,373 | 22 | 59 |
| 842 | PRKARIA | NON | 4 | 17 | R96* | C | T | 0,11 | 10 | 91 |
| 842 | SEMA6D | MIS | 15 | 15 | I406T | T | C | 0,156 | 10 | 64 |
| 842 | PTEN | MIS | 7 | 10 | L220V | C | G | 0,346 | 27 | 78 |

Chrom: cromossomo, AA: aminoácido, Al Ref: alelo referência, Al Alt: alelo alterado, Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, FRA: frameshift, MIS: missense, NON: nonsense

O gene *GATA3* destacou-se por apresentar variantes *frameshift* em três pacientes, 800, 822 e 838, todas referentes a inserção de nucleotídeos. A paciente 822, além de apresentar uma variante deletéria no gene *GATA3*, também foi detectada com uma variante *frameshift* no gene *GRHL2*. Duas pacientes, 805 e 821, apresentaram uma variante *frameshift* cada. A paciente 805 apresentou uma alteração *frameshift* no gene

MEN1, promovida pela deleção de um nucleotídeo, e a paciente 821 foi identificada com uma variante no gene *TP53*, relacionada a inserção de um nucleotídeo.

A paciente 819 apresentou uma variante *nonsense* no gene *NF1* e 5 alterações *missenses* identificadas em diferentes genes; *CACNA1E*, *FAT2*, *NF1* e *AHNAK*. Um total de três alterações foram identificadas na paciente 842, sendo uma *nonsense* no gene *PRKAR1A* e duas *missenses* referentes aos genes *SEMA6D* e *PTEN*. Outras alterações *missenses* patogênicas foram identificadas em duas pacientes, uma variante no gene *PIK3CA* detectada na paciente 809 e uma variante no gene *AKT1* referente a paciente 829.

4.1.2 Variantes gênicas germinativas patogênicas

Dentre as 19 pacientes incluídas neste estudo, 4 apresentaram variantes germinativas patogênicas (811, 821, 830 e 842) em 3 genes relacionados a síndromes hereditárias; duas em *BRCA1*, uma em *BRCA2* e uma em *ATM*. O gene *BRCA1* apresentou duas variantes germinativas patogênicas, uma alteração em *sítio de splicing* na paciente 811 e uma alteração *frameshift* identificada na paciente 821. Em relação aos genes *BRCA2* e *ATM*, uma variante germinativa foi identificada em cada gene, sendo uma alteração *missense* no gene *BRCA2* detectada na paciente 842, e uma alteração em *sítio de splicing* no gene *ATM* referente a paciente 830.

A partir destas considerações, duas pacientes apresentaram alterações somáticas e germinativas patogênicas (842 e 821). A amostra de tumor da paciente 821 apresentou duas alterações patogênicas, uma alteração somática *frameshift* em *TP53* e uma alteração germinativa *frameshift* em *BRCA1*. O total de quatro alterações patogênicas foram detectadas na paciente 842, três variantes somáticas (1 *nonsense* em *PRKAR1A* e 2 *missenses* em *SEMA6D* e *PTEN*) e uma variante germinativa (1 *missense* em *BRCA2*). Na tabela 19 estão descritas as variantes germinativas e somáticas patogênicas das pacientes 821 e 842.

Tabela 19: Variantes patogênicas identificadas nas pacientes 821 e 842

| Paciente | Gene | Variante | Alteração | HGVS c.DNA | Freq. Al | Bases Alt | Bases Ref |
|----------|---------|-------------|-----------|------------|----------|-----------|-----------|
| 821 | BRCA1 | germinativa | FRA | c.5266dupC | 0,513 | 196 | 381 |
| 821 | TP53 | somática | FRA | ND | 0,346 | 27 | 78 |
| 842 | PRKAR1A | somática | NON | ND | 0,110 | 10 | 91 |
| 842 | SEMA6D | somática | MIS | ND | 0,156 | 10 | 64 |
| 842 | PTEN | somática | MIS | ND | 0,346 | 27 | 78 |
| 842 | BRCA2 | germinativa | MIS | c.250C>T | 0,475 | 131 | 146 |

Freq. Al: frequência alélica, Bases Alt, bases alteradas, Bases Ref: bases referência, FRA: frameshift, MIS: missense, NON: nonsense e ND: não determinado

5. Discussão

Neste estudo, 13 genes apresentaram variantes somáticas patogênicas, destes 61% são classificados como genes *drivers*, a maioria relacionados aos processos celulares de motilidade, metabolismo e ciclo celular, e provenientes de estudo do nosso grupo (Encinas et al., 2018). A tabela 3 (suplementar) apresenta as informações sobre os genes com variantes somáticas patogênicas correspondentes à classificação, distribuição nos diferentes estudos e função em processos celulares.

Em 4 de 13 genes com alterações somáticas patogênicas foram detectadas variantes com alterações *frameshift*; *GATA3*, *GRHL2*, *MEN1* e *TP53*.

O gene *GATA3* pertence à família de fatores de transcrição, sendo responsável pela diferenciação de células epiteliais mamárias. Ele é um gene supressor tumoral, frequentemente mutado em câncer de mama, do qual está relacionado ao processo de regulação da transição epitélio-mesenquimal. Neste estudo três pacientes jovens com câncer de mama subtipo luminal apresentaram alterações *frameshift* no gene *GATA3*, sendo todas referentes a inserção de um nucleotídeo. Da mesma maneira na publicação de Kan et al., 2018, quatro pacientes coreanas jovens com o mesmo perfil, apresentaram variantes *frameshift* no gene *GATA3*, que promoveram a inserção de nucleotídeos.

O gene *GRHL2*, assim como o gene *GATA3*, também é um fator de transcrição que participa da regulação da transição epitélio-mesenquimal e diferenciação de células epiteliais. A partir deste estudo foi identificada uma alteração *frameshift* no gene *GRHL2*, promovida pela à inserção de um nucleotídeo referente a uma paciente jovem com câncer de mama subtipo luminal (822). Outra alteração patogênica no gene *GRHL2*, sendo esta *nonsense*, foi detectada em uma pacientes com esse mesmo perfil a partir de um estudo anterior do nosso grupo de pesquisa (Encinas et al., 2018).

O gene *MEN1*, ou menin 1, está associado a neoplasia endócrina múltipla tipo 1, causada por mutações germinativas neste gene. Ele é um supressor tumoral, que em tumores esporádicos promove a proliferação celular e a resistência à terapia endócrina. Neste estudo uma alteração *frameshift* no gene *MEN1* foi detectada em uma paciente jovem com câncer de mama subtipo luminal, referente à deleção de um nucleotídeo (805). Outra alteração *frameshift* no gene *MEN1*, deleção de nucleotídeos,

foi identificada em uma paciente coreana jovem com câncer de mama subtipo luminal (Kan et al., 2018).

O gene *TP53*, frequentemente mutado em câncer de mama, apresentou neste estudo uma alteração *frameshift* com a inserção de um nucleotídeo, detectada em uma paciente jovem com câncer de mama subtipo luminal (821). Da mesma maneira, em pesquisa com pacientes coreanas jovens com este mesmo perfil, uma alteração *frameshift* foi identificada no gene *TP53*, referente a deleção de nucleotídeos, além desta outras 4 alterações *missenses* foram detectadas no gene *TP53* (Kan et al., 2018).

As variantes nonsense detectadas neste estudo são referentes aos genes *NF1* e *PRKAR1A*. O gene *NF1* foi selecionado para este estudo devido a detecção de variantes em pacientes jovens com câncer de mama subtipo triplo negativo. O gene *NF1* mutado promove um distúrbio caracterizado pelo desenvolvimento de tumores cerebrais, cutâneos e em outras partes do corpo. A proteína codificada pelo gene *NF1* participa do processo de transdução de sinal, através da regulação negativa de proteína *RAS*. A partir da análise de variantes somáticas deste estudo foram identificadas duas alterações no gene *NF1*, uma nonsense e outra missense, em uma paciente jovem com câncer de mama subtipo luminal (819).

O gene *PRKARIA* codifica uma proteína quinase associada à enzima cAMP, que promove a regulação do ciclo celular. As alterações no gene *PRKARIA* foram relatadas em neoplasias endócrinas. Duas pacientes jovens com câncer de mama subtipo luminal apresentaram alterações no gene *PRKARIA*, uma alteração nonsense detectada neste estudo na paciente 842 e uma alteração missense identificada em estudo anterior do nosso grupo de pesquisa (Encinas et al., 2018).

Além dos genes com alterações *frameshift* e *nonsenses*, outros oito genes, (*PIK3CA*, *AKT1*, *NF1*, *AHNAK*, *PTEN*, *CACNA1E*, *FAT2* e *SEMA6D*) apresentaram alterações *missenses* consideradas patogênicas, conforme as ferramentas de predição.

O gene fosfatidilinositol4,5bifosfato3quinase, ou *PIK3CA*, codifica uma proteína tirosina quinase que promove o processo de transdução de sinal pela ativação da proteína *AKT*. O gene *PIK3CA* é componente de vias de sinalização que inibe a apoptose e promove a proliferação, motilidade e angiogênese. O gene *PIK3CA* destaca-se por encontrar-se frequentemente mutado em câncer de mama. Uma alteração somática *missense* foi identificada neste estudo em uma paciente jovem com câncer de mama subtipo luminal (809). Da mesma maneira, em um maior número de casos (6 pacientes), foram identificadas variantes *missenses* no gene *PIK3CA* (Kan et al., 2018).

Em pesquisa realizada no estudo anterior do nosso grupo, oito pacientes jovens com câncer de mama subtipo luminal apresentaram variantes *missenses* no gene *PIK3CA* (Encinas et al., 2018).

O gene *AKT* está diretamente associado ao gene *PIK3CA*, uma vez que a desregulação da via de sinalização *PI3K/AKT* promove alterações relacionadas diretamente à promoção da carcinogênese. O gene *AKT* codifica uma proteína tirosina quinase, diretamente relacionada aos processos celulares de proliferação e autofagia. As alterações *missenses* detectadas no gene *AKT1* foram referentes a duas pacientes jovens com câncer de mama subtipo luminal, uma paciente deste estudo (829) e outra paciente de estudo anterior do nosso grupo (Encinas et al., 2018).

O gene *PTEN* é classificado como um supresso tumoral, descrito como um regulador negativo da via de sinalização de *AKT*. As evidências sugerem que esse gene está relacionado ao controle da divisão celular, migração e angiogênese.

O gene *AHNAK*, codifica uma proteína do núcleo celular, componente de vários processos biológicos. No mecanismo de carcinogênese essa proteína, quando mutada, promove a invasão celular e a metástase. Outra função relacionada ao gene *AHNAK* é a sua atuação no processo celular de ubiquitinação. De acordo com os resultados deste estudo, duas variantes *missenses* patogênicas foram detectadas neste gene, sendo ambas identificadas em uma paciente jovem com câncer de mama subtipo luminal A (819). Em outro estudo, uma variante *missense* foi detectada no gene *ANNAK* em uma paciente jovens com câncer de mama subtipo triplo negativo.

Outro gene, o *CACNA1E*, apresenta uma função diferenciada dos demais. Ele é um importante componente da família proteica de canais de cálcio, que participam de variados processos celulares como proliferação, diferenciação, apoptose e metástase. Neste estudo foi detectada uma alteração *missense* no gene *CACNA1E*. Da mesma maneira, em dois estudos prévios também foram identificadas variantes neste gene. No estudo Kan et al., 2018, foi detectada uma alteração em sítio de splicing no gene *CACNA1E*, a partir da amostra de tumor proveniente de uma paciente coreana jovem com câncer de mama subtipo luminal. Na publicação de Encinas et al., 2018, uma alteração *missense* foi identificada em uma paciente com este mesmo perfil.

Ambos os genes *CACNA1E* e *FAT2* estão relacionados ao mecanismo de motilidade celular, conforme o programa Topp Gene. A proteína *FAT2* é um membro da família da caderina, um grupo de proteínas de membrana que promovem a adesão celular e controla o processo de proliferação. Neste estudo, uma alteração *missense* foi

detectada no gene *FAT2*, sendo considerada patogênica em quatro do total de seis preditores.

O gene *SEMA6D* participa da resposta imune e do processo de neurogênese. A região extracelular da proteína *SEMA6D* atua no sistema imunológico através da secreção de citocinas. As alterações neste gene foram recentemente associadas ao processo de metástase. Duas variantes *missenses* foram identificadas no gene *SEMA6D*, ambas em paciente jovem com câncer de mama subtipo luminal, proveniente deste estudo (842) e de pesquisa anterior do nosso grupo (Encinas et al., 2018).

6. Conclusão

Em 9 de 11 pacientes com câncer de mama subtipo luminal foram encontradas variantes somáticas patogênicas em 13 genes potencialmente *drivers*: *GATA3*, *GRHL2*, *MEN1*, *TP53*, *CACNA1*, *FAT2*, *NF1*, *AHNAK*, *PIK3CA*, *AKT1*, *PRKARIA*, *SEMA6D* e *PTEN*.

Duas pacientes destacaram-se neste estudo por apresentar, em amostras de tumores, tanto variantes somáticas como germinativas patogênicas. Destas, uma paciente apresentou 4 variantes (3 somáticas e 1 germinativa), do qual as variantes somáticas referem-se a uma alteração *nonsense* em *PRKARIA* e duas alterações *missenses*, uma em *SEMA6D* e outra em *PTEN*, além de uma alteração germinativa *missense* em *BRCA2*. Em outra paciente foram identificadas duas alterações, ambas *frameshift*, sendo uma variante somática em *TP53* e uma germinativa em *BRCA1*.

Em relação às pacientes com alterações exclusivamente somáticas patogênicas, três participantes apresentaram variantes somáticas em *GATA3*, todas elas *frameshift*, sendo este o gene mais frequentemente alterado nesta coorte de pacientes jovens. Uma das participantes apresentou em concomitância uma variante *frameshift* somática em *GRHL2*. Outra paciente apresentou mutação somática *frameshift* em *MEN1*.

Uma única paciente apresentou 6 alterações somáticas, uma variante *nonsense* no gene *NF1* e 5 alterações *missenses* identificadas em diferentes genes, *CACNA1E*, *FAT2*, *NF1* e *AHNAK*. Outras duas pacientes apresentaram uma alteração *missense* em cada, sendo uma paciente com uma variante no gene *PIK3CA* e outra paciente uma variante no gene *AKT1*.

A maioria das variantes somáticas relatadas, exceto *FAT2* e *NF1*, foram previamente descritas em câncer de mama subtipo luminal de mulheres jovens.

Referências Bibliográficas

- Adzhubei I, Jordan DM, Sunyaev SR, Predicting functional effect of human missense mutations using PolyPhen-2, *Current Protocols in Human Genetics* 2013, Chapter 7: Unit7.20.
- Althubaiti S, Karwath A, Dallol A et al., Ontology-based prediction of cancer driver genes. *Sci Rep.* 2019; 22;9(1):17405.
- Anders CK, Hsu DS, Broadwater et al., Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J Clin Oncol.* 2008; 10;26(20):3324-30.
- Anders CK, Acharya CR, Hsu DS et al., Age-specific differences in oncogenic pathway deregulation seen in human breast tumors. *PLoS One.* 2008; 2;3(1):e1373.
- Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet.* 2003;33 Suppl:238-44. Review. *PubMed* PMID: 12610533.
- Bollet MA, Sigal-Zafrani B, Mazeau V et al., Age remains the first prognostic factor for loco-regional breast cancer recurrence in young (<40 years) women treated with breast conserving surgery first. *Radiother Oncol.* 2007; 82(3):272-80.
- Cancer Genome Atlas N. The cancer genome atlas pan-cancer analysis project. *Nat Genet.* 2013; 45(10): 1113-1120.
- Copson E, Eccles B, Maishman T et al., Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study. *J Natl Cancer Inst.* 2013;105(13):978-88.
- Encinas G, Sabelnykova VY, de Lyra EC, et al. Somatic mutations in early onset luminal breast cancer. *Oncotarget.* 2018; 9(32): 22460-22479.
- Fundação Oncocentro de São Paulo - FOSP. 2015. <http://www.fosp.saude.sp.gov.br/>
- Gajdos C, Tartter PI, Bleiweiss IJ et al., Stage 0 to stage III breast cancer in young women. *J Am Coll Surg,* 2000; 190(5):523-529.
- Garraway, L. & Lander, E. Lessons from the cancer genome. *Cell.* 2013; 153, 17-37.
- Greenman C, Stephens P, Smith R., et al. Patterns of somatic mutation in human cancer genomes. *Nature.* 2007; 446: 153-158.
- GLOBOCAN 2018 (IARC), available from <http://gco.iarc.fr/>; The World Bank Group Indicators; available from <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
- Gray PN, Dunlop CL, Elliott AM et al. Not all next generation sequencing diagnostics are created equal: understanding the nuances of solid tumor assay design for somatic mutation detection. *Cancers,* 2015, 7:1313-1332.

Han W, Kim SW, Park IA et al., Young age: an independent risk factor for disease-free survival in women with operable breast cancer. *BMC Cancer*. 2004; 17;4:82.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144(5):646-74.

Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet*. 2014;15(9):585-98.

Hoogstrat M, Hinrichs JW, Besselink NJ et al. Simultaneous detection of clinically relevant mutations and amplifications for routine cancer pathology. *J Mol Diagn*, 2015; 17: 10-18.

Instituto Nacional de Câncer José Alencar Gomes da Silva. A situação do câncer de mama no Brasil: síntese de dados dos sistemas de informação. Rio de Janeiro: INCA, 2019.

Ioannidis NM, Rothstein JH, Pejaver V et al., REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*. 2016; 6;99(4):877-885.

Kan Z, Ding Y, Kim J, et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. *Nat Commun*. 2018; 30; 9(1):1725.

Kanke Y, Shimomura A, Saito M et al., Gene aberration profile of tumors of adolescent and young adult females. *Oncotarget*. 2017;9(5):6228-6237.

Knudson AG. Chasing the cancer demon. *Annu Rev Genet*. 2000;34:1-19. Review. PubMed PMID: 11092820.

Knudson AG. Cancer genetics. *Am J Med Genet*. 2002 Jul 22;111(1):96-102. Review. PubMed PMID: 12124744.

Liao S, Harmaier RJ, Mc Guire KP et al., The molecular landscape of premenopausal breast cancer. *Breast Cancer Res*. 2015;17:104.

Mitelman F. Recurrent chromosome aberrations in cancer. *Mutat Res*. 2000; 462(2-3): 247-253. Review. PubMed PMID: 10767636.

Nik-Zainal S, Davies H, Staaf J et al., Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016; 534 (7605): 47-54.

Nik-Zainal S, Morganella S. Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clin Cancer Res*. 2017; 1;23(11):2617-2629.

Olivier M, Bouaoun I, Villar S et al., Molecular features of premenopausal breast cancers in Latin American women: Pilot results from the PRECAMA study. *PLoS One*. 2019;14(1):e0210372.

- Park C, Yoo KA, Kim J et al., Integrative molecular profiling identifies a novel cluster of estrogen receptor-positive breast cancer in very young women. *Cancer Sci.*2019; 110(5):1760-1770.
- Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004; 351(27):2817-26.
- Perou CM, Sørlie T, Eisen MB, et al: Molecular portraits of human breast tumours. *Nature.* 2000; 406: 747-752.
- Ribnikar D, Ribeiro JM, Pinto D, et al. Breast cancer under age 40: a different approach. *Curr Treat Options in Oncol.* 2015; 16:16.
- Shihab HA, Gough J, Cooper DN, Day IN, Gaunt TR. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics.* 2013; 15;29(12):1504-10.
- Shyr C, Tarailo-Graovac M, Gottlieb M, et al. FLAGS, frequently mutated genes in public exomes. *BMC Med Genomics.* 2014; 7:64.
- Sorlie T, Perou CM, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA.* 2001; 98(19): 10869-10874
- Stephens PJ, Tarpey PS, Davies H, Loo PV, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature,* 2012; 486 (7403): 400-404.
- Tokheim, CJ, Papadopoulos N, Kinzler, KW et al., Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences.* 2016; 113,14330–35.
- Van't Veer L, Dai H, Van de Vijver MJ, et al., Gene expression profiling predicts clinical outcome of breast cancer. *Nature,* 2002; 415(6871): 530-536.
- Vaser, R., Adusumalli, S., Leng, S. et al. SIFT missense predictions for genomes. *Nat Protoc* 11, 2016; 1-9.
- Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet,* 2013; 14(10): 730-718.
- Weir B, Zhao X, Meyerson M. Somatic alterations in the human cancer genome. *Cancer Cell,* 2004; 6:433-438.
- WHO 2018 - <https://www.who.int/cancer/detection/breastcancer/en/>
- Yap YS, Singh AP, Lim JHC et al., Elucidating therapeutic molecular targets in premenopausal Asian women with recurrent breast cancers. *NPJ Breast Cancer.* 2018; 26;4:19.
- Yap TA, Gerlinger M, Futreal PA, et al. Intra-tumor heterogeneity: seeing the wood for the trees. *Sci Transl Med.* 2012; 4:12

Tabela 1: Cobertura das amostras de tumores das pacientes no painel de genes

| | % de bases de exons com <20 leituras | | | | | | | | | | | | | | | | Mediana (%) de bases de exons do gene por todas as amostras |
|-------------------------------------------------------------|--------------------------------------|------|------|------|------|------|------|------|------|-------|------|------|------|------|------|------|-------------------------------------------------------------|
| | 800 | 805 | 809 | 811 | 812 | 814 | 818 | 819 | 821 | 822 | 823 | 828 | 829 | 837 | 838 | 842 | |
| AHNAK | 9,5 | 6,8 | 1,5 | 7,0 | 12,6 | 32,6 | 5,4 | 9,3 | 10,5 | 4,8 | 2,1 | 6,1 | 11,5 | 5,5 | 6,6 | 7,3 | 6,9 |
| AKT1 | 0,0 | 3,1 | 0,0 | 0,8 | 3,3 | 32,6 | 0,0 | 6,2 | 0,8 | 0,0 | 0,0 | 0,0 | 10,9 | 0,0 | 0,0 | 0,5 | 0,2 |
| ANLN | 35,3 | 38,4 | 0,0 | 29,0 | 21,4 | 68,8 | 52,1 | 24,6 | 43,4 | 68,8 | 0,0 | 23,2 | 36,1 | 68,8 | 0,0 | 23,8 | 32,2 |
| APC | 0,0 | 40,6 | 0,0 | 3,7 | 8,6 | 81,5 | 0,0 | 11,3 | 0,0 | 0,0 | 0,0 | 4,4 | 33,4 | 0,0 | 0,0 | 11,3 | 1,9 |
| ARID1A | 3,2 | 18,0 | 0,3 | 3,3 | 5,4 | 68,8 | 3,3 | 17,2 | 5,0 | 2,5 | 2,5 | 1,7 | 3,1 | 3,6 | 2,9 | 5,6 | 3,3 |
| ATAD2B | 7,6 | 45,2 | 0,0 | 7,8 | 11,1 | 74,3 | 0,2 | 26,6 | 11,4 | 0,0 | 0,0 | 15,4 | 44,3 | 1,7 | 0,0 | 16,0 | 9,4 |
| ATM | 16,1 | 61,8 | 0,0 | 11,3 | 19,4 | 91,3 | 11,7 | 53,2 | 29,3 | 3,3 | 0,0 | 10,1 | 69,5 | 20,8 | 0,0 | 24,4 | 17,8 |
| ATR | 5,2 | 55,7 | 0,0 | 5,8 | 16,4 | 81,5 | 9,9 | 19,8 | 13,8 | 0,0 | 0,0 | 9,1 | 37,3 | 0,0 | 0,0 | 13,6 | 9,5 |
| ATXN1 | 0,3 | 6,4 | 0,0 | 0,0 | 0,7 | 13,2 | 0,0 | 3,0 | 0,2 | 0,0 | 0,0 | 0,0 | 1,2 | 0,0 | 0,0 | 0,6 | 0,1 |
| BAP1 | 0,0 | 18,4 | 0,0 | 3,8 | 4,4 | 70,3 | 0,0 | 15,6 | 25,0 | 0,0 | 0,0 | 0,0 | 8,1 | 0,0 | 0,0 | 9,4 | 1,9 |
| BRCA1 | 41,7 | 36,4 | 0,0 | 15,2 | 18,9 | 58,1 | 7,7 | 20,2 | 13,6 | 0,0 | 0,0 | 10,8 | 41,1 | 0,0 | 0,0 | 23,0 | 14,4 |
| BRCA2 | 11,8 | 72,5 | 0,0 | 6,1 | 25,5 | 88,6 | 0,6 | 24,8 | 20,8 | 0,0 | 0,0 | 12,7 | 77,7 | 0,0 | 0,0 | 31,8 | 12,3 |
| CACNA1E | 11,8 | 11,5 | 0,0 | 1,5 | 3,9 | 39,8 | 2,4 | 19,5 | 5,3 | 0,0 | 0,0 | 0,0 | 9,4 | 2,4 | 0,0 | 8,2 | 3,1 |
| CAMK1G | 0,0 | 4,2 | 0,0 | 0,0 | 6,3 | 45,0 | 0,0 | 16,6 | 9,4 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 2,4 | 0,0 |
| CDH1 | 3,1 | 21,8 | 0,0 | 0,6 | 4,2 | 48,0 | 0,0 | 21,5 | 2,7 | 0,0 | 0,0 | 0,0 | 6,1 | 0,0 | 0,0 | 4,2 | 1,7 |
| CSMD3 | 6,1 | 47,6 | 0,0 | 6,3 | 11,5 | 67,6 | 0,0 | 11,1 | 7,2 | 0,0 | 0,0 | 9,0 | 33,1 | 0,5 | 0,0 | 9,9 | 6,7 |
| CSPP1 | 13,6 | 51,2 | 0,0 | 8,6 | 19,9 | 91,0 | 0,0 | 13,1 | 15,9 | 0,0 | 0,0 | 11,4 | 29,8 | 0,0 | 0,0 | 16,1 | 12,2 |
| DALRD3 | 0,0 | 29,2 | 0,0 | 12,5 | 11,0 | 59,4 | 0,0 | 27,4 | 0,0 | 0,0 | 0,0 | 0,0 | 7,2 | 0,0 | 0,0 | 6,4 | 0,0 |
| DICER1 | 13,4 | 21,4 | 0,3 | 18,4 | 14,2 | 72,3 | 0,8 | 13,7 | 11,3 | 0,6 | 0,2 | 10,9 | 51,4 | 0,7 | 0,4 | 9,9 | 11,1 |
| EP300 | 7,4 | 23,3 | 0,0 | 7,7 | 7,7 | 57,9 | 0,6 | 36,7 | 11,9 | 0,0 | 0,0 | 3,3 | 33,8 | 0,7 | 0,0 | 11,9 | 7,5 |
| ERBB2 | 0,0 | 5,5 | 0,0 | 9,8 | 4,2 | 33,2 | 29,7 | 8,9 | 9,6 | 0,0 | 0,0 | 4,2 | 5,0 | 0,0 | 3,6 | 1,2 | 4,2 |
| ERBB4 | 4,5 | 24,8 | 0,0 | 1,5 | 4,7 | 88,6 | 0,0 | 11,7 | 9,6 | 0,0 | 0,0 | 3,7 | 16,4 | 0,0 | 0,0 | 2,7 | 3,6 |
| ESR1 | 3,7 | 26,5 | 0,0 | 0,2 | 9,0 | 44,4 | 0,5 | 2,2 | 6,2 | 0,0 | 0,0 | 0,0 | 14,7 | 0,0 | 0,0 | 7,7 | 1,3 |
| FAT2 | 2,9 | 37,8 | 0,0 | 2,0 | 1,9 | 68,9 | 2,0 | 19,4 | 5,6 | 0,0 | 0,0 | 0,0 | 7,5 | 0,0 | 0,0 | 2,7 | 2,0 |
| FAT4 | 2,4 | 38,4 | 0,0 | 6,6 | 3,0 | 69,8 | 0,5 | 6,7 | 6,6 | 0,0 | 0,0 | 3,1 | 20,4 | 3,4 | 0,0 | 4,2 | 3,3 |
| FBXW7 | 3,4 | 48,5 | 0,0 | 2,6 | 5,6 | 57,6 | 6,9 | 9,1 | 7,2 | 6,9 | 0,0 | 3,8 | 20,4 | 7,3 | 0,0 | 5,1 | 6,3 |
| GATA3 | 0,0 | 0,0 | 0,0 | 1,3 | 0,8 | 21,6 | 0,0 | 5,8 | 0,0 | 0,0 | 0,0 | 0,0 | 0,8 | 0,0 | 0,0 | 0,4 | 0,0 |
| GRHL2 | 0,0 | 7,5 | 0,0 | 0,0 | 2,1 | 39,0 | 0,0 | 0,3 | 0,0 | 0,0 | 0,0 | 0,0 | 9,9 | 0,0 | 0,0 | 0,0 | 0,0 |
| GRM3 | 2,0 | 10,8 | 0,0 | 1,5 | 2,3 | 28,9 | 0,0 | 3,7 | 5,9 | 0,0 | 0,0 | 3,3 | 7,4 | 0,0 | 0,0 | 2,9 | 2,1 |
| HERC2 | 58,3 | 57,3 | 28,4 | 56,1 | 61,5 | 74,7 | 52,0 | 53,0 | 57,4 | 42,8 | 29,4 | 57,9 | 55,2 | 56,9 | 39,1 | 64,7 | 56,5 |
| HUWE1 | 4,9 | 38,8 | 0,0 | 7,1 | 13,0 | 76,3 | 4,0 | 19,8 | 13,2 | 1,8 | 0,0 | 4,1 | 26,2 | 0,3 | 0,0 | 11,0 | 6,0 |
| LYST | 7,1 | 40,1 | 0,6 | 12,9 | 14,7 | 70,4 | 3,5 | 21,0 | 14,1 | 7,9 | 7,9 | 12,1 | 37,8 | 19,1 | 7,9 | 16,0 | 13,5 |
| MAP2K4 | 47,6 | 65,2 | 0,0 | 58,8 | 44,9 | 91,2 | 32,6 | 50,7 | 39,3 | 34,4 | 0,6 | 29,9 | 65,8 | 32,2 | 58,5 | 29,7 | 42,1 |
| MAP3K1 | 6,4 | 39,1 | 0,0 | 19,5 | 16,8 | 83,4 | 14,6 | 9,5 | 16,1 | 0,0 | 0,0 | 42,1 | 35,5 | 0,0 | 2,2 | 21,7 | 15,4 |
| MED23 | 14,2 | 52,0 | 0,0 | 11,7 | 14,2 | 75,0 | 15,8 | 12,4 | 13,3 | 15,5 | 0,0 | 15,3 | 61,3 | 14,2 | 0,0 | 13,5 | 14,2 |
| MEN1 | 0,0 | 1,2 | 0,0 | 0,0 | 0,0 | 57,3 | 0,0 | 1,2 | 0,0 | 0,0 | 0,0 | 0,0 | 11,5 | 0,0 | 0,0 | 0,0 | 0,0 |
| MET | 5,7 | 24,6 | 0,0 | 13,3 | 8,7 | 66,4 | 0,0 | 7,1 | 9,6 | 0,0 | 0,0 | 7,9 | 17,2 | 0,0 | 0,0 | 8,4 | 7,5 |
| MTHFD2 | 3,7 | 34,7 | 0,0 | 8,1 | 24,0 | 54,7 | 2,9 | 21,3 | 9,4 | 0,5 | 0,0 | 4,4 | 35,9 | 2,5 | 3,3 | 15,5 | 6,2 |
| MTOR | 6,4 | 45,4 | 0,0 | 1,2 | 9,9 | 86,3 | 0,0 | 32,1 | 22,9 | 0,0 | 0,0 | 3,7 | 12,8 | 4,6 | 0,0 | 6,0 | 5,3 |
| NCOA3 | 0,0 | 19,5 | 0,0 | 0,7 | 6,3 | 77,8 | 0,0 | 10,3 | 6,9 | 0,0 | 0,0 | 0,0 | 24,7 | 0,0 | 0,0 | 4,4 | 0,4 |
| NF1 | 29,6 | 39,5 | 24,4 | 26,8 | 24,4 | 71,5 | 32,0 | 29,2 | 22,0 | 27,4 | 25,9 | 31,8 | 31,6 | 26,5 | 24,8 | 31,8 | 28,3 |
| NOTCH1 | 12,4 | 16,6 | 0,0 | 11,5 | 10,3 | 53,7 | 9,0 | 61,8 | 14,6 | 0,0 | 0,0 | 0,4 | 5,5 | 0,0 | 8,1 | 0,0 | 8,5 |
| PARP4 | 60,5 | 74,3 | 51,2 | 55,7 | 56,7 | 87,0 | 50,2 | 57,0 | 58,3 | 41,5 | 43,0 | 60,1 | 60,0 | 46,9 | 41,2 | 51,2 | 56,2 |
| PCDH10 | 0,0 | 42,2 | 0,0 | 0,1 | 0,1 | 51,3 | 0,0 | 13,8 | 0,0 | 0,0 | 0,0 | 0,0 | 5,3 | 0,0 | 0,0 | 0,0 | 0,0 |
| PIK3AP1 | 26,4 | 19,4 | 0,0 | 45,7 | 10,0 | 62,7 | 46,4 | 16,0 | 6,4 | 22,9 | 0,0 | 0,0 | 22,1 | 46,4 | 28,6 | 0,8 | 20,7 |
| PIK3CA | 45,0 | 80,5 | 21,3 | 33,2 | 40,1 | 90,2 | 77,4 | 50,9 | 38,4 | 55,2 | 32,3 | 36,4 | 66,3 | 68,4 | 30,9 | 33,8 | 42,6 |
| PIK3R1 | 5,5 | 26,4 | 0,0 | 8,0 | 21,2 | 69,4 | 0,0 | 15,7 | 9,9 | 0,0 | 0,0 | 16,5 | 45,2 | 0,0 | 0,0 | 14,8 | 8,9 |
| POLD1 | 0,0 | 12,8 | 0,0 | 0,7 | 0,0 | 38,1 | 4,6 | 20,9 | 0,0 | 0,0 | 0,0 | 0,0 | 4,2 | 0,0 | 0,6 | 0,7 | 0,3 |
| PRKAR1A | 0,0 | 18,6 | 0,0 | 0,0 | 9,4 | 70,2 | 0,0 | 2,2 | 0,0 | 0,0 | 0,0 | 0,0 | 42,2 | 0,0 | 0,0 | 3,3 | 0,0 |
| PRKD1 | 10,0 | 21,2 | 0,0 | 6,3 | 8,5 | 56,9 | 24,4 | 15,3 | 21,5 | 2,8 | 0,0 | 8,9 | 29,3 | 0,0 | 4,8 | 15,0 | 9,4 |
| PTEN | 27,6 | 54,5 | 17,0 | 43,4 | 51,5 | 74,6 | 31,6 | 55,6 | 46,2 | 14,4 | 15,3 | 41,3 | 60,5 | 19,9 | 18,7 | 49,8 | 42,3 |
| PTPN13 | 3,8 | 52,9 | 0,0 | 4,3 | 15,9 | 73,3 | 4,3 | 14,4 | 7,5 | 0,0 | 0,0 | 11,4 | 41,6 | 0,5 | 0,0 | 12,9 | 5,9 |
| RAD51 | 11,7 | 40,1 | 10,0 | 4,8 | 13,9 | 70,0 | 8,3 | 7,5 | 14,6 | 8,3 | 4,2 | 0,0 | 31,7 | 14,2 | 18,3 | 5,4 | 10,8 |
| RAD9A | 0,0 | 11,7 | 0,0 | 0,0 | 0,0 | 30,2 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,3 | 0,0 | 0,0 | 0,0 | 0,0 |
| RB1 | 42,3 | 79,0 | 0,0 | 32,7 | 42,0 | 90,7 | 52,7 | 47,7 | 53,8 | 100,0 | 0,0 | 57,4 | 87,2 | 51,8 | 3,2 | 47,6 | 49,7 |
| RSBN1 | 0,3 | 36,9 | 0,0 | 5,0 | 2,6 | 61,3 | 0,0 | 10,2 | 13,0 | 0,0 | 0,0 | 0,0 | 34,6 | 0,0 | 0,0 | 4,2 | 1,4 |
| SEMA6D | 3,8 | 48,6 | 0,0 | 7,7 | 6,5 | 65,7 | 0,0 | 21,1 | 10,5 | 0,0 | 0,0 | 14,2 | 31,3 | 24,0 | 0,0 | 23,6 | 9,1 |
| SETD2 | 0,2 | 30,4 | 0,0 | 21,4 | 15,6 | 65,4 | 0,8 | 15,8 | 17,6 | 0,0 | 0,0 | 15,3 | 15,9 | 0,0 | 0,0 | 15,9 | 15,4 |
| SMARCA4 | 2,2 | 11,5 | 0,0 | 3,8 | 7,4 | 43,7 | 3,8 | 45,8 | 4,1 | 0,0 | 0,0 | 0,0 | 6,8 | 0,0 | 0,0 | 3,5 | 3,7 |
| SMURF2 | 22,7 | 52,5 | 19,1 | 22,8 | 40,0 | 81,3 | 17,8 | 26,9 | 41,5 | 26,6 | 12,5 | 22,7 | 44,8 | 13,5 | 25,1 | 28,8 | 25,8 |
| SPEN | 6,8 | 22,9 | 0,0 | 11,9 | 12,7 | 71,4 | 0,4 | 16,0 | 5,9 | 0,0 | 0,0 | 7,5 | 28,1 | 0,0 | 0,7 | 9,7 | 7,2 |
| TNC | 3,2 | 20,1 | 0,0 | 1,1 | 1,6 | 58,7 | 0,0 | 22,9 | 6,7 | 0,0 | 0,0 | 1,5 | 7,6 | 0,0 | 0,0 | 2,3 | 1,5 |
| TP53 | 0,0 | 14,4 | 0,0 | 0,0 | 2,4 | 58,3 | 0,0 | 10,1 | 8,6 | 0,0 | 0,0 | 0,0 | 26,6 | 0,0 | 0,0 | 15,4 | 0,0 |
| UBR5 | 1,8 | 26,7 | 0,0 | 6,0 | 13,1 | 60,6 | 0,0 | 7,8 | 13,1 | 0,0 | 0,0 | 1,3 | 16,6 | 0,0 | 0,0 | 5,3 | 3,6 |
| Mediana (%) de bases de exons por todos os genes da amostra | 4,7 | 29,8 | 0,0 | 6,3 | 9,9 | 68,2 | 0,7 | 15,9 | 9,7 | 0,0 | 0,0 | 3,9 | 26,4 | 0,0 | 0,0 | 8,9 | |

Tabela 2: Variantes somáticas detectadas nas amostras de tumores das pacientes jovens com câncer de mama subtipo luminal

| Paciente | Gene | HGVS c.DNA | | HGVS proteína | | Aminoácido | Cromossomo | Posição | Codon | Exon | NCBI (1000 genomas) | HOM/HET |
|----------|---------|----------------|------------------|----------------|------------------|------------|------------|-----------|---------|------|---------------------|---------|
| 800 | GATA3 | NM_001002295.1 | c.1224dupG | NP_001002295.1 | p.Pro409Alafs*99 | ND | 10 | 8115874 | ND | 6 | rs752977342 | HET |
| 805 | MEN1 | NM_000244 | ND | ND | ND | ND | 11 | 64577495 | ND | 2 | ND | LowAF |
| 805 | BRCA2 | NM_000059.3 | c.10211G>A | NP_000050.2 | p.Cys3404Tyr | C3404Y | 13 | 32972861 | tGt/tAt | 27 | rs587782614 | HET |
| 805 | HUWE1 | NM_031407.6 | c.9155C>T | NP_113584.3 | p.Ser3052Leu | S3052L | X | 53578092 | tCa/tTa | 65 | ND | HET |
| 805 | RAD51 | NM_001164269 | ND | ND | ND | T76M | 15 | 40994005 | aCg/aTg | 4 | ND | LowAF |
| 809 | PIK3CA | NM_006218.2 | c.1624G>A | NP_006209.2 | p.Glu542Lys | E542K | 3 | 178936082 | Gaa/Aaa | 10 | rs121913273 | HET |
| 811 | CAMK1G | NM_020439.2 | c.1397G>C | NP_065172.1 | p.Arg466Pro | R466P | 1 | 209786186 | cGg/cCg | 12 | rs769916493 | LowAF |
| 811 | PIK3CA | NM_006218.2 | c.2727C>G | NP_006209.2 | p.Phe909Leu | F909L | 3 | 178947852 | ttC/ttG | 19 | ND | HET |
| 819 | SPEN | NM_015001.2 | c.5267C>T | NP_055816.2 | p.Thr1756Ile | T1756I | 1 | 16258002 | aCc/aTc | 11 | ND | HET |
| 819 | CACNA1E | NM_001205294 | ND | ND | ND | L1773M | 1 | 181752824 | Ctg/Atg | 39 | ND | HET |
| 819 | BAP1 | NM_004656.2 | c.1792C>T | NP_004647.1 | p.Pro598Ser | P598S | 3 | 52437252 | Cca/Tca | 14 | ND | HET |
| 819 | FAT2 | NM_001447.2 | c.12880G>A | NP_001438.1 | p.Gly4294Ser | G4294S | 5 | 150885296 | Ggc/Agc | 23 | ND | HET |
| 819 | ANLN | NM_001284301 | ND | ND | ND | A40V | 7 | 36435975 | gCt/gTt | 2 | ND | HET |
| 819 | TNC | NM_002160.3 | c.5222G>A | NP_002151.2 | p.Arg1741Gln | R1741Q | 9 | 117804524 | cGg/cAg | 18 | rs149299073 | HET |
| 819 | NOTCH1 | NM_017617.4 | c.4568G>A | NP_060087.3 | p.Arg1523His | R1523H | 9 | 139399780 | cGt/cAt | 25 | rs367589813 | HET |
| 819 | AHNAK | NM_001346445 | ND | ND | p.Asp1342Ala | D1342A | 11 | 62297864 | gAt/gCt | 5 | ND | HET |
| 819 | AHNAK | NM_001346445 | ND | ND | p.Asp1342Tyr | D1342Y | 11 | 62297865 | Gat/Tat | 5 | rs776554777 | HET |
| 819 | NF1 | NM_000267.3 | ND | NP_000258.1 | p.Leu1183Met | L1183M | 17 | 29560070 | Ctg/Atg | 27 | ND | HET |
| 819 | NF1 | NM_000267.3 | ND | NP_000258.1 | p.Glu1181* | E1181* | 17 | 29560064 | Gaa/Taa | 27 | ND | HET |
| 821 | TP53 | NM_000546 | ND | ND | ND | ND | 17 | 7579447 | ND | 4 | ND | HET |
| 821 | TNC | NM_002160.3 | c.5024G>T | NP_002151.2 | p.Arg1675Met | R1675M | 9 | 117808790 | aGg/aTg | 17 | ND | HET |
| 822 | GATA3 | NM_001002295.1 | c.1307_1308dupCC | NP_001002295.1 | p.Ser437Profs*40 | ND | 10 | 8115955 | ND | 6 | ND | HET |
| 822 | GRHL2 | NM_001330593.1 | c.180_181insT | NP_001317522.1 | p.Lys61* | ND | 8 | 102564954 | ND | 3 | rs373113216 | HET |
| 822 | PIK3CA | NM_006218.2 | c.3140A>G | NP_006209.2 | p.His1047Arg | H1047R | 3 | 178952085 | cAt/cGt | 21 | rs121913279 | HET |
| 828 | AHNAK | NM_001346445 | ND | ND | ND | V3972I | 11 | 62289975 | Gtt/Att | 5 | ND | LowAF |
| 828 | RAD51 | NM_001164269 | ND | ND | ND | V112A | 15 | 40994113 | gTt/gCt | 4 | rs777467455 | HET |
| 829 | AKT1 | NM_001014431 | ND | ND | ND | P423L | 14 | 105237177 | cCa/cTa | 13 | ND | HET |
| 838 | GATA3 | NM_001002295.1 | c.1331dupG | NP_001002295.1 | p.*445Leufs*63 | ND | 10 | 8115979 | ND | 6 | ND | HET |
| 842 | SEMA6D | NM_001198999 | ND | ND | ND | I406T | 15 | 48056954 | aTt/aCt | 15 | ND | HET |
| 842 | PRKAR1A | NM_001276289 | ND | ND | ND | R96* | 17 | 66519005 | Cga/Tga | 4 | rs281864783 | LowAF |
| 842 | PTEN | NM_000314 | ND | ND | ND | L220V | 10 | 89717633 | ND | 7 | ND | HET |

| Paciente | Gene | Al Ref | Al Alt | Freq Al | Base Ref | Base Alt | Fow Ref | Rev Ref | Fow Alt | Rev Alt | Classe Funcional | Efeito | SIFT | PolyPhen | FATHMM |
|----------|---------|--------|--------|---------|----------|----------|---------|---------|---------|---------|------------------|----------|------------------|------------------|------------------|
| 800 | GATA3 | C | CG | 0,419 | 86 | 36 | 34 | 51 | 17 | 19 | frameshift | alto | ND | ND | ND |
| 805 | MEN1 | TC | T | 0.133 | 90 | 12 | 42 | 35 | 0 | 12 | frameshift | alto | ND | ND | ND |
| 805 | BRCA2 | G | A | 0.37 | 27 | 10 | 6 | 12 | 10 | 0 | missense | moderado | benigno (0.25) | deletério (0.72) | benigno (0.00) |
| 805 | HUWE1 | G | A | 0.412 | 51 | 21 | 17 | 14 | 0 | 21 | missense | moderado | benigno (0.11) | benigno (0.00) | benigno (0.36) |
| 805 | RAD51 | C | T | 0.127 | 110 | 14 | 43 | 54 | 8 | 6 | missense | moderado | benigno (0.06) | benigno (0.00) | benigno (0.39) |
| 809 | PIK3CA | G | A | 0.263 | 266 | 70 | 96 | 101 | 35 | 35 | missense | moderado | deletério (0.04) | deletério (0.91) | benigno (0.61) |
| 811 | CAMK1G | G | C | 0.109 | 110 | 12 | 59 | 51 | 13 | 0 | missense | moderado | benigno (0.12) | benigno (0.00) | benigno (0.69) |
| 811 | PIK3CA | C | G | 0.156 | 77 | 12 | 46 | 31 | 0 | 12 | missense | moderado | benigno (0.11) | benigno (0.38) | benigno (0.75) |
| 819 | SPEN | C | T | 0.269 | 52 | 14 | 21 | 18 | 14 | 0 | missense | moderado | benigno (0.12) | benigno (0.10) | benigno (0.08) |
| 819 | CACNA1E | C | A | 0.152 | 105 | 16 | 20 | 70 | 0 | 16 | missense | moderado | deletério (0.01) | deletério (0.92) | deletério (0.96) |
| 819 | BAP1 | G | A | 0.321 | 56 | 18 | 13 | 26 | 0 | 18 | missense | moderado | benigno (0.50) | benigno (0.02) | benigno (0.49) |
| 819 | FAT2 | C | T | 0.368 | 38 | 14 | 13 | 12 | 14 | 0 | missense | moderado | deletério (0.01) | deletério (0.75) | benigno (0.71) |
| 819 | ANLN | C | T | 0.245 | 49 | 12 | 23 | 15 | 0 | 12 | missense | moderado | benigno (0.37) | benigno (0.00) | benigno (0.43) |
| 819 | TNC | C | T | 0.351 | 57 | 20 | 19 | 19 | 20 | 0 | missense | moderado | benigno (0.36) | deletério (0.75) | benigno (0.57) |
| 819 | NOTCH1 | C | T | 0.283 | 53 | 15 | 19 | 20 | 15 | 0 | missense | moderado | benigno (0.11) | benigno (0.00) | benigno (0.64) |
| 819 | AHNAK | T | G | 0.309 | 25 | 81 | 24 | 33 | 0 | 25 | missense | moderado | deletério (0.00) | deletério (0.98) | benigno (0.03) |
| 819 | AHNAK | C | A | 0.305 | 25 | 82 | 25 | 33 | 0 | 25 | missense | moderado | deletério (0.00) | deletério (0.99) | benigno (0.03) |
| 819 | NF1 | C | A | 0.333 | 11 | 33 | 16 | 7 | 0 | 11 | missense | moderado | deletério (0.00) | deletério (0.97) | benigno (0.73) |
| 819 | NF1 | G | T | 0.314 | 11 | 35 | 13 | 12 | 0 | 11 | nonsense | Alto | ND | ND | ND |
| 821 | TP53 | A | AG | 0.346 | 78 | 27 | 48 | 29 | 22 | 5 | frameshift | alto | ND | ND | ND |
| 821 | TNC | C | A | 0.256 | 39 | 10 | 16 | 14 | 8 | 2 | missense | moderado | benigno (0.22) | benigno (0.10) | benigno (0.55) |
| 822 | GATA3 | A | ACC | 0.252 | 310 | 78 | 158 | 151 | 30 | 48 | frameshift | alto | ND | ND | ND |
| 822 | GRHL2 | C | CT | 0.271 | 107 | 29 | 73 | 33 | 21 | 8 | frameshift | alto | ND | ND | ND |
| 822 | PIK3CA | A | G | 0.265 | 170 | 45 | 61 | 65 | 18 | 27 | missense | moderado | benigno (0.11) | benigno (0.08) | benigno (0.80) |
| 828 | AHNAK | C | T | 0.137 | 22 | 161 | 82 | 58 | 7 | 15 | missense | moderado | deletério (0.02) | benigno (0.23) | benigno (0.00) |
| 828 | RAD51 | T | C | 0.228 | 57 | 13 | 10 | 35 | 1 | 12 | missense | moderado | benigno (0.07) | benigno (0.00) | benigno (0.07) |
| 829 | AKT1 | G | A | 0.373 | 59 | 22 | 14 | 24 | 22 | 0 | missense | moderado | deletério (0.01) | deletério (0.98) | benigno (0.62) |
| 838 | GATA3 | T | TG | 0.349 | 192 | 67 | 91 | 100 | 34 | 33 | frameshift | alto | ND | ND | ND |
| 842 | SEMA6D | T | C | 0,156 | 64 | 10 | 38 | 17 | 4 | 6 | missense | moderado | deletério (0.03) | deletério (0.95) | benigno (0.11) |
| 842 | PRKAR1A | C | T | 0,11 | 91 | 10 | 45 | 37 | 9 | 1 | nonsense | alto | ND | ND | ND |
| 842 | PTEN | C | G | 0,346 | 27 | 78 | 31 | 21 | 27 | 0 | missense | moderado | benigno (0.50) | benigno (0.33) | deletério (0.84) |

| Paciente | Gene | REVEL | Mut. Assessor | Mut. Taster | Clinical (Clinvar) | Cosmic Tumor Site | COSMIC ID |
|----------|---------|------------------|------------------|------------------|-------------------------|-------------------------|---------------------------------------------------------|
| 800 | GATA3 | ND | ND | ND | | breast | c.1223_1224insAp.P409fs*>37,c.1223_1224insTp.P409fs*>37 |
| 805 | MEN1 | ND | ND | ND | | | |
| 805 | BRCA2 | benigno (0.32) | benigno (0.45) | benigno (0.08) | | biliary tract | c.10211G>Cp.C3404S |
| 805 | HUWE1 | benigno (0.15) | benigno (0.59) | benigno (0.40) | | | |
| 805 | RAD51 | benigno (0.07) | ND | benigno (0.08) | | | |
| 809 | PIK3CA | deletério (0.74) | deletério (0.44) | benigno (0.81) | | breast, ovary, prostate | |
| 811 | CAMK1G | benigno (0.53) | benigno (0.35) | benigno (0.38) | | | |
| 811 | PIK3CA | deletério (0.86) | benigno (0.41) | deletério (0.58) | | breast, oesophagus | c.2727C>Gp.F909L,c.2727C>Ap.F909L,c.2727C>Tp.F909F |
| 819 | SPEN | benigno (0.23) | benigno (0.60) | benigno (0.08) | | | |
| 819 | CACNA1E | deletério (0.85) | benigno (0.41) | deletério (0.58) | | | |
| 819 | BAP1 | benigno (0.23) | benigno (0.06) | benigno (0.31) | | | |
| 819 | FAT2 | deletério (0.55) | benigno (0.39) | deletério (0.51) | | | |
| 819 | ANLN | benigno (0.02) | benigno (0.24) | benigno (0.08) | | | |
| 819 | TNC | deletério (0.68) | deletério (0.71) | benigno (0.34) | | urinary tract | c.5222G>Ap.R1741Q |
| 819 | NOTCH1 | benigno (0.11) | benigno (0.26) | benigno (0.08) | | | |
| 819 | AHNAK | benigno (0.43) | deletério (0.79) | deletério (0.81) | | | |
| 819 | AHNAK | benigno (0.46) | deletério (0.79) | deletério (0.81) | | | |
| 819 | NF1 | deletério (0.57) | deletério (0.82) | deletério (0.58) | | | |
| 819 | NF1 | ND | ND | deletério (0.81) | | | |
| 821 | TP53 | ND | ND | ND | pathogenic | | |
| 821 | TNC | benigno (0.40) | benigno (0.21) | deletério (0.81) | probable-non-pathogenic | | |
| 822 | GATA3 | ND | ND | ND | | breast, lymphoid tissue | c.1304_1305insCp.S437fs*>9,c.1304_1305insCCp.S437fs*>9 |
| 822 | GRHL2 | ND | ND | ND | | | |
| 822 | PIK3CA | deletério (0.75) | benigno (0.15) | deletério (0.81) | pathogenic | breast, ovary, prostate | c.3140A>Cp.H1047P,c.3140A>Gp.H1047R |
| 828 | AHNAK | benigno (0.09) | benigno (0.34) | benigno (0.08) | | | |
| 828 | RAD51 | benigno (0,09) | ND | benigno (0.08) | | | |
| 829 | AKT1 | deletério (0.84) | deletério (0.99) | deletério (0.81) | | | |
| 838 | GATA3 | ND | ND | ND | | | |
| 842 | SEMA6D | deletério (0.61) | benigno (0.46) | deletério (0,54) | | | |
| 842 | PRKAR1A | ND | ND | ND | | | |
| 842 | PTEN | deletério (0.75) | deletério (0,78) | deletério (0,81) | | | |

Tabela 3: Características dos genes com variantes somáticas patogênicas

| Gene | Classificação do Gene | | | Câncer de Mama LM | | Câncer de Mama TN | | Função Molecular | | | Processo Biológico | | | Transdução de Sinal |
|---------|-----------------------|----------|-----------|---------------------|-----------------|----------------------------|------------------|------------------|---------------|-------------|--------------------|------------|-----------|---------------------|
| | Driver | Oncogene | Supressor | Encinas et al.,2018 | Kan et al.,2018 | Literatura e base de dados | Proteína Quinase | Transcrição | Ubiquitinação | Metabolismo | Ciclo Celular | Motilidade | Cromatina | Reparo do DNA |
| AHNAK | | ■ | | | | ■ | | | ■ | | | | | |
| AKT1 | ■ | ■ | | ■ | | | ■ | | | ■ | ■ | ■ | | |
| CACNA1E | | | | ■ | | | | | | | | ■ | | |
| FAT2 | | ■ | | | | | | | | | | | | |
| GATA3 | ■ | | ■ | ■ | | | | ■ | | ■ | ■ | ■ | ■ | |
| GRHL2 | | | | ■ | | | | | | ■ | | | | |
| MEN1 | ■ | | ■ | | ■ | | | ■ | | ■ | ■ | | ■ | |
| NF1 | ■ | | ■ | | | ■ | ■ | | | | | ■ | | |
| PIK3CA | | ■ | | ■ | | | ■ | | | ■ | | | | |
| PRKAR1A | | | ■ | ■ | | | | | ■ | | ■ | | | |
| PTEN | ■ | | ■ | ■ | | | ■ | | | | ■ | ■ | | |
| SEMA6D | | | | ■ | | | | | | | | | | |
| TP53 | ■ | | ■ | ■ | | | | ■ | ■ | ■ | ■ | ■ | ■ | ■ |