PEDRO ADOLPHO DE MENEZES PACHECO SERIO

Perfil de mutações de câncer de mama triplo negativo em pacientes adultos jovens

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

Programa de Oncologia

Orientadora: Prof.ª Dr.ª Maria Aparecida Azevedo Koike Folgueira

São Paulo

2023

PEDRO ADOLPHO DE MENEZES PACHECO SERIO

Perfil de mutações de câncer de mama triplo negativo em pacientes adultos jovens

Versão original

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

Programa de Oncologia

Orientadora: Prof.ª Dr.ª Maria Aparecida Azevedo Koike Folgueira

São Paulo

2023

Dados Internacionais de Catalogação na Publicação (CIP)

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

Serio, Pedro Adolpho de Menezes Pacheco
Perfil de mutações de câncer de mama triplo negativo em pacientes adultos jovens / Pedro Adolpho de Menezes Pacheco Serio. -- São Paulo, 2023.
Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Oncologia.
Orientadora: Maria Aparecida Azevedo Koike Folgueira.

Descritores: 1.Neoplasias de mama triplo negativas 2.Neoplasias da mama 3.Adulto jovem 4.Variantes somáticas 5.Sequenciamento de nucleotídeos em larga escala

USP/FM/DBD-104/23

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Nome: Serio, Pedro Adolpho de Menezes Pacheco.

Título: Perfil de mutações de câncer de mama triplo negativo em pacientes adultos jovens

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

Programa de Oncologia

Orientadora: Prof.ª Dr.ª Maria Aparecida Azevedo Koike Folgueira

Aprovado em:

Banca examinadora

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

AGRADECIMENTOS

Agradeço à Deus, por toda iluminação que deu e continuar a dar em minha vida, de maneira contínua.

Agradeço ao meu pai, à minha mãe, à minha irmã e à minha noiva que sempre me apoiam em minhas jornadas e me dão forças.

Agradeço à minha orientadora, Profª. Drª. Maria Aparecida Azevedo Koike Folgueira, que sempre me tratou com muito carinho e me guiou ativamente em meus projetos.

Agradeço aos membros de meu grupo de pesquisa: Lucia, Simone, Rosi, Gláucia, Lívia e Vinícius, além de outros que já fizeram parte da equipe. Os quais me ajudaram durante todo percurso do meu doutorado, além de me oferecerem sua amizade e companheirismo.

Agradeço a todos os amigos e colegas que conheci durante meu doutorado, que ajudaram direta ou indiretamente com minha evolução pessoal e/ou profissional.

Agradeço à Raquel Piovezan, que sempre me acudiu de forma atenciosa e gentil em minhas dúvidas e trapalhadas durante os procedimentos burocráticos da pós-graduação.

RESUMO

Serio PAMP. Perfil de mutações de câncer de mama triplo negativo em pacientes adultos jovens [tese]. São Paulo: Faculdade de Medicina da Universidade de São Paulo, Instituto do Câncer do Estado de São Paulo, 2023.

Com o objetivo de melhor caracterizar o perfil de mutações de pacientes adultas jovens com câncer de mama triplo-negativo, construímos painel personalizado de genes, selecionando genes relevantes em câncer de mama e/ou frequentemente alterados em mulheres jovens adultas com câncer de mama. Pacientes foram incluídas no estudo de forma prospectiva e tiveram material genético de seu tumor e sangue analisados por meio de sequenciamento-alvo de DNA. Este painel, constituído de 64 genes, foi utilizado para análise de variantes germinativas em 15 pacientes e somáticas em 12 pacientes. Dentre as 12 pacientes, apenas uma não apresentou qualquer variante somática. Nesta coorte, 75% das pacientes apresentaram mutação somática em *TP53*, todas classificadas como causadoras ou prováveis causadoras de perda de função na proteína. Outros genes relevantes em câncer, como *NF1* e *NOTCH1* também se mostraram alterados, além de outros genes menos frequentemente alterados, como *PTPN13, FBXW7* e *UBR5*. Todas as variantes somáticas em *NOTCH1, CACNA1E* e *PTPN13* estavam associadas com uma mutação em *TP53*. Duas pacientes não apresentavam mutação somática patogênica em *TP53. Em uma delas, NF1* apresentava-se mutado, associado com alterações em *FBXW7, PIK3CA, SPEN* e *UBR5*. Em outra, observou-se uma única variante em *ATR*. Dentre as 15 pacientes que tiveram variantes germinativas analisadas, duas (13%) apresentaram uma mutação germinativa patogênica, ambas no gene *BRCA1*. Adicionalmente, realizamos uma análise exploratória usando os portais COSMIC e cBioPortal, para identificação de alterações moleculares de acordo com a faixa etária. Identificamos *CDH1* e *MAP3K1* com menor predominância de variantes somáticas em pacientes adultas jovens, em comparação com pacientes de idade mais avançada. Os genes *SMURF2* e *PRKAR1A* foram identificados com maior número de amplificações em pacientes adultas jovens com câncer de mama, em comparação com pacientes de idade mais avançada. Em conclusão, o presente trabalho confirma dados prévios que demonstram alta frequência de mutações somáticas em *TP53*, assim como outros genes associados, e contribui com a expansão de conhecimento das alterações somáticas em mulheres jovens com câncer de mama triplo-negativo, que carecem de informação na literatura.

Palavras-chave: Neoplasias de mama triplo negativas. Neoplasias da mama. Adulto jovem. Variantes somáticas. Sequenciamento de nucleotídeos em larga escala.

ABSTRACT

Serio PAMP. Mutational profile of young adult triple-negative breast cancer patients [thesis]. São Paulo: Faculdade de Medicina da Universidade de São Paulo, Instituto do Câncer do Estado de São Paulo, 2023.

To better characterize the mutation profile of young adult female patients with triple-negative breast cancer, we constructed a custom gene panel, selecting relevant genes in breast cancer and/or genes frequently altered in young adult women with breast cancer. Patients were prospectively enrolled in the study and had genetic material from their tumor and blood analyzed through DNA target sequencing. This panel, consisting of 64 genes, was used to analyze germline variants in 15 patients and somatic variants in 12 patients. Among the 12 patients, only one did not present any somatic variant. In this cohort, 75% of patients presented somatic mutation in *TP53*, all classified as causing or likely causing loss of function in the protein. Other relevant genes in cancer, such as *NF1* and *NOTCH1*, were also affected, in addition to other less frequently altered genes, such as *PTPN13*, *FBXW7* and *UBR5*. All somatic variants in *NOTCH1*, *CACNA1E* and *PTPN13* were associated with a mutation in *TP53*. In two patients no pathogenic somatic mutation was detected in *TP53. In one of these patients*, *NF1* was mutated, associated with alterations in *FBXW7*, *PIK3CA*, *SPEN* and *UBR5*. In the other patient, a single variant in *ATR was observed*. Among the 15 patients who had germline variants analyzed, two (13%) were carriers of a pathogenic germline mutation, both in the *BRCA1* gene. Additionally, we performed an exploratory analysis using the COSMIC and cBioPortal portals, to detect molecular alterations according to age group. We identified, a lower predominance of somatic variants in *CDH1* and *MAP3K1* in young adult patients, compared to elderly groups. The *SMURF2* and *PRKAR1A* genes were identified with a higher number of amplifications in young adult patients with breast cancer, compared to elderly groups. In conclusion, the present work confirms previous data that demonstrate a high frequency of somatic mutations in *TP53*, as well as other associated genes, and contributes to the expansion of knowledge of somatic alterations in young women with triple-negative breast cancer, who lack information in the literature.

Keywords: Triple negative breast cancer. Breast cancer. Young adults. Somatic variants. High-throughput nucleotide sequencing.

SUMÁRIO

1 Introdução

Embora a maioria dos cânceres de mama ocorra entre as idades de 60 a 80 anos, a doença é a principal causa de morte por câncer em mulheres adultas jovens (18 a 40 anos). Além disso, enquanto dados indicam diminuição de incidência de câncer de mama no contexto geral, a porcentagem anual de pacientes adultos jovens acometidos pela doença parece não seguir o mesmo padrão, se mantendo constante (1,2)

Alguns estudos demonstraram que mulheres diagnosticadas com câncer em idades menos avançadas podem apresentar pior prognóstico. Relata-se que a sobrevida global e a sobrevida livre de doença em adultos muito jovens (≤35 anos) é menor quando comparada com pacientes pré-menopausadas de idade um pouco mais avançada (de 36 a 49 anos de idade ao diagnóstico), estratificando-se a doença de acordo com o subtipo molecular. Excluindo o subtipo de câncer de mama luminal A, que é conhecido como um subtipo menos agressivo, mulheres mais jovens (≤35 anos ao diagnóstico) com todos os outros subtipos apresentam um pior prognóstico, quando comparadas ao grupo de mulheres com idades mais avançadas. Em outro estudo, foram analisadas 1732 pacientes com câncer de mama triplo-negativo (*triple-negative breast cancer* - TNBC), as quais foram estratificadas em 5 coortes, de acordo o intervalo de idade ao diagnóstico (≤30 anos, 31-40 anos, 41-50 anos, 51-60 anos e >60 anos). Observou-se que as pacientes mais jovens (≤40 anos) apresentam sobrevida global e livre de doença significativamente pior. A mediana de anos de sobrevida global foi de 7 em pacientes mais jovens e de 12 a 14 anos nas faixas etárias mais avançadas (3,4).

No aspecto transcricional, poucos estudos exploraram as diferenças do câncer de mama entre grupos etários. Em um dos estudos, relata-se assinaturas de expressão gênicas relacionadas com células-tronco, fatores de crescimento e sistema imune em pacientes jovens (4,5). Estudos mais recentes também destacam maior representatividade do sistema imune em jovens, sugerindo que tais pacientes podem apresentar um microambiente tumoral mais ativo (6-8). Além disso, relata-se a associação da maior expressão de Tenascina C em jovens adultos com câncer de mama. Tenascina C é um

importante membro componente da matriz extracelular, correlacionado com maior invasividade em câncer de mama (9,10).

Dentre os subtipos de câncer de mama, o câncer de mama triplo-negativo (TNBC), que não expressa receptores hormonais e receptor do fator de crescimento epidérmico humano, apresenta o comportamento mais agressivo, estando associado a maiores taxas de metástase (11). Além disso, alguns estudos demonstram que a proporção de pacientes com TNBC pode ser maior em adultos jovens (~20-30%) quando comparada com idades ao diagnóstico mais avançadas (7,12,13). É importante notar que o TNBC e o câncer de mama tipo basal compartilham uma alta correlação molecular. Enquanto o primeiro é caracterizado por técnicas de imunohistoquímica, o último é classificado por análise de assinatura de expressão gênica. Um dos principais obstáculos no tratamento do TNBC está localizado em suas características moleculares, uma vez que não há expressão de receptores hormonais ou *HER2* para um tratamento alvo, os pacientes são tratados principalmente com taxanos e agentes antraciclínicos. Adicionalmente, o subtipo molecular basal (que apresenta grande sobreposição com o subtipo triplo-negativo) pode ser subdividido em outros subtipos moleculares, caracterizados por assinaturas de expressão gênica que apresentam valor clínico e prognóstico variado, sendo estas: *Basal-like* 1 (BL1) e 2 (BL2), *mesenchymal* (M) e *luminal androgen receptor* (LAR) (14). Em um estudo recente, pesquisadores realizaram o sequenciamento do transcriptoma (RNA-seq) de 237 pacientes com TNBC, sendo 28 destes pacientes, adultos jovens (<40 anos) e observou-se que o subtipo LAR é mais frequente em pacientes de idade mais avançada do que em adultos jovens (8).

Mutação em BRCA1 é outra característica relevantes em TNBC. *BRCA1* e *BRCA2* são genes envolvidos na via de reparo homólogo do DNA e mutações nesses genes estão associadas ao câncer de mama de início precoce e à síndrome hereditária do câncer de mama e ovário (HBOC). De fato, estudos recentes demonstram que pacientes com câncer de mama adultas jovens mais frequentemente apresentam tumores com alterações em *BRCA1* e *BRCA2*, bem como assinatura mutacional relacionada ao fenótipo de deficiência da via de reparo homólogo de DNA (8).

As deficiências nos genes de reparo homólogo do DNA permitem que os pacientes se beneficiem da terapia com compostos de platina, bem como do uso de terapia com inibidor de poli (adenosina difosfato-ribose) polimerase (PARP) (15). De fato, diante de resultados de alguns estudos que demonstraram tempo livre de progressão aumentado em pacientes com TNBC tratados com inibidores de PARP (*PARPi*), *guidelines* como a *National Comprehensive Cancer Network* (NCCN), já indicam o uso de *PARPi* (olaparib e talazoparib) como tratamento de primeira e segunda linha em pacientes com TNBC com tumor não ressecável e/ou doença de estadiamento avançado, cuja mutação germinativa em *BRCA1* e/ou *BRCA2* tenha sido detectada (16).

Outra característica notável do subtipo TNBC é a alta frequência (60%-80%) de mutações somáticas patogênicas no gene *TP53*, um supressor tumoral importante, com múltiplos papéis como participação no reparo de danos ao DNA, controle do ciclo celular e apoptose (17). O mesmo também está envolvido na síndrome de *Li-Fraumeni* (no caso de mutações germinativas), que também é correlacionada com o desenvolvimento de câncer de mama em idades menos avançadas (18,19).

Em concordância com o apontado acima, em um recente estudo publicado por nosso grupo **(APÊNDICE C)**, onde compilamos estudos que sequenciaram tumores de paciente com câncer de mama e foram publicados na literatura e em bancos de dados públicos, observamos que a maioria das pacientes jovens com TNBC apresentava pelo menos um oncogene afetado em associação com pelo menos um gene supressor tumoral. O gene mais frequentemente afetado foi o *TP53*, detectado em 70% dos tumores. Além disso, 20% dos tumores apresentam mutações em genes envolvidos nas vias de sinalização RAS ou *PIK3CA*. Observou-se predominância da assinatura mutacional 3, que está relacionada a defeitos de reparo homólogo (HRD). Apesar disso, foi detectada uma baixa frequência (2%) de mutações somáticas em genes relacionados ao reparo homólogo (20). Além de nosso estudo, outros também demonstram a associação de variantes somáticas patogênicas em *TP53* e assinaturas de defeito na via de HRD (6,21,22)

Há evidências de que a etnia pode influenciar a taxa de incidência de câncer e seu prognóstico (23,24). Apesar disso, a maioria dos estudos de sequenciamento aborda principalmente pacientes brancos caucasianos e não-

hispânicos. Logo, a contribuição desses estudos para populações etnicamente diversas, como o Brasil, pode ser mais baixa. Em um estudo brasileiro, dados de registros hospitalares de 188753 pacientes com câncer de mama foram comparados a dados de 922962 pacientes com câncer de mama listados no *National Cancer Institute* (SEER) e sugeriram uma maior prevalência de casos de adultos jovens no Brasil (odds ratio: 2,2; p<0,001). Comparando pacientes adultas jovens e pacientes com idade superior a 40 anos em São Paulo, os autores também reportam maiores taxas de metástase, menor tempo livre de doença, tumores de graus mais avançados e maior proporção de tumores triplo-negativos no grupo de menor faixa etária. Outro grupo brasileiro analisou dados nacionais de taxas de mortalidade em 19105 adultos jovens e com câncer de mama entre os anos de 1996 e 2017, reportando aumento na mortalidade de câncer de mama em adultas jovens, principalmente nas regiões sul, sudeste, centro-oeste e nordeste (25,26).

Observa-se que o câncer de mama triplo-negativo, especialmente em pacientes adultas jovens, apresenta algumas características peculiares e não bem compreendidas, que diferenciam o comportamento do câncer nesses pacientes com os de idade mais avançada. Ao mesmo tempo, notamos uma baixa representatividade dessa população nos múltiplos artigos de sequenciamento publicados, dificultando a formatação de um perfil. Portanto, nosso principal objetivo foi identificar mutações somáticas em pacientes adultos jovens com câncer de mama triplo-negativo por meio de sequenciamento alvo. Nossos objetivos secundários foram definir um painel de genes personalizado para pacientes jovens com TNBC por meio da coleta de dados da literatura e / ou bancos de dados, investigar as taxas de variantes nesses pacientes e explorar o potencial driver das variantes identificadas por meio de algoritmos de previsão de efeitos variantes e informações de banco de dados complementares.

2 Pacientes e métodos

2.1 Seleção de genes para painel personalizado

Para a seleção dos genes do painel, coletamos primeiramente dados clínicos e de sequenciamento de estudos de câncer de mama no Catálogo de Mutações Somáticas no Câncer (COSMIC; https://cancer.sanger.ac.uk/cosmic), CbioPortal (https://www.cbioportal.org/) e PubMed. Para pesquisa em literatura (PubMed), os termos de pesquisa foram: "*("breast cancer" OR "triple-negative breast cancer" OR TNBC) AND ("whole genome sequencing" OR "whole exome sequencing" OR "NGS" OR "somatic mutations" OR WGS OR WES OR "next generation sequencing")*". Os critérios de inclusão para a seleção de estudos foram estudos que sequenciaram genoma completo (*Whole Genome Sequencing* - WGS) ou exoma (*Whole Exome Sequencing* - WES) de pacientes adultos muito jovens (≤35 anos no diagnóstico) com carcinoma invasivo de mama, com dados clínicos informativos sobre o subtipo de câncer de mama e idade ao diagnóstico.

Identificamos 4 estudos que possuíam dados clínicos e de sequenciamento que se encaixavam em nossos critérios de inclusão, sendo eles: TCGA-BRCA (4 amostras), Nik-Zainal et.al., 2016 (17 amostras), Shah et.al., 2012 (1 amostra) e Kan et.al., 2018 (18 amostras), resultando em um total de 40 amostras de pacientes com TNBC e idade no diagnóstico ≤35 anos (6,17,27,28). Os genes alterados nos 4 estudos destacados foram selecionados seguindo a ordem a seguir:

I - Normalização entre estudos: uma vez que os dados selecionados foram combinados a partir de estudos que aplicaram WES e WGS, excluímos as variantes não codificantes. Variantes silenciosas também foram filtradas, pois essa informação estava ausente em alguns estudos. Os dados referentes às variantes foram complementados com a ferramenta online Oncotator (http://portals.broadinstitute.org/oncotator/; v1.5.3.0), onde informações de posição genômica, previsão de patogenicidade por meio de ferramentas *in silico*, mudança de aminoácidos, entre outros dados, foram adicionados.

II – CGC e CCGD: Todos os genes mutados foram analisados em duas bases de dados: *Cancer Gene Census* (CGC; https://cancer.sanger.ac.uk/census) e *Candidate Cancer Gene Database*

(CCGD) (29). O primeiro é um banco de dados curado onde genes com literatura relacionada ao câncer são catalogados. Os genes são divididos em *Tier 1* e *Tier 2*, sendo a primeira classificação referente a genes que já possuem forte literatura de sua relevância em relação ao câncer, a segunda é dada a genes cuja relação com câncer é mais recente ou menos robusta na literatura. Nós optamos por usar ambas as classificações para a seleção de genes. O segundo banco (CCGD) contém dados sobre múltiplos estudos que investigam a contribuição de certos genes para a tumorigênese de camundongos, com base em estudos de inserção mutacional de *transposons*. Para o último banco de dados, examinamos apenas os genes com classificação mais alta (classificação A), a qual se refere a genes que possuem maior estimativa de serem relevantes para a progressão tumoral. Genes que eram CGC e / ou CCGD (classificação A) foram selecionados.

III – Variantes causadoras de proteínas truncadas: Também selecionamos genes que tinham variantes que geram proteínas truncadas (variantes *frameshift* e *nonsense*) e / ou variantes de sítio de *splicing* com no máximo 3 bases de distância da região codificante (também referida como sítio de *splicing* canônico).

IV - Frequência em câncer de mama: utilizamos o banco de dados *Catalogue of Somatic Mutations in Cancer* (COSMIC) para analisar e filtrar genes com alterações e previamente filtrados (critérios acima), de acordo com sua frequência em câncer de mama. Todos os genes com frequência igual ou superior a 1% foram selecionados. Os scripts para cálculo de frequência foram escritos em R conforme as instruções do COSMIC (https://cancer.sanger.ac.uk/cosmic/help/faq).

IV – Literatura: Após a filtragem de frequência, os genes selecionados foram pesquisados no Pubmed utilizando os seguintes termos: “*Nome do gene*” *AND* “breast cancer” e “*Nome do gene*” *AND “cancer”*. De acordo com estes termos pesquisados, genes que apresentavam literatura robusta em câncer e/ou câncer de mama foram selecionados. Note que os genes CGC não foram filtrados pela literatura, uma vez que sua classificação já segue esse processo.

V – *FLAGS (FrequentLy mutAted GeneS)*: A filtragem final foi feita com base no artigo publicado por Shyr e colegas, no qual os autores catalogam genes que codificam grandes proteínas e/ou com muitos parálogos, ou menor pressão evolucionária, sendo estes nomeados *FLAGS*. Tais genes são de difícil interpretação clínica e possuem alta frequência de variantes em diversos tipos de câncer, além disso, por terem grande sequência codificante, seu sequenciamento pode se mostrar desafiador. Os principais genes da lista classificados como FLAGS e com pouca literatura foram excluídos (30).

VI – Genes frequentemente afetados em pacientes adultas jovens com câncer de mama triplo-negativo: nós também compilamos todas as variantes dos pacientes detectados em nosso levantamento na literatura e extraímos os genes mais frequentemente afetados desta coorte. Os genes selecionados foram processados como indicado nos itens anteriores.

2.2 Seleção de pacientes

Pacientes foram incluídos de forma prospectiva no Instituto do Câncer do Estado de São Paulo (ICESP), São Paulo, Brasil, entre os anos de 2016 e 2021.O número de 40 pacientes foi escolhido para o estudo, por conveniência. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo (CAAE: 54689316.9.0000.006).

Os critérios de inclusão foram: pacientes adultas jovens, do sexo feminino, diagnosticadas com câncer de mama triplo-negativo entre 18 e 40 anos de idade. Pacientes que aceitaram participar do estudo e estavam dentro dos critérios de inclusão assinaram o Termo de Consentimento Livre Esclarecido (TCLE).

2.3 Coleta e extração de amostras

Pacientes que aceitaram participar do estudo tiveram suas respectivas amostras de sangue coletadas e sua respectiva amostra tumoral fixada em formalina e emblocada em parafina (*Formalin-Fixed Paraffin-Embedded* - FFPE) coletada no setor de patologia.

2.3.1 Sangue Periférico

Para cada paciente, uma amostra de sangue foi coletada por enfermeiros do Biobanco de Amostras do Centro de Investigação Translacional em Oncologia. Um total de 8 ml de sangue periférico foi coletado por meio de punção da veia do antebraço em cinco tubos com ácido etilenodiamino tetra-acético (EDTA), de 2 ml cada. As amostras foram transportadas em caixa de isopor com gelo e posteriormente processadas no laboratório para separação do *Buffy coat*.

2.3.2 Material Parafinado

O material parafinado foi recuperado de biópsia ou produto da cirurgia já realizada previamente pela paciente, sem comprometer seu diagnóstico e tratamento.

2.3.3 Extração do DNA de material parafinado

Para cada amostra, uma lâmina corada com HE foi analisada pelo médico-patologista e a região de interesse (com ≥30% de células tumorais) foi demarcada. Com auxílio do Micrótomo (Zeiss), foram obtidos 5 cortes de 20 µm. Utilizando como guia a lâmina de HE previamente analisada, a região de interesse foi raspada e armazenada em microtubos de 1,5 ml. O DNA foi extraído por meio do *QIAamp® DNA FFPE Tissue kit* (Qiagen - 56404). Para a desparafinização do tecido tumoral, foi adicionado 1 ml de xilol (100%) e agitado em vórtex por 10 segundos. Em seguida, as amostras foram centrifugadas a 20.000 g por 2 minutos. Os tubos foram invertidos para a retirada de todo xilol, sendo adicionado 1 ml de etanol (100%) em cada tubo, que a seguir, foram agitados e centrifugados a 20.000g por 2 minutos. Em seguida, os tubos foram invertidos e permaneceram com a tampa aberta à temperatura ambiente por até 30 minutos, para evaporação de todo etanol. A próxima etapa do processo baseia-se na digestão das proteases. Para isso, foram adicionados Buffer ATL e Proteínase K, seguida de homogeneização e incubação a 56 °C por 4 horas e a 90 °C por 30 minutos. Para garantir a remoção de RNA, foi adicionado RNase A (100 mg/µl) em cada tubo, incubando-os por 2 minutos. Foram adicionados *Buffer AL* e etanol (96-100%) e o volume total foi cuidadosamente transferido para colunas. As colunas foram centrifugadas e o precipitado foi descartado. O mesmo processo foi realizado

com o *Buffer AW1* e *Buffer AW2*, onde novamente o precipitado foi descartado. Para remoção de qualquer resíduo, as colunas foram centrifugadas a 20.000g por 3 minutos. Por fim, as colunas foram transferidas para tubos de 1,5 ml, onde foram adicionados 30 µl de *Buffer ATE*, incubadas por 5 minutos a temperatura ambiente e centrifugadas a 20.000g por 1 minuto. As colunas foram removidas e as amostras armazenadas a -20 °C. Uma alíquota de cada amostra foi avaliada quantitativa e qualitativamente por *Nanodrop 1000 Spectrophotometer* e *Qubit (Thermo Fisher Scientific Inc.)*.

2.3.4 Extração de DNA das Células Mononucleares

O DNA foi extraído por meio do *QIAamp® DNA Mini Kit (Qiagen – 51306).* Para cada microtubo de 1,5 ml foram adicionados tampão de digestão contendo proteínase K seguido de incubação a 56 °C por 3 horas. Após adição do *Buffer AL* e incubação a 70 °C por 10 minutos, foi adicionado etanol (96-100%), a mistura foi transferida para as colunas, as quais foram centrifugadas. Após lavagem com *Buffer AW1* e *Buffer AW2*, as colunas foram transferidas para microtubos de 2 ml, centrifugadas e a seguir transferidas para microtubos de 1,5 ml, onde as amostras foram eluidas com 50 µl de água ultrapura. Após incubação por 4 minutos e centrifugação a 6.000 g por 1 minuto, as colunas foram removidas e as amostras armazenadas a -20 °C. Uma alíquota de cada amostra foi avaliada quantitativa e qualitativamente por *Nanodrop 1000 Spectrophotometer* e *Qubit (Thermo Fisher Scientific Inc.)*.

2.4 Construção de biblioteca de DNA

Amostras pareadas de sangue e FFPE que foram extraídas com sucesso e em uma concentração mínima de 10 ng/uL seguiram para o protocolo de preparação de biblioteca de DNA para sequenciamento. O protocolo detalhado da preparação de tais amostras podem ser visualizado no **APÊNDICE A**.

Por fim, as bibliotecas foram combinadas para atingir a mesma equimolaridade (10 nM) em tampão *Low TE* 1x, e foram então sequenciadas no aparelho NextSeq (*NextSeq 500/550 Mid Output Kit* v2.5, 150 ciclos; Illumina) do laboratório de Sequenciamento em Larga Escala (SELA) da Rede Multiusuários do Sistema FMUSP/HC, da Universidade de São Paulo (USP).

2.5 Controle de Qualidade (QC) e processamento de dados

2.5.1 Controle de qualidade durante e após sequenciamento

Durante e após o processo de sequenciamento, por meio da interface da ferramenta BaseSpace (Illumina), diretamente ligada ao equipamento de sequenciamento, verificamos a qualidade da formação de cluster total, por *lanes* e o escore “Q30”, que se refere à estimativa de probabilidade de quantas bases foram chamadas de forma incorreta. Uma pontuação de 30 refere-se a uma probabilidade de 0,01% de erro e quanto maior o score, menor a probabilidade de erro, portanto, é desejável que no mínimo 80% das leituras possuam uma pontuação mínimo de 30. Quanto aos clusters, é desejável que possuam distribuição similar entre as *lanes*, tanto nos ciclos em *forward* quanto em *reverse.* Para o equipamento (NextSeq; Illumina) e o kit de reagentes utilizado (*NextSeq 500/550 Mid Output Kit* v2.5, 150 ciclos) é desejável que a densidade de clusters (K/mm²) fique entre 170 e 220. Uma baixa densidade de clusters indica uma baixa quantidade de DNA, portanto, uma provável baixa cobertura, enquanto uma densidade exacerbada pode resultar na falha de captura da emissão de sinais de fluorescência pela câmera do equipamento.

2.5.2 Processamento de dados após sequenciamento

A trimerização, para remoção de adaptadores e bases de baixa qualidade (*quality trimming*), o alinhamento com genoma referência (BWA MEM; hg19), remoção de duplicatas, chamada de CNV (*Copy Number Variation*) e chamada de variantes foram realizadas por meio do software SureCall (v.4.2.2; Agilent), usando suas configurações base.

2.5.3 Controle de qualidade após alinhamento e chamada de variante

Complementamos a chamada de variante, submetendo os arquivos de chamada de variante (.vcf) ao anotador *Variant Effect Predictor* (VEP, *GRCh37 Archive browser 106; Ensembl*) para dados complementares de ferramentas de predição de efeito de alterações (seção 2.5.4) e frequência alélica populacional (GnomAD v.2.1.1.).

Como controle de qualidade, escolhemos o valor mínimo de 20 leituras (*reads*) totais e relação *reads* totais/*reads* alterados de no mínimo 10%, por variante. Para verificar a qualidade global de *probes* e amostras, mapeamos a

cobertura dos exons de cada gene do painel, por paciente, calculando a mediana de cobertura entre as *probes* de um paciente ou de um gene em todos os pacientes, extraindo assim a estimativa de qualidade global de *probes* e amostras. Consideramos “amostra de qualidade baixa” se a mediana de frequência de todas as *probes* com menos de 20 *reads* em uma mesma amostra era igual ou superior a 20% e “*probe* de baixa qualidade” se a mediana de frequência de uma *probe* com menos de 20 *reads* era igual ou maior que 15% entre todas as amostras. Amostras e *probes* que foram classificadas como de baixa qualidade não foram incluídas nas análises finais.

Como a variabilidade de cobertura e qualidade de regiões-alvo pode variar muito entre amostras (e regiões) nós também criamos uma matriz de qualidade individualizada de exons, de cada gene, por paciente. Todas as variantes detectadas foram verificadas em relação à qualidade de cobertura conforme a paciente, gene e exon que estavam localizadas. Um exemplo da tabela de qualidade pode ser visualizado na **Tabela 1**, que demonstra a porcentagem de região (de cada exon) com menos de 20 *reads* de cobertura. Por exemplo, uma variante detectada no exon 5 do gene *AHNAK*, na amostra tumoral da paciente 701 não seria classificada como de baixa qualidade, mas teria que ser interpretada com cautela. A tabela completa possui 2400 linhas e 95 colunas, e por isso foi disponibilizada de forma integral apenas na versão digital deste manuscrito.

Tabela 1 – Exemplo da tabela de qualidade individualizada de exons, representando os 9 exons do gene *AHNAK*, em 12 amostras tumorais. Os valores numéricos apresentados para cada exon representa a porcentagem de cada região (exon) com cobertura abaixo de 20 *reads*.

Transcrito	ID	exon_1	exon_2	exon_3	exon_4	exon_5	exon_6	exon_7	exon_8	exon_9
AHNAK_NM_001346446...199	605_FFPE	0	0	0	0	2.22	NA	NA	NA	0
AHNAK_NM_001346446...219	608_FFPE	0	0	0	0	8.96	0	0	0	0
AHNAK_NM_001346446...258	611_FFPE	0	0	0	0	9.61	0	0	0	0
AHNAK_NM_001346446...300	635_FFPE	0	0	0	0	1.78	NA	NA	NA	0
AHNAK_NM_001346446...322	700_FFPE	0	0	0	0	1.5	NA	NA	NA	0
AHNAK_NM_001346446...344	701_FFPE	0	0	0	0	13.05	0	0	0	0
AHNAK_NM_024060...345	701_FFPE	0	0	0	0	0	4.58	0	0	0
AHNAK_NM_001346446...455	702_FFPE	0	0	0	0	9.35	0	0	0	0
AHNAK_NM_001346446...539	703_FFPE	0	0	0	0	2.64	NA	NA	NA	0
AHNAK_NM_001346446...559	715_FFPE	0	0	0	0	8.56	0	0	0	0
AHNAK_NM_001346446...598	718_FFPE	0	0	4.58	0	17.5	0	0	0	0
AHNAK_NM_024060...599	718_FFPE	0	0	0	0	1.01	9.74	0	0	0
AHNAK_NM_001346446...735	719_FFPE	0	0	0	0	2.51	0	NA	NA	0
AHNAK_NM_001346446...1	728_FFPE	0	0	0	0	0.93	NA	NA	NA	0
AHNAK_NM_001346446...13	730_FFPE	0	0	0	0	2.66	NA	NA	NA	0

Todas as variantes não-raras (polimorfismos) com frequência alélica (GnomAD) igual ou superior a 1% foram excluídas. Após essa exclusão, verificamos todas as variantes individualmente em relação à sua qualidade de mapeamento, verificando variantes detectadas em sequências de baixa complexidade e/ou em áreas de sequência comum em múltiplos sítios do genoma, significando um possível falso-positivo. Essa verificação foi feita tanto de forma automática, verificando os escores de qualidade de mapeamento, extraídos por meio do software Surecall, como visualmente, por meio de um visualizador de genoma.

2.5.4 Dados complementares em variantes

Todos os genes e variantes detectados foram classificados de acordo com seu potencial como carreador (*driver*). Consideramos como *driver*, genes que estivessem catalogados no *Cancer Gene Census* (CGC; COSMIC v.96; https://cancer.sanger.ac.uk/census) e que tivessem predição de patogenicidade em ferramentas de predição de efeito de variante (detalhado no fim desta seção) e/ou que estivessem catalogados em portais especializados (detalhado nos parágrafos a seguir). Os genes foram classificados como oncogene, gene supressor de tumor (TSG, do inglês *Tumor Supressor Gene*) ou gene de dupla função (oncogene e TSG), seguindo a classificação do CGC.

Os bancos de dados OncoKB (https://www.oncokb.org/; acessado em janeiro de 2022) e TP53 Database (https://tp53.isb-cgc.org/; acessado em janeiro de 2022) foram usados para classificar as variantes somáticas de acordo com seu potencial como drivers, conforme relatado na literatura por meio de estudos funcionais. Complementarmente, nós usamos a ferramenta *CancerVar,* publicada recentemente (https://cancervar.wglab.org/), para classificar a patogenicidade de variantes somáticas (31).

O banco de dados OncoKB contém informação coletada por meio de curadoria de literatura, com evidências clínicas e biológicas do impacto funcional de variantes e genes afetados em diversos tipos de câncer, reportando se a variante ou gene sob investigação tem efeitos oncogênicos e efeitos de perda ou ganho de função. Todas as variantes catalogadas são classificadas em relação ao seu efeito nos genes, como: causadora de

provável perda de função (LLOF, do inglês *likely loss-of-function*), perda de função (LOF, do inglês *loss-of-function*), provável ganho de função (LGOF, do inglês *likely gain-of-function*) ou ganho de função (GOF, do inglês gain-of-function) (32).

O banco de dados TP53 Database compila resultados de estudos que investigaram variantes de TP53 através de ensaios funcionais quantitativos com levedura, biblioteca de plasmídeos para inserção de variantes em *TP53* em células humanas e ensaio de saturação CRISPR-CAS9, também em células humanas (33–35)

CancerVar se trata de uma ferramenta desenvolvida para interpretação de variantes somáticas (pontuais, estruturais, *indels* e CNVs). No estudo, os autores desenvolveram uma classificação de variantes somáticas baseada em diretrizes clínicas, similar à metodologia inserida pela ACMG para classificação de variantes germinativas, juntamente com aplicação de aprendizado de máquina (31).

Variantes *missense* somáticas foram avaliadas quanto ao seu possível efeito funcional por meio de ferramentas *in sílico*. Para isso, utilizamos ferramentas *in sílico* de previsão de impacto, além de ferramentas que utilizam metodologias de compilação de ferramentas de previsão de impacto funcional de variantes, sendo essas:

- Ferramentas de predição de impacto funcional de variantes: FATHMM (36), MutationAssesor (37), MutationTaster (38), PROVEAN (39), Polyphen2 HDIV, Polyphen2 HVAR (40), SIFT e SIFT4G (41).

- Ferramentas de compilação: Definimos ferramentas de compilação como ferramentas de utilizam métodos de integração de dados de múltiplas ferramentas de predição de impacto funcional de variantes, como as citadas acima. As ferramentas utilizadas no presente trabalho foram: REVEL (42), METALR e METASVM (43) e MCAP (44).

Classificamos as variantes *missense* como possivelmente patogênicas (PP) caso elas tivessem predição de patogenicidade em ao menos 4 de 8 ferramentas de predição e 2 de 4 ferramentas de compilação ou em pelo

menos 2 de 8 das ferramentas de predição e 1 de 4 ferramentas de compilação, caso o gene em que ela se encontrasse estivesse catalogado como um gene CGC.

Variantes somáticas geradoras de proteínas truncadas (*frameshift*, *nonsense* ou em sítio de *splicing* canônico) foram consideradas provavelmente patogênicas.

As variantes germinativas foram classificadas como benignas, provavelmente benignas, variantes de significado incerto (VUS, do inglês *Variant of Uncertain Significance*), provavelmente patogênica ou patogênica, seguindo sua classificação no Clinvar e aplicando a classificação sugerida pelo Colégio Americano de Genética Médica (ACMG). Além disso, os dados relatados no estudo publicado por Findlay e colegas foram contabilizados ao aplicar os critérios ACMG para classificação das variantes germinativas no gene *BRCA1* detectadas nas pacientes do presente estudo (45)

2.5.5 Copy Number Variation (CNV)

CNVs somáticas foram detectadas por meio do software SureCall, onde amostras tumorais foram usadas como amostra de investigação, e suas respectivas amostras pareadas de sangue foram usadas como amostra de referência. Para o cálculo de CNVs, o programa considera a cobertura de *reads* de cada intervalo de regiões alvo dos genes do painel (nesse caso, seus exons), computando e normalizando os valores de cobertura entre os pares de amostras. Como *output*, obtemos o gene e a coordenada da região onde ele se encontra com amplificação ou deleção, sendo esse efeito representado pelo valor de razão entre amostras pareadas, além disso, o tamanho da região que representa a CNV também é computado.

Para a chamada de CNVs, usamos as configurações-base de filtragem e controle de qualidade do software, mas também aplicamos alguns controles adicionais para garantir o menor número de falsos-positivos possível, alguns deles inspirados em métodos publicados para controle de qualidade e filtragem de CNVs somáticas e germinativas (46–49). Como primeiro fator, escolhemos apenas variantes com escore de confiança (probabilidade de ser um achado

verdadeiro-positivo) mínimo de 0,97 (97%). A seguir, apenas regiões com no mínimo 20 *reads* de cobertura foram aceitas.

Nesta metodologia, CNVs são inferidas a partir da razão de valores de cobertura de amostra em estudo (tecido tumoral) e amostra referência (sangue). Porém, o principal empecilho nessa condição de comparação está na grande diferença de qualidade entre a amostra controle, que se encontra mais íntegra, portanto, geralmente, sequenciada com maior cobertura, e as amostras tumorais. Diante deste desafio, para controle de possíveis falso-positivos, usamos pontos de cortes personalizados para inferir ganho e deleção, conforme a razão de cobertura mediana global de regiões-alvo das amostras tumorais em relação às amostras referências, e vice-versa.

Semelhantemente ao que foi descrito anteriormente na seção 2.5.3 (métodos), verificamos a variabilidade de cobertura e qualidade de regiões-alvo dos exons onde as CNVs estavam localizadas, por meio da matriz de qualidade individualizada de exons **(Tabela 1)**. CNVs em regiões com baixa qualidade foram classificadas como achados falso-positivos.

2.5.6 Análise exploratória de dados tumorais públicos de pacientes com câncer de mama na pré-menopausa e pacientes de idade mais avançada

A ferramenta Cancer Browser do COSMIC (https://cancer.sanger.ac.uk/cosmic/browse/tissue; v.96; acessado em maio de 2022) e as ferramentas de comparação de coorte do cBioPortal (https://www.cbioportal. org/; acessado em junho de 2022) foram usadas para realizar análises exploratórias, integrando dados de vários estudos de sequenciamento de exoma e genoma de tumores de pacientes com câncer de mama, de modo a comparar dados genômicos entre adultos jovens (diagnosticados até 40 anos) e adultos mais velhos, com 5 faixas etárias de diagnóstico (≤40, 41-50, 51-60, 61-70 e >70).

Os dados do COSMIC foram baixados diretamente do portal (https://cancer.sanger.ac.uk/cosmic/download; acessado em maio de 2022) e processados em R (v. 4.1.2). Os dados do cBioPortal foram analisados principalmente por suas ferramentas de comparação de coorte. Os valores de p obtidos pela análise de dados do COSMIC (Teste Exato de Fisher) foram

ajustados (p.aj.) com a correção de *Bonferroni*, enquanto os resultados obtidos pela análise integrada do cBioPortal (através dos testes *Kruskal-Wallis*, Qui-Quadrado e *T-Student*) foram corrigidos com o procedimento de *Benjamini-Rochberg*.

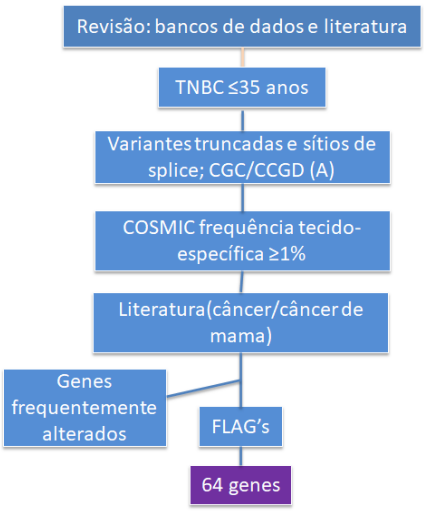
Como a maioria dos estudos não tinha CNV e/ou expressão gênica e/ou dados de expressão proteica, esses tipos de dados foram investigados apenas na coorte de câncer de mama do TCGA (*TCGA Breast Cancer Firehose Legacy Cohort*).

3 Resultados

3.1 Painel personalizado de genes

Foram selecionados 39 genes, utilizando os critérios descritos nos métodos. Foram adicionados outros 25 genes, com base em resultados de análise prévia de amostras de câncer de mama de pacientes adultas jovens (50) **(Figura 1)**. Os 64 genes escolhidos para o painel foram representados por um total de 6928 *probes*, com tamanho total de 489 kbp, representando todas as regiões codificantes dos genes selecionados, além de 10 bases de distância das regiões de extremidade 3' e 5', baseados na sequência referência do genoma humano (GRCh37).

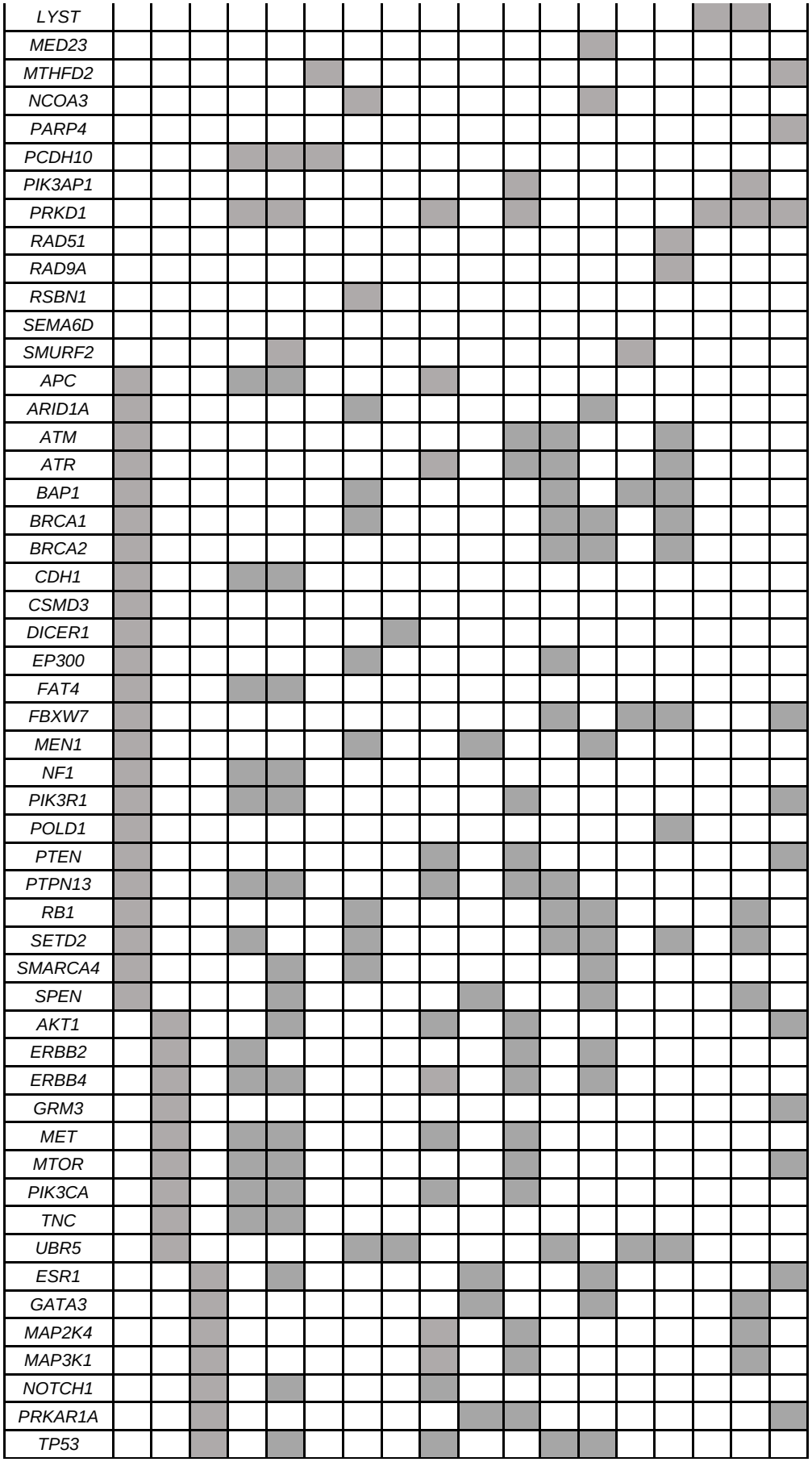
Figura 1 – Esquema para a seleção de genes para sequenciamento-alvo.



A tabela a seguir **(Tabela 2)** contém um resumo das principais funções em que os genes selecionados estão envolvidos. Para isso, usamos o banco de dados GeneCard (https://www.genecards.org/) para coletar informação de principais vias e funções destes genes. A literatura específica de tais genes com sua correlação com câncer de mama e/ou câncer em geral não foi utilizada para a montagem da tabela, pois traria uma classificação de funções muito ampla, dependendo do gene. Os aspectos individuais dos genes e sua relação com câncer serão discutidos mais adiante. Em resumo, 36% (n=23) dos genes do painel são classificados como TSGs (de acordo com a classificação do CGC), 14% (N=9) como oncogenes e 11% (n=7) como função dupla (TSG e oncogene). Os demais genes (n=25) não possuíam classificação no *Cancer Gene Census* (oncogene, gene supressor tumoral, ou ambos), mas possuíam literatura relevante em câncer e/ou alta frequência de alteração em câncer de mama geral e/ou relevância em câncer de mama em mulheres adultas jovens. Cerca de 1/3 dos genes estão envolvidos com migração e citoesqueleto celular e cerca de 1/4 dos genes estão envolvidos com quinases, fatores de transcrição, ciclo celular e morte celular.

Tabela 2 – Resumo das principais funções dos 64 genes selecionados em nosso painel personalizado. TSG: gene supressor tumoral; OG: oncogene; DUAL: função dupla (TSG e OG); RNA: gene que participa de vias de processamento ou controle de RNA; FT: gene codifica fator de transcrição (FT) ou molécula que interage com FTs; Expressão gênica: gene que participa do controle de expressão gênica.

Gene	TSG	OG	DUAL	Citoesqueleto celular	Migração	Canais Iônicos	Cromatina/Histona	RNA	Morte celular	Modulação de expressão Gênica	Atividade de quinase	Ciclo celular	FT	Ubiquitina	Reparo de DNA	Vesículas	Imunidade	Metabolismo
AHNAK				■	■	■												
ANLN				■	■													
ATAD2B							■											
ATXN1							■	■										
CACNA1E				■	■	■			■	■								
CAMK1G						■					■							
CSPP1				■								■						
DALRD3								■										
FAT2				■	■													
GRHL2							■						■					
HERC2														■	■			
HUWE1									■					■	■			


LYST
MED23
MTHFD2
NCOA3
PARP4
PCDH10
PIK3AP1
PRKD1
RAD51
RAD9A
RSBN1
SEMA6D
SMURF2
APC
ARID1A
ATM
ATR
BAP1
BRCA1
BRCA2
CDH1
CSMD3
DICER1
EP300
FAT4
FBXW7
MEN1
NF1
PIK3R1
POLD1
PTEN
PTPN13
RB1
SETD2
SMARCA4
SPEN
AKT1
ERBB2
ERBB4
GRM3
MET
MTOR
PIK3CA
TNC
UBR5
ESR1
GATA3
MAP2K4
MAP3K1
NOTCH1
PRKAR1A
TP53

3.2 Amostras e pacientes

3.2.1 Características das pacientes

As 15 pacientes foram obtidas de pacientes com mediana de idade ao diagnóstico de 36 anos (min: 26; máx: 40) e 40% (6/15) das pacientes possuíam alguma história familiar (parentesco de até 3º grau) de câncer de mama, ovário, próstata ou pâncreas e todas possuíam carcinoma ductal invasivo. A tabela a seguir detalha as informações clínicas das pacientes analisadas no presente estudo **(Tabela 3)**.

Tabela 3 – Dados clínicos das pacientes sequenciadas no estudo. Hist: tipo histológico; CDI: carcinoma ductal invasivo; Cel: celularidade tumoral; HF: história familiar, com história familiar positiva sendo ao menos um parente (parentesco de até 3º grau) com diagnóstico de câncer de mama, ovário, próstata ou pâncreas.

ID	Idade	Hist	Grau	Cel	TNM	HF
605	38	CDI	3	0.3	T2N0	n
608	37	CDI	3	0.3	T4N0	n
610	35	CDI	2	0.3	T1N1	n
611	37	CDI	3	0.3	T3N0	s
635	39	CDI	3	0.4	T2N0	n
700	33	CDI	3	0.8	T2N0	s
701	39	CDI	3	0.6	T3N1	n
702	39	CDI	2	0.3	T3N1	n
703	26	CDI	2	0.4	T3N1	s
712	29	CDI	2	0.4	T4N2	n
715	30	CDI	2	0.3	T4N1	s
718	34	CDI	3	0.7	T3N2	s
719	36	CDI	2	0.7	T4N1	n
728	37	CDI	3	0.6	T3N1	n
730	27	CDI	2	0.9	T3N3	s

3.2.2 Característica das amostras

Incluímos 40 participantes com diagnóstico de câncer de mama entre 18-40 anos, atendidas no ICESP. Todas as pacientes receberam informação sobre os objetivos do estudo e assinaram o Termo de Consentimento Livre e Esclarecido (TCLE). Uma amostra de sangue foi obtida e o bloco de parafina foi solicitado ao serviço de patologia. A extração de DNA das amostras pareadas de tumor fixado em formalina e emblocado em parafina e sangue periférico (*buffy coat*) foi realizada. Dessas, 22 amostras pareadas (22 FFPE + 22 sangue) tiveram qualidade boa o suficiente para prosseguir para a

construção da biblioteca de DNA. A seguir, 7 amostras pareadas (7 FFPE + 7 sangue) foram excluídas durante a construção da biblioteca, pois seu DNA não foi recuperado na etapa de amplificação. Por fim, 15 amostras pareadas (15 amostras tumorais + 15 amostras de sangue) foram sequenciadas **(Tabela 4)**.

Tabela 4 – Concentração e volume final de amostras sequenciadas. FFPE: amostras tumorais fixadas em formalina e emblocadas em parafina; SG: amostras de sangue; C: concentração.

	FFPE	SG	FFPE	SG	FFPE (volume final 50-53uL)		SG (volume final 50-53uL)	
ID	C ng/uL	C ng/uL	Diluição ng/uL (15uL)	Diluição ng/uL (15uL)	Amostra (uL)	Low TE (uL)	Amostra (uL)	Low TE (uL)
605	225	80.8	72.1	NA	2.8	50.2	2.5	50.5
608	15.3	57.4	NA	NA	14	36	5	45
610	41.4	117	NA	NA	6	44	2	48
611	62.4	103	NA	NA	5	45	2	48
635	40	118	NA	NA	6	44	2	48
700	20.8	177	NA	69.8	9.5	43.5	3	50
701	9.45	355	NA	110	21.1	31.9	2	51
702	127.2	103	NA	NA	11	39	2	48
703	11.2	195	NA	166	18	32	2	48
712	31.9	17.2	NA	NA	6.25	46.75	11.5	41.5
715	24.3	27.1	NA	NA	8.2	44.8	7.4	45.6
718	8.07	92	NA	NA	25	28	2	51
719	261	84.4	74.5	NA	2.7	50.3	2.3	50.7
728	13.7	65	NA	NA	15	35	3	47
730	147	63.9	70	NA	4.3	45.7	4.3	45.7

3.2.3 Exclusões após sequenciamento

Após o sequenciamento, 1 amostras de FFPE foi excluída pois possuía baixa qualidade global **(Tabela 5)** (amostras cuja mediana de frequência de todas as *probes* com menos de 20 *reads* em uma mesma amostra era igual ou superior a 20%, como descrito na metodologia), e 2 amostras adicionais de FFPE foram excluídas, pois não foram sequenciadas com sucesso, resultando em 12 amostras tumorais (FFPE) e 15 de sangue, processadas e analisadas como descrito na metodologia **(Tabela 6)**.

Tabela 5 – Qualidade global de amostras e *probes.* Amostras de baixa qualidade (mediana de frequência de todas as *probes* com menos de 20 *reads* em uma mesma amostra era igual ou superior a 20%) e *probes* de baixa qualidade (mediana de frequência de uma *probe* com menos de 20 *reads* era igual ou maior que 15% entre todas as amostras) foram demarcadas em vermelho. Valores acima do limiar mínimo de qualidade global (15%) foram demarcados em rosa, enquanto valores abaixo foram demarcados em verde. MED: mediana.

Genes	605	608	611	635	700	701	702	703	715	718	719	728	730	MED
AHNAK	0.2	1.0	1.1	0.2	0.2	2.0	1.0	0.3	1.0	3.6	0.3	0.1	0.3	0.3
AKT1	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	2.2	5.7	0.0	0.0	0.0	0.0
ANLN	0.6	3.0	2.8	2.1	1.8	31.4	16.4	1.2	1.1	45.6	2.4	0.0	0.0	2.1
APC	0.0	0.0	0.0	0.0	0.0	27.2	5.4	0.0	1.3	73.9	0.0	0.0	0.0	0.0
ARID1A	0.0	0.0	0.2	0.1	0.0	1.9	0.1	0.1	0.2	6.3	0.0	0.0	0.0	0.1
ATAD2B	0.0	0.2	0.1	0.0	0.0	36.8	3.8	0.0	0.0	39.4	0.0	0.0	0.0	0.0
ATM	0.0	1.2	1.3	0.2	0.0	51.8	13.0	0.0	1.0	70.3	0.1	0.0	0.0	0.2
ATR	0.0	0.2	0.0	0.0	0.0	41.1	8.6	0.0	0.1	30.5	0.0	0.0	0.0	0.0
ATXN1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0
BAP1	0.0	0.0	0.0	0.0	0.0	0.4	1.0	0.0	0.0	9.1	0.0	0.0	0.0	0.0
BRCA1	0.0	0.5	2.8	0.0	0.0	43.3	7.1	0.0	2.7	59.0	0.0	0.0	0.0	0.0
BRCA2	0.0	0.1	0.2	0.0	0.0	59.6	21.0	0.0	0.4	75.1	0.0	0.0	0.0	0.0
CACNA1E	0.0	0.0	0.0	0.0	0.0	3.8	0.1	0.0	0.0	12.3	0.0	0.0	0.0	0.0
CAMK1G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	21.3	0.0	0.0	0.0	0.0
CDH1	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.1	18.7	0.0	0.0	0.0	0.0
CSMD3	0.0	1.2	0.5	0.0	0.0	30.5	8.3	0.0	0.0	17.8	0.0	0.0	0.0	0.0
CSPP1	0.0	0.4	1.1	0.0	0.0	26.0	8.0	0.0	0.2	19.2	0.0	0.0	0.0	0.0
DALRD3	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	16.2	0.0	0.0	0.0	0.0
DICER1	0.0	0.2	0.4	0.0	0.0	21.9	2.5	0.0	0.4	32.9	0.0	0.0	0.0	0.0
EP300	0.0	0.0	0.0	0.0	0.0	15.7	2.4	0.0	0.0	29.9	0.0	0.0	0.0	0.0
ERBB2	0.0	0.0	0.0	0.0	0.0	0.3	0.5	0.0	0.3	25.3	0.0	0.0	0.0	0.0
ERBB4	0.0	0.2	1.8	0.0	0.0	15.6	0.8	0.0	0.0	26.2	0.0	0.0	0.0	0.0
ESR1	0.0	0.0	0.0	0.0	0.0	2.0	0.3	0.0	0.0	7.2	0.0	0.0	0.0	0.0
FAT2	0.0	0.0	0.2	0.0	0.0	1.1	0.7	0.0	0.0	29.6	0.0	0.0	0.0	0.0
FAT4	0.0	0.3	0.8	0.0	0.0	18.4	3.7	0.0	0.0	30.2	0.0	0.0	0.0	0.0
FBXW7	0.2	0.3	0.6	0.3	0.2	18.7	2.3	0.1	0.2	39.8	0.2	0.0	0.0	0.2
GATA3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.7	0.0	0.0	0.0	0.0
GRHL2	0.0	0.0	0.0	0.0	0.0	6.3	0.0	0.0	0.0	2.8	0.0	0.0	0.0	0.0
GRM3	0.0	0.3	0.0	0.0	0.0	3.5	0.5	0.0	0.0	12.3	0.0	0.0	0.0	0.0
HERC2	15.3	25.5	28.3	12.8	13.2	37.0	31.0	11.8	24.7	53.0	12.8	5.3	6.4	15.3
HUWE1	0.0	0.0	0.0	0.0	0.0	11.2	1.7	0.0	0.0	66.7	0.0	0.0	0.0	0.0
LYST	0.1	0.3	0.1	0.0	0.0	28.6	3.6	0.3	0.1	29.3	0.1	0.0	0.0	0.1
MAP2K4	0.5	7.0	9.3	2.8	2.6	52.6	18.1	0.2	13.6	73.2	0.0	0.0	2.1	2.8
MAP3K1	0.0	0.0	0.1	0.0	0.0	34.8	6.9	0.0	0.0	75.7	0.0	0.0	0.0	0.0
MED23	0.3	1.6	1.6	1.0	1.3	28.4	4.5	0.4	1.1	39.8	0.0	0.0	0.0	1.1
MEN1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.8	0.0	0.0	0.0	0.0
MET	0.0	0.0	0.0	0.0	0.0	12.2	2.0	0.0	0.0	34.5	0.0	0.0	0.0	0.0
MTHFD2	0.4	1.1	0.7	0.1	0.0	20.5	4.1	0.2	0.7	37.5	0.2	0.0	0.1	0.4
MTOR	0.0	0.0	0.0	0.0	0.0	3.9	0.0	0.0	0.0	27.0	0.0	0.0	0.0	0.0

Genes	605	608	611	635	700	701	702	703	715	718	719	728	730	MED
NCOA3	0.0	0.0	0.0	0.0	0.0	19.9	2.2	0.0	0.0	34.2	0.0	0.0	0.0	0.0
NF1	4.8	7.7	7.8	5.3	3.2	42.2	17.1	4.5	10.6	63.1	4.9	3.3	4.0	5.3
NOTCH1	0.0	0.0	0.2	0.0	0.0	0.6	0.4	0.0	0.2	17.0	0.0	0.0	0.0	0.0
PARP4	18.7	30.5	29.0	16.7	17.1	50.2	38.1	13.9	21.8	61.2	14.1	10.2	9.5	18.7
PCDH10	0.0	0.0	0.0	0.0	0.0	4.1	0.0	0.0	0.1	24.1	0.0	0.0	0.0	0.0
PIK3AP1	0.0	0.8	0.9	1.2	0.8	8.1	0.6	0.0	0.8	37.4	0.0	0.0	0.0	0.8
PIK3CA	9.3	15.0	12.0	5.0	7.0	56.7	20.6	7.7	13.3	46.3	9.5	1.8	0.9	9.5
PIK3R1	0.0	0.0	0.0	0.0	0.0	26.0	9.3	0.0	0.0	52.6	0.0	0.0	0.0	0.0
POLD1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0
PRKAR1A	0.0	0.0	0.0	0.0	0.0	14.8	2.7	0.0	0.0	35.2	0.0	0.0	0.0	0.0
PRKD1	0.0	0.0	0.0	0.0	0.0	23.7	6.0	0.0	1.7	56.1	0.0	0.0	0.0	0.0
PTEN	8.6	15.1	28.6	9.3	9.6	68.4	48.1	7.0	18.4	75.7	9.5	5.5	5.5	9.6
PTPN13	0.0	0.1	0.9	0.0	0.0	38.0	5.5	0.0	0.8	37.1	0.0	0.0	0.0	0.0
RAD51	0.8	0.8	2.0	0.0	1.1	14.1	1.9	0.9	0.0	36.9	0.3	0.1	0.0	0.8
RAD9A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	0.0	0.0	0.0	0.0
RB1	3.6	5.2	9.4	3.8	3.6	75.2	32.3	3.6	4.5	87.3	3.6	3.6	0.0	3.8
RSBN1	0.0	0.0	0.0	0.0	0.0	18.1	1.8	0.0	0.0	46.7	0.0	0.0	0.0	0.0
SEMA6D	0.0	0.1	4.2	0.0	0.0	22.1	3.8	0.0	0.0	45.1	0.0	0.0	0.0	0.0
SETD2	0.0	0.0	0.0	0.0	0.0	16.8	5.0	0.0	0.0	24.6	0.0	0.0	0.0	0.0
SMARCA4	0.0	0.0	0.0	0.0	0.0	1.3	0.1	0.0	0.0	5.3	0.0	0.0	0.0	0.0
SMURF2	2.8	4.2	4.5	2.3	2.1	42.7	15.2	1.8	3.5	52.0	2.2	0.6	1.3	2.8
SPEN	0.0	0.0	0.0	0.0	0.0	9.7	4.0	0.0	0.0	29.1	0.0	0.0	0.0	0.0
TNC	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.0	0.0	5.6	0.0	0.0	0.0	0.0
TP53	0.0	0.0	0.0	0.0	0.0	14.8	2.1	0.0	0.0	39.4	0.0	0.0	0.0	0.0
UBR5	0.0	0.0	0.0	0.0	0.0	20.0	3.6	0.0	0.0	10.7	0.0	0.0	0.0	0.0
MED	0.0	0.0	0.0	0.0	0.0	16.3	2.4	0.0	0.0	30.3	0.0	0.0	0.0	

Tabela 6 – Resumo de qualidade de amostras tumorais e sangue após sequenciamento. Mediana *reads*: mediana de *reads* por região sequenciada em amostras tumorais e sangue; Resultado: qualidade da amostra após sequenciamento, com "OK" representando sucesso, FALHA_SEQ representando falha ao sequenciar e BAIXA QUAL representando exclusão de amostra em análises posteriores devido à baixa qualidade global, como descrito em métodos. As duas últimas colunas representam a divisão (razão) da mediana de *reads* de amostras tumorais contra suas respectivas amostras pareadas de sangue, e vice-versa.

	Tumor (FFPE)		Sangue		Razão	
ID	Mediana *reads*	Resultado	Mediana *reads*	Resultado	FFPE/SG	SG/FFPE
605	154	OK	344	OK	0.44767442	2.23376623
608	67	OK	390	OK	0.17179487	5.82089552
610	3	FALHA_SEQ	361	OK	0.00831025	120.333333
611	66	OK	456	OK	0.14473684	6.90909091
635	180	OK	352	OK	0.51136364	1.95555556
700	194	OK	398	OK	0.48743719	2.05154639
701	33	OK	342	OK	0.09649123	10.3636364
702	55	OK	338	OK	0.16272189	6.14545455
703	199	OK	125	OK	1.592	0.6281407
712	3	FALHA_SEQ	294	OK	0.01020408	98
715	78	OK	350	OK	0.22285714	4.48717949
718	17	BAIXA QUAL	258	OK	0.06589147	15.1764706
719	196	OK	361	OK	0.54293629	1.84183674
728	273	OK	167	OK	1.63473054	0.61172161
730	337	OK	92	OK	3.66304348	0.27299703
Mediana	167	-	344	-	-	-

Dois genes foram excluídos de análises posteriores, sendo esses *HERC2* e *PARP4*, pois estavam com baixa qualidade em todas as amostras analisadas **(Tabela 5)**, demonstrando um provável problema em suas respectivas *probes.*

Para análises posteriores, excluímos uma variante *missense* somática no gene *DICER1*, e outra em *RAD51* na amostra tumoral da paciente 701, devido às mesmas se encontrarem em região de baixa complexidade, além de outra variante *missense* somática em *PIK3CA*, devido à sua baixa qualidade de mapeamento. De modo similar, 2 variantes *missense* somáticas em *AHNAK* e *RAD51* na amostra tumoral da paciente 702 também foram excluídas, pois se encontravam com baixa qualidade de mapeamento e em região de repetição de bases, respectivamente. Além disso, **(Figuras 2-4)**.

Figura 2 – Visualização gráfica de uma variante *missense* somática detectada no gene *RAD51*, na amostra tumoral 701. A variante se encontra em uma região de baixa complexidade, representada por bases nitrogenadas em letras minúsculas (retângulo vermelho).

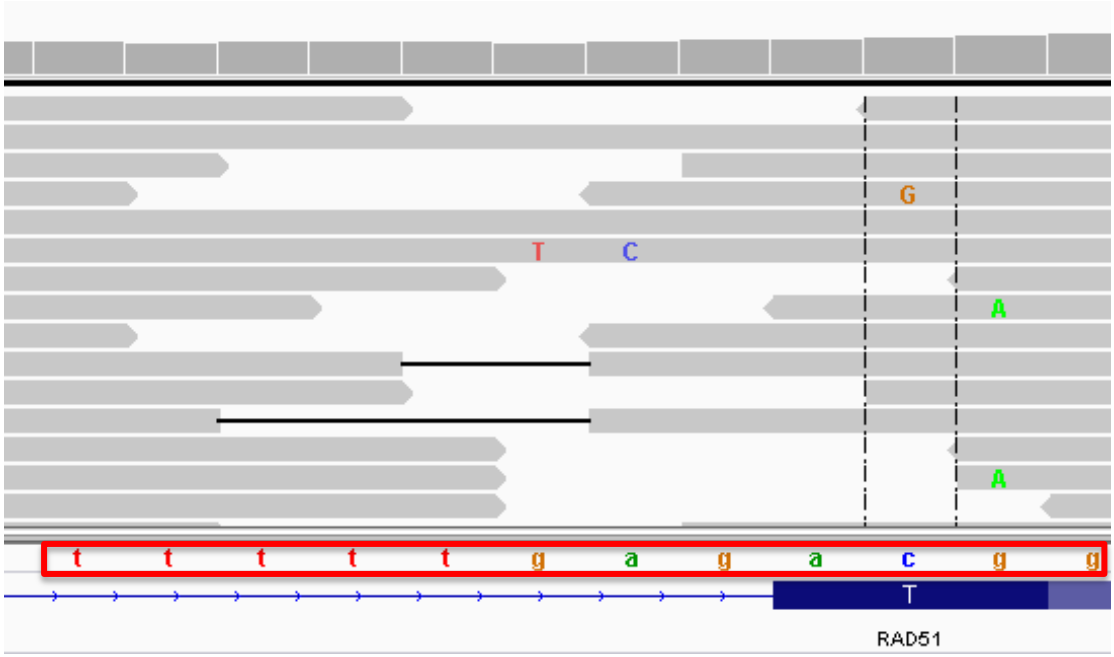


Figura 3 – Visualização gráfica de uma variante *missense* somática detectada no gene *DICER1*, na amostra tumoral 701. A variante se encontra em uma região de baixa complexidade, representada por bases nitrogenadas em letras minúsculas (retângulo vermelho).

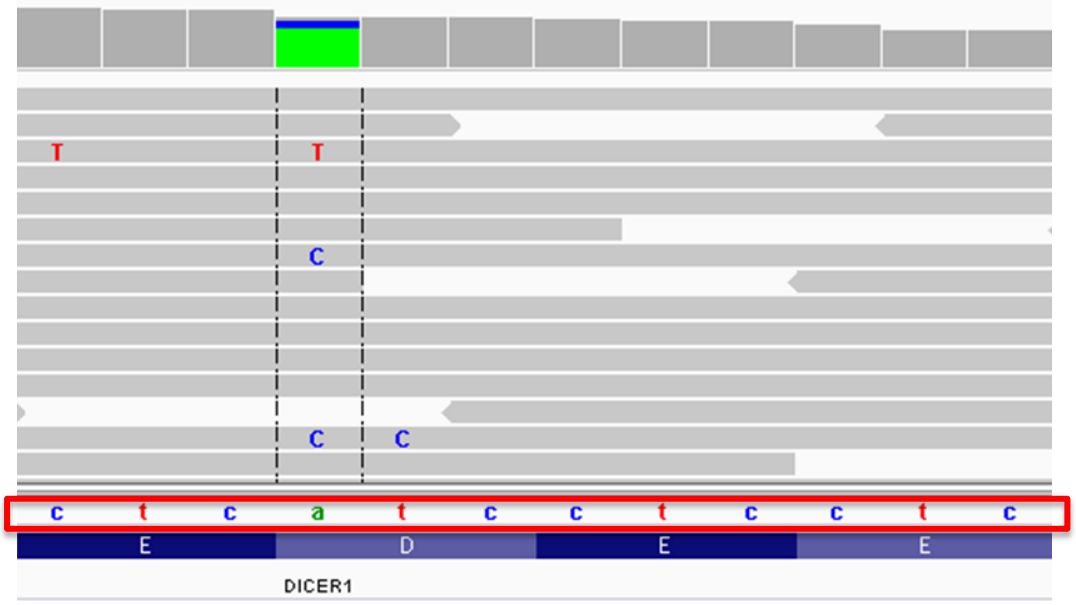
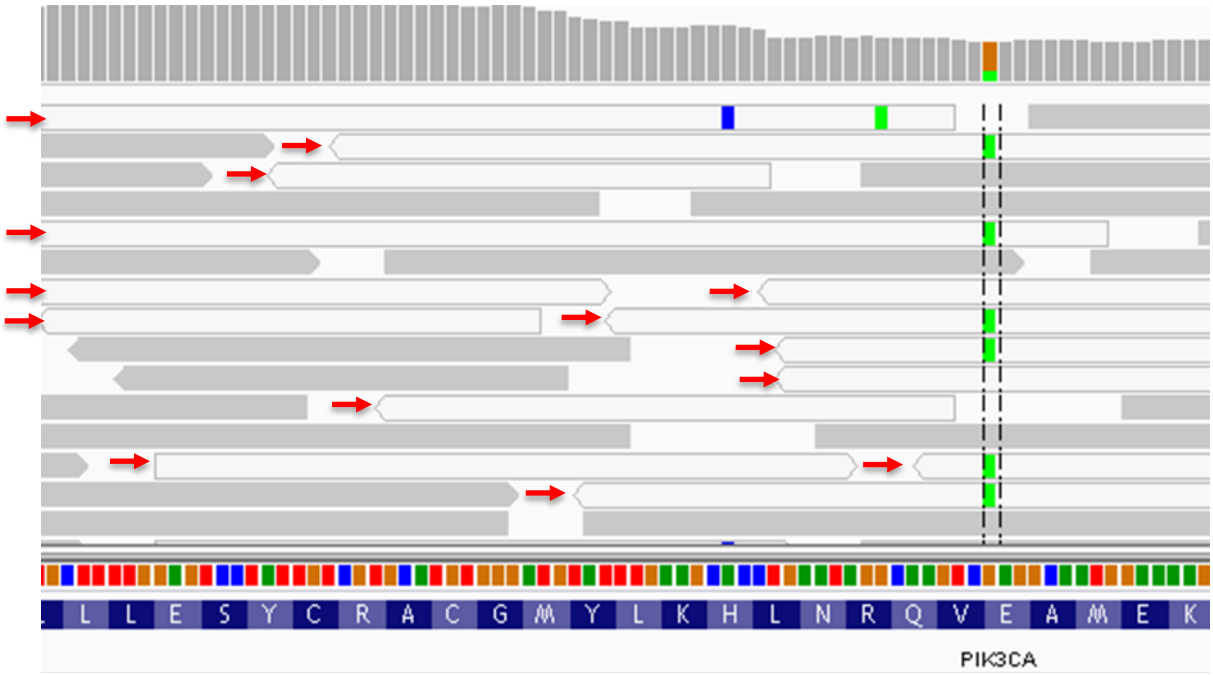


Figura 4 – Visualização gráfica de uma variante *missense* somática detectada no gene *PIK3CA*, na amostra tumoral 701. A variante se encontra em uma região de baixa qualidade de mapeamento, representada por *reads* com cor de preenchimento branca (setas vermelhas).



3.3 Escopo geral de variantes

Identificamos 20 genes diferentes com variantes somáticas entre todas as pacientes avaliadas, sendo eles: *ARID1A, ATR, ATXN1, BRCA1, CACNA1E, CAMK1G, CSMD3, FAT2, FBXW7, GATA3, MAP2K4, MET, NF1, NOTCH1, PIK3CA, PTEN, PTPN13, SPEN, TP53* e *UBR5.*

Detectamos 32 genes diferentes com variantes germinativas entre todas as pacientes avaliadas, sendo eles: *AHNAK, ANLN, APC, ARID1A, ATAD2B, ATM, ATR, ATXN1, BAP1, BRCA1, BRCA2, CSMD3, CSPP1, DICER1, ESR1, FAT2, FAT4, LYST, MAP3K1, MED23, MET, MTHFD2, MTOR, NCOA3, NOTCH1, PCDH10, POLD1, PRKD1, PTPN13, SETD2, SPEN* e *TNC.*

No total, dentre todas as pacientes analisadas, detectamos variantes em 41 dos 64 (64%) genes do painel.

3.4 Variantes germinativas patogênicas

Das 15 amostras de sangue analisadas no presente estudo, apenas 2 apresentaram variantes germinativas patogênicas ou provavelmente patogênicas dentre os 64 genes investigados **(Tabela 7)**. Ambas as variantes

foram identificadas no gene *BRCA1*, sendo uma localizada na paciente 610, do tipo inserção *frameshift* (c.5266dupC) e classificada como patogênica, de acordo com critérios ACMG. A segunda se trata de uma variante *missense* (c.245T>G), identificada na paciente 635, previamente classificada como VUS, porém reclassificada como provavelmente patogênica após ter sido reportada (ensaio funcional) como variante com potencial de causar perda de função em BRCA1 (45). Nenhuma das pacientes portadoras de mutações em BRCA1 relataram história familiar (parentesco de até 3º grau) de câncer de mama, ovário, próstata ou pâncreas.

Tabela 7 – Resumo das duas variantes patogênicas detectadas no gene BRCA1. Tipo: tipo de variante; Clinvar: classificação da variante de acordo com depósitos no Clinvar; OncoKB: efeito da variante, de acordo com curadoria do OncoKB; ACMG: classificação da variante de acordos com critérios estabelecidos pela ACMG. *: variante reclassificada devido ao ensaio funcional realizado por Findlay e colegas (Findlay, 2018).

	610	635
Gene	*BRCA1*	*BRCA1*
Tipo	Inserção *frameshift*	*Missense*
Aminoácido	p.G1777fs	p.L82R
cDNA	c.5266dupC	c.245T>G
Frequência alélica	0.464	0.738
N reads alterados	143	152
N reads totais	308	206
Clinvar	Patogênica	VUS
OncoKB	Provável perda de função	NA
ACMG	Patogênica	Provavelmente Patogênica
Critérios_ACMG	PP5;PS3;PVS1;PM2	PS3;PM2;PP3*

Dentre os 64 genes que integram nosso painel, sete são genes com relevância clínica e listados nas diretrizes "*Genetic/Familial High-Risk Assessment in breast ovarian and pancreatic cancer (v 2.2022)*" da National Cancer Comprehensive Network (NCCN): *ATM, BRCA1, BRCA2, CDH1, NF1, PTEN* e *TP53*.

3.5 Variantes germinativas de significado incerto (VUS)

Seis das quinze pacientes apresentaram VUS germinativas nos genes: *APC, ATM, ATXN1, BRCA2, CSPP1* e *MET* **(Tabela 8)**.

Tabela 8 – Variantes germinativas de significado incerto (VUS) detectadas dentre as 15 amostras de sangue sequenciadas com sucesso. Tipo: tipo de variante; Aminoácido: troca de aminoácidos; REF: base referência (no genoma referência); ALT: base alterada (no tumor sequenciado); Fita: variante se encontra na fita *forward* (1) ou *reverse* (-1).

ID	Gene	rs	Posição	Tipo	Aminoácido	REF	ALT	Fita
635	*MET*	rs757883355	7:116339379-116339379	missense	V81F	G	T	1
700	*BRCA2*	rs80358732	13:32913588-32913588	missense	D1699G	A	G	1
702	*ATM*	rs587781352	11:108137923-108137923	missense	D831G	A	G	1
702	*CSPP1*	rs756423026	8:68007858-68007858	missense	R281W	C	T	1
703	*APC*	rs754783550	5:112177107-112177107	missense	D1939V	A	T	1
728	*APC*	rs111866410	5:112175651-112175651	missense	K1454E	A	G	1
730	*ATXN1*	rs139918586	6:16326864-16326864	missense	V560M	G	A	-1

3.6 Variantes somáticas

3.6.1 Genes comumente afetados

A mediana de variantes somáticas detectadas entre as 12 amostras tumorais sequenciadas com sucesso foi 2. Dentre essas, uma amostra tumoral (703) não apresentou nenhuma variante somática dentre os 64 genes analisados em nosso painel.

O gene mais frequentemente afetado foi *TP53*, afetado em 75% (9/12) das amostras, seguido de *NF1* (3/15). Os genes *CACNA1E, NOTCH1* e *PTPN13* foram identificados em duas amostras, cada, enquanto os demais genes apareceram apenas em uma das amostras tumorais, cada **(Tabela 9)**. Todas as variantes somáticas em *NOTCH1, CACNA1E* e *PTPN13* estavam associadas com uma mutação em *TP53*. Em dois pacientes sem mutação somática em *TP53*, *NF1* apresentava-se mutado, associado com alterações em *FBXW7, PIK3CA, SPEN* e *UBR5*. A outra amostra com *TP53* tipo selvagem apresentou uma única variante em *ATR*.

Tabela 9 – Variantes somáticas detectadas dentre as 12 amostras tumorais sequenciadas com sucesso. Tipo: tipo de variante; Aminoácido: troca de aminoácidos; REF: base referência (no genoma referência); ALT: base alterada (no tumor sequenciado); Fita: variante se encontra na fita *forward* (1) ou *reverse* (-1); ONCOKB: classificação do efeito da variante, de acordo com curadoria do OncoKB; CancerVar: classificação da variante, de acordo com a ferramenta CancerVar; TP53_DB: classificação de variantes no gene TP53, de acordo com resultados de ensaios funcionais depositados no bancos de dados *TP53 Database;* PRED: número de ferramentas de predição de efeito de variantes (n total=8) no qual a variante foi classificada como patogênica/danosa. PRED_COMP: número de ferramentas que utilizam métodos de integração de dados de múltiplas ferramentas de predição de impacto funcional de variantes (n total=4) no qual a variante foi classificada como patogênica/danosa.

ID	Gene	rs	Posição	Tipo	Aminoácido	REF	ALT	Fita	ONCOKB	Cancer_var	TP53_DB	PRED	PRED_COMP
605	NOTCH1	rs779613930	9:139391014-139391014	stop_codon	Q2393*	C	T	-1	LLOF	NA	NA	NA	NA
605	TP53	rs1057520002	17:7577560-7577560	missense	S241P	T	C	-1	NA	TIER_1_STRONG	LOF	8	4
608	ATR	.	3:142279283-142279283	missense	D455Y	G	T	-1	NA	TIER_3_UNCERTAIN	NA	6	1
611	ATXN1	NA	6:16328499:16328634	CNV_GANHO	NA	NA	NA	NA	NA	NA	NA	NA	NA
611	BRCA1	rs80357710	17:41215950-41215952	stop_codon	C1718-1719*	TG	-	-1	LLOF	NA	NA	NA	NA
611	CSMD3	.	8:113253991-113253991	missense	D3476H	G	C	-1	NA	TIER_4_BENIGN	NA	2	0
611	PTPN13	.	4:87666147-87666147	stop_codon	S839*	C	A	1	NA	NA	NA	NA	NA
611	TP53	rs876660548	17:7579414-7579414	stop_codon	W91*	G	A	-1	LLOF	NA	LOF	NA	NA
635	CACNA1E	rs201622587	1:181765920-181765920	missense	A2109S	G	T	1	NA	ND	NA	2	3
635	CAMK1G	NA	1:209781075:209781383	CNV_GANHO	NA	NA	NA	NA	NA	NA	NA	NA	NA
635	MET	.	7:116340204-116340204	missense	P356T	C	A	1	NA	TIER_3_UNCERTAIN	NA	6	0
635	MET	.	7:116395452-116395452	missense	T582I	C	T	1	NA	TIER_3_UNCERTAIN	NA	7	4
635	TP53	rs121912651	17:7577539-7577539	missense	R248W	C	T	-1	LOF	TIER_1_STRONG	LOF	8	4
700	GATA3	NA	10:8105966:8106132	CNV_GANHO	NA	NA	NA	NA	LGOF	NA	NA	NA	NA
700	TP53	rs1057519977	17:7578507-7578507	stop_codon	C141*	C	A	-1	LLOF	NA	not_LOF	NA	NA
701	FBXW7	NA	4:153269751:153269914	CNV_PERDA	NA	NA	NA	NA	LLOF	NA	NA	NA	NA
701	NF1	.	17:29559152-29559152	missense	P1087T	C	A	1	NA	TIER_3_UNCERTAIN	NA	7	1
701	PIK3CA	rs3729687	3:178938877-178938877	missense	E707K	G	A	1	NA	TIER_2_POTENTIAL	NA	7	4
701	SPEN	.	1:16259906-16259906	missense	H2391Y	C	T	1	NA	TIER_4_BENIGN	NA	1	0
701	UBR5	.	8:103300382-103300382	splice_donor	-	-	-	-1	NA	NA	NA	NA	NA
702	NF1	NA	17:29548870:29549083	CNV_PERDA	NA	NA	NA	NA	LOF	NA	NA	NA	NA
702	NOTCH1	rs373806373	9:139391932-139391932	missense	R2087W	C	T	-1	NA	TIER_2_POTENTIAL	NA	7	2
702	TP53	rs1057520004	17:7578202-7578202	missense	V216E	T	A	-1	NA	TIER_1_STRONG	LOF	8	4
715	PTPN13	NA	4:87622918:87623052	CNV_PERDA	NA	NA	NA	NA	NA	NA	NA	NA	NA
715	TP53	rs764735889	17:7576897-7576897	stop_codon	Q317*	C	T	-1	LLOF	NA	LOF	NA	NA
719	CACNA1E	NA	1:181767460:181767887	CNV_GANHO	NA	NA	NA	NA	NA	NA	NA	NA	NA
719	TP53	rs121913344	17:7577022-7577022	stop_codon	R306*	C	T	-1	LLOF	NA	LOF	NA	NA
728	ARID1A	rs1266385064	1:27023007-27023007	inframe_insertion	E38A	-	GGC	1	NA	NA	NA	NA	NA
728	NF1	.	17:29527569-29527569	missense	S340A	T	G	1	NA	TIER_3_UNCERTAIN	NA	2	1
728	NF1	.	17:29553568-29553568	missense	A706V	C	T	1	NA	TIER_3_UNCERTAIN	NA	6	1
728	NF1	.	17:29667634-29667638	frameshift	L2345-2346X	TAGA	-	1	LLOF	NA	NA	NA	NA
728	TP53	rs1321845532	17:7574003-7574003	stop_codon	R342*	C	T	-1	LLOF	NA	LOF	NA	NA
730	FAT2	rs148551207	5:150922928-150922928	missense	A2587V	C	T	-1	NA	TIER_4_BENIGN	NA	6	1
730	MAP2K4	.	17:12013737-12013737	missense	H227Y	C	T	1	NA	TIER_3_UNCERTAIN	NA	7	4
730	PTEN	rs1114167622	10:89712017-89712017	splice_donor	-	-	-	1	LLOF	NA	NA	NA	NA
730	TP53	rs863224451	17:7577114-7577114	missense	C275Y	G	A	-1	LLOF	TIER_1_STRONG	LOF	8	4

3.7 Genes drivers

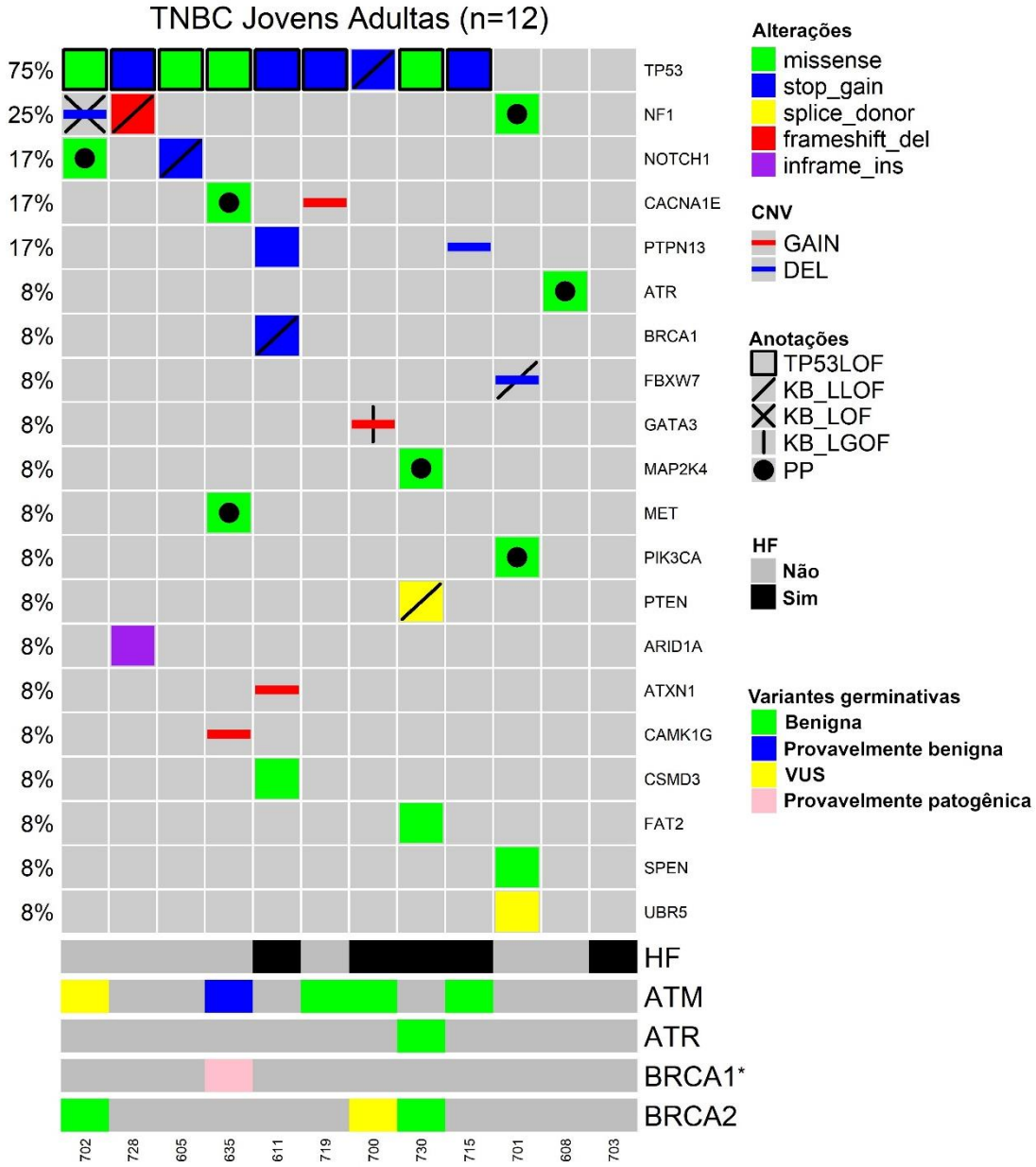
Como descrito na metodologia, consideramos genes e variantes como *drivers* caso fossem catalogados no CGC e que tivessem predição de patogenicidade em ao menos 2 de 8 ferramentas de predição de efeito de variante e 1 de 4 ferramentas de compilação e/ou tivessem evidência de efeito funcional em câncer (OncoKB e TP53 Database). Dentre as 12 amostras tumorais sequenciadas com sucesso, 10 possuíam ao menos um gene com potencial driver. O gene driver com maior frequência de alterações foi *TP53*, onde todas suas variantes somáticas detectadas foram classificadas como prováveis causadoras de perda de função (LLOF) ou causadoras de perda de função (LOF). Duas variantes de *CACNA1E* foram detectadas, em uma das amostras tumorais o gene foi identificado com uma variante *missense* classificada como possivelmente patogênica (PP), enquanto outro tumor possuía ganho de cópias do gene. Por fim, o gene *PTPN13* apresentou uma variante causadora de ganho de códon de parada em uma amostra tumoral, e outra amostra apresentou perda de cópias do gene.

Genes potencialmente drivers também foram identificados nos TSGs BRCA1, FBXW7 e PTEN. Nenhum oncogene com evidência funcional de patogenicidade foi identificado.

Uma amostra tumoral apresentou uma variante *missense* somática possivelmente patogênica (PP) no oncogene *MET*. Outra amostra de TNBC apresentou uma variante *missense* PP em *PIK3CA* e *UBR5* (ambos classificados como oncogenes).

A maioria das amostras possuía ao menos dois genes drivers, representados por uma combinação de um gene de dupla função (oncogene e TSG), como *TP53*, associado a um TSG, oncogene ou outro gene de dupla função. A amostra 608 não apresentou drivers, apenas uma variante *missense* em *ATR* sem evidência funcional na literatura ou previsão de dano em ferramentas *in silico*. A imagem a seguir possui o resumo de todas as variantes somáticas identificadas, assim como suas anotações complementares **(Figura 5)**.

Figura 5 – Oncoplot de variantes somáticas detectadas nos tumores sequenciados no presente estudo. Linhas representam genes e colunas representam amostras tumorais. HF: história familiar; ATM, ATR, BRCA1: classificação de variantes germinativas detectadas em ATM, ATR e BRCA1, de acordo com critérios de classificação da ACMG. TP53LOF: Variante causa perda de função (TP53 Database); KB_LLOF: Provável perda de função (OncoKB); KB_LOF: perda de função (OncoKB); KB_LGOF: Provável ganho de função (OncoKB); PP: Possivelmente patogênica. *: Uma das pacientes (610) com variante germinativa patogênica em BRCA1 não é demonstrada, pois sua amostra tumoral não foi sequenciada com sucesso (Tabela 6).



3.8 Análise exploratória

Usamos dados curados dos portais de dados COSMIC e cBioPortal para explorar a relevância de alterações nos 64 genes avaliados no presente trabalho, em outros tumores. Os dados de tumores de pacientes foram comparados entre adultos jovens (diagnosticados até 40 anos) e adultos mais velhos, com 5 faixas etárias de diagnóstico (≤40, 41-50, 51-60, 61-70 e >70) para comparar as taxas de mutação. Também investigamos CNV, expressão gênica, expressão proteica e metilação entre grupos etários na coorte *TCGA Breast Cancer Firehose Legacy Cohort* (TCGA-BRCA), dividindo amostras tumorais conforme descrito acima.

Primeiro exploramos se existiam diferenças na frequência de variantes pontuais e truncadas (*missense, inframe, nonsense, frameshift* e sítio de *splicing* canônico) entre pacientes jovens e idosos nos genes que integram o painel personalizado do presente estudo (mesmo aqueles em que nenhuma variante foi detectada entre as amostras sequenciadas neste estudo). Acessamos os dados de câncer de mama disponíveis na coorte *COSMIC Mutation Data (Genome-Screens)* (https://cancer.sanger.ac.uk/cosmic/download; acessado em maio de 2022), totalizando 2.672 amostras de vários estudos. Em seguida, excluímos amostras provenientes de relatos de casos e estudos em linhagens celulares, bem como amostras de carcinoma adenóide, acínico, neuroendócrino, metaplásico e não-primário, e amostras sem informação de idade ao diagnóstico, finalizando com dados de 184 adultos jovens (≤40 anoso), 265 pacientes diagnosticados entre 41 e 50 anos, 316 pacientes entre 51 e 60, 307 pacientes entre 61 e 70 e 225 diagnosticados com 71 anos ou mais, compreendendo 7 estudos diferentes (16,21–26). Nenhuma distinção de subtipo de tumor foi feita, uma vez que esse dado estava ausente em alguns estudos.

Em seguida, criamos uma tabela de contingência para comparar a frequência de amostras afetadas entre adultos jovens (≤40) e idosos estratificados em faixas etárias (41-50, 51-60, 61-70 e >70) para cada um dos 64 genes de nosso painel, para verificar se algum desses genes apresentava maior frequência de variantes pontuais e truncadas no grupo mais jovem. Entre os 64 genes de nosso painel, nenhuma das comparações de faixa etária

revelou uma frequência diferencial de variantes pontuais e truncadas no grupo de adultos jovens. No entanto, observamos maior frequência de mutações em *CDH1* em todos os grupos de idosos em comparação com adultos jovens (p aj.<0,001). Da mesma forma, *MAP3K1* apresentou maior frequência de mutação em quase todos os grupos de idosos (excluindo 41-50) em comparação com o grupo de adultos jovens (p aj.<0,05).

Maior frequência de variantes em *TP53* (p aj.<0,05) foi identificada nos grupos de idosos com idades menos avançadas (41-50 e 51-60) em comparação com o grupo de adultos jovens, mas o mesmo não foi visualizado em grupos de faixa etária mais avançada (61-70 e >70). O inverso foi observado em *PIK3CA*, onde apenas o grupo com idade mais avançada (>70) apresentou maior frequência de variantes (p aj.=0,003) em comparação ao grupo de adultos jovens.

Em seguida, usamos a plataforma de ferramentas de comparação de coorte disponíveis no cBioPortal para realizar uma análise semelhante, desta vez com a coorte isolada TCGA-BRCA (câncer de mama), comparando a expressão de genes e proteínas, CNVs e perfil de metilação entre adultos jovens e pacientes idosos. A coorte foi composta por 76 adultos jovens e 992 idosos portadores de carcinoma primário de mama, sendo estes divididos nas seguintes faixas etárias: 41-50 (n=230), 51-60 (n=273), 61-70 (280) e >70 (n=209) (21). Nenhuma distinção de subtipo de tumor foi feita, uma vez que esses dados estavam faltando em alguns pacientes e outras divisões de subgrupos enfraqueceriam o poder da análise estatística.

Ao comparar adultos jovens com todos os grupos de idosos (41-50, 51-60, 61-70 e >70), o grupo adulto jovem apresentou maior frequência de ganho de número de cópias em *SMURF2* e *PRKAR1A* (p aj.<0,05) e foram observados com maior expressão gênica (p aj.<0,05) em adultos jovens, quando comparados às faixas etárias mais avançadas (61-70 e >70). Por fim, observamos maior (p aj.<0,05) expressão gênica e proteica de *CDH1* em adultos jovens quando comparados a todos os subgrupos de idade mais avançada. Além disso, observamos maior expressão gênica de alguns genes em adultos jovens quando comparada ao grupo >70, sendo estes: *ARID1A, ANLN, ATAD2B, FAT2, FAT4, FBXW7, MET, MTOR, PARP4, PIK3CA* e *RAD51*.

4 Discussão

Realizamos a análise de variantes somáticas em 12 amostras tumorais e variantes germinativas em 15 amostras de sangue, provenientes de pacientes adultas jovens com câncer de mama triplo-negativo.

Dentre as amostras tumorais, o gene mais frequentemente alterado foi *TP53*, seguido pelo *NF1*. Identificamos 2 de 15 pacientes com mutação germinativa patogênica ou provavelmente patogênica em *BRCA1* (13%), o que se assemelha com achados de estudos anteriores, incluindo um publicado por nosso grupo, em que detectamos 16% de mutações patogênicas de *BRCA1* e *BRCA2* em 79 tumores luminais de adultos jovens (50–52). Nenhuma das duas pacientes apresentavam história familiar de câncer de mama, ovário, próstata e pâncreas.

Considerando as variantes somáticas, a maioria das variantes em *TP53* foram classificadas como LLOF ou LOF. Extensivos estudos de caracterização molecular de subtipos de câncer de mama demonstraram que de fato, variantes em *TP53* são comuns em TNBC, enquanto o subtipo luminal, por exemplo, apresenta maior predominância de variantes somáticas em *PIK3CA* e *GATA3*. De fato, apenas um tumor (701) apresentou uma variante em *PIK3CA*, que foi classificada como possivelmente patogênica (classificada como danosa em 11/12 ferramentas *in-silico*). O mesmo tumor também tinha uma variante somática possivelmente patogênica (PP) no oncogene *UBR5* e ganho de cópias no gene supressor de tumor *FBXW7*, evento este classificado como LLOF. A amostra de tumor do paciente 700 foi a única amostra tumoral com uma variante somática em *GATA3*, na qual a variante (ganho de cópias) foi relatada como provável GOF. Este evento é interessante pois, geralmente, a maioria das variantes somáticas em *GATA3* detectadas em pacientes com câncer de mama são caracterizadas por causarem a perda de função da proteína. Embora o *GATA3* seja citado principalmente como um TSG, ainda há discussão na literatura e relatos de que os papéis do *GATA3* também podem influenciar vias oncogênicas (53,54). Portanto, a classificação das mutações prejudiciais no gene é uma tarefa desafiadora.

Entre todas as variantes somáticas detectadas em *TP53*, apenas uma amostra (700) não apresentou evidência direta de causar perda de função, de acordo com ensaios funcionais reportados na literatura. No entanto, é provável que a variante cause a transcrição de uma proteína truncada no domínio do sítio de ligação de DNA da proteína, onde variantes de inserção e deleção e até mesmo variantes que causam apenas uma única alteração de nucleotídeo (*missense*) provavelmente causarão perda de função (55).

Detectado com variantes somáticas em duas amostras tumorais, o gene *PTPN13* codifica um membro da família da proteína tirosina fosfatase (PTP) e é classificado como um gene supressor de tumor, com papéis na regulação da sinalização de apoptose. Estudos mostram que a disfunção do gene em modelos TNBC in vivo e in vitro leva a um aumento de proliferação e invasão de células malignas (56). Curiosamente, uma amostra de tumor (611) tinha uma variante de stop-códon em *PTPN13*, localizada no domínio FERM-c. Embora, até onde sabemos, não haja nenhum estudo que investigue o impacto funcional dessa variante específica, estudos mostram que a disfunção dos domínios PDZ (interação proteína-proteína) e da proteína tirosina-fosfatase em *PTPN13* interromperia suas principais funções antitumorais (57). Além disso, o gene se encontra frequentemente inibido epigeneticamente no câncer de mama (58). Esses fatos destacam a relevância desse gene no câncer de mama, embora não necessariamente exclusivamente em adultos jovens, pois não se mostrou altamente alterado no presente estudo nem diferencialmente alterado, expresso ou metilado entre faixas etárias em nossa análise exploratória.

Embora a maioria das amostras apresentasse pelo menos uma variante em um oncogene ou TSG com evidência de disfunção, algumas amostras não apresentaram um perfil mais claro de mutagênese. Por exemplo, a amostra tumoral 701 apresenta uma variante somática *missense* no gene *PIK3CA* (E707K), que não possui classificação específica na literatura quanto à sua patogenicidade, apesar de ser classificada como VUS no contexto germinativo. No entanto, a variante é classificada como patogênica em quase todas as ferramentas individuais e de compilação de predição de efeito (7/8 e 4/4, respectivamente). Além disso, a variante é classificada como *Tier 2*

(possivelmente patogênica) pelo método de classificação de variantes somáticas desenvolvido por Li e colegas (31). Da mesma forma, identificamos uma variante somática *missense* em *NOTCH1* (R2087W) na amostra tumoral 702, que não possui classificação específica na literatura quanto à sua patogenicidade, apesar de ser classificada como VUS no contexto germinativo (ACMG). No entanto, estudos indicam que variantes *upstream* do domínio de transativação da proteína podem causar perda de função (59). Além disso, a variante é classificada como patogênica em quase todas as ferramentas individuais e de compilação de previsão de efeitos (7/8 e 2/4, respectivamente) e é classificada como *Tier 2* (possivelmente patogênica) pelo método de classificação de variantes somáticas desenvolvido por Li e colegas. Finalmente, identificamos uma variante *missense* somática possivelmente patogênica no gene *MAP2K4* no tumor 730. Embora a mesma amostra apresentasse uma mutação de perda de função em *TP53* e uma variante LLOF no gene *PTEN*, *MAP2K4* não pode ser ignorado, pois estudos sugerem que mutações disruptivas no gene podem sensibilizar células malignas aos inibidores de *MEK* (60).

Por fim, realizamos análises exploratórias usando ferramentas de grandes bancos de dados de câncer para verificar se os genes investigados em nosso painel poderiam separar faixas etárias de acordo com seus perfis moleculares. Os genes *SMURF2* e *PRKAR1A* foram observados com maior frequência de amplificação em pacientes adultas jovens com câncer de mama quando comparadas às faixas etárias mais avançadas (BRCA-TCGA). Enquanto nenhuma variante foi detectada nestes genes nas amostras investigadas no presente estudo, outro estudo de nosso grupo, no qual o mesmo painel de genes foi utilizado em amostras tumorais de pacientes jovens com câncer de mama luminal demonstrou um ganho de CNV em *SMURF2* em uma das amostras (artigo submetido para publicação) **(APÊNDICE B)**. Embora o *SMURF2* ainda não esteja catalogado no CGC, alguns autores mostraram que o gene pode influenciar a tumorigênese de alguns tipos de câncer (61). Da mesma forma, em outro estudo do nosso grupo, no qual investigamos variantes somáticas em pacientes com câncer de mama luminal em adultos jovens, variantes em *PRKAR1A* e *SMURF2* foram identificadas como possíveis

causadores do câncer (50). O *SMURF2* também apresentou maior expressão gênica no grupo de adultos jovens BRCA-TCGA, mostrando que o gene pode ter um papel relevante neste grupo, embora seja claramente necessária uma investigação mais aprofundada. *CDH1* foi identificado com maior expressão gênica e proteica em tumores de pacientes adultos jovens, indicando que a modulação de sua expressão pode afetar sua atividade proteica neste grupo de pacientes, o que possivelmente poderia afetar os processos de transição epitélio-mesenquima. Esse gene é mais frequentemente afetado por mutações somáticas pontuais e truncadas, com potencial inativador em tumores de pacientes idosos. De fato, outros estudos relataram mutações de *CDH1* como um marcador em tumores de pacientes com idades mais avançadas, bem como normalmente associados a tumores lobulares, que são infrequentes em pacientes mais jovens (62).

O presente estudo apresenta algumas limitações, como o pequeno tamanho da amostra e as análises de um número limitado de genes. Porém, considerando grandes coortes de câncer de mama, como o estudo TCGA-BRCA (17), em que aproximadamente 80 amostras tumorais eram de pacientes jovens, concluímos que o presente projeto contribuiu com um número significativo de tumores, dada a baixa representatividade geral dessa população em outras estudos. Outro ponto positivo é que este painel customizado foi dirigido para detecção de mutações somáticas em câncer de mama, com foco em câncer de mama de pacientes jovens. Além disso, a maioria dos estudos relata dados de pacientes com ascendência europeia, portanto, acreditamos que este é um dado relevante, proveniente de uma população étnica miscigenada.

5 Conclusão

Diante da detecção de variantes na maioria das pacientes que tiveram seu tumor sequenciado com sucesso (11/12), concluímos que o painel construído possui boa qualidade e alcance, sendo adequado para a investigação de drivers e possíveis drivers em pacientes adultas jovens com câncer de mama triplo-negativo.

Por meio do sequenciamento-alvo do tumor e sangue de pacientes adultas jovens, foi possível identificar uma alta prevalência de mutações somáticas em *TP53* e destacar a presença de outros genes drivers e potencialmente drivers, como *NF1, NOTCH1* e *PTPN13*. Além disso, detectamos a ocorrência de 2 mutações germinativas patogênicas em *BRCA1*, dentre as 15 pacientes com sangue sequenciado.

Por fim, podemos concluir que o presente trabalho contribui para o desenvolvimento de conhecimento relacionado às características deste grupo seleto e pouco estudados de pacientes.

6 Referências

1. Johnson RH, Chien FL, Bleyer A. Incidence of breast cancer with distant involvement among women in the United States, 1976 to 2009. JAMA. 2013 Feb 27;309(8):800–5.
2. Miller KD, Fidler-Benaoudia M, Keegan TH, Hipp HS, Jemal A, Siegel RL. Cancer statistics for adolescents and young adults, 2020. CA Cancer J Clin. 2020 Nov;70(6):443–59.
3. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. Ann Oncol. 2010 Oct;21(10):1974–81.
4. Liedtke C, Rody A, Gluz O, Baumann K, Beyer D, Kohls E-B, et al. The prognostic impact of age in different molecular subtypes of breast cancer. Breast Cancer Res Treat. 2015 Aug;152(3):667–73.
5. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? J Clin Oncol. 2011 Jan 1;29(1):e18-20.
6. Kan Z, Ding Y, Kim J, Jung HH, Chung W, Lal S, et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. Nat Commun. 2018 Apr 30;9(1):1725.
7. Azim HA, Partridge AH. Biology of breast cancer in young women. Breast Cancer Res. 2014 Aug 27;16(4):427.
8. Aine M, Boyaci C, Hartman J, Häkkinen J, Mitra S, Campos AB, et al. Molecular analyses of triple-negative breast cancer in the young and elderly. Breast Cancer Res. 2021 Feb 10;23(1):20.
9. Sun Z, Velázquez-Quesada I, Murdamoothoo D, Ahowesso C, Yilmaz A, Spenlé C, et al. Tenascin-C increases lung metastasis by impacting blood vessel invasions. Matrix Biol. 2019 Oct;83:26–47.
10. Oskarsson T, Acharyya S, Zhang XH-F, Vanharanta S, Tavazoie SF, Morris PG, et al. Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. Nat Med. 2011 Jun 26;17(7):867–74.
11. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010 Nov 11;363(20):1938–48.
12. Keegan THM, DeRouen MC, Press DJ, Kurian AW, Clarke CA. Occurrence of breast cancer subtypes in adolescent and young adult women. Breast Cancer Res. 2012 Mar 27;14(2):R55.
13. Kim J, Hong S, Lee JJ, Won Y-J, Lee ES, Kang H-S, et al. Analysis of the tumor characteristics in young age breast cancer patients using collaborative stage data of the Korea Central Cancer Registry. Breast Cancer Res Treat. 2021 Feb 18;
14. Lehmann BD, Jovanović B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of Triple-Negative Breast Cancer Molecular Subtypes:

Implications for Neoadjuvant Chemotherapy Selection. PLoS ONE. 2016 Jun 16;11(6):e0157368.

15. Sæther NH, Skuja E, Irmejs A, Maksimenko J, Miklasevics E, Purkalne G, et al. Platinum-based neoadjuvant chemotherapy in BRCA1-positive breast cancer: a retrospective cohort analysis and literature review. Hered Cancer Clin Pract. 2018 Apr 27;16:9.
16. Robson M, Im SA, Senkus E, Xu B, Domchek SM, Masuda N, et al. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. N Engl J Med. 2017 Aug 10;377(6):523–33.
17. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct 4;490(7418):61–70.
18. Soussi T, Wiman KG. TP53: an oncogene in disguise. Cell Death Differ. 2015 Aug;22(8):1239–49.
19. Olivier M, Langerød A, Carrieri P, Bergh J, Klaar S, Eyfjord J, et al. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. Clin Cancer Res. 2006 Feb 15;12(4):1157–67.
20. Serio PAMP, de Lima Pereira GF, Katayama MLH, Roela RA, Maistro S, Folgueira MAAK. Somatic Mutational Profile of High-Grade Serous Ovarian Carcinoma and Triple-Negative Breast Carcinoma in Young and Elderly Patients: Similarities and Divergences. Cells. 2021 Dec 20;10(12).
21. Midha MK, Huang Y-F, Yang H-H, Fan T-C, Chang N-C, Chen T-H, et al. Comprehensive cohort analysis of mutational spectrum in early onset breast cancer patients. Cancers (Basel). 2020 Jul 28;12(8).
22. Mealey NE, O'Sullivan DE, Pader J, Ruan Y, Wang E, Quan ML, et al. Mutational landscape differences between young-onset and older-onset breast cancer patients. BMC Cancer. 2020 Mar 12;20(1):212.
23. Ooi SL, Martinez ME, Li CI. Disparities in breast cancer characteristics and outcomes by race/ethnicity. Breast Cancer Res Treat. 2011 Jun;127(3):729–38.
24. Ansari-Pour N, Zheng Y, Yoshimatsu TF, Sanni A, Ajani M, Reynier J-B, et al. Whole-genome analysis of Nigerian patients with breast cancer reveals ethnic-driven somatic evolution and distinct genomic subtypes. Nat Commun. 2021 Nov 26;12(1):6946.
25. Silva JDDE, de Oliveira RR, da Silva MT, Carvalho MD de B, Pedroso RB, Pelloso SM. Breast cancer mortality in young women in brazil. Front Oncol. 2020;10:569933.
26. Orlandini LF, Antonio MV do N, Espreafico CR, Bosquesi PL, Poli-Neto OB, de Andrade JM, et al. Epidemiological analyses reveal a high incidence of breast cancer in young women in brazil. JCO Glob Oncol. 2021 Jan;7:81–8.
27. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016 Jun 2;534(7605):47–54.
28. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and

mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012 Apr 4;486(7403):395–9.

29. Abbott KL, Nyre ET, Abrahante J, Ho Y-Y, Isaksson Vogel R, Starr TK. The Candidate Cancer Gene Database: a database of cancer driver genes from forward genetic screens in mice. Nucleic Acids Res. 2015 Jan;43(Database issue):D844-8.

30. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJY, van Karnebeek C, Wasserman WW. FLAGS, frequently mutated genes in public exomes. BMC Med Genomics. 2014 Dec 3;7:64.

31. Li Q, Ren Z, Cao K, Li MM, Wang K, Zhou Y. CancerVar: An artificial intelligence-empowered platform for clinical interpretation of somatic mutations in cancer. Sci Adv. 2022 May 6;8(18):eabj1624.

32. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: A precision oncology knowledge base. JCO Precis Oncol. 2017 Jul;2017.

33. Monti P, Perfumo C, Bisio A, Ciribilli Y, Menichini P, Russo D, et al. Dominant-negative features of mutant TP53 in germline carriers have limited impact on cancer outcomes. Mol Cancer Res. 2011 Mar;9(3):271–9.

34. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet. 2018 Oct;50(10):1381–7.

35. Kotler E, Shani O, Goldfeld G, Lotan-Pompan M, Tarcic O, Gershoni A, et al. A systematic p53 mutation library links differential functional impact to cancer mutation pattern and evolutionary conservation. Mol Cell. 2018 Jul 5;71(1):178-190.e8.

36. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013 Jan;34(1):57–65.

37. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011 Sep 1;39(17):e118.

38. Steinhaus R, Proft S, Schuelke M, Cooper DN, Schwarz JM, Seelow D. MutationTaster2021. Nucleic Acids Res. 2021 Jul 2;49(W1):W446–51.

39. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE. 2012 Oct 8;7(10):e46688.

40. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248–9.

41. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016 Jan;11(1):1–9.

42. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti

S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016 Oct 6;99(4):877–85.

43. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015 Apr 15;24(8):2125–37.

44. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016 Dec;48(12):1581–6.

45. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018 Oct;562(7726):217–22.

46. Chiang T, Liu X, Wu T-J, Hu J, Sedlazeck FJ, White S, et al. Atlas-CNV: a validated approach to call single-exon CNVs in the eMERGESeq gene panel. Genet Med. 2019 Sep;21(9):2135–44.

47. Soong D, Stratford J, Avet-Loiseau H, Bahlis N, Davies F, Dispenzieri A, et al. CNV Radar: an improved method for somatic copy number alteration characterization in oncology. BMC Bioinformatics. 2020 Mar 6;21(1):98.

48. Suvakov M, Panda A, Diesh C, Holmes I, Abyzov A. CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. Gigascience. 2021 Nov 18;10(11).

49. Moreno-Cabrera JM, Del Valle J, Castellanos E, Feliubadaló L, Pineda M, Serra E, et al. CNVfilteR: an R/bioconductor package to identify false positives produced by germline NGS CNV detection tools. Bioinformatics. 2021 May 13;

50. Encinas G, Sabelnykova VY, de Lyra EC, Hirata Katayama ML, Maistro S, de Vasconcellos Valle PWM, et al. Somatic mutations in early onset luminal breast cancer. Oncotarget. 2018 Apr 27;9(32):22460–79.

51. Carraro DM, Koike Folgueira MAA, Garcia Lisboa BC, Ribeiro Olivieri EH, Vitorino Krepischi AC, de Carvalho AF, et al. Comprehensive analysis of BRCA1, BRCA2 and TP53 germline mutation and tumor characterization: a portrait of early-onset breast cancer in Brazil. PLoS ONE. 2013 Mar 1;8(3):e57581.

52. Copson ER, Maishman TC, Tapper WJ, Cutress RI, Greville-Heygate S, Altman DG, et al. Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. Lancet Oncol. 2018 Feb;19(2):169–80.

53. Takaku M, Grimm SA, Wade PA. GATA3 in breast cancer: tumor suppressor or oncogene? Gene Expr. 2015;16(4):163–8.

54. Mair B, Konopka T, Kerzendorfer C, Sleiman K, Salic S, Serra V, et al. Gain- and Loss-of-Function Mutations in the Breast Cancer Gene GATA3 Result in Differential Drug Sensitivity. PLoS Genet. 2016 Sep

2;12(9):e1006279.

55. Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. Why are there hotspot mutations in the TP53 gene in human cancers? Cell Death Differ. 2018 Jan;25(1):154–60.

56. Hamyeh M, Bernex F, Larive RM, Naldi A, Urbach S, Simony-Lafontaine J, et al. PTPN13 induces cell junction stabilization and inhibits mammary tumor invasiveness. Theranostics. 2020 Jan 1;10(3):1016–32.

57. Freiss G, Chalbos D. PTPN13/PTPL1: an important regulator of tumor aggressiveness. Anticancer Agents Med Chem. 2011 Jan;11(1):78–88.

58. Ying J, Li H, Cui Y, Wong AHY, Langford C, Tao Q. Epigenetic disruption of two proapoptotic genes MAPK10/JNK3 and PTPN13/FAP-1 in multiple lymphomas and carcinomas through hypermethylation of a common bidirectional promoter. Leukemia. 2006 Jun;20(6):1173–5.

59. Gerhardt DM, Pajcini KV, D'altri T, Tu L, Jain R, Xu L, et al. The Notch1 transcriptional activation domain is required for development and reveals a novel role for Notch1 signaling in fetal hematopoietic stem cells. Genes Dev. 2014 Mar 15;28(6):576–93.

60. Xue Z, Vis DJ, Bruna A, Sustic T, van Wageningen S, Batra AS, et al. MAP3K1 and MAP2K4 mutations are associated with sensitivity to MEK inhibitors in multiple cancer models. Cell Res. 2018 Jul;28(7):719–29.

61. David D, Jagadeeshan S, Hariharan R, Nair AS, Pillai RM. Smurf2 E3 ubiquitin ligase modulates proliferation and invasiveness of breast cancer cells in a CNKSR2 dependent manner. Cell Div. 2014 Aug 31;9:2.

62. Encinas G, Maistro S, Pasini FS, Katayama MLH, Brentani MM, Bock GH de, et al. Somatic mutations in breast and serous ovarian cancer young patients: a systematic review and meta-analysis. Rev Assoc Med Bras. 2015;61(5):474–83.

APÊNDICE A - Procedimentos Operacionais Padronizados (POP) para utilização do SureSelectXT HS Target Enrichment System

Para a extração de material parafinado ou sangue, siga as instruções do manual da empresa fornecedora.

Observação: Em nosso laboratório temos usado os kits:

- QIAmp DNA Mini Kit (para amostras emblocadas em parafina e para sangue);
- QIAmp DNA Blood Mini Kit (para sangue);
- QIAmp DNA FFPE Mini Kit (para amostras emblocadas em parafina);

Observação: Para sangue, um maior aproveitamento pode ser obtido por meio de uso do QIAmp DNA Blood Midi Kit.

Após a extração:

- Verificar a qualidade das amostras por meio dos fatores do Nanodrop (260/280 e 260/230);
- Verificar concentração das amostras por meio de QUBIT ou técnica similar (considerar concentração obtida no QUBIT ao invés da concentração obtida no nanodrop, por ser mais precisa);

Preparação da amostra para fragmentação:

O protocolo de montagem de biblioteca exige um input mínimo de 10ng e um máximo de 200ng. Aconselha-se que sempre se use o input máximo.

ATENÇÃO: Os valores de 10-200ng se referem à QUANTIDADE de DNA e não à concentração.

No nosso laboratório, utilizamos o equipamento Covaris, que fornece fragmentação por meio da técnica de sonicação. Também é possível usar fragmentação enzimática

Para a aplicação das amostras no Covaris, as mesmas devem ser preparadas diluindo 200ng de material genômico em um volume final de 50uL de 1X Low TE Buffer (10 mM Tris-HCl, pH 7.5-8.0, 0.1 mM EDTA), em microtubos DNA LoBind de 1,5mL.

Dica: Para amostras que já possuem 200ng ou mais, ou seja, na qual o volume de amostra ao ser diluído para o Covaris seria igual ou inferior a 1 uL, indica-se que se dilua tal amostra de modo a deixá-la com concentração de cerca de 70ng, permitindo o uso de cerca de 3uL da mesma. Dessa forma, eliminam-se os riscos ao se pipetar volumes pequenos.

Dispomos do serviço de fragmentação do Laboratório de Sequenciamento em Larga Escala (SELA) da Rede Premium da Faculdade de Medicina da Universidade de São Paulo (FMUSP). As configurações do equipamento Covaris para a fragmentação são seguidas conforme apontado no manual SureSelectXT HS Target Enrichment System (páginas 26-28). Ao fim, as amostras devem ser armazenadas em freezer.

Observação: Nas instruções a seguir, os termos placas ou microtubos serão usados em relação às amostras a serem preparadas. Tanto placas de PCR quanto microtubos podem ser usados durante todo o protocolo e a escolha de um ou de outro pode ser feita de acordo com a disponibilidade e disposição de equipamentos e desejo de quem está seguindo o presente protocolo. As etapas em que o uso de placas é substituído pelo uso de microtubos e vice-versa se baseiam na experiência prática do grupo que desenvolveu este POP e, portanto, é opcional.

Preparação da biblioteca:

1. **Reparo e ligação de caldas-poli-A nas extremidades das fitas de DNA.**

A fragmentação do DNA causa perda de bases em suas extremidades. Essa fase serve principalmente para consertar esse problema, garantido o pareamento correto nas próximas etapas. A adição dos fragmentos polo-A serve para garantir a boa atividade da polimerase nas reações de PCR.

Verifique se você possui todos os seguintes reagentes, nas seguintes condições:

Tabela 1

Nome	Armazenar	descongelar	homogeneizar
End Repair-A Tailing Buffer (tampa amarela ou garrafa)	dentro do kit (-20°C)	descongelar em gelo (pelo menos 20 minutos antes do uso)	vortexar
Ligation Buffer (tampa roxa ou garrafa)	dentro do kit (-20°C)	descongelar em gelo (pelo menos 20 minutos antes do uso)	vortexar
End Repair-A Tailing Enzyme Mix (tampa laranja)	dentro do kit (-20°C)	botar no gelo apenas quando for usar (manter no freezer)	inversão
T4 DNA Ligase (tampa azul)	dentro do kit (-20°C)	botar no gelo apenas quando for usar (manter no freezer)	inversão
Adaptor Oligo Mix (tampa branca)	dentro do kit (-20°C)	manter no gelo	vortexar

A - Prepare o **Ligation master mix** de antemão, para que ele alcance temperatura ambiente durante os demais processos:

Tabela 2

Nome	Volume por amostra
Ligation Buffer (tampa roxa ou garrafa)	23uL
T4 DNA Ligase (tampa azul)	2uL

Observações:

- Não se esqueça de adicionar um erro na sua multiplicação;
- Vortexar o Ligation Buffer por 15 segundos;
- Os dois reagentes são viscosos, pipetar lentamente;
- Após a mistura, vortexar.

- Dica: deixe esse mix em um lugar isolado da bancada. Apesar de ser o primeiro mix a ser preparado, ele não será o primeiro a ser usado, cuidado para não se confundir.

B - Prepare o termociclador na seguinte programação:

Tabela 3

	temperatura (°C)	tempo (min)
Passo 1	20	15
Passo 2	72	15
Passo 3	4	hold

Observações:

- Caso o aparelho requira a seleção do volume, **configure-o para 70uL.**
- Dê início à corrida e espere o lid alcançar 20°C, então imediatamente pause a corrida e adicione a placa.

C - A seguir, prepare o **End Repair/dA-Tailing master mix** de acordo com a tabela a seguir:

Tabela 4

Nome	Volume por amostra
End Repair-A Tailing Buffer (tampa amarela ou garrafa)	16uL
End Repair-A Tailing Enzyme Mix (tampa laranja)	4uL

Observações:

- Não se esqueça de adicionar um erro na sua multiplicação;
- Homogeneizar com up and down.

D - Distribua suas amostras nos poços em sua placa de PCR. **Caso amostras pareadas** (tumor/sangue) estejam sendo usadas indica-se que cada tipo seja posicionado em paralelo em **colunas diferentes.**

E - Adicione **20uL do End Repair/dA-Tailing master mix** (não confundir com o primeiro mix preparado nesse protocolo, o Ligation master mix) a cada poço contendo suas amostras, misturando com **up and down cerca de 20x**, **sele** sua placa, **centrifugue** brevemente.

F - Insira a placa no termociclador pré-configurado e dê início à corrida.

G - Terminada a corrida, mantenha a placa em gelo e configure o termociclador:

Tabela 5

	temperatura (°C)	tempo (min)
Passo 1	20	30
Passo 2	4	hold

- Caso o aparelho requira a seleção do volume, **configure-o para 100uL**.
- Dê início à corrida e espere o lid alcançar 20°C, então imediatamente pause a corrida.

H - Retire sua placa do gelo e adicione **25uL de Ligation master mix** (o primeiro mix preparado nesse protocolo) a cada poço contendo suas amostras, misture com **up and down por cerca de 10x**.

I - Adicione **5uL do Adaptor Oligo Mix** a cada poço contendo suas amostras, misture com **up and down por cerca de 20x**. Sele a placa, centrifugue brevemente, adicione sua placa no termociclador e inicie a corrida.

Observação: Após a corrida de PCR, se desejado, o protocolo pode ser parado nessa etapa para que se continue no próximo dia, armazenando a placa selada e envolta em alumínio em geladeira ou freezer.

2. Primeira purificação das amostras por meio dos beads AMPure XP

Observções:

- As beads devem ser usadas em temperatura ambiente;
- Nunca congele as beads;

A - Prepare **400uL de etanol 70%** para cada amostra, adicionando erro no cálculo.

ATENÇÃO: ao se preparar o álcool 70%, preste muita atenção para que a correta proporção seja misturada. A mistura errada pode gerar perda parcial ou total do material genômico durante essa etapa de purificação.

B - **Vortexar** o frasco contendo as beads até que se ele apresente homogeneidade em sua cor. Verifique o fundo do frasco, observando se não existem beads concentradas no local.

C - Separe alíquota dos **beads** em microtubo com volume de **80uL multiplicado pela quantidade de amostras** mais erro.

D - Adicione **80uL da alíquota dos beads** nos poços de **cada amostra**, misturando com **up and down por cerca de 20x.**

Dica: indica-se que o tubo com a alíquota dos beads seja vortexado em intervalos de tempo curtos durante a adição dele nos poços, pois as beads começam a se concentrar no fundo do tubo em pouco tempo.

E - **Incube** as amostras em temperatura ambiente por **5 minutos**

F - Posicione a placa na base de **separação magnética**, espere de **5 a 10 minutos**, até total separação das beads.

G - Sem retirar a placa da base magnética, **retire e descarte o sobrenadante** (P200), tomando cuidado para não atingir as beads.

H - Ainda mantendo a placa na base magnética, **adicione 200uL de álcool 70%** em cada poço, espere 1 minuto para total separação e remova o álcool.

I - Repita os dois últimos passos (**G e H**) mais uma vez.

Dica: nós utilizamos uma cuba de plástico descartável para adicionar o álcool por meio de uma pipeta multicanal.

J - **Sele** a placa, retire-a da base magnética e **centrifugue-a** brevemente. **Retorne-a à base magnética**, retire o selo, **aguarde cerca de 30 segundos** e **remova todo resíduo** no fundo do poço com uma P20.

K - **Sem selá-la**, transfira a placa para um **termobloco a 37°C** e mantenha a placa **por 1 ou 2 minutos** ou até o pellet com as beads estiver visivelmente seco.

ATENÇÃO: Em dias de temperatura elevada ou baixa esse tempo pode diminuir ou aumentar, respectivamente. Fique atento ao estado dos pellets, não é aconselhado que ele seja seco demasiadamente, pois isso pode acarretar a diminuição da eficiência da eluição.

L - Adicione **35uL de água** ultrapura em cada poço com os beads, homogeneizando com **up and down por cerca de 15x**. Sele a placa e **centrifuge** brevemente. Por fim, incube a placa em temperatura ambiente por 2 minutos.

M - Posicione a placa na **base magnética** e aguarde **5 minutos** ou até a total separação da água e dos beads.

Observação: A ligação dos beads com os fragmentos de DNA funciona por meio de carga. A adição de água irá, portanto, causar a dissociação dos fragmentos das beads, sendo assim, o material genômico estará presente na água após a separação. Tome cuidado para não se confundir, a água NÃO deve ser descartada.

N - **Mantendo a placa na base magnética**, transfira o **eluente para uma nova placa**. A placa com os beads pode ser descartada.

ATENÇÃO:

- Se você está trabalhando com **dois tipos de amostras diferentes** (ex: FFPE e sangue) você deve transferir **cada tipo de amostra para uma placa diferente.** Não esqueça de identificar as placas.
- É provável que não seja possível capturar o volume total (35uL), por isso indica-se o uso de uma P20, coletando o máximo de sobrenadante possível.

3. Amplificando a biblioteca com os adaptadores

Para os passos a seguir você deve possuir os seguintes reagentes nas seguintes condições:

Tabela 6

Nome	Armazenar	descongelar	homogeneizar
Herculase II Fusion DNA Polymerase (tampa vermelha)	dentro do kit (-20°C)	descongelar e manter em gelo	up and down
5× Herculase II Reaction Buffer (tampa transparente)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar
100 mM dNTP Mix (tampa verde)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar
Forward Primer (tampa marrom)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar
SureSelect XT HS Index Primers A01 through H04 (tubos com tampa preta)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar

ATENÇÃO: Os index primers são utilizados para que cada amostra possa ser identificada após o pooling de amostras (etapas finais desse protocolo). Você deve escolher e **ANOTAR** qual index (cada tubo tem uma identificação diferente) você irá aplicar a qual amostra. Um mesmo index **NUNCA** deve ser aplicado para amostras diferentes de uma mesma biblioteca para sequenciamento.

A - Configure 2 (DOIS) termocicladores caso você esteja trabalhando com amostras de qualidade/tipo diferente (ex: FFPE e sangue), o número de ciclos no **Passo 2** no termociclador dependerá do tipo (sangue:DNA intacto; FFPE:DNA de menor qualidade) e da quantidade de material genômico utilizado. Veja tabelas a seguir para configurar corretamente os termocicladores:

Tabela 7

	n ciclos	temperatura	tempo

		(°C)	(min)
Passo 1	1	98	2
Passo 2	olhar tabela a seguir	98	30s
		60	30s
		72	1
Passo 3	1	72	5
Passo 4	1	4	hold

Para escolher o número de ciclos do Passo 2, olhar tabela a seguir:

Tabela 8

Qualidade DNA	Quantidade	N ciclos
DNA intacto (ex: sangue)	100 a 200ng	8
	50ng	9
	10ng	11
FFPE	100 a 200ng	11
	50ng	12
	10ng	14

Observações:

- Caso o aparelho requira a seleção do volume**, configure-o para 50uL**.
- Dê início à corrida e espere o lid alcançar 98°C, então imediatamente pause a corrida.
- Cuidado: o lid estará em alta temperatura.

B - Prepare o pre-capture PCR reaction mix de acordo com a tabela a seguir:

Tabela 9

Reagente	volume
5× Herculase II Reaction Buffer (tampa transparente)	10uL
100 mM dNTP Mix (tampa verde)	0.5uL
Forward Primer (tampa marrom)	2uL
Herculase II Fusion DNA Polymerase (tampa vermelha)	1uL

Observações:

- Não se esqueça de adicionar um erro na sua multiplicação;
- Com excessão da Herculase II Fusion DNA Polymerase (tampa vermelha), indica-se vortexar todos os reagentes antes de sua adição.

- Homogeneizar o mix por vortex.

C - Adicione **13,5uL do mix** (tabela 9) a cada poço com suas amostras;

D - Adicione **2uL dos index primers** selecionados para cada amostra. Homogeneizar por **up and down cerca de 15x**, **selar** as placas, **centrifugar** brevemente;

E - Adicione cada placa no seu termociclador e inicie a corrida.

4. **Segunda purificação das amostras por meio dos beads AMPure XP**

Observções:

- As beads devem ser usadas em temperatura ambiente;
- Nunca congele as beads;

A - Prepare **400uL de etanol 70%** para cada amostra, adicionando erro no cálculo.

ATENÇÃO: ao se preparar o álcool 70%, preste muita atenção para que a correta proporção seja misturada. A mistura errada pode gerar perda parcial ou total do material genômico durante essa etapa de purificação.

B - **Vortexar** o frasco contendo as beads até que se ele apresente homogeneidade em sua cor. Verifique o fundo do frasco, observando se não existem beads concentradas no local.

C - Separe **alíquota dos beads** em microtubo com volume de **50uL** multiplicado pela quantidade de amostras mais erro.

D - Adicione **50uL da alíquota dos beads** nos poços de cada amostra, misturando com **up and down por cerca de 20x.**

Dica: indica-se que o tubo com a alíquota dos beads seja vortexado em intervalos de tempo curtos durante a adição dele nos poços, pois as beads começam a se concentrar no fundo do tubo em pouco tempo.

E - **Incube as amostras** em temperatura ambiente por **5 minutos**

F - Posicione a placa na base de **separação magnética**, espere de **5 a 10 minutos**, até total separação das beads.

G - **Sem retirar a placa da base magnética**, retire e descarte o sobrenadante (P200) separado dos beads, tomando cuidado para não atingir as beads.

H - **Ainda mantendo a placa na base magnética**, adicione **200uL de álcool 70%** em cada poço, espere **1 minuto** para total separação **e remova o álcool.**

I - Repita o **passo H** mais uma vez.

Dica: nós utilizamos uma cuba de plástico descartável para adicionar o álcool por meio de uma pipeta multicanal.

J - **Sele a placa**, **retire-a da base magnética** e **centrifugue-a brevemente**. **Retorne-a à base magnética, retire o selo, aguarde cerca de 30 segundos** e **remova todo resíduo no fundo do poço com uma P20**.

K - **Sem selá-la**, transfira a placa para um termobloco a **37°C** e mantenha a placa por **1 ou 2 minutos ou até o pellet com as beads estiver visivelmente seco.**

ATENÇÃO: Em dias de temperatura elevada ou baixa esse tempo pode diminuir ou aumentar, respectivamente. Fique atento ao estado dos pellets, não é aconselhado que ele seja seco demasiadamente, pois isso pode acarretar a diminuição da eficiência da eluição.

L - Adicione **15uL de água ultrapura** em cada poço com os beads, homogeneizando com **up and down por cerca de 15x**. **Sele a placa e centrifuge brevemente.** Por fim, **incube** a placa em temperatura ambiente por **2 minutos**.

M - Posicione a placa na **base magnética** e aguarde **2-3 minutos** ou até a total separação da água e dos beads.

Observação: A ligação dos beads com os fragmentos de DNA funciona por meio de carga. A adição de água irá, portanto, causar a dissociação dos fragmentos das beads, sendo assim, o material genômico estará presente na água após a separação. Tome cuidado para não se confundir, a água NÃO deve ser descartada.

N - A seguir, aconselhamos a transferir o conteúdo de cada well para **microtubo de PCR**, pois isso facilitará alguns dos passos a seguir.

5. Verificando qualidade e quantidade nas amostras

Essa análise pode ser realizada por meio dos equipamentos 2100 Bionalyzer ou Agilent Tapestation. Em nosso laboratório nós usamos a metodologia e o equipamento Agilent Tapestation, seguindo o manual do fornecedor.

Os tubos usados no equipamento são unidos em colunas, portanto, não se esqueça de identificar e anotar qual das extremidades da coluna é a parte inicial ou final.

Anote em uma das extremidades do strip e na parede de um dos tubos da extremidade qual o lado inicial ou final.

Em cada tubo deve ser adicionado **3uL do Tapestation D1000 sample buffer** e **1uL de sua respectiva amostra**. Tampar os tubos como demonstrado na Figura 1 e vortexe as amostras no vórtex próprio para eles por 1 minuto. Centrifugue brevemente. Verifique se o mix se apresenta nos tubos de maneira homogênea. A figura a seguir exemplifica como o mix NÃO deve ser apresentado:

Após a corrida, o resultado da análise poderá ser visualizado no software próprio para o equipamento. Na visualização das curvas de densidade dos tamanhos dos fragmentos de DNA, amostras de boa qualidade (ex: sangue) devem possuir picos entre 300 e 400pb e 200 a 400pb para FFPE. O software também fornecerá um relatório com a concentração da amostra em ng por uL, dado esse que será usado na etapa a seguir.

Observação: Caso desejado, o protocolo pode ser pausado aqui e as amostras podem ser armazenadas em geladeira ou freezer. **GERALMENTE PARAMOS AQUI NO PRIMEIRO DIA.**

6. Hibridização e captura

A seguir iremos hibridizar os genes-alvo do nosso painel de genes com o DNA de nossas amostras, seguidamente capturando o material genômico com beads de streptavidina.

Para os passos a seguir você deve possuir os seguintes reagentes nas seguintes condições:

Tabela 10

Nome	Armazenar	descongelar	homogeneizar
SureSelect XT HS and XT Low Input Blocker Mix (tampa azul)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar
SureSelect RNase Block (tampa roxa)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar
SureSelect Fast Hybridization Buffer (garrafa)	dentro do kit (-20°C)	descongelar e manter em temperatura ambiente	vortexar
Biblioteca (probes para captura)	freezer -80°C	descongelar e manter em gelo	vortexar

Configure o termociclador:

Tabela 11

	n ciclos	temperatura (°C)	tempo (min)
Passo 1	1	95	5
Passo 2	1	65	10
Passo 3	1	65	1
-	1	65	hold
Passo 4	60	65	1
		37	3s
Passo 5	1	65	hold

Observações:

- Caso o aparelho requira a seleção do volume, configure-o para **30uL**.
- Dê início à corrida e espere o lid alcançar 95°C, então imediatamente pause a corrida.
- Cuidado: o lid estará em alta temperatura.
- Repare que o Passo 3 possui um hold. Isso se deve ao fato de que, durante o **Passo 3, a corrida deve ser pausada** para que reagentes adicionais sejam incluídos aos tubos (será detalhado a seguir). Após a adição dos reagentes, basta pular esse Passo para que a corrida continue normalmente.
- Caso seja desejado, as amostras podem ser mantidas no termociclador overnight em até 16h. Para isso, basta trocar a temperatura do Passo 5 por 21°C.

Partindo da concentração obtida por meio de análise no Tapestation, cada amostra deve ser preparada (novamente, **aconselhamos o uso de tubos de PCR ao invés de placas**, nessa etapa) de forma a ter de **500 a 1000ng de material genômico (quantidade) em volume final de 12uL**, podendo esse ser completado com água ultrapura. **Aconselha-se o uso de quantidade máxima (1000ng), se possível.**

Adicione **5uL de SureSelect XT HS and XT Low Input Blocker Mix (tampa azul)** em cada amostra. **Vortexar e centrifugar rapidamente.**

Adicione os tubos no **termociclador** e inicie a corrida.

Enquanto a corrida acontece, você terá cerca de **15 minutos para preparar os reagentes adicionais**. Não se preocupe, as amostras podem ficar no Passo 3 por alguns minutos adicionais.

Prepare a solução **de 25% de SureSelect RNase Block** como indicado na tabela a seguir:

Tabela 12

Reagente	Volume por amostra
SureSelect RNase Block (tampa roxa)	0,5uL
Água ultra pura	1,5uL

Observações:

- Não se esqueça de adicionar um erro na sua multiplicação;
- Vortexar rapidamente.

Prepare o **Capture Library Hybridization Mix** como descrito nas tabelas a seguir.

ATENÇÃO: as tabelas a seguir (13 e 14) demonstram 2 possibilidades de mix a serem feitos. O mix a ser escolhido dependerá do **tamanho de sua biblioteca**. É extremamente importante que se faça o mix apropriado para o tamanho de sua biblioteca.

Para bibliotecas com 3Mb ou mais:

Tabela 13

Reagente	volume
Solução de 25% de SureSelect RNase Block (tabela 12)	2uL
Probe (com design ≥3Mb)	5uL
SureSelect Fast Hybridization Buffer	6uL

Para bibliotecas com menos de 3Mb:

Tabela 14

Reagente	volume
Solução de 25% de SureSelect RNase Block (tabela 12)	2uL
Probe (com design <3Mb)	2uL
SureSelect Fast Hybridization Buffer	6uL
Água ultra pura	3uL

Observações:

- Não se esqueça de adicionar um erro na sua multiplicação;
- Vortexar rapidamente.

Tendo o termociclador alcançado o Passo 3, **SEM tirar os tubos do aparelho,** adicione **13uL do Capture Library Hybridization Mix** em cada amostra, misturando com **up and down 10x**;

Vortexar as amostras, centrifugar e recolocá-las no termociclador. Prosseguir com a corrida.

7. Preparação das beads magnéticas com estreptavidina

A etapa a seguir deve ser preparada enquanto as amostras estão no termociclador.

Verifique se você possui os seguintes reagentes nas seguintes condições:

Tabela 15

Nome	Armazenar	homogeneizar
SureSelect Binding Buffer	temperatura ambiente	vortexar
SureSelect Wash Buffer 1	temperatura ambiente	vortexar
SureSelect Wash Buffer 2	temperatura ambiente	vortexar
Dynabeads MyOne Streptavidin T1	geladeira	vortexar

Observções:

- As beads devem ser usadas em temperatura ambiente;
- Nunca congele as beads;
- Nessa etapa, **utilizamos bases magnéticas para microtubos de 1,5mL**. Caso opte por usar placa, o volume dos beads pode ser distribuído pelos poços.

A - Vortexar o frasco contendo as **Dynabeads MyOne Streptavidin T1** até que se ele apresente homogeneidade em sua cor. Verifique o fundo do frasco, observando se não existem beads concentradas no local.

B - Separe alíquota dos beads em microtubo com volume de **50uL multiplicado pela quantidade de amostras mais erro** (provavelmente será necessário dividir o volume em 2 microtubos ou mais).

Dica: indica-se que o tubo com a alíquota dos beads seja vortexado em intervalos de tempo curtos durante seu uso, pois as beads começam a se concentrar no fundo do tubo em pouco tempo.

C - Adicione 200uL do SureSelect Binding Buffer multiplicado pela quantidade de amostras.

Exemplo: suponha que você está trabalhando com 10 amostras. Você necessitará de 500uL de Dynabeads MyOne Streptavidin T1 e 2000uL de SureSelect Binding Buffer, resultando num volume final de 2500uL. No nosso laboratório, como usamos bases magnética para microtubos de 1,5mL nessa etapa, geralmente faz-se necessário dividir o volume total em diferentes microtubos, como por exemplo nesse caso, em 2 microtubos serão preparados 250 uL de Dynabeads MyOne Streptavidin T1 + 1250uL de SureSelect Binding Buffer (mais erro) em cada.

D - Coloque o(s) microtubo(s) no **separador magnético e espere por 5 minutos** ou até total separação dos beads, então **descarte o sobrenadante**.

E - Repita os últimos passos **(C e D) mais 2 vezes** (totalizando 3).

F - **Ressuspenda as beads em 200uL** de **SureSelect Binding Buffer** multiplicado pela quantidade de amostras.

G - Quando o programa no termociclador terminar, transfira as amostras para temperatura ambiente (aguarde que esfriem se necessário).

H - **Distribua 200uL dos beads recém preparadas** para **microtubos low bind** de 1,5uL de acordo com o número de amostras sendo trabalhadas, **identifique os tubos** conforme a identificação de suas amostras.

I - Transfira todo o conteúdo de cada uma de **suas amostras (cerca de 30uL)** para seu respectivo micrutubo com as beads. **Vortexar.**

J - **Incube** suas amostras misturadas com as beads por **30 minutos** a temperatura ambiente em termobloco e agitação de **1400-1800 rpm.**

K - Enquanto isso, **prepare o SureSelect Wash Buffer 2 pré-aquecido** seguindo os seguintes passos:

- Prepare **alíquotas do SureSelect Wash Buffer 2 em microtubos ou placa igual ao número de suas amostras multiplicado por 1200uL (+ erro).**
- Mantenha essas aliquotas em **70°C** (ex: termobloco para tubos).

L - **Quando a incubação de 30 minutos acabar, centrifugue** brevemente os tubos com suas amostras + beads.

M - Coloque os tubos com suas amostras + beads no **separador magnético** e espere até **total separação dos beads** e então **descarte todo o sobrenadante**.

N - **Retire os tubos do separador magnético, ressuspenda os beads em 200uL de SureSelect Wash Buffer 1com up and down por cerca de 20x.**

O - **Recoloque os tubos no separador, aguarde até total separação das beads (cerca de 1 minuto) e então descarte o sobrenadante.**

P - A seguir, será demonstrado o passo-a-passo para a aplicação do SureSelect Wash Buffer 2 pré-aquecido. **ATENÇÃO**, todas as amostras que passarem por esses passos devem ser colocadas a **70°C** no fim deles:

- **Ressuspender as beads em 200uL de SureSelect Wash Buffer 2 pré-aquecido e homogeneizar por vórtex e centrifugar brevemente** (cuidado para não formar pellets nos beads, se necessário use o vórtex mais uma vez para garantir que os beads estão em suspensão).
- **Incube as amostras a 70°C por 5 minutos.**
- **Bote as amostras no separador e aguarde até total separação das beads (cerca de 1 minuto)**
- **Repita os 3 últimos passos mais 5 vezes** (totalizando 6).

Q - Após total separação das beads, **retire todo o sobrenadante e adicione 25uL de água ultra pura. Homogeneíze com up and down por cerca de 8x.**

R - Transfira o conteúdo de cada amostra para seu respectivo **tubo de PCR.**

Observação: diferentemente das beads usadas anteriormente, o DNA continuará retido nas Dynabeads MyOne Streptavidin T1 após a eluição, sendo usadas nessa forma nos passos seguintes.

8. Amplifique as bibliotecas amplificadas

Verifique se você possui os seguintes reagentes nas seguintes condições:

Tabela 16

Nome	Armazenar	descongelar	homogeneizar
Herculase II Fusion DNA Polymerase (tampa vermelha)	dentro do kit (-20°C)	descongelar e manter em gelo	up and down
5× Herculase II Reaction Buffer (tampa transparente)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar
100 mM dNTP Mix (tampa verde)	dentro do kit (-20°C)	descongelar e manter em gelo	vortexar
SureSelect Post-Capture Primer Mix (clear cap)	dentro do kit (-20 ou -80°C)	descongelar e manter em gelo	vortexar

Veja as tabelas **(17 e 18)** a seguir para configurar corretamente o termociclador:

Tabela 17

	n ciclos	temperatura (°C)	tempo (min)
Passo 1	1	98	2
Passo 2	verificar tabela 18	98	30s
		60	30s
		72	1
Passo 3	1	72	5
Passo 4	1	4	hold

Observações:

- Caso o aparelho requira a seleção do volume, configure-o para 50uL.
- Dê início à corrida e espere o lid alcançar 98°C, então imediatamente pause a corrida.

Para escolher o número de ciclos no Passo 2 da tabela 17:

Tabela 18

Tamanho/Descrição biblioteca	n ciclos
Probes <0.2 Mb	14
Probes 0.2–3 Mb (inclui ClearSeq Comp Cancer)	12
Probes 3–5 Mb	10
Probes >5 Mb (inclui Human All Exon V6, Human All ExonV7 e Clinical Research Exome V2)	9

A - Prepare o mix de reação como descrito a seguir:

Tabela 19

Reagente	volume
Água ultra pura	12,5uL
5× Herculase II Reaction Buffer(tampa transparente)	10uL
Herculase II Fusion DNAPolymerase (tampa vermelha)	1uL
100 mM dNTP Mix (tampa verde)	0,5uL
SureSelect Post-Capture Primer Mix (tampa transparente)	1uL

B - Adicione **25uL do mix de PCR** em cada uma de suas amostras com a mistura dos beads. Homogeneizar por up and down.

ATENÇÃO: não vortexar nessa etapa.

C - Posicione suas amostras com o mix de PCR no termociclador configurado e inicie a corrida.

ATENÇÃO: após o uso do termociclador, os beads podem ser separados da amostra, adicionando as amostras em separador magnético, porém, esse passo é opcional e nós não o realizamos em nosso laboratório. Sendo assim, a próxima etapa, que se trata de purificação com as beads AMPure, ocorrerá junto com as Dynabeads MyOne Streptavidin T1 que já estão adicionadas nas amostras.

9. Terceira purificação das amostras por meio dos beads AMPure XP

Observções:

- As beads devem ser usadas em temperatura ambiente;
- Nunca congele as beads;

A - Prepare **400uL de etanol 70%** para cada amostra, adicionando erro no cálculo.

ATENÇÃO: ao se preparar o álcool 70%, preste muita atenção para que a correta proporção seja misturada. A mistura errada pode gerar perda parcial ou total do material genômico durante essa etapa de purificação.

B - **Vortexar** o frasco contendo as beads até que se ele apresente homogeneidade em sua cor. Verifique o fundo do frasco, observando se não existem beads concentradas no local.

C - Separe **alíquota dos beads** em microtubo com volume de **50uL** multiplicado pela quantidade de amostras mais erro.

D - Adicione **50uL da alíquota dos beads** em tubo low bind respectivo de cada amostra

E - Adicione suas amostras a cada um dos seus respectivos microtubos com as beads, misturando com **up and down por cerca de 20x.**

Dica: indica-se que o tubo com a alíquota dos beads seja vortexado em intervalos de tempo curtos durante a adição dele nos poços, pois as beads começam a se concentrar no fundo do tubo em pouco tempo.

F - **Incube as amostras** em temperatura ambiente por **5 minutos**

G - Posicione a placa na base de **separação magnética**, espere de **5 a 10 minutos**, até total separação das beads.

H - **Sem retirar os tubos da base magnética**, retire e descarte o sobrenadante (P200) separado dos beads, tomando cuidado para não atingir as beads.

I - **Ainda mantendo os tubos na base magnética**, adicione **200uL de álcool 70%** em cada poço, espere **1 minuto** para total separação **e remova o álcool.**

J - Repita o **passo H** mais uma vez.

K – **Retire os tubos da base magnética e centrifugue-a brevemente**. **Retorne-os à base magnética, aguarde cerca de 30 segundos e remova todo resíduo com uma P20.**

L – **Sem tampar**, transfira os tubos para um termobloco a **37°C** e mantenha-as por **1 ou 2 minutos** ou até que o pellet com as beads estiver visivelmente seco.

ATENÇÃO: Em dias de temperatura elevada ou baixa esse tempo pode diminuir ou aumentar, respectivamente. Fique atento ao estado dos pellets, não é aconselhado que ele seja seco demasiadamente, pois isso pode acarretar a diminuição da eficiência da eluição.

M - Adicione **25uL de água ultrapura** em cada microtubo com os beads, homogeneizando com **up and down por cerca de 15x. Centrifugue brevemente.** Por fim, **incube as amostras em temperatura ambiente por 2 minutos.**

N - Posicione as amostras na **base magnética e aguarde 2-3 minutos** ou até a total separação da água e dos beads.

Observação: A ligação dos beads com os fragmentos de DNA funciona por meio de carga. A adição de água irá, portanto, causar a dissociação dos fragmentos das beads, sendo assim, o material genômico estará presente na água após a separação. Tome cuidado para não se confundir, a água NÃO deve ser descartada.

O - A seguir, aconselhamos a transferir o conteúdo de cada microtubo para **novos microtubos low bind**.

10. Verificando qualidade e quantidade nas amostras

Essa análise pode ser realizada por meio dos equipamentos 2100 Bionalyzer ou Agilent Tapestation. Em nosso laboratório nós usamos a metodologia e o equipamento Agilent Tapestation, seguindo o manual do fornecedor.

Os tubos usados no equipamento são unidos em colunas, portanto, não se esqueça de identificar e anotar qual das extremidades da coluna é a parte inicial ou final.

Anote em uma das extremidades do strip e na parede de um dos tubos da extremidade qual o lado inicial ou final.

Em cada tubo deve ser adicionado **2uL do Tapestation High Sensitivity D1000 sample buffer** e **2uL de sua respectiva amostra**. Tampar os tubos como demonstrado na Figura 1 e vortexe as amostras no vórtex próprio para eles por 1 minuto. Centrifugue brevemente. Verifique se o mix se apresenta nos tubos de maneira homogênea. A figura a seguir exemplifica como o mix NÃO deve ser apresentado:

Após a corrida, o resultado da análise poderá ser visualizado no software próprio para o equipamento. Na visualização das curvas de densidade dos tamanhos dos fragmentos de DNA, amostras devem possuir picos entre 200 a 400pb. O software também fornecerá um relatório com a concentração da amostra

Observação: Caso desejado, o protocolo pode ser pausado aqui e as amostras podem ser armazenadas em geladeira ou freezer.

Apêndice B

Custom target-sequencing in triple-negative and luminal breast cancer from young Brazilian patients

Pedro Adolpho de Menezes Pacheco Serio[1], Daniela Marques Saccaro[1], Ana Carolina Ribeiro Chaves de Gouvêa[1], Giselly Encinas[2], Simone Maistro[1], Gláucia Fernanda de Lima Pereira[1], Larissa Dias de Souza[1], Viviane Jennifer da Silva[1], Maria Lucia Hirata Katayama[1], Maria Aparecida Azevedo Koike Folgueira[1]

Affiliation:

1 - Centro de Investigação Translacional em Oncologia, Departamento de Radiologia e Oncologia, Instituto do Cancer do Estado de Sao Paulo, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo 01246-000, Brazil.

2 – Agilent Technologies.

Authors' contribution

Pedro Adolpho de Menezes Pacheco Serio (https://orcid.org/0000-0001-9498-3658): Conceived the study, wrote the original draft, included patients, performed laboratorial and formal analysis and data curation, revised and approved the final manuscript;

Daniela Marques Saccaro (https://orcid.org/0000-0002-1668-7295): Conceived the study, included patients, performed laboratorial and formal analysis, revised and approved the final manuscript;

Ana Carolina Ribeiro Chaves de Gouvea (https://orcid.org/0000-0003-4363-6669): Included patients, revised and approved the final manuscript;

Giselly Encinas (https://orcid.org/0000-0002-2647-7400): Conceived the study, performed formal analysis, revised and approved the final manuscript;

Simone Maistro (https://orcid.org/0000-0002-0937-2391): Included patients, performed laboratorial analysis, revised and approved the final manuscript;

Gláucia Fernanda de Lima Pereira (https://orcid.org/0000-0002-8265-2808): Included patients, revised and approved the final manuscript;

Larissa Dias de Souza (https://orcid.org/0000-0003-4018-0926): Performed laboratorial and formal analysis, revised and approved the final manuscript;

Viviane Jennifer da Silva (https://orcid.org/0000-0002-5934-011X): Performed laboratorial and formal analysis, revised and approved the final manuscript;

Maria Lucia Hirata Katayama (https://orcid.org/0000-0002-2672-8222): Conceived the study, included patients, performed laboratorial and formal analysis, revised and approved the final manuscript;

Maria Aparecida Azevedo Koike Folgueira (https://orcid.org/0000-0001-9269-8052): Conceived the study, wrote the original draft, revised and approved the final manuscript;

Highlights:

- Tumors from young adults with breast cancer present different somatic variant profiles, according to its subtype.

- This data provides additional evidence that *PTPN13*, in TNBC and *SMURF2* and *PRKAR1A*, in luminal samples, are potential drivers in young patients.

Abstract

Objectives: To identify somatic mutations in tumors from young adults with triple-negative and luminal breast cancer, through targeted sequencing and to explore the driver potential of the identified genes and its variants. **Methods:** A customized gene panel was assembled based on data from previous sequencing studies in breast cancer from young women. Triple-negative and luminal tumors and paired blood samples from young breast cancer patients were sequenced and variants were investigated according to its driver potential. Additionally, we performed an exploratory analysis using large curated databases to evaluate the frequency of somatic mutations in this gene panel in tumors stratified by age groups (every 10 years). **Results:** A total of 28 young women had their tumoral tissue and blood samples sequenced. Using a customized panel consisting of 64 genes, we could detect potential driver genes in 10/12 TNBC tumor samples and 10/16 luminal samples. Among triple-negative breast cancer patients (n=12), the most frequent driver gene was *TP53,* and a potential driver gene was *PTPN13*. In luminal samples (n=16), *PIK3CA* and *GATA3* were the main drivers, and potential drivers were *SMURF2*

and *PRKAR1A*. The exploratory analysis also indicated the potential role of *SMURF2* in luminal BC development in young patients. **Conclusions:** Our data confirm that some drivers are more common in a specific breast cancer subtype from young patients, such as *TP53* in TNBC and *PIK3CA* and *GATA3* in luminal samples. These results also provide additional evidence that *PTPN13*, in TNBC and *SMURF2* and *PRKAR1A*, in luminal samples, are potential drivers in this age group.

Keywords: young adults, breast cancer, driver genes.

1. Introduction

Breast cancer is predominantly diagnosed in women with advanced age. Despite that, a population-based study carried out in the USA showed that breast cancer is the most common cancer in young women aged 30 to 39 years and the most lethal cancer type in this age range. This study also revealed disparities represented by a 14% higher breast cancer incidence rate in Black adolescent and young adults (AYA), mainly of triple-negative subtype, as compared with Caucasian AYA. In addition, death rates were also higher in Black AYAs than in Caucasian AYAs, reflecting a poorer prognosis associated with tumor aggressiveness and inequities in access to cancer care (Miller et al., 2020).

In Brazil, a work evaluating hospitals registry data from 188,753 Brazilian breast cancer patients against 922,962 breast cancer patients listed in the *National Cancer Institute* (SEER), suggested a higher prevalence of breast cancer in young patients in Brazil (odds ratio: 2.2). Comparing young and older patients from Sao Paulo State, in Brazil, more advanced tumors, higher rate of metastasis and shorter disease-specific survival were observed in the younger group, mainly in the triple-negative (TNBC) and luminal B tumors. Another Brazilian group analyzed public national mortality rates in 19105 young adult breast cancer women between 1996 and 2017 and reported an increase in breast cancer mortality in young adults in most Brazilian regions (1,2).

Other studies have shown that young patients may have a worse prognosis compared with older patients. Very young women (≤35 years), diagnosed with luminal B, HER2 and triple-negative breast cancer subtypes were shown to have a reduced overall survival and/or disease-free survival than their older counterparts. Similar results were reported in a large cohort of 1,732 triple-negative patients, where the median overall survival was 7 years for younger patients and 12 to 14 years in older age groups (3,4).

Some authors investigated the transcriptional tumor profile from young adult patients and reported enriched gene expression signatures related to stem cells, growth factors and immune system (4,5). More recent studies highlight the importance of the immune system in young women, suggesting that such patients may have a more active tumor microenvironment (TME) (6–8). Following these findings, some studies have also reported specific TME signatures in young adults with breast cancer (7,9).

There are indications that the TNBC subtype may present a higher number of cases in young adult patients than other subtypes (7,10,11). Among the TNBC subtypes, gene expression analysis suggests that LAR (luminal androgen receptor) subtype is less frequent in younger patients (8), showing that age may directly or indirectly influence the molecular profile (8). In addition, it was also shown that luminal tumors, mainly luminal B tumors, may have its prognosis influenced by age (12).

Our group recently explored the mutational profile of both TN and luminal breast cancer subtypes in young adults. In young TNBC patients, the median number of driver genes (following the genes cataloged at the Cancer Gene Census) in younger patients was three, which was twice lower compared to the elderly group. The majority (72%) of TNBC samples presented at least one affected oncogene in association with at least one tumor suppressor gene. The gene most frequently affected in young TNBC samples was *TP53*, detected in 70% of the tumors, followed by other oncogenes and tumor suppressor genes, such as *PIK3CA*, *KMT2C*, *and NF1*. Other potential driver tumor suppressor genes with likely loss-of-function variants were identified, such as *PHF6*, *GRIN2A*, *PIK3R1* and *MED12*. In addition, 20% of tumors had mutations in genes involved in the *RAS* or *PIK3CA* signaling pathways. There was a predominance of the Mutational Signature 3 in young patients, which is related

to homologous repair defects (HRD) (13). Other studies have also demonstrated an association of pathogenic somatic variants in *TP53* and mutational signatures in the HRD pathway in young patients (6,14,15).

In luminal samples from young women, the median mutation rate was 1.9/Mbp per patient sample and the most frequent event was C to T base transitions. Affected driver genes were *PIK3CA, TP53, PRKAR1A, SMURF2, POLD1* and *GATA3* and new identified potential drivers (genes which were not cataloged at CGC), were *GRHL2, PIK3AP1, CACNA1E* and *SEMA6D*. There was a predominance of mutational Signature 1 (age-related) (16).

Thus, breast cancer in young adults may present some peculiar and not well understood characteristics, which may differentiate their behavior from patients at more advanced ages. Therefore, our main goal was to identify somatic mutations in young adults with triple-negative and luminal breast cancer through targeted sequencing and to explore the driver potential of the identified variants.

Patients and methods

Patients

Patients were prospectively included at Instituto do Câncer do Estado de São Paulo (ICESP), São Paulo, Brazil, between the years 2016 and 2021. The study was approved by the Research Ethics Committee of the Faculdade de Medicina da Universidade de São Paulo and followed in accordance to the principles of the Helsinki Declaration.

Inclusion criteria were young female patients diagnosed with triple-negative or luminal (HER2 negative) breast cancer, aged 18 to 40 years at diagnosis, without previous cancer treatment, who had Formalin-Fixed Paraffin-Embedded (FFPE) tumor samples collected during biopsy or breast surgery available for the study. Patients who agreed to participate in the study and met the inclusion criteria, signed the informed consent, and donated a peripheral blood sample.

Customized gene panel

The genes chosen to integrate our customized panel comprises genes frequently altered in young adults with breast cancer, as well as oncogenes and tumor suppressor genes frequently affected in cancer in general. We also used specialized databases, such as the Catalog of Somatic Mutations in Cancer (COSMIC) to select genes by their frequency in breast cancer and relevance in literature (13,16). In summary, we:

I – selected studies in which the tumor of young adult breast cancer patients was sequenced;

II – selected genes that were altered in these studies and were cataloged in the Cancer Gene Census and/or with strong literature and/or with frequent truncated (frameshift, stop gain, canonical splice site) variants;

III – filtered the previously selected genes by its specific breast cancer frequency (>1%) in COSMIC, as described in the “Calculate Mutation Frequencies” in https://cancer.sanger.ac.uk/cosmic/help/faq;

IV – excluded genes in which a higher occurrence rate could be biased by its protein size or paralog numbers (17).

The final panel comprised a total of 6928 probes and size of 489 kbp, representing all coding regions of 64 genes and 10 bases of distance from the 3' and 5' extremities (GRCh37). Following the classification in the curated database Cancer Gene Census (**Supplementary Methods**), **Table 1** shows all the genes sequenced in our personalized panel, classified as tumor suppressor gene (TSG), oncogene (OG) and dual role (TSG or OG), as well as those previously shown to be mutated in breast cancer, but not yet reported as a cancer causing gene in cancer gene census database (CGC).

DNA extraction and sequencing

Patients who agreed to participate in the study had their respective tumor embedded in paraffin (Formalin-Fixed Paraffin-Embedded - FFPE) and peripheral blood samples DNA extracted using the *QIAamp DNA FFPE Tissue* (Qiagen - 56404) and *QIAamp DNA Mini Kit* (Qiagen – 51306), respectively, following manufacturer protocol.

Library preparation was conducted as described in the *SureSelectXT HS Target Enrichment System* for Illumina Multiplexed Sequencing Platforms protocol (Agilent) and sequenced on the NextSeq device (NextSeq 500/550 Mid Output Kit v2.5, 150 cycles; Illumina).

The median coverage of target areas was higher than 330 reads for all blood samples and 167 reads and 78 reads for TNBC and luminal tumor samples, respectively.

Data processing after sequencing

Trimerization, to remove low quality bases and adapters (*quality trimming*), alignment with the reference genome (BWA MEM; hg19), removal of duplicates, somatic CNV (Copy Number Variation) and variant call were performed using the SureCall software (v.4.2.2; Agilent) base settings.

Complementary data and analysis

All genes and variants detected were classified according to its driver potential. Genes were classified as oncogene, tumor suppressor gene (TSG) or dual-role gene (oncogene and TSG), following the classification of the curated Cancer Gene Census (CGC; COSMIC v.96; https://cancer.sanger.ac.uk/census) database.

OncoKB (https://www.oncokb.org/; Accessed January 2022) and *TP53* Database (https://tp53.isb-cgc.org/; Accessed January 2022) were used to classify variants according to its curated driver potential, as reported in the literature by functional studies. Based on both databases, variants were classified as causing likely loss-of-function (L-LOF), loss-of-function (LOF), likely gain-of-function (L-GOF) or gain-of-function (GOF). Complementarily, we used the recently published *CancerVar* tool (https://cancervar.wglab.org/) which integrates clinical guidelines and deep learning to classify the pathogenicity of somatic variants (18).

Germline variants were classified as benign, likely-benign, uncertain significance (VUS), likely-pathogenic or pathogenic, following the classification

method proposed by ACMG (19). Also, data reported in the study published by Findlay and colleagues (20) was accounted for when applying ACMG classification criteria to the *BRCA1* germline variants.

Missense variants in genes cataloged in CGC but with no report about its functional effect in cancer (OncoKB and TP53 Database) or missense variants in genes not cataloged in CGC and with no reported functional effect were classified according to its damaging scores in in-silico variant effect prediction tools and in in-silico tools that use methodologies for compiling functional impact prediction results. Missense variants were classified as possibly pathogenic (PP) if they were predicted to be damaging in at least 4 out of 8 prediction tools and 2 out of 4 compilation tools or in at least 2 out of 8 prediction tools and 1 out of 4 compilation tools if the gene in which it was found was cataloged as a CGC gene **(Supplementary methods)**.

***In-silico* analysis of tumor data from pre-menopausal and elderly breast cancer patients**

The Cancer Browser Tool from COSMIC (https://cancer.sanger.ac.uk/cosmic/browse/tissue; v.96; Accessed May 2022) and the cohort comparison tools from the cBioPortal (https://www.cbioportal.org/; Accessed June 2022) were used to perform exploratory analyses, integrating data from multiple breast cancer exome and genome sequencing studies. Data from COSMIC was directly downloaded from the portal (https://cancer.sanger.ac.uk/cosmic/download; Accessed: May 2022) and processed in R (v. 4.1.2). cBioPortal data was mostly analyzed by its cohort comparison tools. The p-values obtained from COSMIC analysis (*Fisher's Exact Test*) were adjusted (adj.p) with the *Bonferroni* correction, while the results obtained by the cBioPortal built-in analysis (through *Kruskal-Wallis, Chi-Squared* and *Student's t-test*) workflow were corrected with the *Benjamini-Rochberg* procedure.

Since most of the studies did not have CNV and/or gene expression and/or protein expression data, these types of data were only investigated in the TCGA breast cancer cohort.

Detailed quality control, data processing, filtering and complementary annotations are described in the **Supplementary Methods**.

Results

Characteristics of the patients

In total, 28 young women were analyzed, including 12 TNBC, 13 luminal B and 3 luminal A breast cancer patients. All patients were diagnosed with invasive ductal carcinoma, with a median age at diagnosis of 33 years. Most tumors were histological grade 3 (61%) and 2 (35%). Among the patients, 71% (20/28) reported a family history (FH) of any cancer (until third-degree relatives), including 32% (9/28) who had a positive FH of at least one relative (until third-degree relatives) with a diagnosis of breast, ovary, pancreas, or prostate cancer (**Table 2**).

Germline variants

Among the 64 genes chosen to integrate our panel, seven are known clinically actionable genes, listed in the National Cancer Comprehensive Network (NCCN) guidelines for Genetic/Familial High-Risk Assessment in breast ovarian and pancreatic cancer (v 2.2022): *ATM, BRCA1, BRCA2, CDH1, NF1, PTEN* and *TP53*.

Among TNBC patients, considering the seven genes reported in the NCCN panel and applying ACMG classification criteria, a likely-pathogenic variant, *BRCA1* (c.245T>G), was detected in one patient, who did not report family history of cancer. Half of the TNBC patients presented a VUS (variant of uncertain significance), which were identified in *APC, ATM, BRCA2* and in other genes not reported in the NCCN panel: in *ATXN1, CSPP1* and *MET* (Supp. Table 1; Supp. Fig. 1).

Three luminal breast cancer patients presented a pathogenic germline variant in genes listed in the NCCN panel: two patients presented one pathogenic variant in *BRCA1*, with one frameshift (c.5266dup; p.Q1777fs) and a splice-site variant (c.441+2T>A). Another patient presented a stop-codon gain

in *BRCA2* (c.250C>T). None of these three patients reported family history of cancer. The VUS (variant of uncertain significance) rate in luminal patients was 37.5% (6/16), reported in *APC, ATM, CDH1* and in other genes not reported in the NCCN panel: *FAT4, MTOR, NOTCH1* and *POLD1* (Supp. Table 1; Supp. Fig. 1).

Profile of somatic variants in TNBC and luminal samples: general view

The median number of somatic variants per sample was 2 for TNBC cases and 1 for luminal cases. In one TNBC sample (703) and 4 luminal samples (731, 812, 818 and 823) neither somatic single nucleotide variant nor indels were detected, although somatic CNVs were detected in the tumor sample 818.

Considering both TN and luminal samples, somatic variants were detected in a total of 35 different genes: *AHNAK, AKT1, ANLN, ARID1A, ATR, ATXN1,BAP1, BRCA1, CACNA1E, CAMK1G, CDH1, CSMD3, CSPP1, ERBB2, FAT2, FBXW7, GATA3, GRHL2, MAP2K4, MET, NCOA3, NF1, NOTCH1, PIK3CA, PRKAR1A, PTEN, PTPN13, RAD51, SEMA6D, SMARCA4, SMURF2, SPEN, TNC, TP53* and *UBR5.* Among all samples, irrespective of the subtype, *TP53* was altered in 10 out of 28 tumor samples (36%), followed by *GATA3* in 5 samples (18%) and *PIK3CA*, in 4 samples (14%) **(Figure 1; Table 3; Supp. Table 2)**.

Among TNBC patients, *TP53* was the most frequently affected gene, altered in almost all patients (9/12), followed by *NF1* (3/12). All *TP53* variants were classified as likely (L-LOF) or loss-of-function (LOF) causing, according to curated databases (OncoKB and/or *TP53* Database). Also, the tumor suppressor gene (TSG) *PTPN13* was exclusively altered in TNBC patients (2/12), although none of the variants had a loss or gain-of-function (GOF) effect **(Figure 1; Table 3; Supp. Table 2)**.

The most frequently affected genes among the luminal samples were *GATA3* (5/16) and *PIK3CA* (4/16). All *PIK3CA* variants detected in luminal tumors were reported as GOF causing in literature. Only one luminal breast

cancer patient had a variant in *TP53*. The genes *AHNAK* and *TNC* were exclusively altered in luminal tumors (3/16; 2/16), however, neither variant was cataloged as loss or gain-of-function causing, or as possibly pathogenic **(Figure 1; Table 3; Supp. Table 2).**

Potential driver genes

The main goal was to identify known somatic driver genes as well as new potential drivers, the former reported and the latter, not yet reported as CGC cataloged genes or affected by variants with functional consequences (OncoKB, *TP53* database). This analysis was performed specifically in each BC subtype, separately.

Among TNBC, 10/12 samples had at least one potential driver gene. The most frequent driver was *TP53*, (9/12), in which all detected variants were classified as likely (L-LOF) or loss-of-function (LOF) causing (Figure 1; Supp. Table 2). The second most frequent driver was *NF1* (3/12), in which 2 variants were reported as causing functional consequence (L-LOF and LOF) and one missense variant had no functional evidence, although it was classified as a possibly pathogenic variant (according to the classification of variant effect prediction tools) **(Figure 1; Table 3; Supp. Table 2)**.

Potential driver genes were also identified in the TSGs *BRCA1*, *FBXW7* and *PTEN*. No oncogene with functional evidence of pathogenicity was identified, although one patient presented a possibly pathogenic missense variant in the oncogene *MET* and other TNBC sample present a possibly pathogenic missense variant in *PIK3CA* and *UBR5* (both classified as oncogenes).

Most samples presented two potential driver genes, represented by one dual role (*TP53*) associated with one TSG, one oncogene or other dual-role gene **(Table 3)**. Sample 608 presented no driver genes, only a missense variant in ATR with no functional evidence in literature or prediction of damage in *in-silico* tools.

In luminal samples, the most frequent driver genes were *PIK3CA* (4/16) and *GATA3* (3/16), in which all detected variants had reported functional evidence (OncoKB). Two patients had a likely-LOF and a LOF variant in *NF1*, and another patient had a L-LOF variant in *CDH1*. Although two patients also had a variant in *NOTCH1*, only one of the variants had functional evidence of being deleterious in the literature **(Figure 1; Table 3; Supp. Table 3)**. Possibly pathogenic missense variants (*in-silico* variant effect prediction tools) were identified in the TSG *PTEN* and in the *AKT1* oncogene. A possibly pathogenic variant was also detected in *CACNA1E*.

Exploratory analysis

We used COSMIC and cBioPortal curated data to explore the relevance of alterations in the 64 genes evaluated in the present work, in other tumors. These databases were used to compile data from premenopausal and elderly patients from multiple breast cancer exome and genome sequencing projects. Patient tumor data was compared between young adults (diagnosed until 40 years old) and older adults, with 5 diagnosis age groups (≤40, 41-50, 51-60, 61-70 and >70) to compare mutation rates. We also investigated CNV, gene expression, protein expression and methylation among age groups in the TCGA Breast Cancer Firehose Legacy Cohort (TCGA-BRCA) cohort, dividing tumor samples as described above.

At first, we explored whether there were differences in the frequency of point and truncated variants (missense, inframe, nonsense, frameshift and splice-site) between young and older patients in the genes integrating the personalized panel of the present study (even those where no variant was detected among the samples sequenced in the present study). We accessed breast cancer data available at COSMIC Mutation Data (Genome Screens) cohort (https://cancer.sanger.ac.uk/cosmic/download; Accessed: May 2022), comprising 2672 samples from multiple studies. Next, we excluded samples from case-report and cell-line studies, as well as samples of adenoid, acinic, neuroendocrine, metaplastic and non-primary carcinoma, and samples with no information of age at diagnosis, finishing with data from 184 young adults, 265

patients diagnosed between 41 and 50 years old, 316 patients between 51 and 60, 307 patients between 61 and 70, and 225 diagnosed at 71 or higher, comprising 7 different studies (16,21–26). No tumor subtype distinction was done, since this data was missing in some studies.

We then created a contingency table to compare the frequency of affected samples between young adults (≤40) and elderly patients stratified in age groups (41-50, 51-60, 61-70 and >70) for each of the 64 genes of our panel, to verify if any of those genes presented higher frequency of point and truncated variants in the younger group. Among the 64 genes of our panel, none of the age group comparisons revealed a differential frequency of point and truncated variants in the young adult group. However, *CDH1* point and truncated variants were more frequently altered in all elderly groups compared to young adults (adj.p<0.001). Similarly, *MAP3K1* presented higher mutation frequency in almost all the elderly groups (excluding 41-50) in comparison to the young adults group (adj.p<0.05).

A higher frequency of variants in *TP53* (adj.p<0.05) was identified in the elderly groups with less advanced ages (41-50 and 51-60) in comparison with the young adult group, but the same was not visualized in the age groups with the most advanced ages at diagnosis (61-70 and >70). The inverse was observed in *PIK3CA*, where only the group with the most advanced age (>70) presented higher variant frequency (adj.p=0.003) in comparison to the young adults group.

Next, we used the cohort comparison workflow from cBioPortal to perform a similar analysis, this time with the isolated TCGA-BRCA (breast cancer) cohort, comparing gene and protein expression, copy-number variation and methylation profile between young adults and elderly patients. The cohort comprised 76 young adults and 992 elderly primary breast carcinoma patients, with the latter being divided in the following age groups: 41-50 (n=230), 51-60 (n=273), 61-70 (280) and >70 (n=209) (21). No tumor subtype distinction was done since this data was missing in some patients and further subgroup divisions would weaken the power of statistical analysis.

When comparing young adults with all the elderly groups (41-50, 51-60, 61-70 and >70), young adult group presented higher frequency of copy-number gain in *SMURF2* and *PRKAR1A* (adj.p<0.05) and were observed with higher gene expression (adj.p<0.05) in young adults, when compared to the most advanced age groups (61-70 and >70). Finally, we observed a higher (adj.p<0.05) *CDH1* gene expression and protein expression in young adults when compared to all the elderly subgroups. In addition, a higher gene expression in young adults when compared to the >70 group was observed in the following genes: *ARID1A, ANLN, ATAD2B, FAT2, FAT4, FBXW7, MET, MTOR, PARP4, PIK3CA* and *RAD51*.

Discussion

We conducted targeted-sequencing analysis in the tumor and paired blood samples of 28 young adult breast cancer patients, mainly focusing on the somatic variants in TNBC and luminal samples, separately. The most frequently altered gene in TNBC was *TP53*, followed by *NF1*, while *PIK3CA* and *GATA3* were the most affected genes in luminal tumors.

We identified 4 out of 28 patients with a likely or pathogenic germline mutation in *BRCA1* or *BRCA2* (14%), which goes in accordance with other published articles, including one published by our group in which we detected 16% of pathogenic *BRCA1* and *BRCA2* mutations in 79 young adults luminal tumors (10,16,27). None of the four patients presented a family history of breast, ovary, prostate or pancreatic cancer.

Considering somatic variants, most *TP53* variants were classified as a L-LOF or LOF causing variant and almost exclusively detected in TNBC patients. Similarly, *PIK3CA* and *GATA3* mutations were mostly detected in luminal tumors, which goes in accordance with the frequencies reported in large breast cancer molecular profiling publications (21,25).

The tumor from patient 701 was the only TNBC case with a detected somatic *PIK3CA* variant, which was classified as possibly pathogenic (damaging in 11/12 tools). The same tumor also had a possibly pathogenic

somatic variant detected in the *UBR5* oncogene and a L-LOF CNV gain in the *FBXW7* tumor suppressor gene.

The tumor sample from patient 837 was the only luminal sample with a *GATA3* somatic variant in which the variant (a CNV gain) was reported as likely GOF causing, while in other luminal samples *GATA3* variants were reported as L-LOF. Although *GATA3* is mainly cited as a TSG, there is still discussion in the literature and reports that *GATA3* roles may also influence oncogenic pathways (28,29). Thus, classification of the impairing mutations in *GATA3* is a challenging task.

Among all the somatic variants detected in *TP53*, only two samples (tumor samples 700 and 821) had no direct evidence to cause loss-of-function in the literature. However, both variants are likely to induce the transcription of a truncated protein at the protein's DNA binding site domain, where indel variants and even variants that cause only a single nucleotide change (missense) are likely to cause loss-of-function (30).

The alteration in *PTPN13* gene was exclusively detected in TNBC tumors. The gene encodes a member of the protein tyrosine phosphatase (PTP) family and is classified as a tumor suppressor gene, with roles in apoptosis signaling regulation. Studies show that *PTPN13* dysfunction in *in-vivo* and *in-vitro* TNBC models lead to enhanced malignant growth and invasiveness (31). Interestingly, one tumor sample (611) had a stop-codon variant in *PTPN13*, localized at the FERM-c domain. Although to our knowledge there is no study that investigated the functional impact of this specific variant, studies show that the disruption of PDZ (protein-protein interaction) and protein tyrosine-phosphatase domains in the protein encoded by *PTPN13* would disrupt its main antitumoral functions (32). Additionally, the gene is frequently epigenetically inhibited in breast cancer (33). These facts highlight the relevance of this gene in breast cancer, although not necessarily exclusively in young adults, as it is not shown to be highly altered in the present study nor differentially altered, expressed, or methylated between age groups in our exploratory analysis.

Although most samples presented at least one variant in an oncogene or TSG with functionality disruption evidence, some samples did not present a clearer profile of mutagenesis. Interestingly, the tumor sample 818 only presented somatic CNV gain in the *CSPP1* and *NCOA3* genes, which encodes proteins with main roles in cell-cycle and transcriptional regulation, respectively. This gene was previously shown to be involved in cancer, however, it is not yet classified as TSG or oncogenes in breast cancer. Some studies show that the overexpression of both genes can assist the progression of multiple cancer types, including breast cancer (34,35). Despite that, our exploratory analysis did not show evidence of differential expression between age groups in these genes.

Some other variants detected in the tumors from the participants of the present study deserve more detailed functional study. For example, the tumor sample from patient 701 has a somatic missense variant in the *PIK3CA* gene (E707K), which has no specific classification in the literature regarding its pathogenicity, despite being classified as VUS in the germline context. However, the variant is classified as pathogenic in almost all individual tools and compilation of effect prediction tools (7/8 and 4/4, respectively). Furthermore, the variant is classified as Tier 2 (possibly pathogenic) by the somatic variant classification method developed by Li and colleagues (18). Similarly, we identified a somatic missense variant in *NOTCH1* (R2087W) in the tumor sample from patient 702, which has no specific classification in the literature regarding its pathogenicity, despite being classified as VUS in the germline context (ACMG). However, studies indicate that upstream variants of the transactivation domain of the encoded protein can cause loss-of-function (36). Additionally, the variant is classified as pathogenic in almost all individual and compilation of effect prediction tools (7/8 and 2/4, respectively). Furthermore, the variant is classified as Tier 2 (possibly pathogenic) by the somatic variant classification method developed by Li and colleagues. Finally, we identified a possibly pathogenic somatic missense variant in the *MAP2K4* gene in the tumor from patient 730. Although the same sample presented a loss-of-function TP53 mutation and a likely-loss-of-function variant in the *PTEN* gene, the variant in

MAP2K4 cannot be ignored. Studies suggest that disruptive mutations in the gene can sensibilize malignant cells to MEK inhibitors (37).

Finally, we conducted exploratory analysis using tools from large cancer databases to verify if the genes investigated in our panel could separate age groups according to its molecular profiles. Our observation confirms previous reported results showing an inverse relation between age, *TP53* and *PIK3CA*. When comparing young adults with elderly groups, while *TP53* variant frequency is higher in age groups 41-50 and 51-60, a higher frequency in *PIK3CA* is only observed when we compare tumors from young with tumors from patients with the most advanced age at diagnosis (>70).

Interestingly, we identified *SMURF2* and *PRKAR1A* with higher amplification frequency in young adult breast cancer patients when compared to elderly age groups (BRCA-TCGA). Moreover, a CNV gain in *SMURF2* was detected in one of the TNBC samples analyzed in the present study. Although *SMURF2* is not cataloged in CGC yet, some authors showed that the gene can influence the tumorigenesis of some cancer types (38). Similarly, in another study from our group in which we investigated somatic variants in young adult luminal breast cancer patients, variants in *PRKAR1A* and *SMURF2* were identified as possible cancer drivers (16). SMURF2 also presented higher gene expression in the BRCA-TCGA young adult group, showing that the gene can have a relevant role in this group, although further investigation is clearly needed.

CDH1 was identified with higher gene and protein expression in tumors from young adult patients, indicating that the modulation of its expression may notably affect its protein activity in this group of patients, which could possibly affect epithelial-mesenchymal transition processes. This gene is more frequently affected by point and truncating somatic mutations, with inactivating potential in tumors from elderly patients. In fact, other studies reported *CDH1* mutations as a marker in tumors from patients with more advanced ages, as well as normally associated with lobular tumors, which are infrequent in younger patients (39).

The present report presents some limitations, such as the small sample size and the analyses of a limited number of genes. Considering large breast cancer cohorts, such as the TCGA-BRCA study, in which approximately 80 tumor samples were from young patients, we conclude that the present project has contributed with a significant number of tumors, given the poor general representation of this population in other studies. Besides that, most studies report data from patients with European ancestry, therefore, we believe this is relevant data reporting on tumors from an admixed ethnic population.

Conclusion

Our data confirm that some drivers are more common in a specific breast cancer subtype from young patients, such as *TP53* in TNBC and *PIK3CA* and *GATA3* in luminal samples. These results also provide additional evidence that genes such as *PTPN13*, in TNBC and *SMURF2* and *PRKAR1A*, in luminal samples, are potential drivers in this age group.

Conflict of interests

Giselly Encinas is an employee of Agilent Technologies. All other authors declare no conflict of interest.

Funding

Pedro Adolpho de Menezes Pacheco Serio and Daniela Marques Saccaro received a PhD scholarship grant by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Simone Maistro received a postdoctoral scholarship by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, PVE #029/2012), Giselly Encinas received a PhD scholarship grant by São Paulo Research Foundation (FAPESP, #2011/09572-1). Larissa Dias de Souza and Viviane Jennifer da Silva received a specialization grant by Secretaria de Estado da Saúde do Estado de São Paulo. São Paulo Research Foundation (FAPESP, grant #2012/12306-4) supported this work.

References

1. Silva JDDE, de Oliveira RR, da Silva MT, Carvalho MD de B, Pedroso RB, Pelloso SM. Breast cancer mortality in young women in brazil. Front

Oncol. 2020;10:569933.

2. Orlandini LF, Antonio MV do N, Espreafico CR, Bosquesi PL, Poli-Neto OB, de Andrade JM, et al. Epidemiological analyses reveal a high incidence of breast cancer in young women in brazil. JCO Glob Oncol. 2021 Jan;7:81–8.

3. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. Ann Oncol. 2010 Oct;21(10):1974–81.

4. Liedtke C, Rody A, Gluz O, Baumann K, Beyer D, Kohls E-B, et al. The prognostic impact of age in different molecular subtypes of breast cancer. Breast Cancer Res Treat. 2015 Aug;152(3):667–73.

5. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? J Clin Oncol. 2011 Jan 1;29(1):e18-20.

6. Kan Z, Ding Y, Kim J, Jung HH, Chung W, Lal S, et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. Nat Commun. 2018 Apr 30;9(1):1725.

7. Azim HA, Partridge AH. Biology of breast cancer in young women. Breast Cancer Res. 2014 Aug 27;16(4):427.

8. Aine M, Boyaci C, Hartman J, Häkkinen J, Mitra S, Campos AB, et al. Molecular analyses of triple-negative breast cancer in the young and elderly. Breast Cancer Res. 2021 Feb 10;23(1):20.

9. Azim HA, Michiels S, Bedard PL, Singhal SK, Criscitiello C, Ignatiadis M, et al. Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. Clin Cancer Res. 2012 Mar 1;18(5):1341–51.

10. Carraro DM, Koike Folgueira MAA, Garcia Lisboa BC, Ribeiro Olivieri EH, Vitorino Krepischi AC, de Carvalho AF, et al. Comprehensive analysis of BRCA1, BRCA2 and TP53 germline mutation and tumor characterization: a portrait of early-onset breast cancer in Brazil. PLoS ONE. 2013 Mar 1;8(3):e57581.

11. Kim J, Hong S, Lee JJ, Won Y-J, Lee ES, Kang H-S, et al. Analysis of the tumor characteristics in young age breast cancer patients using collaborative stage data of the Korea Central Cancer Registry. Breast Cancer Res Treat. 2021 Feb 18;

12. Keegan THM, Press DJ, Tao L, DeRouen MC, Kurian AW, Clarke CA, et al. Impact of breast cancer subtypes on 3-year survival among adolescent and young adult women. Breast Cancer Res. 2013;15(5):R95.

13. Serio PA de MP, de Lima Pereira GF, Katayama MLH, Roela RA, Maistro S, Folgueira MAAK. Somatic Mutational Profile of High-Grade Serous Ovarian Carcinoma and Triple-Negative Breast Carcinoma in Young and Elderly Patients: Similarities and Divergences. Cells. 2021 Dec 20;10(12).

14. Midha MK, Huang Y-F, Yang H-H, Fan T-C, Chang N-C, Chen T-H, et al. Comprehensive cohort analysis of mutational spectrum in early onset breast cancer patients. Cancers (Basel). 2020 Jul 28;12(8).

15. Mealey NE, O'Sullivan DE, Pader J, Ruan Y, Wang E, Quan ML, et al. Mutational landscape differences between young-onset and older-onset breast cancer patients. BMC Cancer. 2020 Mar 12;20(1):212.

16. Encinas G, Sabelnykova VY, de Lyra EC, Hirata Katayama ML, Maistro S, de Vasconcellos Valle PWM, et al. Somatic mutations in early onset luminal breast cancer. Oncotarget. 2018 Apr 27;9(32):22460–79.

17. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJY, van Karnebeek C, Wasserman WW. FLAGS, frequently mutated genes in public exomes. BMC Med Genomics. 2014 Dec 3;7:64.

18. Li Q, Ren Z, Cao K, Li MM, Wang K, Zhou Y. CancerVar: An artificial intelligence-empowered platform for clinical interpretation of somatic mutations in cancer. Sci Adv. 2022 May 6;8(18):eabj1624.

19. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405–24.

20. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018 Oct;562(7726):217–22.

21. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct 4;490(7418):61–70.

22. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012 Apr 4;486(7403):395–9.

23. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012 May 25;149(5):979–93.

24. Vaca-Paniagua F, Alvarez-Gomez RM, Maldonado-Martínez HA, Pérez-Plasencia C, Fragoso-Ontiveros V, Lasa-Gonsebatt F, et al. Revealing the Molecular Portrait of Triple Negative Breast Tumors in an Understudied Population through Omics Analysis of Formalin-Fixed and Paraffin-Embedded Tissues. PLoS ONE. 2015 May 11;10(5):e0126762.

25. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016 Jun 2;534(7605):47–54.

26. Korea NCC. National Cancer Center Korea - BRCA-KR (ICGC). Available from: https://dcc.icgc.org/projects/BRCA-KR

27. Copson ER, Maishman TC, Tapper WJ, Cutress RI, Greville-Heygate S, Altman DG, et al. Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. Lancet Oncol. 2018 Feb;19(2):169–80.

28. Takaku M, Grimm SA, Wade PA. GATA3 in breast cancer: tumor suppressor or oncogene? Gene Expr. 2015;16(4):163–8.

29. Mair B, Konopka T, Kerzendorfer C, Sleiman K, Salic S, Serra V, et al. Gain- and Loss-of-Function Mutations in the Breast Cancer Gene GATA3 Result in Differential Drug Sensitivity. PLoS Genet. 2016 Sep 2;12(9):e1006279.

30. Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. Why are there hotspot mutations in the TP53 gene in human cancers? Cell Death Differ. 2018 Jan;25(1):154–60.

31. Hamyeh M, Bernex F, Larive RM, Naldi A, Urbach S, Simony-Lafontaine J, et al. PTPN13 induces cell junction stabilization and inhibits mammary tumor invasiveness. Theranostics. 2020 Jan 1;10(3):1016–32.

32. Freiss G, Chalbos D. PTPN13/PTPL1: an important regulator of tumor aggressiveness. Anticancer Agents Med Chem. 2011 Jan;11(1):78–88.

33. Ying J, Li H, Cui Y, Wong AHY, Langford C, Tao Q. Epigenetic disruption of two proapoptotic genes MAPK10/JNK3 and PTPN13/FAP-1 in multiple lymphomas and carcinomas through hypermethylation of a common bidirectional promoter. Leukemia. 2006 Jun;20(6):1173–5.

34. Sternemalm J, Russnes HG, Zhao X, Risberg B, Nord S, Caldas C, et al. Nuclear CSPP1 expression defined subtypes of basal-like breast cancer. Br J Cancer. 2014 Jul 15;111(2):326–38.

35. Gupta A, Hossain MM, Miller N, Kerin M, Callagy G, Gupta S. NCOA3 coactivator is a transcriptional target of XBP1 and regulates PERK-eIF2α-ATF4 signalling in breast cancer. Oncogene. 2016 Nov 10;35(45):5860–71.

36. Gerhardt DM, Pajcini KV, D'altri T, Tu L, Jain R, Xu L, et al. The Notch1 transcriptional activation domain is required for development and reveals a novel role for Notch1 signaling in fetal hematopoietic stem cells. Genes Dev. 2014 Mar 15;28(6):576–93.

37. Xue Z, Vis DJ, Bruna A, Sustic T, van Wageningen S, Batra AS, et al.

MAP3K1 and MAP2K4 mutations are associated with sensitivity to MEK inhibitors in multiple cancer models. Cell Res. 2018 Jul;28(7):719–29.

38. David D, Jagadeeshan S, Hariharan R, Nair AS, Pillai RM. Smurf2 E3 ubiquitin ligase modulates proliferation and invasiveness of breast cancer cells in a CNKSR2 dependent manner. Cell Div. 2014 Aug 31;9:2.

39. Encinas G, Maistro S, Pasini FS, Katayama MLH, Brentani MM, Bock GH de, et al. Somatic mutations in breast and serous ovarian cancer young patients: a systematic review and meta-analysis. Rev Assoc Med Bras. 2015;61(5):474–83.

Figure 1

Click here to access/download;Figure;Figure 1.jpg

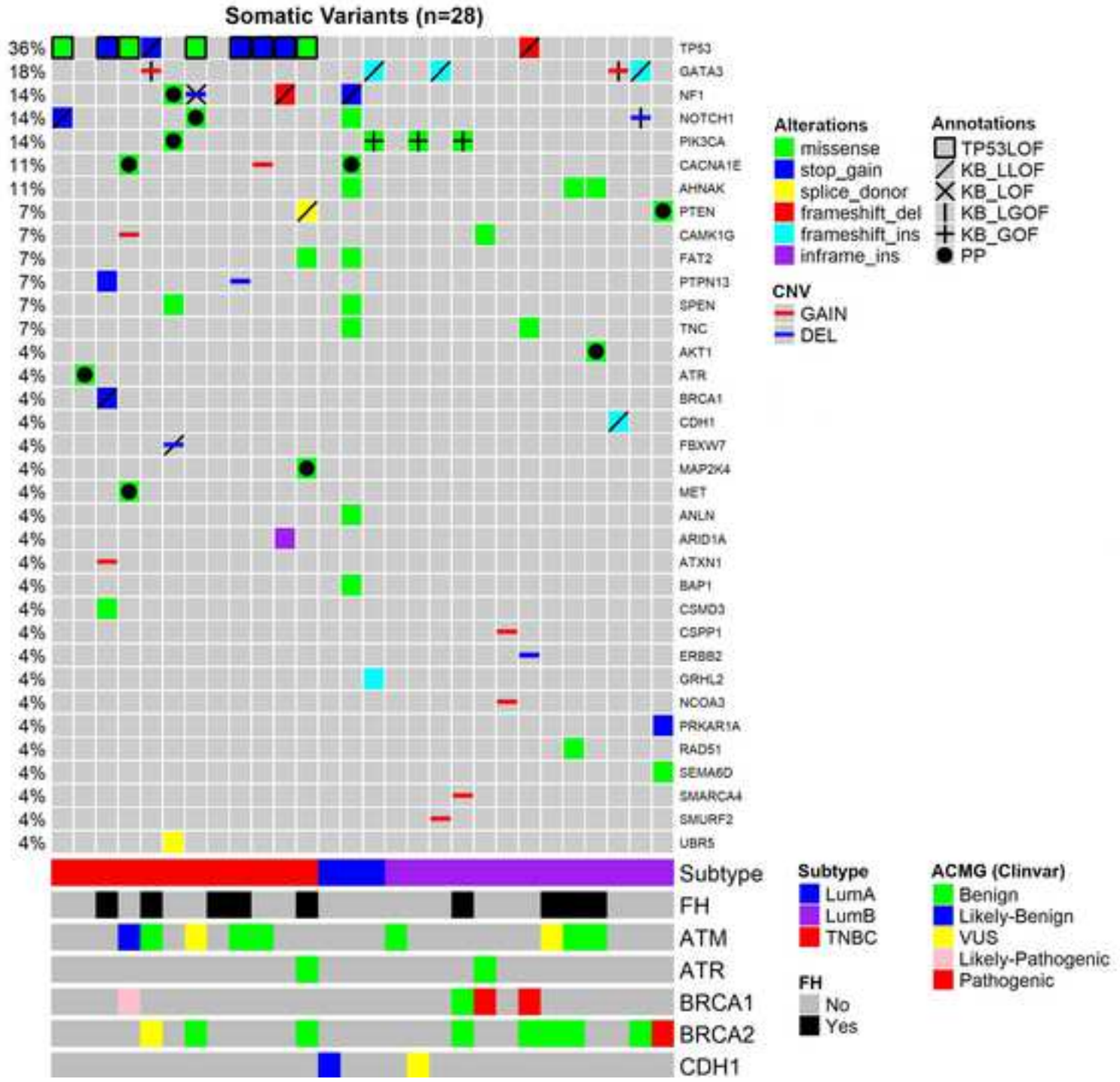


Table 1. Genes integrating the personalized gene-panel, classified according to the Cancer Gene Census.

Non_CGC	TSG	OG	Dual
AHNAK	APC	AKT1	ESR1
ANLN	ARID1A	ERBB2	GATA3
ATAD2B	ATM	ERBB4	MAP2K4
ATXN1	ATR	GRM3	MAP3K1
CACNA1E	BAP1	MET	NOTCH1
CAMK1G	BRCA1	MTOR	PRKAR1A
CSPP1	BRCA2	PIK3CA	TP53
DALRD3	CDH1	TNC	-
FAT2	CSMD3	UBR5	-
GRHL2	DICER1	-	-
HERC2	EP300	-	-
HUWE1	FAT4	-	-
LYST	FBXW7	-	-
MED23	MEN1	-	-
MTHFD2	NF1	-	-
NCOA3	PIK3R1	-	-
PARP4	POLD1	-	-
PCDH10	PTEN	-	-
PIK3AP1	PTPN13	-	-
PRKD1	RB1	-	-
RAD51	SETD2	-	-
RAD9A	SMARCA4	-	-
RSBN1	SPEN	-	-
SEMA6D	-	-	-
SMURF2	-	-	-

Table 2. Clinical data summary. LP: likely-pathogenic; P: pathogenic; VUS: variant of uncertain significance; LB: likely-benign; B: benign; WT: wild-type. The variants are classified following the ACMG germline classification criteria and exported from Clinvar. Family history: family history of at least one relative (until third-degree relatives) with a diagnosis of breast, ovary, pancreas, or prostate cancer

	Patients (n=28)		
	TNBC (n=12)	Luminal n=16)	p-value
Age (median)	36	33	ns
Tumor Grade n(%)			
I	0 (0%)	1 (6%)	
II	5 (42%)	12 (75%)	ns
III	7 (58%)	3 (19%)	
Family History n(%)	5 (42%)	4 (25%)	ns
BRCA1/2 Status n(%)			
LP/P	1 (8%)	3 (19%)	ns
VUS	1 (8%)	0 (10%)	

Table 3. Somatic variants detected in 28 young adult breast cancer patients, classified a
tumor supresspr gene; OG: oncogene; Dual: dual-role gene (TSG and OG); GAIN: CNV g
loss-of-function; +: likely gain-of-function; ++: gain-of-function; PP: possibly pathogenic
prediction tools).

ID	Subtype	Non_CGC	TSG
605	TNBC	-	-
608	TNBC	-	ATR
611	TNBC	ATXN1 (GAIN)	BRCA1-, CSMD3, PTPN13
635	TNBC	CACNA1E, CAMK1G (GAIN)	-
700	TNBC	-	-
701	TNBC	-	FBXW7-(DEL), NF1, SPEN,
702	TNBC	-	NF1--(DEL)
703	TNBC	-	-
715	TNBC	-	PTPN13 (DEL)
719	TNBC	CACNA1E (GAIN)	-
728	TNBC	-	ARID1A, NF1-
730	TNBC	FAT2	PTEN-
731	LumB	-	-
732	LumB	-	-
800	LumB	SMURF2 (GAIN)	-
809	LumB	-	SMARCA4 (GAIN)
811	LumB	CAMK1G	-
812	LumA	-	-
818	LumB	CSPP1 (GAIN), NCOA3 (GAIN)	-
819	LumA	AHNAK, ANLN, CACNA1E (PP), FAT2	BAP1, NF1-, SPEN
821	LumB	-	-
822	LumA	GRHL2	NF1--(DEL)
823	LumB	-	-
828	LumB	AHNAK, RAD51	-
829	LumB	AHNAK	-
837	LumB	-	CDH1-
838	LumB	-	-
842	LumB	SEMA6D	PTEN (PP)

according to the Cancer Gene Census (CGC). TSG:
ain; DEL: CNV deletion; -: likely loss-of-function; --:
missense variant (according to variant effect

OG	Dual
-	TP53--, NOTCH1-
-	-
-	TP53--
MET (PP)	TP53--
-	TP53-, GATA3+ (GAIN)
PIK3CA (PP), UBR5 (PP)	-
-	TP53--, NOTCH1 (PP)
-	-
-	TP53--
-	TP53--
-	TP53--
-	TP53--, MAP2K4 (PP)
-	-
PIK3CA++	-
-	GATA3-
PIK3CA++	-
-	-
-	-
-	-
TNC	NOTCH1
ERBB2 (DEL), TNC	TP53-
PIK3CA++	GATA3-
-	-
-	-
AKT1 (PP)	-
-	GATA3+(GAIN)
-	NOTCH1+(DEL), GATA3-
-	PRKAR1A

Legends: Figures and Tables

Tables

Table 1. **Genes integrating the personalized gene-panel, classified according to the Cancer Gene Census.**

Table 2. **Clinical data summary.** LP: likely-pathogenic; P: pathogenic; VUS: variant of uncertain significance; LB: likely-benign; B: benign; WT: wild-type. The variants are classified following the ACMG germline classification criteria and exported from Clinvar. Family history: family history of at least one relative (until third-degree relatives) with a diagnosis of breast, ovary, pancreas, or prostate cancer

Table 3. **Somatic variants detected in 28 young adult breast cancer patients, classified according to the Cancer Gene Census (CGC).** TSG: tumor suppressor gene; OG: oncogene; Dual: dual-role gene (TSG and OG); GAIN: CNV gain; DEL: CNV deletion; -: likely loss-of-function; --: loss-of-function; +: likely gain-of-function; ++: gain-of-function; PP: possibly pathogenic missense variant (according to variant effect prediction tools).

Supplementary Table 1. **Germline variants from the 28 young adult breast cancer patients.** ID: sample ID; Findley: BRCA1 variant functional classification according to Findley et. al. (2018); Clinvar: Classification of variants according to functional studies, B: benign, VUS: variant of uncertain significance, LB: likely-benign, LP: likely-pathogenic, P: pathogenic.

Supplementary Table 2. **Somatic variants from young adult breast cancer patients.** ID: sample ID; ONCOKB: variant is reported at OncoKB as likely -loss-of-function (L-LOF), loss-of-function (LOF), likely gain-of-function (L-GOF) or gain-of-function causing, or not reported in OncoKB (NA); Cancer Var: classification of somatic variants based on the tool developed by Li et. al.(2022); TP53_DB: TP53 somatic variants classification based on data from functional studies reported at TP53 Database; PRED: number of variant effect prediction tools (max: 8) in which the variant was classified as damaging; COMP_PRED: number of variant effect prediction compilation tools (max: 4) in which the variant was classified as damaging;

Figures

Figure 1. **Oncoplot of the somatic variants detected in the 28 young adult breast cancer patients.** Each column is a patient, and each line represents a gene. Annotations: TP53 variants classified as causing LOF (TP53 Database), variants classified according to curated studies in OncoKB (KB_LLOF, KB_LOF, KB_LGOF, KB_GOF) and variants classified as possibly pathogenic (PP), according to variant effect prediction tools. FH: family history of at least one

relative (until third-degree relatives) with a diagnosis of breast, ovary, pancreas, or prostate cancer. ACMG: classification of germline variants.

Supplementary Figure 1. **Oncoplot of the germline variants detected in the 28 young adult breast cancer patients.** Each column is a patient, and each line represents a gene. Alterations: classification of variants according to ACMG (exported from Clinvar), B: benign, LB: likely-benign, NA: variant was not reported/investigated, VUS: variant of uncertain significance, LP: likely-pathogenic, P: pathogenic. FH: family history of at least one relative (until third-degree relatives) with a diagnosis of breast, ovary, pancreas, or prostate cancer.

Supplementary methods

Custom target-sequencing in triple-negative and luminal breast cancer from young Brazilian patients

Pedro Adolpho de Menezes Pacheco Serio[1], Daniela Marques Saccaro[1], Ana Carolina Ribeiro Chaves de Gouvêa[1], Giselly Encinas[2], Simone Maistro[1], Gláucia Fernanda de Lima Pereira[1], Larissa Dias de Souza[1], Viviane Jennifer da Silva[1], Maria Lucia Hirata Katayama[1], Maria Aparecida Azevedo Koike Folgueira[1]

Sample collection and extraction

Patients who agreed to participate in the study had their respective blood samples collected and their respective tumor samples embedded in paraffin (Formalin-Fixed Paraffin-Embedded - FFPE) accessed through the pathology department of ICESP.

DNA extraction

For each sample, a slide stained with HE was analyzed by the pathologist and the region of interest (with ≥30% of tumor cells) was demarcated. With the aid of a Microtome (Zeiss), 5 cuts of 20 µm were obtained. Using the previously analyzed HE slide as a guide, the region of interest was scraped and stored in 1.5 ml tubes. DNA was extracted using QIAamp® DNA FFPE Tissue (Qiagen - 56404), following manufacturer protocol. For blood samples, DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen – 51306), following manufacturer protocol.

DNA Library Construction

Paired blood and FFPE samples that were successfully extracted and at a minimum concentration of 10 ng/uL were processed as described in the SureSelectXT HS Target Enrichment System for Illumina Multiplexed Sequencing Platforms protocol (Agilent).

Finally, the libraries were combined to reach the same equimolarity (10 nM) in 1x Low TE buffer, and then sequenced on the NextSeq device (NextSeq 500/550 Mid Output Kit v2.5, 150 cycles; Illumina).

Quality Control (QC) and data processing

I - Quality control

Clusterization density and Q30 scores were checked during and after the sequencing process, through the interface of the BaseSpace tool (Illumina), directly linked to the sequencing equipment, where a density between 170 and 220 K/mm² (NextSeq 500/550 Mid Output Kit v2.5, 150 cycles) and a Q30 in at least 80% of the total reads of each sample was considered as good quality.

III - Quality control after alignment and variant

As a quality control, we chose a minimum value of 20 total reads and a ratio of total reads/altered reads of at least 10%, per variant. As the variability of coverage and quality of target regions can vary greatly between samples and regions in target sequencing, all detected variants were checked for quality of coverage according to the patient, gene and exon they were located. To verify the global quality of probes and samples, we checked the coverage of the exons of each gene in the panel, per patient, calculating the median coverage between probes from a patient or from a gene in all patients, thus extracting the estimate of global quality of probes and samples.

We excluded variants mapped in low-complexity regions and/or variants with low mapping quality. In addition, 2 genes were excluded from further analysis, these being HERC2 and PARP4, as they presented low coverage in all analyzed samples. All non-rare variants (polymorphisms) with an allele frequency (GnomAD) equal to or greater than 1% were filtered out.

IV - Complementary data on variants

All variants were classified as to their pathogenic potential using the following tools, databases and articles:

1 - Cancer Gene Census (CGC; COSMIC v.96): Cancer Gene Census (CGC; https://cancer.sanger.ac.uk/census) is a curated database where genes with cancer-related literature are cataloged. Genes are divided into Tier 1 and Tier 2, with the first classification referring to genes that already have a strong literature on their relevance to cancer, the second being given to genes whose relationship with cancer is more recent or less robust in the literature. We chose to use both classifications for gene selection.

2 - Potentially pathogenic variants: truncated somatic variants (frameshift, nonsense and canonical splice-site ±3);

3 - Possibly pathogenic variants: somatic missense variants evaluated using tools to predict the functional impact of variants, namely: FATHMM (1), MutationAssesor (2), MutationTaster (3), PROVEAN (4), Polyphen2 HDIV, Polyphen2 HVAR (5), SIFT and SIFT4G (6). We also use tools that use methodologies for compiling functional impact prediction tools for variants, namely: REVEL (7), METALR and METASVM (8) and MCAP (9).

Finally, we classified missense variants as possibly pathogenic (PP) if they were predicted to be pathogenic in at least 4 out of 8 prediction tools and 2 out of 4 compilation tools or in at least 2 out of 8 prediction tools and 1 out of 4 compilation tools, if the gene in which it was found was cataloged as a CGC gene;

4 - Clinical and biological evidence: All detected somatic variants were explored through the OncoKB database (https://www.oncokb.org/). The database contains clinical and biological evidence of the functional impact of variants and affected genes, curated through the literature. Finally, the bank reports whether the variant or gene under investigation has oncogenic effects and effect of loss or gain of function (10).

5 - TP53: TP53 variants were classified according to the results of functional studies deposited in The TP53 Database (https://tp53.isb-cgc.org). The database compiles results of studies that investigated TP53 variants through quantitative assays with yeast, plasmid library for insertion of TP53 variants in in vivo assay in human cells, and CRISPR-CAS9 saturation assay, also in human cells (11–13);

6 - BRCA1: BRCA1 variants were classified according to a functional study whose results were obtained through a CRISPR-Cas9 panel (14);

7 - Germline variants were verified for their pathogenicity by consulting their classification on Clinvar and applying the classification suggested by the American College of Medical Genetics (ACMG);

8 - Finally, all somatic missense variants were annotated with the recently published tool by Li and colleagues (15), named CancerVar. CancerVar is a tool developed for the interpretation of somatic variants (specific, structural, indels and CNVs). In the study, the authors developed a classification of somatic variants based on fixed rules, similar to the methodology inserted by ACMG for classification of germline variants, together with artificial intelligence training.

V - Copy Number Variation (CNV)

Somatic CNVs were detected using SureCall (Agilent) software, where tumor samples were used as investigational samples, and their respective paired blood samples were used as reference samples. For the calculation of CNVs, the program considers the coverage per base of each range of target regions of the panel genes (in this case, their exons), computing and normalizing the coverage values between the pairs of samples. As an output, we obtain the gene and the coordinate of the region where it is amplification or deletion, this effect being represented by the value of the ratio between paired

samples, in addition, the size of the region that represents the CNV is also computed.

For calling CNVs, we used most of the software's base filtering and quality control settings. Since the coverage of some blood samples was higher than in tumor samples, we used a strict filter for CNV detection, to ensure as few false-positives as possible. Only variants with a confidence score (probability of being a true-positive finding) of at least 0.97 (97%) and only regions with at least 20 reads of coverage were accepted, we also corrected the detection to consider the coverage fold-changes between blood and tumor paired samples.

References

1. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013 Jan;34(1):57–65.
2. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011 Sep 1;39(17):e118.
3. Steinhaus R, Proft S, Schuelke M, Cooper DN, Schwarz JM, Seelow D. MutationTaster2021. Nucleic Acids Res. 2021 Jul 2;49(W1):W446–51.
4. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE. 2012 Oct 8;7(10):e46688.
5. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248–9.
6. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016 Jan;11(1):1–9.
7. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016 Oct 6;99(4):877–85.
8. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015 Apr 15;24(8):2125–37.
9. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016 Dec;48(12):1581–6.

10. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: A precision oncology knowledge base. JCO Precis Oncol. 2017 Jul;2017.
11. Monti P, Perfumo C, Bisio A, Ciribilli Y, Menichini P, Russo D, et al. Dominant-negative features of mutant TP53 in germline carriers have limited impact on cancer outcomes. Mol Cancer Res. 2011 Mar;9(3):271–9.
12. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet. 2018 Oct;50(10):1381–7.
13. Kotler E, Shani O, Goldfeld G, Lotan-Pompan M, Tarcic O, Gershoni A, et al. A systematic p53 mutation library links differential functional impact to cancer mutation pattern and evolutionary conservation. Mol Cell. 2018 Jul 5;71(1):178-190.e8.
14. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018 Oct;562(7726):217–22.
15. Li Q, Ren Z, Cao K, Li MM, Wang K, Zhou Y. CancerVar: An artificial intelligence-empowered platform for clinical interpretation of somatic mutations in cancer. Sci Adv. 2022 May 6;8(18):eabj1624.

Supplementary Tables

Supplementary Table 1. Germline variants from the 28 young adult breast cancer patients. ID: sample ID; Findley: BRCA1 variant functional classification according to Findley et. al. (2018); Clinvar: Classification of variants according to functional studies, B: benign, VUS: variant of uncertain significance, LB: likelybenign, LP: likely-pathogenic, P: pathogenic.

Subtype	SYMBOL	rs	Location	Consequence	Findley	Amino acid	REF	ALT	STRAND	1000g	gnomAD	Clinvar
TNBC	AHNAK	rs11828907	11:62288978-62288978	missense	-	D4304G	A	G	-1	0.0124	0.002478	B
TNBC	AHNAK	rs535365338	11:62292687-62292687	missense	-	E3068K	G	A	-1	0.002	2.78E-05	-
TNBC	AHNAK	rs111966997	11:62285664-62285664	missense	-	P5409S	C	T	-1	0.026	0.0004057	-
TNBC	AHNAK	rs11824660	11:62294573-62294573	missense	-	P2439L	C	T	-1	0.0383	0.008839	-
TNBC	AHNAK	rs111578362	11:62300280-62300280	missense	-	K537Q	A	C	-1	0.026	0.0004139	-
TNBC	AHNAK	rs115000832	11:62296998-62296998	missense	-	G1631S	G	A	-1	0.088	0.001885	B
TNBC	APC	rs754783550	5:112177107-112177107	missense	-	D1939V	A	T	1	-	4.00E-06	VUS
TNBC	APC	rs111866410	5:112175651-112175651	missense	-	K1454E	A	G	1	0.008	0.0005971	VUS
TNBC	ARID1A	rs61756316	1:27102188-27102188	missense	-	N1705S	A	G	1	0.0258	0.006984	LB
TNBC	ARID1A	rs61756316	1:27102188-27102188	missense	-	N1705S	A	G	1	0.0258	0.006984	LB
TNBC	ARID1A	rs61756316	1:27102188-27102188	missense	-	N1705S	A	G	1	0.0258	0.006984	LB
TNBC	ATAD2B	rs201642875	2:23980848-23980848	missense	-	T1173M	C	T	-1	0.012	0.002551	-
TNBC	ATAD2B	rs10210982	2:24118705-24118705	missense	-	S118P	T	C	-1	0.0162	0.003161	-
TNBC	ATM	rs2227922	11:108123551-108123551	missense	-	P604S	C	T	1	0.026	0.003193	B
TNBC	ATM	rs4986761	11:108124761-108124761	missense	-	S707P	T	C	1	0.044	0.007958	B
TNBC	ATM	rs2227922	11:108123551-108123551	missense	-	P604S	C	T	1	0.026	0.003193	B
TNBC	ATM	rs587781352	11:108137923-108137923	missense	-	D831G	A	G	1	-	3.98E-06	VUS
TNBC	ATM	rs3218695	11:108129778-108129778	missense	-	D814E	C	A	1	0.0104	0.001912	B
TNBC	ATM	rs3218673	11:108138045-108138045	missense	-	P872S	C	T	1	0.0164	0.003532	B
TNBC	ATM	rs4988111	11:108198391-108198391	missense	-	L2332P	T	C	1	0.072	0.001621	LB
TNBC	ATR	rs28910273	3:142188337-142188337	missense	-	Y2132D	T	G	-1	0.002	0.003153	LB
TNBC	ATR	rs1402949922	3:142279186-142279186	missense	-	I487T	T	C	-1	-	3.98E-06	-
TNBC	ATR	rs33972295	3:142178118-142178118	missense	-	P2434A	C	G	-1	0.0126	0.003215	B
TNBC	ATXN1	rs34265178	6:16327143-16327143	missense	-	G467S	G	A	-1	0.018	0.0005509	-
TNBC	ATXN1	rs139918586	6:16326864-16326864	missense	-	V560M	G	A	-1	-	0.0001752	VUS
TNBC	BAP1	rs79014342	3:52437258-52437258	missense	-	S596G	A	G	-1	0.027	0.005954	B
TNBC	BRCA1	.	17:41256941-41256941	missense	LOF	L82R	T	G	-1	-	NA	LP
TNBC	BRCA2	rs80358732	13:32913588-32913588	missense	-	D1699G	A	G	1	0.002	2.19E-05	VUS
TNBC	BRCA2	rs11571769	13:32953550-32953550	missense	-	A2951T	G	A	1	0.01	0.00955	B
TNBC	BRCA2	rs28897701	13:32893369-32893369	missense	-	A75P	G	C	1	0.002	0.0002586	B
TNBC	BRCA2	rs11571769	13:32953550-32953550	missense	-	A2951T	G	A	1	0.01	0.00955	B

TNBC	CSMD3	rs368012163	8:113599372-113599372	missense	-	S1270R	A	C	-1	-	5.58E-05	-
TNBC	CSPP1	rs756423026	8:68007858-68007858	missense	-	R281W	C	T	1	-	0.0001166	VUS
TNBC	DICER1	.	14:95570124-95570124	missense	-	N1203K	T	G	-1	-	NA	-
TNBC	ESR1	rs200075329	6:152129399-152129399	missense	-	S118P	T	C	1	0.0016	0.004392	B
TNBC	FAT2	rs34493925	5:150924527-150924527	missense	-	G2054A	G	C	-1	0.0098	0.002767	B
TNBC	FAT2	rs76037457	5:150925143-150925143	missense	-	V1849I	G	A	-1	0.0082	0.002513	B
TNBC	FAT2	rs61743237	5:150930192-150930192	missense	-	D1513N	G	A	-1	0.0082	0.002478	B
TNBC	FAT2	rs61743236	5:150930360-150930360	missense	-	V1457I	G	A	-1	0.0074	0.002151	B
TNBC	FAT2	rs61743254	5:150932879-150932879	missense	-	W1339R	T	C	-1	0.0074	0.002149	B
TNBC	FAT2	rs35640822	5:150933984-150933984	missense	-	L1295P	T	C	-1	0.0074	0.002145	B
TNBC	FAT2	rs116310150	5:150946029-150946029	missense	-	P822S	C	T	-1	0.0038	0.001248	-
TNBC	FAT2	rs200068080	5:150897269-150897269	missense	-	R3792Q	G	A	-1	-	0.0001909	-
TNBC	FAT4	rs1163197653	4:126336783-126336783	missense	-	T2222S	C	G	1	-	3.98E-06	-
TNBC	LYST	rs74641549	1:235963637-235963637	missense	-	D1330A	A	C	-1	0.0443	0.009912	B
TNBC	LYST	rs147221131	1:235827874-235827874	missense	-	V3696I	G	A	-1	0.0002	0.0006055	LB
TNBC	LYST	rs77091385	1:235972432-235972432	missense	-	Q562H	G	C	-1	0.012	0.003787	LB
TNBC	MAP3K1	rs916555277	5:56178599-56178599	missense	-	I1191N	T	A	1	-	NA	-
TNBC	MED23	rs139398917	6:131910690-131910690	missense	-	S1285T	G	C	-1	0.01	0.0001154	-
TNBC	MET	rs35225896	7:116340086-116340086	missense	-	I316M	A	G	1	0.054	0.001627	B
TNBC	MET	rs757883355	7:116339379-116339379	missense	-	V81F	G	T	1	-	8.03E-06	VUS
TNBC	NCOA3	.	20:46252708-46252708	missense	-	E46G	A	G	1	-	NA	-
TNBC	PCDH10	rs1176501665	4:134072200-134072200	missense	-	G302V	G	T	1	-	NA	-
TNBC	PRKD1	rs45471692	14:30396622-30396622	missense	-	P33S	C	T	-1	-	0	B
TNBC	PTPN13	rs61750815	4:87666225-87666225	missense	-	H865R	A	G	1	0.002	0.002849	LB
TNBC	SETD2	rs9311404	3:47163824-47163824	missense	-	V768L	G	C	-1	0.026	0.0005611	B
TNBC	SPEN	rs115323964	1:16255205-16255205	missense	-	R824C	C	T	1	0.054	0.001174	-
TNBC	SPEN	rs77870948	1:16260513-16260513	missense	-	N2593S	A	G	1	0.395	0.008844	-
TNBC	SPEN	rs146173073	1:16247386-16247386	missense	-	G553S	G	A	1	0.024	0.0007093	B
TNBC	TNC	rs142544129	9:117819530-117819530	missense	-	H1494R	A	G	-1	-	5.57E-05	-
TNBC	TNC	rs75198247	9:117786265-117786265	missense	-	N2161S	A	G	-1	0.046	0.001359	B
LUM	AHNAK	rs115693058	11:62297175-62297175	missense	-	M1572V	A	G	-1	0.014	0.001948	B
LUM	AHNAK	.	11:62285402-62285402	missense	-	G5496E	G	A	-1	0.016	0.002671	B
LUM	AHNAK	rs141151380	11:62285402-62285402	missense	-	G5496E	G	A	-1	0.016	0.002671	B
LUM	AHNAK	rs116243978	11:62286165-62286165	missense	-	G5242R	G	C	-1	0.046	0.007204	B
LUM	AHNAK	rs201059372	11:62287616-62287616	missense	-	K4758R	A	G	-1	-	0.0002307	-
LUM	AHNAK	rs11231129	11:62292321-62292321	missense	-	V3190I	G	A	-1	0.01	0.0002545	-
LUM	AHNAK	rs7112976	11:62295180-62295180	missense	-	P2237T	C	A	-1	0.01	0.0002664	-
LUM	AHNAK	rs11828907	11:62288978-62288978	missense	-	D4304G	A	G	-1	0.124	0.002478	B

LUM	AHNAK	rs118246660	11:62294573-62294573	missense	-	P2439L	C	T	-1	0.0383	0.008839	-
LUM	AHNAK	rs118246660	11:62294573-62294573	missense	-	P2439L	C	T	-1	0.0383	0.008839	-
LUM	AHNAK	rs149928120	11:62289818-62289818	missense	-	P4024R	C	G	-1	0.0006	8.75E-05	B
LUM	AHNAK	rs118246660	11:62294573-62294573	missense	-	P2439L	C	T	-1	0.0383	0.008839	-
LUM	AHNAK	rs115937732	11:62291129-62291129	missense	-	N3587S	A	G	-1	0.0122	0.002231	B
LUM	AHNAK	rs148347501	11:62291585-62291585	missense	-	I3435T	T	C	-1	0.002	0.0002587	LB
LUM	AHNAK	rs148663483	11:62292980-62292980	missense	-	L2970W	T	G	-1	0.0122	0.002175	B
LUM	APC	rs944674770	5:112176648-112176648	missense	-	R1786T	G	C	1	-	3.99E-06	VUS
LUM	APC	rs1370102446	5:112174750-112174753	inframe_deletion	-	E1154-	GAA	-	1	-	0.001008	VUS
LUM	ATAD2B	rs10210982	2:24118705-24118705	missense	-	S118P	T	C	-1	0.0162	0.003161	-
LUM	ATM	rs369903995	11:108160494-108160494	missense	-	V1468I	G	A	1	0.002	8.37E-05	VUS
LUM	ATM	rs3092857	11:108143299-108143299	missense	-	M1040V	A	G	1	0.0142	0.003072	B
LUM	ATM	rs3092857	11:108143299-108143299	missense	-	M1040V	A	G	1	0.0142	0.003072	B
LUM	ATM	rs56009889	11:108196896-108196896	missense	-	L2307F	C	T	1	-	0.001412	B
LUM	ATR	rs33972295	3:142178118-142178118	missense	-	P2434A	C	G	-1	0.0126	0.003215	B
LUM	BAP1	rs35448940	3:52437206-52437206	missense	-	T613M	C	T	-1	0.0042	0.0009228	B
LUM	BAP1	rs79014342	3:52437258-52437258	missense	-	S596G	A	G	-1	0.027	0.005954	B
LUM	BRCA1	rs56082113	17:41245090-41245090	missense	NA	K820E	A	G	-1	0.0104	0.002401	B
LUM	BRCA1	rs397509173	17:41256137-41256137	splice_donor	NA	---			-1	-	NA	P
LUM	BRCA1	rs1217805587	17:41209079-41209079	frameshift	NA	Q1777X	-	C	-1	-	0.0001829	P
LUM	BRCA2	rs11571769	13:32953550-32953550	missense	-	A2951T	G	A	1	0.01	0.009955	B
LUM	BRCA2	rs56248502	13:32912582-32912582	missense	-	I1364L	A	C	1	0.044	0.001243	B
LUM	BRCA2	rs144862123	13:32911418-32911418	missense	-	S976T	T	A	1	0.024	0.0004819	B
LUM	BRCA2	rs11571656	13:32911419-32911419	missense	-	S976F	C	T	1	0.024	0.0004811	B
LUM	BRCA2	rs55969723	13:32912679-32912679	missense	-	Q1396R	A	G	1	0.002	0.000414	B
LUM	BRCA2	rs11571769	13:32953550-32953550	missense	-	A2951T	G	A	1	0.01	0.009955	B
LUM	BRCA2	rs80358408	13:32906738-32906738	missense	-	P375S	C	T	1	0.001	2.80E-05	B
LUM	BRCA2	rs80358515	13:32893396-32893396	stop_gained	-	Q84*	C	T	1	-	NA	P
LUM	CACNA1E	rs202202209	1:181701810-181701810	missense	-	R863Q	G	A	1	0.003	0.001201	B
LUM	CDH1	rs121964872	16:68867265-68867265	missense	-	S838G	A	G	1	-	4.37E-05	LB
LUM	CDH1	rs763203357	16:68863612-68863612	missense	-	R784H	G	A	1	-	3.98E-06	VUS
LUM	CSMD3	rs150862620	8:113668398-113668398	missense	-	K997E	A	G	-1	0.012	0.0006726	-
LUM	CSMD3	rs80277352	8:113418822-113418822	missense	-	I1914V	A	G	-1	0.0116	0.003027	B
LUM	CSMD3	rs76916857	8:113504760-113504760	missense	-	D1746N	G	A	-1	0.0046	0.001683	B
LUM	FAT2	rs57774012	5:150891768-150891768	missense	-	T3955P	A	C	-1	0.0166	0.003552	B
LUM	FAT2	rs60050170	5:150891806-150891806	missense	-	T3942S	C	G	-1	0.0166	0.003561	B
LUM	FAT2	rs116401802	5:150925883-150925883	missense	-	F1602S	T	C	-1	0.003	0.0008404	B
LUM	FAT2	rs114551196	5:150934209-150934209	missense	-	P1220H	C	A	-1	0.0036	0.0006795	-

LUM	FAT4	.	4:126373873-126373873	missense	-	I3901T	T	C	1	-	NA	-
LUM	FAT4	rs145639192	4:126320038-126320038	missense	-	I1759V	A	G	1	0.0004	0.0008059	VUS
LUM	FAT4	rs114637892	4:126373594-126373594	missense	-	H3808R	A	G	1	0.082	0.002054	B
LUM	FAT4	rs202188213	4:126240990-126240990	missense	-	V1142M	G	A	1	-	0.0003328	VUS
LUM	FAT4	rs115822434	4:126400922-126400922	missense	-	T4167I	C	T	1	0.004	0.002574	B
LUM	FAT4	rs28515675	4:126389832-126389832	missense	-	R4022Q	G	A	1	0.132	0.001953	B
LUM	HUWE1	rs142176360	X:53586416-53586416	missense	-	R2605H	G	A	-1	-	5.47E-06	-
LUM	LYST	rs143223086	1:235969077-235969077	missense	-	S1120I	G	T	-1	0.03	0.000773	B
LUM	MAP2K4	rs940014479	17:11924303-11924303	missense	-	V34F	G	T	1	-	0.000014	-
LUM	MET	rs370529693	7:116415124-116415124	missense	-	P1091L	C	T	1	0.002	0.0001243	LB
LUM	MTOR	rs141877007	1:11303235-11303235	missense	-	V450I	G	A	-1	-	1.19E-05	VUS
LUM	NCOA3	rs551807396	20:46277797-46277797	missense	-	T1199A	A	G	1	0.002	2.39E-05	-
LUM	NCOA3	rs2230781	20:46264805-46264805	missense	-	P559S	C	T	1	0.192	0.004208	B
LUM	NOTCH1	rs111309246	9:139391200-139391200	missense	-	A2331T	G	A	-1	0.076	0.001335	B
LUM	NOTCH1	rs188270459	9:139390543-139390543	missense	-	I2550V	A	G	-1	0.004	0.0002464	LB
LUM	NOTCH1	rs80340744	9:139401302-139401302	missense	-	P1256L	C	T	-1	0.072	0.001512	B
LUM	NOTCH1	rs150737112	9:139417343-139417343	missense	-	R234H	G	A	-1	0.004	0.0005159	VUS
LUM	POLD1	.	19:50902196-50902196	missense	-	R30W	C	T	1	0.038	0.008208	B
LUM	POLD1	rs3218772	19:50902196-50902196	missense	-	R30W	C	T	1	0.038	0.008208	B
LUM	POLD1	rs914238978	19:50905485-50905485	missense	-	G205S	G	A	1	-	1.31E-05	VUS
LUM	POLD1	rs914238978	19:50905485-50905485	missense	-	G205S	G	A	1	-	1.31E-05	VUS
LUM	POLD1	rs1726803	19:50905310-50905310	missense	-	S173N	G	A	1	0.316	0.007704	B
LUM	PTPN13	rs34226837	4:87735618-87735618	missense	-	I2463V	A	G	1	0.188	0.003839	-
LUM	PTPN13	rs372924378	4:87610176-87610176	missense	-	H127N	C	A	1	-	5.63E-05	-
LUM	PTPN13	rs114206680	4:87684234-87684234	missense	-	K1303R	A	G	1	0.022	0.0005866	-
LUM	PTPN13	rs34226837	4:87735618-87735618	missense	-	I2463V	A	G	1	0.188	0.003839	-
LUM	RAD9A	rs568076402	11:67164800-67164803	inframe_deletion	-	E342-	GAG	-	1	-	0.0006759	-
LUM	SETD2	rs780288575	3:47162089-47162089	missense	-	G1346V	G	T	-1	-	3.98E-06	-
LUM	SPEN	rs564799201	1:16262685-16262685	missense	-	P3317L	C	T	1	-	4.19E-06	-
LUM	SPEN	rs61749275	1:16258756-16258756	missense	-	D2007E	T	A	1	0.018	0.004591	-
LUM	TNC	rs144032672	9:117849382-117849382	missense	-	G210S	G	A	-1	0.012	0.001297	LB

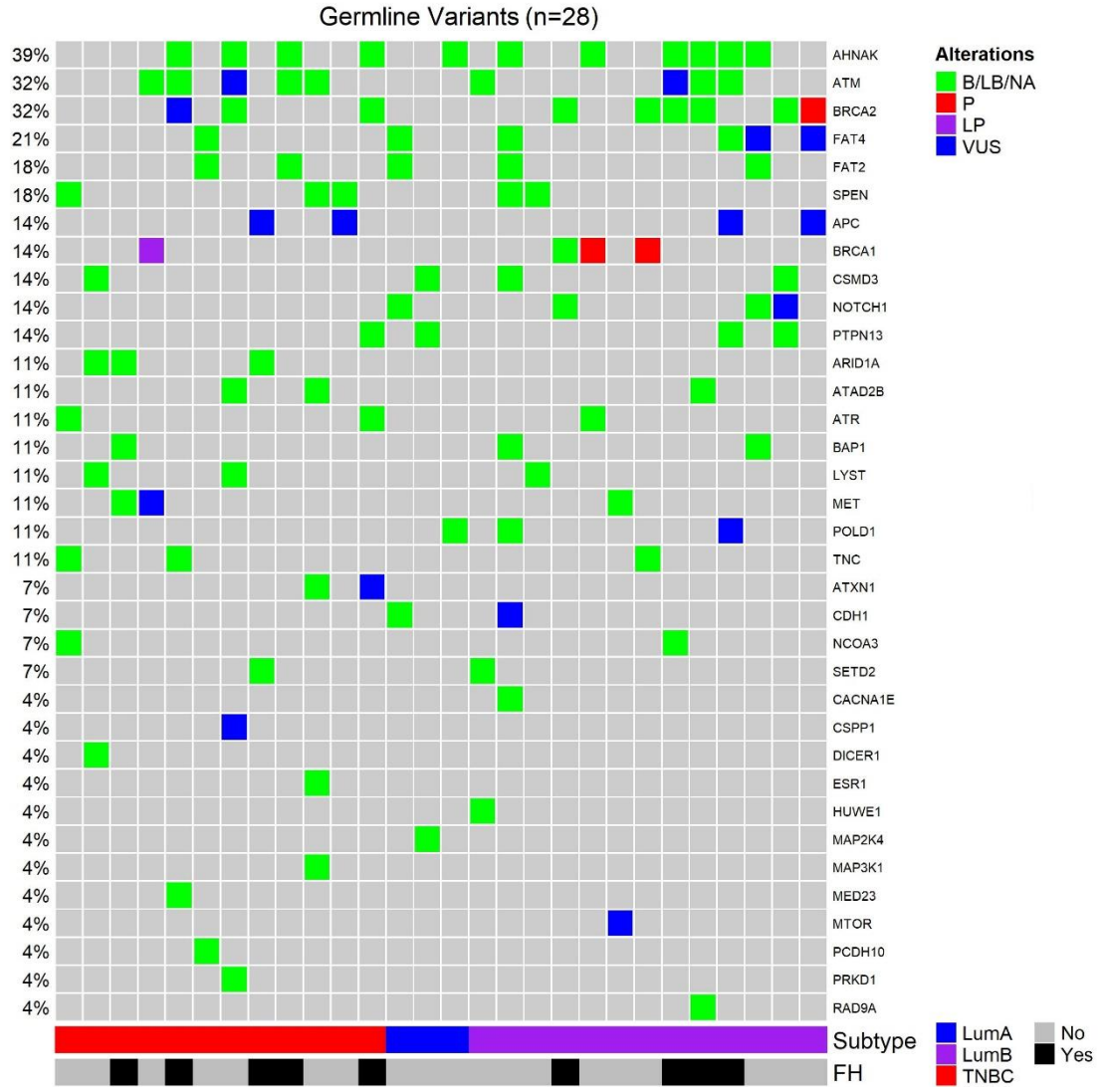
Supplementary Table 2. Somatic variants from young adult breast cancer patients. ID: sample ID; ONCOKB: variant is reported at OncoKB as likely loss of- function (L-LOF), loss-of-function (LOF), likely gain-of-function (L-GOF) or gain-of-function causing, or not reported in OncoKB (NA); Cancer Var: classification of somatic variants based on the tool developed by Li et. al.(2022); TP53_DB: TP53 somatic variants classification based on data from functional studies reported at TP53 Database; PRED: number of variant effect prediction tools (max: 8) in which the variant was classified as damaging; COMP_PRED: number of variant effect prediction compilation tools (max: 4) in which the variant was classified as damaging;

Subtype	ID	SYMBOL	rs	Location	Consequence	Aminoacid	REF	ALT	STRAND	ONCOKB	Cancer_var	TP53_DB	PRED	COMP_PRED
TNBC	605	NOTCH1	rs779613930	9:139391014-139391014	stop_gained	Q2393*	C	T	-1	LLOF	NA	NA	NA	NA
TNBC	605	TP53	rs1057520002	17:7577560-7577560	missense	S241P	T	C	-1	NA	TIER_1_STRONG	LOF	8	4
TNBC	608	ATR	.	3:142279283-142279283	missense	D455Y	G	T	-1	NA	TIER_3_UNCERTAIN	NA	6	1
TNBC	611	BRCA1	rs80357710	17:41215950-41215952	stop_gained	C1718-1719*	TG	-	-1	LLOF	NA	NA	NA	NA
TNBC	611	CSMD3	.	8:113253991-113253991	missense	D3476H	G	C	-1	NA	TIER_4_BENIGN	NA	2	0
TNBC	611	PTPN13	.	4:87666147-87666147	stop_gained	S839*	C	A	1	NA	NA	NA	NA	NA
TNBC	611	TP53	rs876660548	17:7579414-7579414	stop_gained	W91*	G	A	-1	LLOF	NA	LOF	NA	NA
TNBC	635	CACNA1E	rs201622587	1:181765920-181765920	missense	A2109S	G	T	1	NA	ND	NA	2	3
TNBC	635	MET	.	7:116340204-116340204	missense	P356T	C	A	1	NA	TIER_3_UNCERTAIN	NA	6	0
TNBC	635	MET	.	7:116395452-116395452	missense	T582I	C	T	1	NA	TIER_3_UNCERTAIN	NA	7	4
TNBC	635	TP53	rs121912651	17:7577539-7577539	missense	R248W	C	T	-1	LOF	TIER_1_STRONG	LOF	8	4
TNBC	700	TP53	rs1057519977	17:7578507-7578507	stop_gained	C141*	C	A	-1	LLOF	NA	not_LOF	NA	NA
TNBC	701	NF1	.	17:29559152-29559152	missense	P1087T	C	A	1	NA	TIER_3_UNCERTAIN	NA	7	1
TNBC	701	PIK3CA	rs3729687	3:178938877-178938877	missense	E707K	G	A	1	NA	TIER_2_POTENTIAL	NA	7	4
TNBC	701	SPEN	.	1:16259906-16259906	missense	H2391Y	C	T	1	NA	TIER_4_BENIGN	NA	1	0
TNBC	701	UBR5	.	8:103300382-103300382	splice_donor	---	-	-	-1	NA	NA	NA	NA	NA
TNBC	702	NOTCH1	rs373806373	9:139391932-139391932	missense	R2087W	C	T	-1	NA	TIER_2_POTENTIAL	NA	7	2
TNBC	702	TP53	rs1057520004	17:7578202-7578202	missense	V216E	T	A	-1	NA	TIER_1_STRONG	LOF	8	4
TNBC	715	TP53	rs764735889	17:7576897-7576897	stop_gained	Q317*	C	T	-1	LLOF	NA	LOF	NA	NA
TNBC	719	TP53	rs121913344	17:7577022-7577022	stop_gained	R306*	C	T	-1	LLOF	NA	LOF	NA	NA
TNBC	728	ARID1A	rs1266385064	1:27023007-27023007	inframe_insertion	E38A	-	GGC	1	NA				
TNBC	728	NF1	.	17:29527569-29527569	missense	S340A	T	G	1	NA	TIER_3_UNCERTAIN	NA	2	1
TNBC	728	NF1	.	17:29553568-29553568	missense	A706V	C	T	1	NA	TIER_3_UNCERTAIN	NA	6	1
TNBC	728	NF1	.	17:29667634-29667638	frameshift	L2345-2346X	TAGA	-	1	LLOF	NA	NA	NA	NA
TNBC	728	TP53	rs1321845532	17:7574003-7574003	stop_gained	R342*	C	T	-1	LLOF	NA	LOF	NA	NA
TNBC	73	FAT2	rs148551207	5:150922928-150922928	missense	A2587V	C	T	-1	NA	TIER_4_BENIGN	NA	6	1

	0													
TNBC	730	MAP2K4	.	17:12013737-12013737	missense	H227Y	C	T	1	NA	TIER_3_UNCERTAIN	NA	7	4
TNBC	730	PTEN	rs1114167622	10:89712017-89712017	splice_donor	---	-	-	1	LLOF	NA	NA	NA	NA
TNBC	730	TP53	rs863224451	17:7577114-7577114	missense	C275Y	G	A	-1	LLOF	TIER_1_STRONG	LOF	8	4
LUM	732	PIK3CA	rs121913279	3:178952085-178952085	missense	H1047R	A	G	1	GOF	TIER_2_POTENTIAL	NA	3	1
LUM	800	GATA3	rs757297342	10:8115874-8115874	frameshift	S408X	-	G	1	LLOF	NA	NA	NA	NA
LUM	809	PIK3CA	rs121913273	3:178936082-178936082	missense	E542K	G	A	1	GOF	TIER_2_POTENTIAL	NA	4	1
LUM	811	CAMK1G	rs769916493	1:209786186-209786186	missense	R466P	G	C	1	NA	NA	NA	1	0
LUM	819	AHNAK	.	11:62297864-62297864	missense	D1342A	A	C	-1	NA	TIER_4_BENIGN	NA	7	1
LUM	819	AHNAK	rs776554777	11:62297865-62297865	missense	D1342Y	G	T	-1	NA	TIER_3_UNCERTAIN	NA	7	1
LUM	819	ANLN	.	7:36435975-36435975	missense	A40V	C	T	1	NA	NA	NA	0	0
LUM	819	BAP1	.	3:52437252-52437252	missense	P598S	C	T	-1	NA	TIER_3_UNCERTAIN	NA	0	0
LUM	819	CACNA1E	.	1:181752824-181752824	missense	L1792M	C	A	1	NA	NA	NA	6	4
LUM	819	FAT2	rs1482787673	5:150885296-150885296	missense	G4294S	G	A	-1	NA	TIER_4_BENIGN	NA	5	1
LUM	819	NF1	.	17:29560064-29560064	stop_gained	E1181*	G	T	1	LLOF	NA	NA	NA	NA
LUM	819	NOTCH1	rs367589813	9:139399780-139399780	missense	R1523H	G	A	-1	NA	TIER_3_UNCERTAIN	NA	0	1
LUM	819	SPEN	.	1:16258002-16258002	missense	T1756I	C	T	1	NA	TIER_4_BENIGN	NA	1	0
LUM	819	TNC	rs149299073	9:117804524-117804524	missense	R1741Q	G	A	-1	NA	TIER_3_UNCERTAIN	NA	4	0
LUM	821	TNC	.	9:117808790-117808790	missense	R1675M	G	T	-1	NA	TIER_4_BENIGN	NA	1	0
LUM	821	TP53	.	17:7579447-7579447	frameshift	P80X	-	C	-1	LLOF	NA	NA	NA	NA
LUM	822	GATA3	.	10:8115955-8115955	frameshift	H435X	-	CC	1	LLOF	NA	NA	NA	NA
LUM	822	GRHL2	.	8:102564954-102564954	frameshift	-76-77X	-	T	1	NA	NA	NA	NA	NA
LUM	822	PIK3CA	.	3:178952085-178952085	missense	H1047R	A	G	1	GOF	TIER_2_POTENTIAL	NA	3	1
LUM	828	AHNAK	.	11:62289975-62289975	missense	V3972I	G	A	-1	NA	TIER_4_BENIGN	NA	0	0
LUM	828	RAD51	rs777467455	15:40994113-40994113	missense	V112A	T	C	1	NA	TIER_4_BENIGN	NA	1	0
LUM	829	AHNAK	.	11:62292484-62292484	missense	N3135K	T	G	-1	NA	TIER_4_BENIGN	NA	0	0
LUM	829	AKT1	.	14:105237177-105237177	missense	P423L	C	T	-1	NA	TIER_3_UNCERTAIN	NA	7	4
LUM	837	CDH1	rs766222121	16:68863566-68863566	frameshift	L769X	-	TG	1	LLOF	NA	NA	NA	NA
LUM	838	GATA3	.	10:8115979-8115979	frameshift	M443X	-	G	1	LLOF	NA	NA	NA	NA
LUM	842	PRKAR1A	rs281864783	17:66519005-66519005	stop_gained	R96*	C	T	1	NA	NA	NA	NA	NA
LUM	842	PTEN	.	10:89717633-89717633	missense	L220V	C	G	1	NA	TIER_3_UNCERTAIN	NA	4	2
LUM	842	SEMA6D	.	15:48056954-48056954	missense	I406T	T	C	1	NA	NA	NA	6	0

Supplementary Figures

Supplementary Figure 1. Oncoplot of the germline variants detected in the 28 young adult breast cancer patients. Each column is a patient, and each line represents a gene. Alterations: classification of variants according to ACMG (exported from Clinvar), B: benign, LB: likely-benign, NA: variant was not reported/investigated, VUS: variant of uncertain significance, LP: likely pathogenic, P: pathogenic. FH: family history of at least one relative (until third degree relatives) with a diagnosis of breast, ovary, pancreas, or prostate cancer.



Apêndice C

cells

MDPI

Article

Somatic Mutational Profile of High-Grade Serous Ovarian Carcinoma and Triple-Negative Breast Carcinoma in Young and Elderly Patients: Similarities and Divergences

Pedro Adolpho de Menezes Pacheco Serio, Gláucia Fernanda de Lima Pereira, Maria Lucia Hirata Katayama, Rosimeire Aparecida Roela, Simone Maistro and Maria Aparecida Azevedo Koike Folgueira *

Centro de Investigação Translacional em Oncologia, Departamento de Radiologia e Oncologia, Instituto do Cancer do Estado de Sao Paulo, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo 01246-000, Brazil; pedro.serio@fm.usp.br (P.A.d.M.P.S.); glaucialimap@usp.br (G.F.d.L.P.); maria.katayama@fm.usp.br (M.L.H.K.); r.roela@fm.usp.br (R.A.R.); simone.maistro@hc.fm.usp.br (S.M.)

* Correspondence: maria.folgueira@fm.usp.br

Abstract: Background: Triple-negative breast cancer (TNBC) and High-Grade Serous Ovarian Cancer (HGSOC) are aggressive malignancies that share similarities; however, different ages of onset may reflect distinct tumor behaviors. Thus, our aim was to compare somatic mutations in potential driver genes in 109 TNBC and 81 HGSOC from young (Y $\leq$ 40 years) and elderly (E $\geq$ 75 years) patients. Methods: Open access mutational data (WGS or WES) were collected for TNBC and HGSOC patients. Potential driver genes were those that were present in the Cancer Gene Census—CGC, the Candidate Cancer Gene Database—CCGD, or OncoKB and those that were considered pathogenic in variant effect prediction tools. Results: Mutational signature 3 (homologous repair defects) was the only gene that was represented in all four subgroups. The median number of mutated CGCs per sample was similar in HGSOC (Y:3 vs. E:4), but it was higher in elderly TNBC than it was in young TNBC (Y:3 vs. E:6). At least 90% of the samples from TNBC and HGSOC from Y and E patients presented at least one known affected TSG. Besides *TP53*, which was mutated in 67–83% of the samples, the affected TSG in *TP53* wild-type samples were *NF1* (yHGSOC and yTNBC), *PHF6* (eHGSOC and yTNBC), *PTEN*, *PIK3R1* and *ZHFX3* (yTNBC), *KMT2C*, *ARID1B*, *TBX3*, and *ATM* (eTNBC). A few samples only presented one affected oncogene (but no TSG): *KRAS* and *TSHR* in eHGSOC and *RAC1* and *PREX2* (a regulator of *RAC1*) in yTNBC. At least ⅔ of the tumors presented mutated oncogenes associated with tumor suppressor genes; the Ras and/or PIK3CA signaling pathways were altered in 15% HGSOC and 20–35% TNBC (Y vs. E); DNA repair genes were mutated in 19–33% of the HGSOC tumors but were more frequently mutated in E-TNBC (56%). However, in HGSOC, 9.5% and 3.3% of the young and elderly patients, respectively, did not present any tumors with an affected CGC nor did 4.65% and none of the young and elderly TNBC patients. Conclusion: Most HGSOC and TNBC from young and elderly patients present an affected TSG, mainly TP53, as well as mutational signature 3; however, a few tumors only present an affected oncogene or no affected cancer-causing genes.

Keywords: triple-negative breast cancer; high-grade serous ovarian carcinoma; somatic; young adult

Citation: Serio, P.A.d.M.P.; de Lima Pereira, G.F.; Katayama, M.L.H.; Roela, R.A.; Maistro, S.; Folgueira, M.A.A.K. Somatic Mutational Profile of High-Grade Serous Ovarian Carcinoma and Triple-Negative Breast Carcinoma in Young and Elderly Patients: Similarities and Divergences. *Cells* **2021**, *10*, 3586. https://doi.org/10.3390/cells10123586

Academic Editor: Frank Schnütgen

Received: 4 August 2021
Accepted: 23 November 2021
Published: 20 December 2021

1. Introduction

Most cancers, including breast and ovarian cancers, are mainly detected in elderly people. As a matter of fact, SEER data indicate that the peak age incidences for breast and ovarian cancers are between 65–74 and 75+, respectively. Besides that, breast cancer is the main cause of death among young women (18–40 years old). In young people who are below 40 years of age, the incidence trends for breast cancer increased in the period of 2000–2016, while it was stable or decreasing for the age groups $\geq$40 years during the same period. Among breast cancer subtypes, triple-negative breast cancer (TNBC) is relatively

more frequent in younger patients than it is in older patients [1]. For TNBC, the overall and disease-free survival tends to be compromised in young adults compared to elderly patients [2,3].

Although the most common type of ovarian cancer in all age groups is epithelial cancer, specifically in young women, ovarian carcinoma is not common and is surpassed by borderline tumors as well as by germ cell tumors [4,5]. For those young women who are diagnosed with ovarian carcinoma, the most prevalent subtype is low-grade serous carcinoma, which is in contrast with older women, in whom the most prevalent subtype is high-grade serous ovarian carcinoma (HGSOC) [6].

In young women who are diagnosed with ovarian carcinoma at age 40 or below, the serous subtype is detected in 14.6% to 34% of the cases [7–10]; high-grade carcinoma is detected in 19% of cases; and metastatic disease is detected in 34.8% of the cases. This is in clear contrast with older patients, in whom metastatic and high-grade disease is detected in 77.5% and 67.7% of the cases, respectively [11]. These observations, which are associated with differences in disease management and secondary comorbidities in older patients, translate in a survival advantage for very young women, with the 5-year disease-specific survival estimates at 78.8% vs. 35.3% for younger and older women, respectively [11].

The main risk factor for cancer development is older age, which is due to the accumulation of DNA mutations that occur during a person's lifetime. Cancer diagnosis at younger ages is not expected, excluding the diagnosis of cancers that are associated with hereditary syndromes [1].

Around 5% of all breast carcinoma patients and 12 to 18% of breast cancer in young patients are associated with Hereditary Breast and Ovarian Cancer Syndrome, which are caused by *BRCA1* or *BRCA2* mutations [12,13]. In TNBC specifically, a higher fraction of patients, varying from 9 to 18%, are BRCA1 (more frequently), but are also BRCA2 mutation carriers [14,15]. Although not frequently, *BRCA1* somatic variants are also seen in 1–6% of TNBC tumors [16–18].

In ovarian carcinoma patients, more frequently, *BRCA1* mutations, but also *BRCA2* germline mutations are detected in 18 to 23% of cases [19–21]. In concordance, among young patients with ovarian carcinoma who are below 40 years of age, 22% are *BRCA1* mutation carriers [22]. In addition, *BRCA1* somatic variants may be detected in 8–10% of the tumor samples [19,23].

TNBC and HGSOC are both aggressive malignancies that share some similarities, such as "BRCAness", which is defined as a defect in double-strand break repair by homologous recombination repair (HRR), which mimics the loss of function of *BRCA1* or *BRCA2* [24]. This characteristic enables patients to benefit from platinum therapy as well as from poly (ADP-ribose) polymerase (PARP) inhibitor therapy [25–28]. In fact, PARP inhibitors have shown encouraging results in the treatment of TNBC and ovarian carcinoma [29–31]. In addition, both TNBC and HGSOC tumor samples have been shown to be highly affected by somatic *TP53* mutations.

In summary, cancer in younger age groups is not common and may exhibit differences from the cancer that occurs in older age groups. Hence, one of our aims was to compare the somatic mutations in TNBC and HGSOC from young and elderly patients. On the other hand, TNBC and HGSOC may share some similarities in the processes that occur during tumorigenesis. Thus, another aim was to further compare the somatic characteristics of these two types of cancer, as well as to evaluate which were the potential cancer driver genes in tumors from young and elderly HGSOC and TNBC patients.

2. Methods

This study focused on somatic mutations. Studies focusing on molecular investigation and that were based on sequencing technology were searched in the Catalogue of Somatic Mutations in Cancer (COSMIC; https://cancer.sanger.ac.uk/cosmic; accesed on 1 May 2020), CbioPortal (https://www.cbioportal.org/; accessed on 1 May 2020), and PubMed.

The inclusion criteria were I—the diagnosis of TNBC or HGSOC; II—young (≤40 y); and III—the availability of whole-genome (whole genome sequencing—WGS) or exome (whole-exome sequencing—WES) data.

For TNBC, data from the following manuscripts were evaluated: Shah SP. et al., *Nature*, 2012; the Cancer Genome Atlas Network, *Nature*, 2012; and Nik-Zainal S. et al., *Nature*, 2016 (identified in CbioPortal); Kan Z. et al., *Nature Communications*, 2018 (identified in literature); and Banerji S. et al., *Nature*, 2012 (identified in COSMIC) (n = 86 patients). These studies matched the inclusion criteria [25,32–35]. Mutational and clinical data from two of the five studies (Cancer Genome Atlas Network, *Nature*, 2012; Nik-Zainal S. et. al., *Nature*, 2016) were downloaded from the International Cancer Genome Consortium—ICGC data portal (https://dcc.icgc.org/; release date: 1 January 2020). The other studies had their data downloaded from the supplementary data from their respective original articles.

For HGSOC, only one study was evaluated (Cancer Genome Atlas Network, *Nature*, 2011) since this was the only one to fulfill the inclusion criteria [36] at the time of analysis.

The studies that were chosen according to the inclusion criteria were also used to select data from patients who were aged ≥75 year for further comparison between age groups.

Since the variant data were obtained from both WES and WGS sequencing, we first analyzed the normalized data by excluding non-coding variants. Silent variants were also filtered out, as this information was absent in some studies. Variant data were annotated with Oncotator (http://portals.broadinstitute.org/oncotator/; v1.5.3.0; accessed on 1 May 2020) to determine the genomic position, pathogenicity prediction through in silico tools, amino acid change, and other additional data.

For the interpretation of the gene variants, previous literature was consulted for relevant criteria (Sukhai MA. et. al., 2016 and Li MM. et. al., 2017) [37,38], but since the somatic data were mainly classified according to clinical actionability, which could limit the present analysis, we adopted a less restrictive classification model at first in order to explore possible driver genes in the cancer types that were evaluated. The classification criteria are detailed below.

For a gene presenting a variant, we collected data from/about:

1—Cancer Gene Census database (CGC; v.89), which is a curated repository of mutated genes that are causally implicated in cancer. Both tier 1 and tier 2 genes were considered to be relevant.

2—Candidate Cancer Gene Database (CCGD) [39] which contains data about multiple studies that have investigated the contribution of certain genes to mice tumorigenesis based on mutational insertion transposons studies. For the latter database, we only considered those genes within the higher rank classification (rank A).

3—Potentially pathogenic variants: truncated somatic variants (frameshift, nonsense, and canonical splice-site ±3).

4—Possibly pathogenic variants: missense somatic variants were then assessed using variant functional impact prediction tools: SIFT (https://sift.bii.a-star.edu.sg/, accessed on 1 May 2020); Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/, accessed on 1 May 2020); FATHMM (http://fathmm.biocompute.org.uk/, accessed on 1 May 2020); Mutation Taster (http://www.mutationtaster.org/); and Mutation Assessor (http://mutationassessor.org/r3/), accessed on 1 May 2020. Missense variants were considered to be possibly pathogenic if they accomplished one of the following items: I—the variant was classified as pathogenic in 3 out of 5 predictors; II—the missense variant was placed in a CGC gene and was classified as pathogenic in at least 1 of 5 of the prediction tools

5—Clinical and biological evidence: Finally, we explored all the variants from the CGC oncogenes and TSGs using the OncoKB Database (https://www.oncokb.org/; accessed on 2 November 2021). The database contains curated clinical and biological evidence at the variant and gene level and reports oncogenic and mutation effects from functional and clinical studies as well as biomarkers with data approved by the FDA and other professional guidelines with compelling evidence. Variants were classified as: I—likely loss-of-function if there was functional evidence in the literature; II—likely gain-of-function if there was

functional evidence in the literature; III—loss-of-function if there was functional evidence in literature and/or clinical evidence and/or other therapeutic/diagnostic/prognostic/FDA levels (as described in https://www.oncokb.org/levels; accessed on 2 November 2021); IV—gain-of-function if there was functional evidence in the literature and/or clinical evidence and/or others therapeutic/diagnostic/prognostic/FDA levels (as described in https://www.oncokb.org/levels; accessed on 2 November 2021); V—likely neutral if there was biological and/or clinical evidence in the literature; VI—inconclusive if functional and/or clinical assays were inconclusive; VII—not reported if the gene or variant was not found or curated in the literature [40].

6—DNA repair related genes: A list of genes involved in DNA repair pathways was assembled according to the following published manuscripts: Lange SS, Takata K, Wood RD, *Nature Reviews Cancer*, 2011; Wood RD, Mitchell M, Sgouros JG, Lindahl T, *Science*, 2001; and Chae et al., *Oncotarget*, 2016. This list of genes was used to look for affected genes that are involved in the different DNA repair pathways [41–43].

The Maftools 2.2.0 package [44] and built-in R packages were used to generate frequency summaries for variant types and mutational signatures. The trinucleotideMatrix, estimateSignatures and extractSignatures functions from the Maftools package were used for the mutational signatures analysis based on the Alexandrov publication [45]. The methodology is available at https://github.com/PoisonAlien/maftools (accessed on 1 June 2020).

Gene ontology analyses were performed through g:Profiler (https://biit.cs.ut.ee/gprofiler_archive2/r1760_e93_eg40/web/; accessed on 1 June 2020). Redundant gene ontology terms were filtered out with Revigo (http://revigo.irb.hr/; accessed on 1 June 2020) [46,47].

Descriptive and inferential analyses for tumor and age comparison were conducted in R.v.3.5.0 (Rstudio v.1.1.453). The Mann–Whitney U test was used to compare continuous variables between the groups and Fisher's Exact test for the categorical ones.

Germline data were not evaluated because they were not available as open access data.

3. Results

Mutational data were retrieved from the literature and public databases for 21 young and 60 elderly HGSOC patients as well as for 86 young and 23 elderly TNBC patients.

Among the HGSOC patients, the median age at diagnosis was 37 and 78 years for the young and elderly groups, respectively. Most patients were diagnosed at clinical stages III and IV (95% young vs. 88% elderly). For TNBC patients, the median age at diagnosis was 36 and 80 years for the young and elderly age groups, respectively, and around half of the patients in both age groups presented with disease clinical stage III.

At first, we focused on the number and profile of genes that were affected by non-synonymous coding variants, independent of whether they were neutral or potentially pathogenic/possibly pathogenic. Subsequently, we analyzed the number and profile of driver mutations in genes that were catalogued in the Cancer Gene Census (CGC) database. We also looked forward to identifying possibly putative drivers, which were represented by affected genes that were catalogued in the Candidate Cancer Gene Database (CCGD) with supportive cancer-related literature.

3.1. Base Substitutions

For HGSOC as well as for TNBC, the most prevalent base substitutions in both age groups were C>T (HGSOC: 37% young vs. 41% elderly; TNBC: 27% young vs. 35% elderly) (Figure 1A,B; Tables 1 and S1). There was no difference in the frequency of transitions (TI) between young and elderly HGSOC patients, as well as in transversions (TV). In TNBC, the frequency of TV in young patients was higher than in the elderly ones. Also, young patients had a higher frequency of TV in comparison to TI (TI: 38%; TV: 62%) (Table S2).

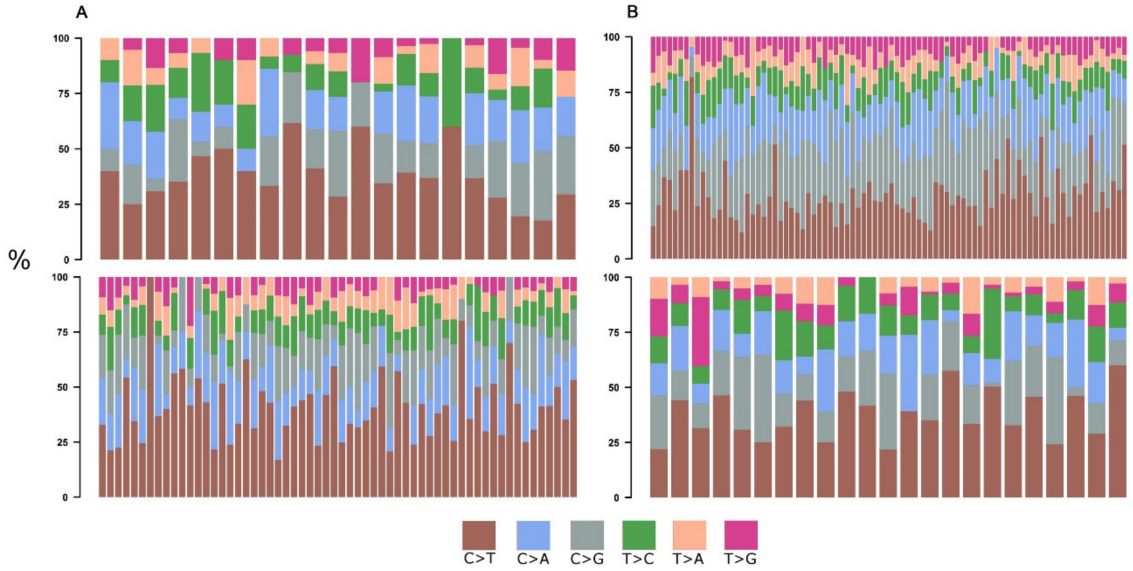


Figure 1. Single base substitutions. (**A**) Distribution of single base substitutions across young (**top**) and elderly (**bottom**) HGSOC patients. (**B**) Distribution of single base substitutions across young (**top**) and elderly (**bottom**) TNBC patients.

3.2. Mutational Signatures

In HGSOC, the most frequently detected mutational signatures in the young and elderly cohorts of patients were signatures 1 (spontaneous deamination of 5-methylcytosine due to age), 3 (defects in double-strand break repair by homologous recombination), and 5 (unknown etiology). Hence, HGSOC tended to present a more homogeneous pattern of mutational signatures in both age groups (Figures 2 and S1).

On the other hand, in TNBC, the most frequent mutational signatures in young and elderly patients seemed more heterogeneous, as only signature 3 appeared in both age groups. The other most frequent mutational signatures that were identified in young people were signatures 1 and 13, and in elderly people, the most frequently occurring signatures were signatures 2 and 6 (Figure 2; Table 1; Figure S1). It is interesting to observe that the APOBEC cytidine deaminase signatures were significantly detected in both young (signature 13) and elderly people (signature 2), but following different strand *bias*, the APOBEC signature in young patients was correlated with transversions, and the one in elderly patients had transitions, which may explain the previously mentioned higher frequency of transversions in TNBC from the young age group. On the other hand, signature 1, which is related to age, was present in tumors from young people, and signature 6, which is related to defective DNA mismatch repair, was present in tumors from elderly people.

Hence, in both the HGSOC and TNBC samples, the only mutational signature that was significantly detected in young as well as in elderly patients was signature 3, indicating that BRCAness is an important feature in cancers occurring at both extremes of age and in both tumor types. In contrast, each tumor type presented an exclusive mutational signature pattern, regardless of age: mutational signature 5 was more characteristic of HGSOC and APOBEC-related signatures of TNBC.

Table 1. Cancer and age groups summaries and comparisons.

		HGSOC			TNBC			Young Patients: TNBC vs. HGSOC	Elderly Patients: TNBC vs. HGSOC
		Young	**Elderly**	***p*-Value**	**Young**	**Elderly**	***p*-Value**	***p*-Value**	***p*-Value**
	Median C>T substitutions (%)	37	41	ns	27	35	0.00055	0.00084	ns
	Median *n*. of coding variants/sample	37	44	ns	50.5	61	ns	0.014	0.046
	Median *n*. of MS variants/sample	32	37.5	ns	42.5	53	ns	0.011	0.046
	Median *n*. of potentially pathogenic (FS/SS/NS)/possibly pathogenic variants/sample	14	13.5	ns	17.5	21	ns	ns	0.016
	Median *n*. CGC/sample	3	4	ns	3	6	0.00013	ns	0.001
	Median *n*. CCGD/sample	4	6	ns	6	7	0.044	ns	0.025
Mutational Signatures (%)	1—Age	33.3	41.7	ns	23.3	-	-	-	-
	2—APOBEC (C>T)	-	-			47.8	-	-	-
	3—HR	42.9	25	0.011	50	17.4	1.63×10^{-6}	ns	ns
	5—Unknown	23.8	33.3	ns	-	-	-	-	-
	6—MMR	-	-	-	-	34.8	-	-	-
	13—APOBEC (C>G)	-	-	-	26.7	-	-	-	-
CGC	CGC combinations (OG and TSG)	14/21 (66.6%)	51/60 (85%)	ns	62/86 (72.1%)	23/23 (100%)	0.0032	ns	ns
	CGC alone (OG or TSG)	5/21 (23.8%)	6/60 (10%)	ns	12/86 (14%)	0/23 (0%)	ns	ns	ns
	DNA repair potentially/possibly	4/21 (19%)	19/60 (32%)	ns	13/86 (15%)	12/23 (52%)	0.0005	ns	ns
	Ras	3/21 (14.3%)	9/60 (15%)	ns	17/86 (19.7%)	8/23 (34.8%)	ns	ns	ns

MS: missense; HR: homologous repair; MMR: mismatch repair; ns: no significance; OG: oncogene; TSG: tumor suppressor gene; FS: frameshift; SS: splice-site; NS: nonsense.

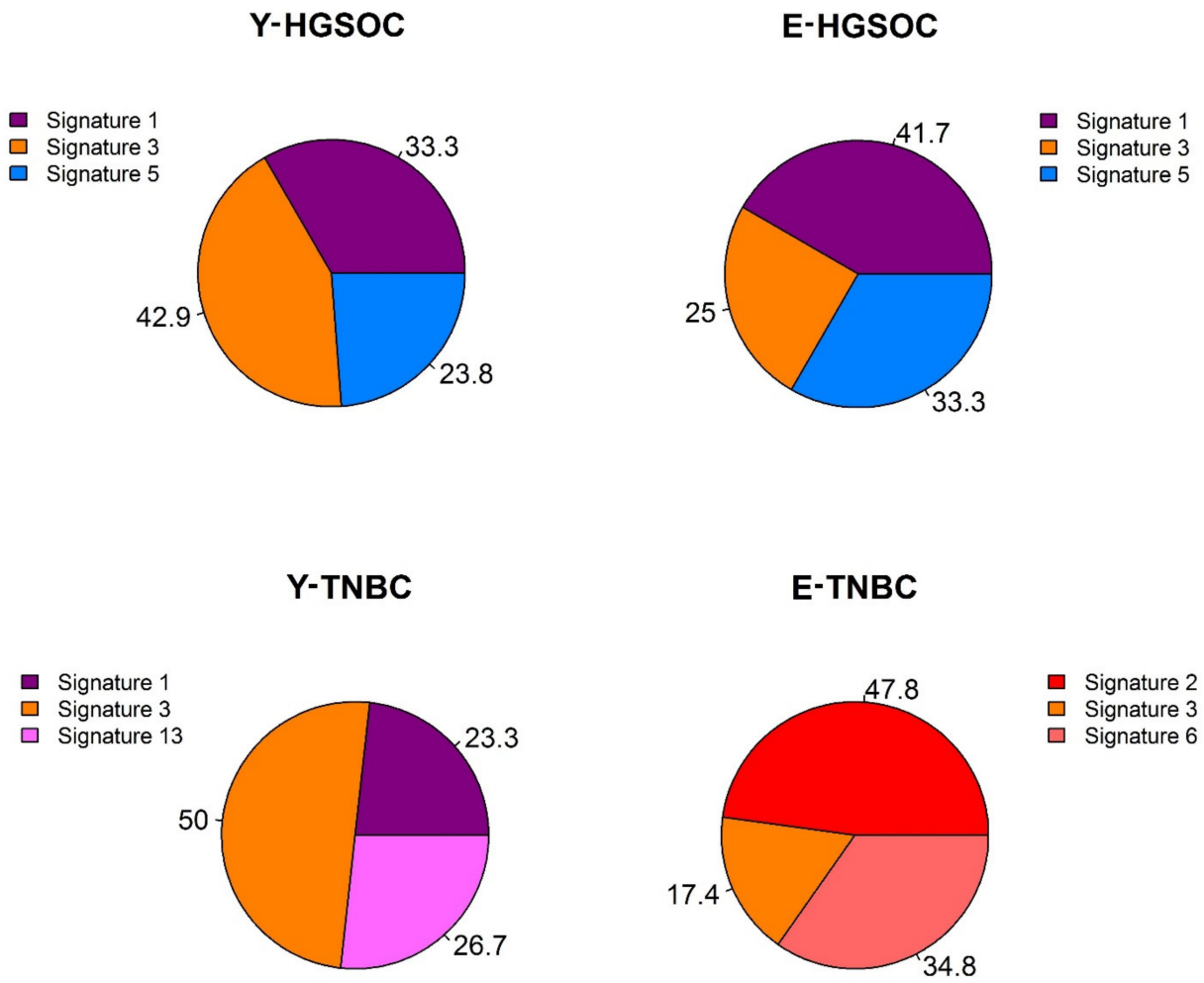


Figure 2. Mutational signatures: The pizza plots show the distribution (%) of the signatures across age groups in both cancer types. Y-HGSOC: young HGSOC patients; E-HGSOC: elderly HGSOC patients; Y-TNBC: young TNBC patients; E-TNBC: elderly TNBC patients.

3.3. Number of (Non-Synonymous) Coding Variants

In the HGSOC samples, the median number of non-synonymous coding variants per sample was 37 for young and 44 for elderly women; for TNBC, it was 50.5 for young and 61 for elderly women. There were no significant differences when the samples from the young and elderly patients from both tumor types were compared. However, when comparing the TNBC and HGSOC samples from the same age groups, TNBC presented a higher number of coding variants compared to HGSOC (Table 1).

Compared to the whole cohort of TCGA samples (including all ages and tumor subtypes), both the young and elderly HGSOC groups showed a lower number of coding variants per sample (TCGA OV cohort vs. young HGSOC, $p = 4.3 \times 10^{-6}$; TCGA OV cohort vs. elderly HGSOC, $p = 1.8 \times 10^{-5}$) (Figure S2; Table S3). In contrast to HGSOC, both young and elderly TNBC patients showed a higher number of coding mutations when compared to the complete TCGA Breast Cancer (BRCA) cohort (TCGA BRCA cohort vs. young TNBC, $p = 0.019$; TCGA BRCA cohort vs. elderly TNBC, $p = 0.0091$) (Figure S2; Table S3).

We also evaluated the number of coding variants per sample according to *TP53* status. In HGSOC, 7 out of 21 (33%) tumors from young patients and 10 out of 60 (16%) tumors from elderly patients were positive for *TP53*wt, but no difference in coding variants per

sample according to TP53 status was observed ($p = 0.12$) (Figure 3A,B; Table S4). For TNBC, the patients with affected *TP53* showed a higher number of variants per sample ($p = 0.012$) (Figure 3C,D; Table S4).

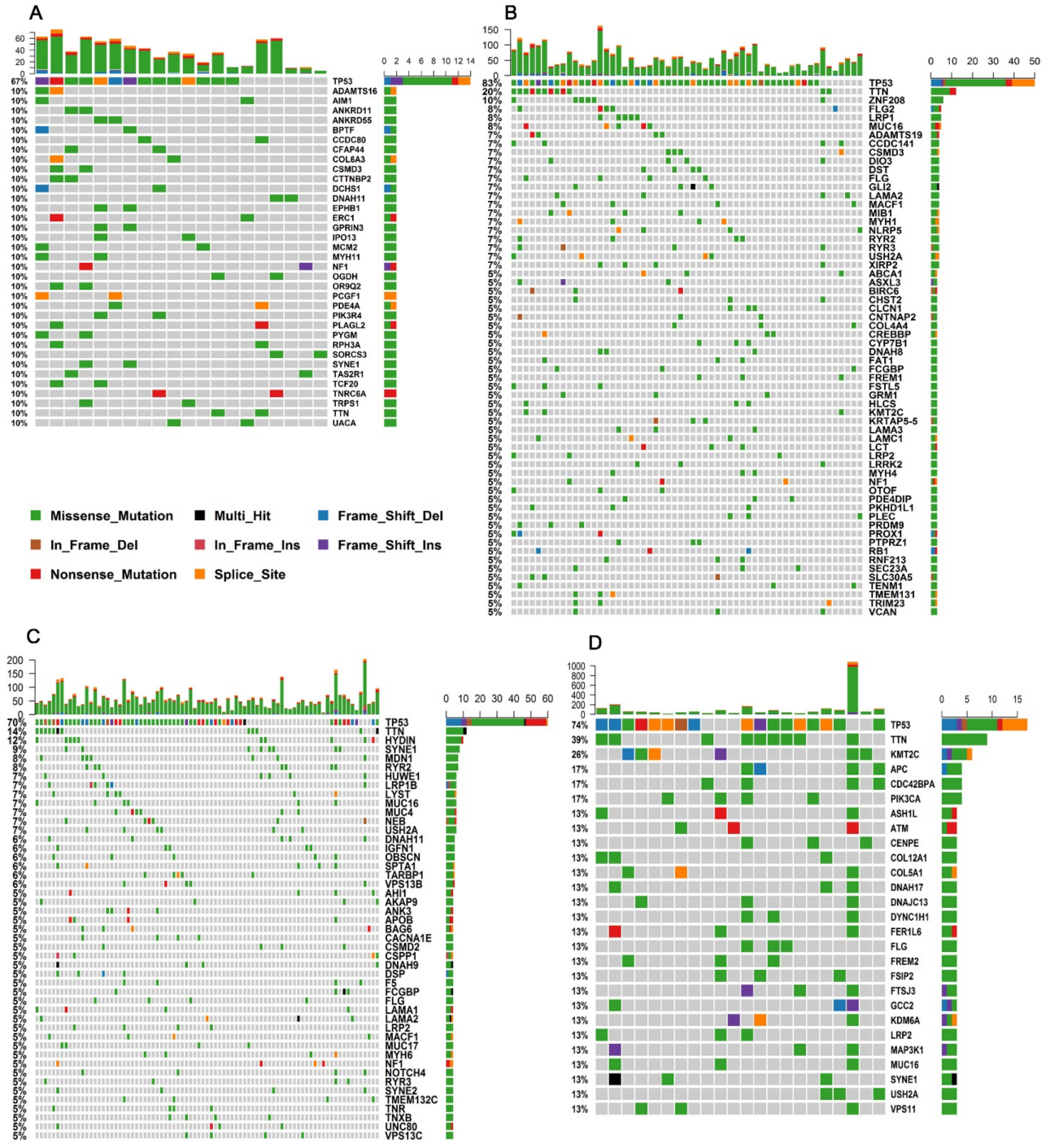


Figure 3. Most frequently affected genes: (**A**) Young HGSOC patients (20 samples shown); (**B**) elderly HGSOC patients; (**C**) young TNBC patients; (**D**) elderly TNBC patients. Each column represents a patient, and each line represents a gene.

Considering the median number of CGC per sample, there was no difference between the young and elderly HGSOC patients ($n = 3$, young vs. $n = 4$, elderly; ns). In contrast, a higher number of CGC per sample was detected in the elderly patients in the TNBC group ($n = 3$, young vs. $n = 6$, elderly; $p = 0.00013$) (Table 1).

3.4. Genes Most Commonly Affected

For this analysis, all genes presenting a coding variant, independent of the gene function or the effect of the variant, were considered (Table S4).

In total, 190 samples from young and elderly HGSOC and TNBC were analyzed. The 50 most commonly affected genes were present in 93% (177/190) of all of the samples (affected in at least six samples of all four groups added together), including 18 genes (36%) that were shown to be affected at least once in each one of the four groups: *TP53*, *TTN*, *SYNE1*, *USH2A*, *CSMD3*, *MACF1*, *MUC17*, *TARBP1*, *TENM1*, *VPS13B*, *DYNC1H1*, *F5*, *FSIP2*, *HMCN1*, *MXRA5*, *PLEC*, *DNAH8*, *and UNC79* (Figure S3).

As expected, the most frequently altered gene was *TP53*, which was affected in 67% and 83% of the young and elderly HGSOC patients and in 70% and 74% of the young and elderly TNBC samples, respectively.

The second most frequently affected gene was *TTN* (18.4% of samples), followed by *MUC16* and *USH2A* (7.4%; 14 out of 190 samples). *TTN* and *MUC16* code for two of the longest known proteins in the human genome, with 35,992 and 14,508 amino acids, respectively, and are frequently affected in a high percentage of tumor types, i.e., 24.49% (*TTN*) and 15.21% (MUC16) (COSMIC database, available at https://cancer.sanger.ac.uk/cosmic/gene/analysis; accessed on 27 May 2020).

TTN, *MUC16,* and *USH2A,* as well as another 23 genes (out of the list of the top 50 genes), integrate a list of FrequentLy mutAted GeneS, termed FLAGS. These genes present longer protein-coding sequences and a greater number of paralogs. They also display less evolutionarily selective pressure than expected [48]. Although FLAGS have been more frequently associated with disease causality than expected for protein-coding genes in general, their functional impact should be interpreted with great care because somatic mutations without functional consequences may occur during cell division and may represent passenger mutations that do not contribute to cancer development [49]. This observation may be applied to 11 out of 26 FLAGS as well as to another three genes that were not originally classified as FLAGs (*ZNF208*, *FSIP2*, *MXRA5*), as they mainly presented missense variants with non-pathogenic effects in all effect prediction tests.

Although its long length, which is consistent with FLAGS, *MUC16* is also listed in the CGC database (Tier 2). *MUC16* codes for a mucin, which is commonly shed in ovarian cancer and less frequently in other tumors of epithelial origin. *MUC16* may transform the NIH3T3 mouse fibroblast cell line and may thus, be considered an oncogene [50].

In fact, in the list of the top 50 most frequently mutated genes from the present series, only ten genes were classified in the Cancer Gene Census database: *TP53*, *MUC16*, *CSMD3*, *KMT2C*, *NF1*, *LRP1B*, *PIK3CA*, *ATM*, *BIRC6,* and *MUC4*.

The top 20 most frequently affected genes as listed in the Cancer Gene Census (CGC) considering the 190 samples from young and elderly HGSOC and TNBC appear on Table 2.

In HGSOC, besides *TP53*, other CGCs that were affected in samples from both young and elderly patients were TSGs, such as *CSMD3* (CUB and Sushi multiple domains 3), *NF1* (neurofibromatosis type 1 gene), *CLTCL1* (clathrin; heavy polypeptide-like 1), and *RB1* (retinoblastoma); however, there were no differences in their frequencies between age groups (Table 2).

In the young cohort of HGSOC, besides *TP53*, another 35 genes were recurrently altered in at least two out of 20 samples (10%) (Figure 3A). One of these genes was *PCGF1*, or polycomb group ring finger 1. It is interesting to point out that the same potentially pathogenic variant (canonical splice site) was found in both young patients. *PCGF1* integrates the *BCL6* corepressor (BCOR) complex, which is involved in transcriptional gene silencing. Inactivating somatic mutations in BCOR were detected in patients with acute myeloid leukemia (AML) and other cancers, suggesting that it might function as a TSG [51].

Table 2. Top 20 most commonly affected oncogenes (OG) and tumor suppressor genes (TSG) across HGSOC and TNBC (young -Y and elderly -Y). Bold: significant *p*-value.

		HGSOC			TNBC				
Gene	**Role in Cancer**	**Y (*n* = 21)**	**E (*n* = 60)**	***p*-value**	**Y (*n* = 86)**	**E (*n* = 23)**	***p*-Value**	**Y-HGSOC vs. Y-TNBC (*p*-value)**	**E-HGSOC vs. E-TNBC (*p*-value)**
TP53	OG/TSG	14	50	ns	60	17	ns	ns	ns
MUC16	OG	-	5	ns	9	6	ns	ns	ns
CSMD3	TSG	2	4	ns	3	2	ns	ns	ns
PIK3CA	OG	-	1	ns	4	6	ns	ns	ns
KMT2C	TSG	-	3	ns	1	6	**0.006**	ns	ns
NF1	TSG	2	3	ns	4	-	ns	ns	ns
ATM	TSG	-	3	ns	2	3	ns	ns	ns
BIRC6	OG	-	3	ns	3	2	ns	ns	ns
LRP1B	TSG	-	2	ns	6	-	ns	ns	ns
MUC4	OG	-	-	ns	6	1	ns	ns	ns
CLTCL1	TSG	1	2	ns	2	1	ns	ns	ns
BRAF	OG	-	1	ns	2	2	ns	ns	ns
BRCA2	TSG	-	2	ns	2	1	ns	ns	ns
CREBBP	OG/TSG	-	3	ns	1	1	ns	ns	ns
GRM3	OG	-	2	ns	3	-	ns	ns	ns
MED12	TSG	-	1	ns	2	2	ns	ns	ns
MET	OG	-	1	ns	3	1	ns	ns	ns
NOTCH1	OG/TSG	-	1	ns	2	2	ns	ns	ns
RB1	TSG	1	3	ns	1	-	ns	ns	ns
ZFHX3	TSG	-	2	ns	2	1	ns	ns	ns

In elderly HGSOC samples, other CGCs that were frequently altered were *MUC16* (detected in 8.3% of the tumors) and *BIRC6* (5%), which are classified as oncogenes, as well as *KMT2C* and *ATM* (5% each), which are classified as TSGs (Figure 3B).

Turning to the top most frequently affected CGCs in TNBC, besides *TP53*, another 15 genes were altered in samples from both age groups, including *MUC16*, *PIK3CA*, *BIRC6*, and *MUC4*, which can be classified as oncogenes, as well as *CSMD3*, *KMT2C*, and *ATM*, which can be classified as TSG, among others (Table 2, Figure 3C,D). Interestingly, *KMT2C* was more frequently altered in elderly patients, i.e., 26% (6/22) than it was in young patients, i.e., 1% (1/86) ($p = 0.006$). *KMT2C* (lysine (K)-specific methyltransferase 2C) promotes the methylation of histone H3 and regulates gene transcription, and even though it is described as TSG in medulloblastoma in COSMIC (https://cancer.sanger.ac.uk/cosmic/census?tier=all#cl_search; accessed on 25 May 2020) in TNBC specifically, it is characterized as both a tumor suppressor and an oncogene. Accordingly, in other studies, *KMT2C* alterations were correlated with more advanced ages (>50 yo) in breast cancer patients [52] and were described as being frequently altered in TNBC [52–56]. Nonetheless, among the top 20 CGCs, it is interesting to observe that with the exception of 1 gene (*MUC4*), all of the other 19 genes that were most frequently altered in the TNBC tumors from young patients were also affected in the HGSOC tumors from elderly patients (Table 2). This observation probably reflects the larger number of samples that was evaluated for these age groups, which allowed the characterization of a larger number of recurrently altered genes (Y TNBC $n = 86$; E HGSOC: $n = 60$).

In TNBC, after *TP53*, the most frequently altered gene in both age groups was *TTN*, which was altered in 14% and 39% of young and elderly patients, respectively.

Although many different genes were altered in the whole cohort, as determined in the previous section describing the most commonly affected genes, not all of them were affected by potentially pathogenic/possibly pathogenic variants. Figure 4 (also, Figure S3 and Table S7) shows the top 50 genes that are the most frequently mutated in tumors from the four groups according to their classification as being potentially pathogenic, possibly pathogenic, or benign; for the latter category, genes other than CGC were not considered

to be pathogenic in variant effect prediction tests. Examples of these genes are *MUC17*, *ZNF208*, *FSIP2*, and *MXRA5*.

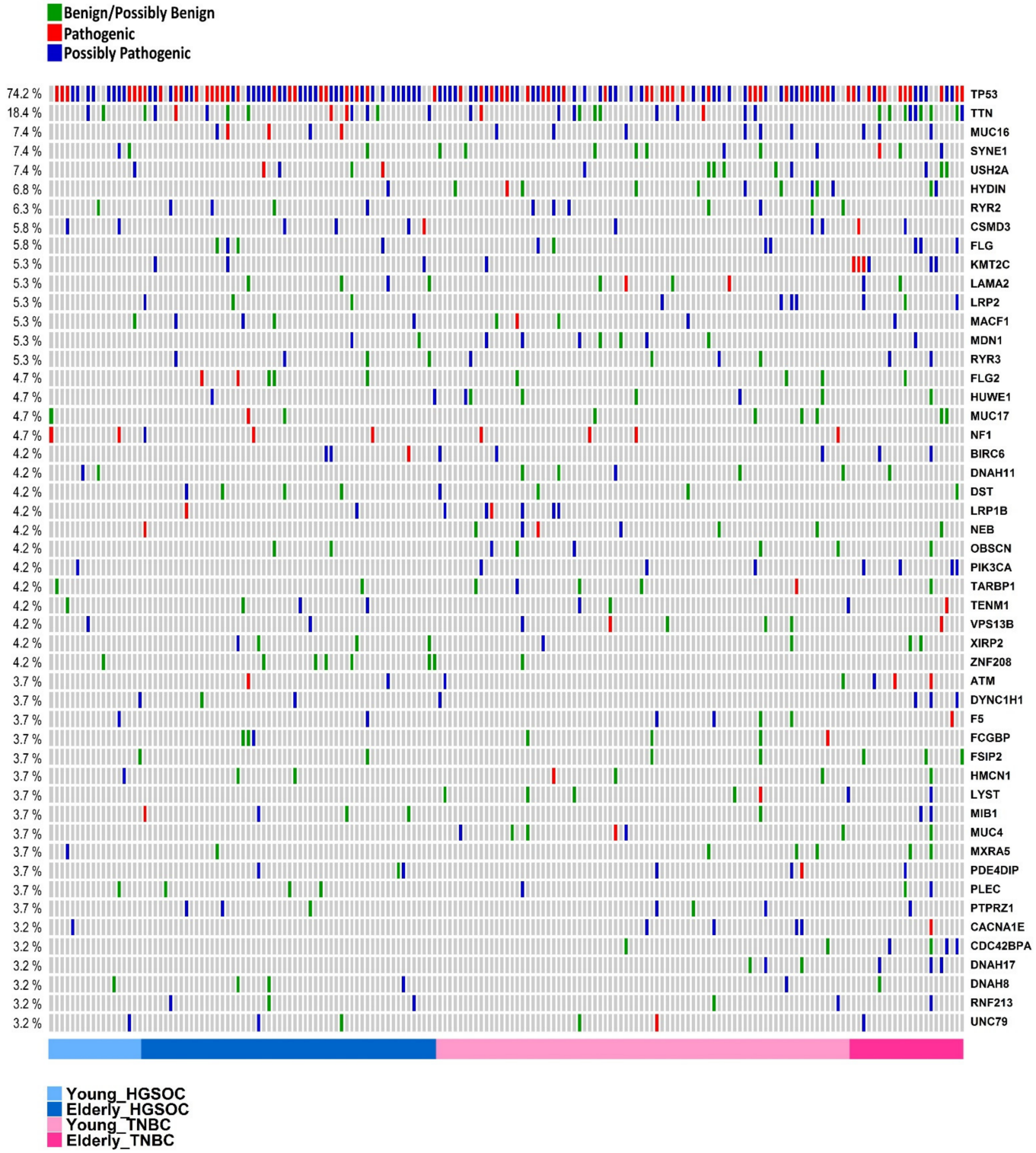


Figure 4. Frequently affected genes considering all samples of HGSOC and TNBC (young -Y and elderly -Y) classified according to their pathogenicity. Combined top 50 most frequently affected genes classified as potentially pathogenic (red), possibly pathogenic (blue) as described in methods, and benign (green). Each column represents a patient, and each line represents a gene (177 samples shown).

On the other hand, 8 of the top 50 (16%) genes were CGCs, where all of the mutations that were detected were potentially/possibly pathogenic, including *TP53*, *MUC16*, *CSMD3*, *KMT2C*, *NF1*, *BIRC6*, *LRP1B*, and *PIK3CA*.

3.5. Cancer Driver Genes

Our main goal was to evaluate what the potential cancer driver genes in the tumors from young and elderly HGSOC and TNBC patients were. We assumed that cancer driver genes or cancer-causing genes were those genes that were catalogued in the Cancer Gene Census database (CGC) (accessed May 2020) that were affected by potentially pathogenic (nonsense, frameshift, and canonical splice-site) or possibly pathogenic (missense variants predicted as pathogenic in at least one out of five variant effect prediction tools) variants (Tables S4–S6).

Most *TP53* variants were present in the DNA binding domain (DBD) in the four groups of samples (Y HGSOC: 78% vs. E HGSOC: 86%; Y TNBC: 88% vs. E TNBC: 75%). Some mutations were in commonly affected hotspots of the DBD, such as the *TP53* variant *p*.R175H, which was detected in eight samples (E HGSOC: $n = 1$ (2%); Y TNBC: $n = 6$ (10%); E TNBC $n = 1$ (6%)) and *p*.R248W, which was detected in six samples (E HGSOC $n = 4$ (8%); Y TNBC: $n = 1$ (1%); E TNBC: $n = 1$ (6%). A third *TP53* variant, that was exclusively observed in young TNBC patients ($n = 4$ or 6%), was p.R213*, a nonsense variant. All of the abovementioned variants were previously reported as germline variants and were shown to be related to Li–Fraumeni syndrome [57–59] (Figure S4).

We then evaluated which were the cancer driver genes for each specific tumor sample were, based on the above assumptions. All of the affected oncogenes and tumor suppressor genes classified as CGCs tier 1 and tier 2 appear on Table 3. For those patients with no affected oncogene or TSG, we assumed that the genes that were catalogued in the Candidate Cancer Gene Database (CCGD) and with supportive cancer-causing related literature as putative drivers.

Table 3. Oncogenes and TSGs altered on samples of HGSOC and TNBC from young and elderly patients and their mutational effects.

a—Young HGSOC			
Sample_ID	**OG**	**TSG**	**OG/TSG**
TCGA-09–1664-01	**KRAS** $^{++}$	**NF1** $^{-}$	-
TCGA-13–0792-01	CTNND2 $^{?}$	CIITA $^{?}$	**TP53** $^{-}$
TCGA-13–0793-01	MECOM $^{?}$	BAZ1A $^{?}$	**TP53** $^{-}$**, DAXX** $^{-}$
TCGA-13–0884-01	ACVR1 $^{?}$, TRRAP $^{?}$	CSMD3 $^{?}$, SLC34A2 $^{?}$	BIRC3 *, **TP53** $^{-}$
TCGA-24–1105-01	-	-	**TP53** $^{-}$
TCGA-24–1416-01	**PIK3CA** $^{++}$	-	**TP53** $^{-}$
TCGA-25–1328-01	-	-	-
TCGA-25–2404-01	-	ZMYM3 $^{?}$	**TP53** $^{-}$
TCGA-25–2408-01	-	CDKN1B $^{?}$	-
TCGA-29–1688-01	HIP1	-	**TP53** $^{-}$
TCGA-29–1769-01	SETDB1 $^{?}$	-	RAD21 $^{?}$
TCGA-29–2436-01	-	KAT6B $^{?}$	-
TCGA-36–2530-01	ERBB2 $^{?}$	DNMT3A $^{?}$, IKBK	-
TCGA-36–2537-01	-	-	**TP53** $^{-}$
TCGA-36–2538-01	-	ERCC3 $^{?}$, RB1-	**TP53** $^{-}$
TCGA-36–2540-01	-	-	-
TCGA-59–2363-01	-	CLTCL1 $^{?}$, CSMD3 $^{?}$, **NF1** $^{-}$	**TP53** $^{-}$

Table 3. *Cont.*

a—Young HGSOC			
Sample_ID	**OG**	**TSG**	**OG/TSG**
TCGA-61–1725-01	JAK3 [?]	RAD17 [?]	**TP53** [-]
TCGA-61–2008-01	-	BLM [?], CNTNAP2 [?]	**TP53** [-]
TCGA-61–2109-01	CHD4 [?]	FAT1 [?]	**TP53** [-]
TCGA-61–2611-02	ACKR3 [?]	-	**TP53** [-]
b—Elderly HGSOC			
Sample_ID	**OG**	**TSG**	**OG/TSG**
TCGA-04–1331-01	DDIT3 [?]	**BRCA2** [-], KDM5C [?], LATS1 [?], NF1 [?]	TBL1XR1 [?], **TP53** [-]
TCGA-04–1337-01	-	CDH1 [?], FBXW7 [−]	ELF4 [?], **TP53** [-]
TCGA-04–1338-01	MTOR [?]	CDH11 [?], FAT1, KMT2C [?]	**CREBBP** [-], PABPC1 [?], **TP53** [-]
TCGA-04–1341-01	CTNND2 [?], GRM3 [?], PDGFRA [?], PREX2 [?], TCF7L2 [?]	CDH10 [?], FANCA [?], ATR [?]	**TP53** [-]
TCGA-04–1342-01	KIT [?]	LZTR1 [?], SMARCA4 [?]	-
TCGA-04–1347-01	BRAF [?], GRM3 [?]	DROSHA [?], FAT1 [?]	**TP53** [-]
TCGA-04–1351-01	-	-	-
TCGA-04–1365-01	AFF3 [?]	**ZFHX3** [-]	**TP53** [-]
TCGA-04–1517-01	-	-	**TP53** [-]
TCGA-04–1652-01	-	LRP1B [?], PTPRT [?]	**TP53** [-]
TCGA-09–0364-01	-	CNOT3 [?], CNTNAP2 [?]	**TP53** [-]
TCGA-09–1661-01	DDIT3 [?]	PBMR1 [?], TNFAIP3 [?]	**TP53** [-]
TCGA-09–1672-01	**KRAS** [++]	-	-
TCGA-09–1674-01	LPP [?], PIK3CB [?]	**HNF1A** [-], PALB2, **RB1** [-], SETD2 [?]	**TP53** [-]
TCGA-09–2044-01	CTNNB1 [?], MET [?]	PTPN13 [?], RMI2 [?], **TSC1** [-]	**TP53** [-]
TCGA-10–0933-01	MUC16 [?]	**RB1-**	RHOA [?], **TP53** [-]
TCGA-10–0938-01	NUP98 [?], SND1 [?]	-	**TP53** [-]
TCGA-13–0755-01	MUC16 [?]	ARID2 [?], FEN1 [?], GRIN2A [?], KMT2C [?]	**TP53** [-]
TCGA-13–0802-01	-	-	-
TCGA-13–0888-01	CRTC1 [?], MYCN [?], RAP1GDS1 [?], SOX2 [?]	BARD1 [?], EBF1 [?], EP300 [?], USP44 [?], ZFHX3 [?]	**TP53** [-]
TCGA-13–0889-01	ERBB2 [?]	AMER1 [?], ATRX [?]	**TP53** [-]
TCGA-13–1411-01	-	CLTCL1 [?], DICER1 [?], SMARCB1 [?]	**TP53** [-]
TCGA-13–1481-01	ARHGAP5 [?], GLI1 [?], MUC16 [?], RARA [?]	**BRCA2** [-]	**TP53** [-]
TCGA-13–1507-01	ATF1 [?], SETDB1 [?]	MLH1 [?], PTPRT [?]	**TP53** [-]
TCGA-20–0991-01	MACC1 [?]	-	**TP53** [-]
TCGA-20–1686-01	WAS [?]	CSMD3 [?], SLC34A2 [?]	IRS4 [?], **TP53** [-]
TCGA-23–1116-01	-	-	**TP53** [-]

Table 3. *Cont.*

b—Elderly HGSOC			
Sample_ID	**OG**	**TSG**	**OG/TSG**
TCGA-23–2641-01	-	-	NOTCH2 [?], **TP53** [-]
TCGA-24–0966-01	-	LZTR1 [?], N4BP2 [?]	CREBBP [?], TET1 [?], **TP53** [-]
TCGA-24–0982-01	-	SMC1A [?]	**TP53** [-]
TCGA-24–1422-01	CHD4 [?], MUC16 [?]	XPC [?]	**TP53** [-]
TCGA-24–1552-01	-	-	**TP53** [-]
TCGA-24–1849-01	BRD4 [?]	GPC5 [?], **RB1** [-]	BTK [?], **TP53** [-]
TCGA-24–2030-01	-	NDRG1 [?]	BIRC6 [?], **TP53** [-]
TCGA-24–2033-01	AFF4 [?], GNAS [?]	DNM2 [?], STAG1 [?]	BIRC6 [?], CBLC [?], EZH2 [?], **TP53** [-]
TCGA-24–2261-01	-	CSMD3 [?]	**TP53** [-]
TCGA-25–1325-01	MUC16 [?]	FAT4 [?]	**TP53** [-]
TCGA-25–1329-01	MTOR [*]	-	**TP53** [-]
TCGA-25–1634-01	EGFR [?]	EIF3E [?]	**TBX3** [-]
TCGA-25–2392-01	FGFR4 [?], SRC [?]	**CDK12** [-], USP44 [?]	**TP53** [-]
TCGA-25–2393-01	-	MED12 [?], RNF43 [?]	BCORL1 [?]
TCGA-25–2399-01	CACNA1D [?]	**ATM** [-], NCOR1 [?]	**TP53** [-]
TCGA-25–2400-01	WAS [?]	LPR1B [?]	**TP53** [-]
TCGA-29–1702-01	KIT	SLC34A2 [?]	**TP53** [-]
TCGA-29–1761-01	-	CNTNAP2 [?], **FANCA** [-], PRMD2, SMC1A [?]	**TP53** [-]
TCGA-29–1771-01	KMT2A [?]	**NF1** [-]	**TP53** [-]
TCGA-29–1774-01	-	DROSHA [?], SIRPA [?]	-
TCGA-29–1778-01	KMT2A [?]	-	**TP53** [-]
TCGA-29–2429-01	PSIP1 [?], ROS1 [?]	ATM [?], **FLCN** [-]	-
TCGA-31–1950-01	PAX3 [?]	CCD6 [?], CTCTCL1 [?]	CREBBP [?], **TP53** [-]
TCGA-36–1575-01	-	KDM5C [?]	**TP53** [-]
TCGA-36–1576-01	TSHR [?]	-	-
TCGA-36–2543-01	RET, UBR5 [?]	NF1 [-]	**TP53** [-]
TCGA-42–2587-01	AFF4 [?], FCRL4 [?], POU2AF1 [?]	CDC73 [?], WNK2	BCL9L [?], ESR1 [?], **TP53** [-]
TCGA-59–2352-01	A1CF [?], SIX1 [?], ZEB1 [?]	CCDC6 [?], CNTNAP2 [?], CSMD3 [?], PRDM1 [?]	BIRC6 [?], **TP53** [-]
TCGA-61–1730-01	ALK [?], PLCG1 [?]	FAT1 [?]	**TP53** [-]
TCGA-61–1741-01	CHD4 [?]	-	**DAXX** [-], **TP53** [-]
TCGA-61–1899-01	CDH10 [?], PIK3CB [?]	CSMD3 [?], KMT2C [?]	-
TCGA-61–2012-01	BCL6 [?], SGK1 [?]	**PHF6** [-], POLG [?]	NOTCH1 [?], TBX3 [?]
TCGA-OY-A56Q-01	-	-	**TP53** [-]

Table 3. *Cont.*

c—Young TNBC				
Sample_ID	**OG**	**TSG**	**OG/TSG**	**Putative Drivers**
BB01_044	-	-	BIRC6 [?], KMT2D [?], **TP53** [-]	-
BB01_074	BCL9 [?], CCND3 [?]	ATM [?], DICER1 [?], LRP1B [?]	**CREBBP** [-], **TP53** [-]	-
BB01_109	ETV1	-	**TP53** [-]	-
BB01_126	RET [?]	-	**TP53** [-]	-
BR067	ACK3 [?], ETV4 [?], GRM3 [?], MUC4 [?]	-	STAT5B [?], **TP53** [-]	-
BR078	-	-	-	HERC1, HUWE1
BR080	BRAF [?]	NRG1 [?]	**TP53** [-]	-
BR088	-	-	-	KIF13B
BR091	-	-	**TP53** [-]	-
BR097	NR4A3 [?], **PIK3CA** [++], PTPN11 [?]	**DICER1** [−], NF1 [-]	FOXO3 [?], **TP53** [-]	-
BR100	MET [?], PIK3CB [?]	KMT2C [?], LRP1B [?]	**TP53** [-]	-
BR105	RANBP2 [?], UBR5 [?]	LRP1B [?]	**TP53** [-]	-
BR108	MUC16 [?]	**SMARCA4** [-]	BIRC6 [?], **TP53** [-]	-
BR121	AFF3 [?], BRAF [?]	ATRX [?]	-	-
BR145	-	ASXL2 [?]	**TP53** [-]	-
BR164	UBR5 [?]	-	**TP53** [-]	-
BR176	ERBB2 [?], MUC4	CLTCL1 [?]	**TP53** [-]	-
BR200	PREX2 [?], TRRAP [?]	LRIG3 [?]	ELF4 [?], **TP53** [-]	-
BR255	CHST11 [?], DNM2 [?]	BARD1 [?], ETV6 [?], FHIT [?], LPR1B [?], POLG [?], TET2	-	-
BR301	MUC4	CTLCL1 [?]	LEF1 [?], **TP53** [-]	-
BR313	USP6	-	**TP53** [-]	-
BR367	STIL [?]	-	**TP53** [-]	-
BR393	-	ZFHX3 [?], ZMYM3 [?]	TET1 [?], **TP53** [-]	-
BR395	-	-	**TP53** [-]	-
BR-M-045	CARD11, HLF [?], MUC16 [?]	LRP1B [?]	**TP53-**	-
BR-V-022	-	CDH1, LRP1B [?]	**TP53** [-]	-
BR-V-051	-	-	**TP53** [-]	-
BR-V-070	-	-	NOTCH1 [?]	-
PD11326a	-	AXIN2, KAT6B	**TP53** [-]	-
PD13627a	-	PTPRC [?]	-	-
PD18024a	-	-	**TP53** [-]	-
PD22036a	-	**NF1** [-]	-	-
PD22358a	CDH17 [?]	POLE [?], SETD2 [?]	-	-
PD23554a	FLI1	BAZ1A [?]	KMT2D [?], **TP53** [-]	-
PD23563a	IL6ST [?]	SPOP [?]	**TP53** [-]	-
PD23566a	HIST1H3B [?], MET [?]	**PTPRD** [-]	FES [?], **TP53** [-]	-

Table 3. *Cont.*

c—Young TNBC				
Sample_ID	**OG**	**TSG**	**OG/TSG**	**Putative Drivers**
PD24182a	MUC4 [?], TRIM27	CDX2 [?], CSMD3 [?]	**TP53** [-]	-
PD24186a	RAC1 [?]	-	-	-
PD24191a	CACNA1D [?], MUC16 [?], MUC4 [?]	**PTEN** [-]	-	-
PD24196	-	FAT4 [?], PTPN13	**TP53** [-]	-
PD24202a	PPM1D [?]	**NF1** [-]	FOXO3 [?]	-
PD24337a	XPO1 [?]	CDH11 [?], PTPRD [?]	**TP53** [-]	-
PD3905a	**PIK3CA** [++]	-	**TP53** [-]	-
PD4005a	-	ERCC5, PTPN6 [?]	TBL1XR1 [?], **TP53** [-]	-
PD4006a	-	-	-	MYT1, NFKB1
PD4107a	ABL2, POU2AF1 [?]	ARHGEF12 [?]	NTRK1 [?], **TP53** [-]	-
PD4833a	UBR5 [?]	-	BCL9L [?], **TP53** [-]	-
PD4836a	-	ARID1B [?], LARP4B [?]	**TP53** [-]	-
PD5930a	-	**PHF6** [-]	RECQL4 [?]	-
PD5945a	-	-	NFE2L2 [?], **TP53** [-]	-
PD6406a	PREX2 [?]	-	-	-
PD6411a	PREX2	-	**TP53** [-]	-
PD6413a	-	TET2 [?]	-	-
PD6722a	CSF3R [?]	FAZ [?], RB1-	**TP53** [-]	-
PD9004a	SSX1, TNC	CDC73 [?], DNMT3A [?], GRIN2A [?]	**TP53** [-]	-
PD9595a	CSF1R [?]	-	BTK [?], **TP53** [-]	-
PD9696a	BCL11A [?]	WNK2 [?]	CUX1 [?], **TP53** [-]	-
SA083	-	**ZFHX3** [-]	-	-
SA097	BRD4 [?], PRDM16 [?]	CCDC6	**TP53** [-]	-
SA208	-	**PIK3R1** [-]	-	-
SA220	FOXP1 [?], NFATC2 [?]	ATR [?],ROBO2, SMARCA4 [?]	-	-
SA231	-	-	-	MBD2
SA235	-	BAP1 [?]	-	-
SA236	FLT3 [?], MUC16 [?]	ARID2 [?], IGF2BP2 [?]	**TP53** [-]	-
TCGA-A1-A0SP-01	-	ARID1A [?], ZMYM3 [?]	**TP53** [-]	-
TCGA-A2-A04P-01	ERBB3, GRM3 [?], **PIK3CA** [++], SETDB1 [?]	ACVR2A [?], BRCA2 [?], ROBO2 [?], TNFAIP3 [?]	IRS4 [?], **TP53** [-]	-
TCGA-A2-A0CM-01	KAT6A [?]	CDH11 [?], MGMT [?]	-	-
TCGA-A2-A3XU-01	-	-	-	-
TCGA-AO-A124–01	CCR7 [?], CSF3R [?], MAP2K2 [?], MUC16 [?]	BRCA2 [?], **CDKN2A-**, EXT2 [?], FANCG [?]	**TP53** [-]	-
TCGA-AO-A129–01	BCL11A [?], IL6ST [?], USP6 [?]	DDX3X [?], KAT6B [?]	MAP3K13 [?], **TP53** [-]	-
TCGA-AO-A12F-01	-	-	-	TAF1

Table 3. *Cont.*

c—Young TNBC				
Sample_ID	**OG**	**TSG**	**OG/TSG**	**Putative Drivers**
TCGA-AR-A0TU-01	AKT3 [?], ETV1 [?]	**GRIN2A [-], NF2 [-]**	-	-
TCGA-AR-A0U1–01	-	BCOR [?], RSPO2 [?]	**TP53 [-]**	-
TCGA-B6-A0IQ-01	-	SMC1A [?]	**TP53 [-]**	-
TCGA-B6-A0RS-01	A1CF [?], CHD4 [?], MET [?], MUC16 [?]	MED12 [?], MUTYH [?], ZMYM3 [?]	**TP53 [-]**	-
TCGA-B6-A0RT-01	GRM3 [?], MUC16 [?]	-	**TP53 [-]**	-
TCGA-B6-A0RU-01	ARHGAP5 [?]	-	POLQ [?], **TP53 [-]**	-
TCGA-B6-A0WX-01	-	SLC34A2 [?], ZNRF3 [?]	EPAS1, **TP53 [-]**	-
TCGA-BH-A0BL-01	HLF [?]	CSMD3 [?], **LZTR1 [-]**	**TP53 [-]**	-
TCGA-BH-A0E0–01	ALK [?]	-	NOTCH1 [?],**RHOA [-]**, **TP53 [-]**	-
TCGA-D8-A27F-01	MDM4	**BAP1 [-]**, CSMD3 [?]	BIRC6 [?], **TP53-**	-
TCGA-E2-A14N-01	MAPK1 [?], ROS1 [?]	ABI1 [?], PTCH1 [?]	**TP53 [-]**	-
TCGA-E2-A1L7–01	SGK1 [?]	GPC5 [?], PTPRB [?]	**TP53 [-]**	-
TCGA-E9-A3QA-01	-	**NF1-**, POT1	-	-
TCGA-OL-A5RW-01	AFF3, CTNND2 [?], HNRNPA2B1 [?], MUC4, RAF1 [?]	ATM, AXIN1 [?], CBLB [?], **MED12 [-]**	IRS4 [?]	-
TCGA-OL-A66I-01	RANBP2 [?]	CYLD [?]	**TP53 [-]**	-
d—Elderly TNBC				
Sample_ID	**OG**	**TSG**	**OG/TSG**	
PD10011a	MPL [?]	**ASXL1 [-], KMT2C [-]**	**TP53 [-]**	
PD13298a	RET	CSMD3 [?], **KMT2C [-]**, PTPRB [?]	NOTCH2 [?], **TP53 [-]**	
PD24333a	**BRAF [++], KRAS [++]**, MUC16 [?], PDGFRA [?], **PIK3CA [++]**, ROS1 [?]	**KMT2C [-]**	RAD21 [?]	
PD6047a	-	KMT2C [?]	**TP53 [-]**	
PD7067a	FLI1 [?], TAL1 [?]	ATM [?], **BRCA1 [-]**	POLQ [?], **TP53 [-]**	
PD8982a	CALR [?], MUC16 [?], TRRAP [?]	ARID1B [?], CDH10 [?], CLTC [?], MED12 [?], SPEN [?]	BIRC6 [?], **MAP3K1 [-]**, NOTCH1 [?], POLQ [?], **TP53 [-]**	
PD9575a	BCL3 [?], **ERBB3 [+]**	**PTEN-**, PTPN13 [?]	**TP53 [-]**	
PD9584a	H3F3A [?], **HRAS [++]**	-	TBX3 [-]	
SA031	BRAF [?], MET [?], NRAS [?]	DDX3X [?], STAG2 [?]	BCORL1	
SA052	-	**ATM [-]**	**KDM6A [-], NOTCH1 [-]**	
SA056	ARGHAP5 [?], ERBB4 [?], **PIK3CA [++]**	-	**TP53 [-]**	
SA106	FCRL4, HIF1A	**BRCA2-**, CAMTA1, CSMD3 [?], FLCN [?], SPEN [?]	BCL9L [?], **TP53 [-]**	
TCGA-A2-A1G1–01	-	-	**CREBBP [-]**, **TP53 [-]**, MAP3K1 [?]	
TCGA-AC-A2BK-01	-	MYH9 [?], PTPRD [?]	**KMT2D [-], TP53 [-]**	

Table 3. *Cont.*

d—Elderly TNBC			
Sample_ID	**OG**	**TSG**	**OG/TSG**
TCGA-AR-A1AJ-01	-	**PTEN** $^{-}$	**TP53** $^{-}$
TCGA-BH-A0WA-01	IL6ST $^{?}$	AXIN1 $^{?}$, BRCA1 $^{-}$	TP53 $^{-}$
TCGA-BH-A18G-01	ARHGAP5 $^{?}$, BCL9 $^{?}$, ERBB3 $^{?}$, MAML2 $^{?}$, MUC16 $^{?}$, MUC4, TRIM27 $^{?}$, ZEB1	ARHGAP26 $^{?}$, ARHGEF10L $^{?}$, **ARID1B** $^{-}$, ASXL2 $^{?}$, **ATM** $^{-}$, ATRX $^{?}$, BCOR, CARS $^{?}$, CCNB1IP1 $^{?}$, CDH10 $^{?}$, CLTCL1 $^{?}$, CNOT3, CYLD $^{?}$, DICER1 $^{?}$, EP300 $^{?}$, ERCC3 $^{?}$, ERCC5 $^{?}$, ETV6 $^{?}$, IGH2BP2 $^{?}$, KDM5C $^{?}$, KMT2C $^{?}$, MLF1 $^{?}$, MYH9 $^{?}$, **PPP2R1A** $^{-}$, TET2, ZFHX3 $^{?}$	FES $^{?}$, KDM6A $^{?}$, KMT2D $^{?}$, MAP3K1 $^{?}$, QKI $^{?}$
TCGA-BH-A1F0–01	SGK1 $^{?}$, STAT3 $^{?}$	ASXL2 $^{?}$, GPC5 $^{?}$, KMT2C $^{?}$	-
TCGA-BH-A1FC-01	CTNNA2 $^{?}$, DDX5 $^{?}$, TCF7L2 $^{?}$, TEC $^{?}$	CLTC $^{?}$, NCOR1 $^{?}$	**TP53** $^{-}$
TCGA-C8-A12K-01	-	ATP2B3 $^{?}$, IKZF1 $^{?}$, STK11 $^{?}$	FES $^{?}$, IRS4 $^{?}$, **TP53** $^{-}$
TCGA-C8-A131–01	**KRAS** $^{++}$, **PIK3CA** $^{++}$	NTHL1 $^{?}$, PTPRD $^{?}$, SETD2 $^{?}$	**TP53** $^{-}$
TCGA-D8-A1JK-01	CTNNA2 $^{?}$, KATA $^{?}$, **PIK3CA** $^{++}$, **STAT3** $^{+}$, TNC $^{?}$	ASXL1 $^{?}$, FANCD2 $^{?}$, FAT4 $^{?}$, NCOR1 $^{?}$	**TP53** $^{-}$
TCGA-E2-A1LK-01	CACNA1D $^{?}$	EP300 $^{?}$, LARP4B $^{?}$, MED12 $^{?}$	**KDM6A** $^{-}$, **TP53** $^{-}$

Oncogenes and TSG (in accordance with Cancer Gene Census—Cosmic—available at https://cancer.sanger.ac.uk/census, accessed on 1 May 2020) altered on samples of young and elderly HGSOC and TNBC patients. OG: oncogene; TSG: tumor supressor gene; OG/TSG: genes with dual role. Green-highlighted genes: genes with benign variants. Red-highlighted genes: Ras pathway-related gene; purple-highlighted gene: RB pathway-related gene. Variant effect (OncoKB): $^{++}$: gain-of-function; $^{+}$: likely gain-of-function; $^{-}$: loss-of-function; $^{-}$: likely loss-of-function; *: likely neutral; $^{?}$: variant or gene not curated or not found in the literature. Bold: genes with annotated likely loss and gain-of-function or just loss or gain-of-function in OncoKB.

For the HGSOC samples, almost all variants in the CGC genes were considered to be potentially pathogenic/possibly pathogenic (based on the described criteria), except for variants in eight genes, *HIP1*, *IKBK*, *FAT1*, *PALB2*, *KIT*, *PRMD2*, *RET*, and *WNK2*, each one detected separately in one single sample. Looking for TNBC samples, the missense variants in 32 CGC genes were considered to be non-pathogenic, in accordance with the criteria exposed above, with focus on *MUC4*, which was considered non-pathogenic in four different tumors. Another gene, *CDH1*, which is considered a cancer driver for hereditary lobular breast cancer, presented a non-pathogenic variant in this cohort of TNBC samples. To further assign the mutation effect of the variants in CGC oncogenes and TSGs for the selected series of samples, we explored each variant (oncogenes and TSGs) in the OncoKB database. Detailed data are reported in Table 3.

In all four groups (HGSOC and TNBC young and elderly), most of the samples presented at least one affected oncogene in association with at least one affected tumor suppressor gene, which were represented by combinations of variants in OG and TSG, OG and a dual role gene (OG or TSG), TSG and a dual role gene, or even in at least 2 dual role genes (one might assume the role of OG and the other of TSG). This condition was verified in 66.6% and 85.0% of the young and elderly HGSOC cases as well as in 72.1% and 100% of the samples from young and elderly TNBC, respectively (Table 1). Among the associations were genes such as *KRAS* and *NF1* (Y-HGSOC), *ERBB2* and *DNMT3A* (Y-HGSOC), *PIK3CA* and *TP53* (Y-TNBC and E-TNBC), among others (Table 3).

In the young HGSOC cohort, three samples (14.3%) presented variants in the genes that were involved in the Ras and or PIK3CA signaling pathways, including *KRAS* and *PIK3CA* (each one in one single sample) and *NF1* (altered in two samples, one of them, in association with *KRAS*). In two of these samples, there was an association with *TP53* mutations. In addition, in three samples, the DNA repair genes *RAD17*, *BLM*, and *ERCC3*, were mutated, which occurred in concomitance with the *TP53* mutations. In five tumors (23.8%), only one potentially pathogenic/possibly pathogenic CGC, such as *CDKN1B* or *KAT6B*, which are considered to be TSGs (each one in a single tumor), was mutated, as was *TP53* (in three different samples), which is considered to have a dual role gene, in accordance with CGC database (https://cancer.sanger.ac.uk/cosmic/census?tier=all#cl_search; accessed on 22 May 2020) [60] (Table 3).

In elderly HGSOC, mutations in the genes that are involved in the Ras and or PIK3CA signaling pathways were detected in nine samples (15%), including the genes *KRAS* and *BRAF* (each one in 1 tumor); *PIK3CB* and *MTOR* (each one in 2 tumors); and *NF1* (in three different samples). Multiple gene mutation associations were detected. In seven tumors (11.6%) there was an association of Ras and/or PIK3CA defective signaling with mutated *TP53*. Another mutation combination was observed between *RB1* and *TP53,* which was detected in three different samples (5%). In addition, in another 12 samples from elderly HGSOC (20%), a DNA repair gene was mutated; such affected genes included *ATR*, *CDK12*, *MLH1*, *POLG*, *FEN1*, *BARD1,* and *XPC*, each one found in a single sample; *ATM*, *BRCA2,* and *FANCA*, each one found in two different samples. In one sample, there was a concomitant mutation in *FANCA* and *ATR*, and in other ten samples (16.6% of the total samples), *TP53* was concomitantly mutated with one of these genes. *MUC16* was mutated in five samples (8.3%) always in concomitance with multiple other CGCs, including TSGs, as well as *TP53*. In four tumors, a growth factor receptor such as *EGFR*, *ERBB2*, *FGFR4*, and *PDGFRA*, was mutated, in combination with other mutations in TSGs, generally in concomitance with *TP53*. Finally, in four samples, the only mutated gene was *TP53,* and in another two samples, only one oncogene was mutated (*KRAS* or *TSHR*). Nonetheless, in two samples (3.4%) mutations, not one defective CGC was detected (Table 3).

As described above, most HGSOC and TNBC patients carry a defective TP53, regardless of age onset (Y: 66.7%; E: 81.7%). Among those who do not, 23.8% of young and 6.67% of elderly patients carry alterations in other TSGs. Most of these genes are involved in epigenetic regulation (*KMT2C*, *KA6T6B*, *DNMT3,* and *DROSHA*), transcriptional regulation (*MED12*, *LZTR1*, *PHF6*, *SMARCA4*, and *TBX3*), and DNA repair (*POLG*, *ATM*, *RAD21*). Among these TSGs, *FLCN*, *PHF6,* and *TBX3* (found in elderly patients) were the only genes whose variants were reported by OncoKB as presenting likely loss-of-function effects, which might indicate new putative drivers.

Considering TNBC from young patients, most samples (90.7%) presented at least one affected TSG according to the CGC database. Besides *TP53* (mutated in 69.8% of the samples), other gene variants that had already been classified with likely loss-of-function in the OncoKB database were *NF1* (affected in three samples or 3.5%) and *PTEN*, *PHF6*, *ZFHX3*, *PIK3R1*, *GRIN2A*, *NF2,* and *MED12* (each one affected in a single sample). The other genes cited as TSG in Table 3 are considered as TSG in the CGC database and are affected by potential damaging alterations, including *ATRX*, *BARD1+POLG; NOTCH1; PTPRC; POLE+SETD2; TRT2; ATR + SMARCA4; BAP1; CDH1 + MGMT; and AXIN1 + CBLB* (each gene, alone or in the combination shown, affected in one sample). Despite that, in eight samples (9.3%), there were no affected TSG. Moreover, in 12 samples (13.9%), only 1 CGC gene was mutated, including 2 oncogenes, PREX*1* and *RAC1*; 5 TSGs, *BAP1*, *NF1*, *PIK3R1*, *PTPRC,* and *ZFHX3*, and 2 dual role genes, *NOTCH1* and *TP53*, the latter, in three samples. Besides that, in six tumors (6.9%), a well-established cancer driver gene was not detected (Table 3).

Among TNBC from young patients, 18 out of 86 (20.9%) tumors presented mutations in genes involved in the Ras and/or PIK3CA signaling pathways, such as *AKT3*, *BRAF*, *MAP2K2*, *MAPK1*, *NF1*, *NF2*, *PTEN*, *PIK3CA*, *PIK3CB*, *PIK3R1*, *RAC1*, *RAF1,* and *RANBP2.*

NF1, *PIK3CA*, and *BRAF* were mutated in four, three, and two samples, respectively. In addition, there was an association between mutated OGs and TSG, such as *AKT3* and *NF2*, and PIK3CA and NF1 in two samples. Besides that, NF1 was the sole mutated CGC in two samples, and *PIK3R1* and *RAC1* were the only mutated CGC, each one detected in a single sample. In nine of these samples, gene mutations involved in the Ras pathway were in concomitance with *TP53* mutations.

Mutations in the DNA repair genes, such as *BRCA2*, *ATM*, *BAP1*, *BARD1*, *FANCG*, *MGMT*, *MUTYH*, *POLE*, *POLG*, *POLQ*, *and RECQL4* were detected in nine tumors. *BAP1* and *BRCA2* were mutated in two tumors, in association with *FANCG* in one of the tumors. *BARD1* and *POLG* were concomitantly mutated in one tumor. In addition, in five tumors, a defective DNA repair gene was concomitant with *TP53* mutation.

Among the elderly TNBC patients, all 23 samples presented more than one affected CGC, represented by at least one mutated oncogene and one TSG. In addition, all the tumors presented gene variants classified as likely loss of function in the OncoKB database, except for one tumor. Besides TP53 (affected in 73.9% of the samples), other affected TSG were *KMT2C*, *TBX3*, and *ATM*. A dysfunction in the Ras and/or PI3K pathways was observed in 8 out of 23 samples (34.8%), among which five (21.7%) were associated with *TP53* mutation. The Ras family members were mutated in four tumors (*KRAS*, $n = 2$; *HRAS* and *NRAS*, $n = 1$); *PIK3CA* in four tumors (in concomitance with *KRAS* in two tumors), and *PTEN* in another two tumors. In addition, *BRCA1* was mutated in two tumors, and *BRCA2* was found in one tumor, all three in concomitance with *TP53* (13% of the tumors). Interestingly, all of the *KRAS* variants from both tumor types were located at the same position, a known hotspot (p.G12V) (Table 3).

Ontology analysis was performed to verify which biological processes or molecular functions were enriched for the genes that were found to harbor potentially pathogenic variants (nonsense, frameshift or canonical splice site) (HGSOC—young: 115 genes vs. elderly: 435 genes; TNBC—young: 671 genes vs. elderly: 363 genes) or through variant effect prediction tools (HGSOC—young: 156 genes vs. elderly: 538 genes; TNBC—young: 935 genes vs. elderly: 568 genes) in both age groups of HGSOC and TNBC. Processes that were enriched in the two age groups from both cancer types were cell adhesion and motility, the extracellular matrix, and ion channels (Tables S8–S11).

4. DNA Repair

We then carefully searched for genes that were directly or indirectly related to DNA repair. This analysis included DNA repair genes that were not currently considered a CGC gene, thus expanding the view of the previously conducted analyses of cancer driver genes (Figure 4). Excluding *TP53* mutations, which were the most frequent mutations found in both age groups for both HGSOC and TNBC by far, potentially or possibly pathogenic variants in DNA repair were detected in 4 out of 21 tumors from young HGSOC patients, i.e., 19% and in 20 out of 60 patients (33%) in the elderly cohort, including genes involved in homologous recombination repair, such as *RAD50* in young patients as well as *BRCA2*, *FANCA*, and *BARD1*, in elderly patients. Three tumors from elderly patients, but none from young patients, presented potentially pathogenic/possibly pathogenic mutations In additional two genes, besides *TP53*: *BARD1* and *CLK2*; *FANCA* and *TOPBP1;* or *FANCA* and *ATR*. Most *TP53*wt tumors did not present any potentially pathogenic or possibly pathogenic variants in their DNA repair genes, except *ATM*, *POLG*, and *RFC1*, each of which was in three different tumors from elderly patients (Figure 5A,B; Table S4).

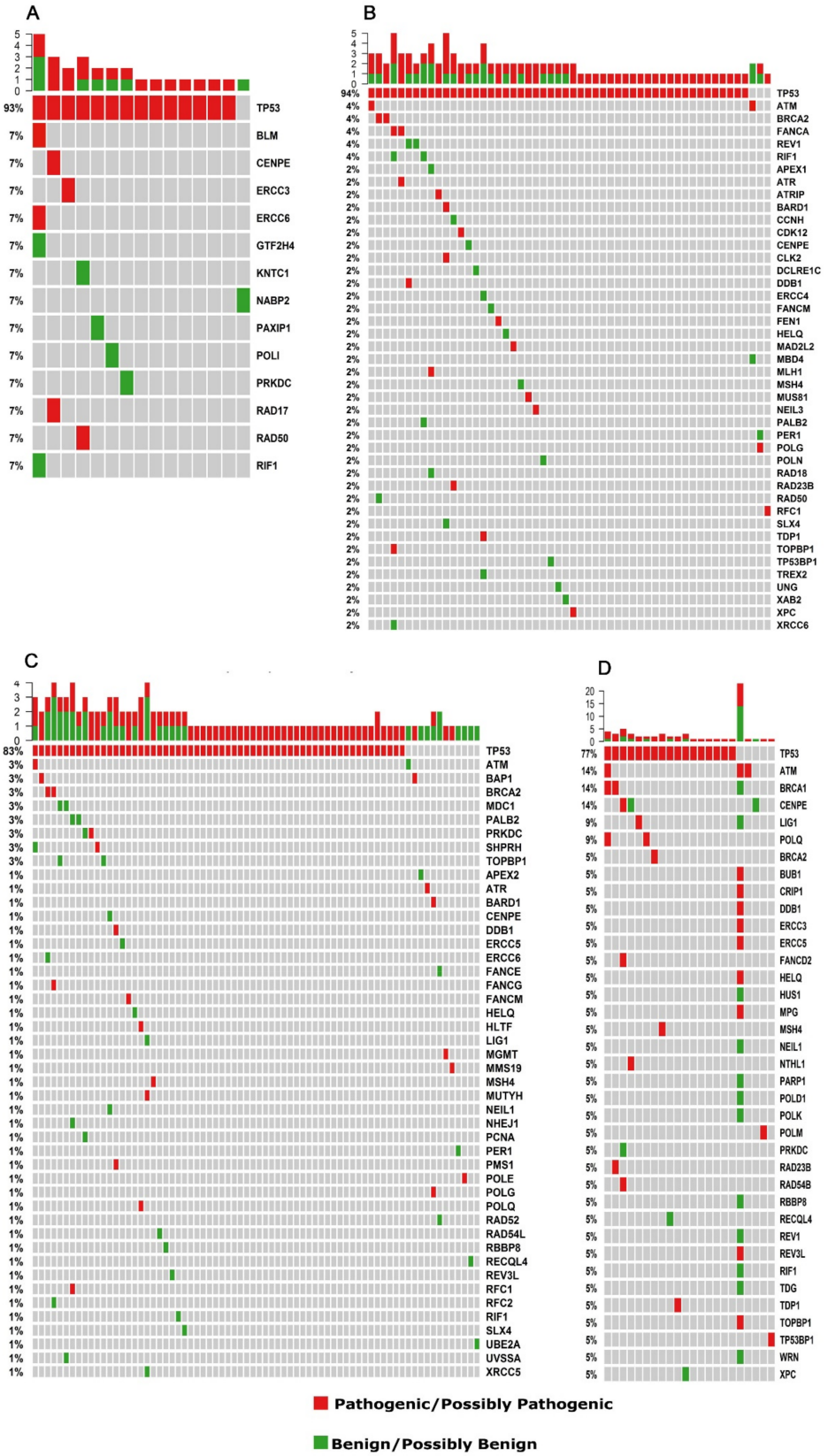


Figure 5. Affected DNA-repair related genes. (**A**) Young HGSOC patients; (**B**) elderly HGSOC patients; (**C**) young TNBC patients; (**D**) elderly TNBC patients. Each column represents a patient, and each line represents a gene. Potentially and possibly pathogenic variants are shown in red, and potentially and possibly benign variants are shown in green.

In TNBC, excluding *TP53*, potentially pathogenic or possibly pathogenic variants were present in 21 (24%) and 13 (56%) of tumors from young and elderly patients, including HRR genes from 3.5% and 21.7% of the tumors from young (*BRCA2*, *BARD1*) and elderly (*BRCA1*, *BRCA2*, *HELQ*, *RAD54B*) patients, respectively. Samples with concomitant DNA repair variants other than *TP53*, occurred in four samples from young patients (*BRCA2* and *FANCG; HLTF* and *POLQ*; *POLG* and *BARD1; PMS1* and *DDB1*) and in four samples from elderly patients (*ATM*, *BRCA1* and *POLQ*; *BRCA1* and *RAD23B*; *CENPE*, *FANCD2*, and *RAD54B)*, including one sample with 10 potentially pathogenic/possibly pathogenic DNA repair genes (Figure 5C,D; Table S4).

5. Discussion

We compared two tumor types that share similar molecular features. As seen in other HGSOC and TNBC studies, *TP53* was the most commonly mutated gene in both age groups [23,34,61,62]. The previously cited studies showed a similar *TP53* frequency (~70–90%) compared to the ratio that we observed in this study. In the present series, the frequency of transversions in young TNBC patients was 62%, which was higher when compared to the elderly patients as well as to HGSOC from both age groups, which may be associated with the presence of mutational signature 13. However, this result should be reevaluated because the analysis was not based on raw sequencing data, as described in the Methods section. Mealey and colleagues reported a higher frequency of transversions in younger breast cancer patients, but this difference was discrete (Tv: 53%; Ti: 47%) and was thus probably not statistically significant (no statistical analysis of this type was reported). Chen and colleagues reported a similar proportion of transitions and transversions in young TNBC patients (~50%/50%) [63,64].

The median number of coding variants was higher in TNBC compared to HGSOC from both age groups, and the median number of potentially pathogenic/possibly pathogenic variants per sample was higher in TNBC from elderly patients compared to the number of pathogenic/possibly pathogenic variants found in other groups. We could not find any other studies in young and elderly TNBC and HGSOC patients reporting similar results in terms of variant classification, although Mealey and colleagues and Azim and colleagues reported a higher number of coding variants in elderly breast cancer patients compared to young age groups [64,65]. As previously reported, somatic mutations other than in *TP53* are rare in HGSOC [36,66]. In all four groups, the majority of the samples presented at least one affected oncogene in association with at least one affected tumor suppressor gene (we also considered these as a combination: an oncogene or TSG with a dual-role gene or two or more dual-role genes). All affected elderly TNBC patients presented an association between OG and TSG.

The predominance of mutational signature 3, which is related to homologous repair defects, was previously reported in both cancer types. In breast cancer samples, mutational signature 3 was also reported to be the most frequent in both young and elderly patients [34,64] as well as in HGSOC in general [67,68]. In addition, a similar signature pattern (1, 3, and 13) was described by Chen and colleagues in young TNBC patients, similar to what has been reported in the present manuscript [63]. Despite that, we found a low frequency of somatic mutations in homologous repair-related genes in both breast and ovarian cancers, although this is not unexpected. In accordance with this, there are studies showing that most alterations in HR-related genes are germline, epigenetic, or CNVs [23,25,36,61].

Independent of age at diagnosis, signature 3 was also the most prevalent in HGSOC patients. In ovarian cancer, signature 3 is the most prevalent in both primary tumors as well as in distant metastatic sites [66,68]. The present results are also in accordance with Yang and colleagues (2018), who explored the somatic characteristics in HGSOC patients and reported that the majority of young patients were mainly represented by mutational signature 3 and that the most commonly altered gene was *TP53* [69].

Multiple studies suggest that the Ras pathway may influence the behavior of TNBC and HGSOC. For instance, Zhu and colleagues recently reported a correlation between Ras/ERK upregulation and chemoresistance in TNBC [70,71]. In the present series, we could not find any differences in the mutation rate of the Ras signaling pathway-related genes while comparing age and cancer type groups. Even though alterations in the RAS and PIK3CA pathways are mostly related to low-grade serous ovarian carcinomas [72–74], we were able to detect mutations in those genes in 14.3% and 15% of young and elderly HGSOC, respectively. *NF1*, which was altered in 5%, and 9.5% of young and elderly patients in the currently presented cohort, usually appears as one of most mutated genes in HGSOC analysis and occurs at higher frequencies of 16%-17% when age is not taken into account [62,75]. However, *NF1* may also be altered through structural variants [66], which were not assessed in this study.

Among all of the genes, *KMT2C* was the only one to be more frequently mutated in TNBC from elderly patients than it was in TNBC from young patients, in the present analysis. Similarly, Azim and colleagues reported a higher frequency of *KMT2C* variants in elderly breast cancer patients as well as higher frequencies of *KMT2D*, another member of the lysine methyltransferase family [65,76]. Alterations affecting the lysine-transferase family genes were also observed in high-grade ovarian cancer patients, (*KMT2A* = 2; *KMT2C* = 3), which is in accordance with a prior study (Zhang, 2021), where alterations in *KMT2C* were specifically identified in elderly patients, suggesting that these might be biomarkers that are related to the patient's age [62,75].

The *CSMD3* gene, a TSG in ovarian cancer (COSMIC; https://cancer.sanger.ac.uk/census; accessed on 1 May 2020), was the second most frequently pathogenically affected gene in HGSOC patients, besides *TP53,* in both age groups. *CSMD3* was also the gene most frequently affected by CNVs in a group of patients with no residual disease after surgery in a HGSOC series [77]. The gene is also in the top 10 mutated genes in the TNBC in the present analysis.

MUC16 mutations were among the most frequent, both in ovarian and breast cancer samples in the present study. *MUC16*, which encodes for Mucin 16, that contains the CA-125 epitope, a known ovarian cancer marker, was only identified among elderly but not in young HGSOC (8.3%) patients. A similar frequency (6%) was reported recently by Zhang, 2021, but this study did not consider age groups. For TNBC, *MUC16* was identified in 7% of the young patients and 13% of the elderly patients, with no difference between the age and cancer groups [62].

Tumor suppressor genes were the most frequently affected group of known cancer-causing genes in HGSOC, in accordance with what has been reported previously. Even with the high prevalence of *TP53* mutations, it was shown that these tumors present disruptions in another TSG that are either caused by somatic or, more commonly, by structural events [66]. In accordance with this, we observed the presence of mutations in other TSG in the 5 out of 7 and in the 5 out of 11 HGSOC tumors harboring wild-type *TP53* from young and elderly patients, respectively. Likewise, in accordance with that, a CRISPR- analysis in Ovarian Epithelium stem-cells showed that the disruption not of *TP53* alone (which is necessary in many cases, but not sufficient) but the combined disruption of tumor suppressor genes, especially *TP53*, *PTEN,* and *RB1,* are sufficient to induce transformation [78].

Breast cancer behavior has not always been associated with age at diagnosis. In fact, in a recently published study, Aine and colleagues investigated the genetic, transcriptional, and immune characteristics of young TNBC patients. The authors did not find a relevant connection between young patients and TNBC behavior, and they reported the possibility of the differential outcome in this cancer type being influenced by specific genetic signatures rather than by age itself. Although the previous results were strengthened with complementary analysis, the sample size was a limitation, as seen in the present study. The abovementioned statements refer to the need for larger cohort sizes to efficiently investigate the particularities of heterogeneous tumor types in young adult patients [79].

A limitation of our analysis is the absence of complementary molecular data, such as the copy number variation and methylation, which are known to play an important role in ovarian and breast cancer development and progression. Since multiple studies were used, the absence of raw data may result in different variant call pipelines. Moreover, the low number of samples in young high-grade serous ovarian cancer and elderly triple-negative breast cancer patients might limit the power of some of the statistical analysis.

The strength of this study is the individualized search of oncogenes and tumor suppressor genes alone or in combination in tumors from women who are 40 years of age apart and that are related to the dichotomized analysis in well-defined age groups to establish differences between young adults who are aged 18 to 40 years old and elderly women who are aged 70 years old or more, as well as the individualized pathogenicity analysis of each variant.

In summary, mutational signature 3 is more frequently detected in tumors from young patients from both HGSOC and TNBC compared to elderly patients with the same tumor type. In addition, TNBC from young and elderly patients presents a differential mutation signature pattern and TNBC from young women presents a higher mutation rate than it does in elderly women. We also noticed that the frequency of the alterations in cancer-causing genes that were directly involved in the cancer (oncogenes and TSG) was observed in most age groups in both cancer types, although some young adults in both cancer types had no identified gene representing a direct and strong role in their cancer.

The present analysis may contribute to the understanding of these types of tumors; however, further analysis integrating a more representative sample size and other "omics" portraits are necessary to validate the age relevance in different cancer types.

6. Conclusions

In conclusion, HGSOC and TNBC are very similar. The median number of mutated CGCs is three in young patients and four and six in elderly, and the main mutational signature is signature 3. At least ⅔ of the tumors presented at least one mutated oncogene in association with tumor suppressor genes.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cells10123586/s1, Figure S1. Mutational signatures. Figure S2. Number of non-synonymous coding variants per sample. Figure S3. Overlapped frequently affected genes classified according to its variant type. Figure S4. TP53 variants. Table S1. Summaries of variant types and frequencies and clinical data of the samples selected in the article. Table S2. Comparison between TI and TV across age groups and cancer types. Table S3. Comparison between the different cancer cohorts in the TCGA project and the cohorts in this study. Table S4. Mutational data collected from multiple studies and datasets for both tumor types that were analyzed. Table S5. Genes catalogued at the Cancer Gene Census and altered in the samples selected for this study. Table S6. Number of HGSOC and TNBC young and elderly patients with tumors presenting a cancer gene census. Table S7. Most frequently altered genes in all summed groups. Genes in bold are classified as dual roles in cancer (oncogene and TSG); genes highlighted in dark blue are classified as TSG. Table S8. HGSOC young patients: gene ontology (GO) aspect (molecular function, MF; biological process, BP) enriched in genes affected with potentially pathogenic or possibly pathogenic variants. Table S9. HGSOC elderly patients: gene ontology (GO) aspect (molecular function, MF; biological process, BP) enriched in genes affected with potentially pathogenic or possibly pathogenic variants. BP: Biological process; MF: molecular function. Table S10. TNBC young patients: gene ontology (GO) aspect (molecular function, MF; biological process, BP) enriched in genes affected with potentially pathogenic or possibly pathogenic variants. Table S11. TNBC elderly patients: Gene ontology (GO) aspect (molecular function, MF; biological process, BP) enriched in genes affected with potentially pathogenic or possibly pathogenic variants.

Author Contributions: P.A.d.M.P.S. and G.F.d.L.P. contributed equally to the paper. P.A.d.M.P.S., G.F.d.L.P. and M.A.A.K.F. conceived the study and wrote the original draft. P.A.d.M.P.S., G.F.d.L.P., M.L.H.K., S.M., R.A.R. and M.A.A.K.F. completed the data curation and investigation. P.A.d.M.P.S. conducted the formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Mutational and clinical data from two of the five studies (Cancer Genome Atlas Network, *Nature*, 2012; Nik-Zainal S. et. al., *Nature*, 2016) were downloaded from the International Cancer Genome Consortium—ICGC data portal (https://dcc.icgc.org/; realease date: January 2020). The other studies had their data downloaded from the supplementary data from their respective original articles [25,32–35].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Howlader, N.; Noone, A.M.; Krapcho, M.; Miller, D.; Brest, A.; Yu, M.; Ruhl, J.; Tatalovich, Z.; Mariotto, A.; Lewis, D.R.; et al. Cancer Statistics Review, 1975–2016—SEER Statistics. Available online: https://seer.cancer.gov/csr/1975_2016/ (accessed on 9 March 2020).
2. Cancello, G.; Maisonneuve, P.; Rotmensz, N.; Viale, G.; Mastropasqua, M.G.; Pruneri, G.; Veronesi, P.; Torrisi, R.; Montagna, E.; Luini, A.; et al. Prognosis and Adjuvant Treatment Effects in Selected Breast Cancer Subtypes of Very Young Women (<35 Years) with Operable Breast Cancer. *Ann. Oncol.* **2010**, *21*, 1974–1981. [CrossRef] [PubMed]
3. Liedtke, C.; Rody, A.; Gluz, O.; Baumann, K.; Beyer, D.; Kohls, E.-B.; Lausen, K.; Hanker, L.; Holtrich, U.; Becker, S.; et al. The Prognostic Impact of Age in Different Molecular Subtypes of Breast Cancer. *Breast Cancer Res. Treat.* **2015**, *152*, 667–673. [CrossRef] [PubMed]
4. Torre, L.A.; Trabert, B.; DeSantis, C.E.; Miller, K.D.; Samimi, G.; Runowicz, C.D.; Gaudet, M.M.; Jemal, A.; Siegel, R.L. Ovarian Cancer Statistics, 2018. *CA Cancer J. Clin.* **2018**, *68*, 284–296. [CrossRef] [PubMed]
5. Lockley, M.; Stoneham, S.J.; Olson, T.A. Ovarian Cancer in Adolescents and Young Adults. *Pediatr. Blood Cancer* **2019**, *66*, e27512. [CrossRef] [PubMed]
6. Gockley, A.; Melamed, A.; Bregar, A.J.; Clemmer, J.T.; Birrer, M.; Schorge, J.O.; Del Carmen, M.G.; Rauh-Hain, J.A. Outcomes of Women with High-Grade and Low-Grade Advanced-Stage Serous Epithelial Ovarian Cancer. *Obstet. Gynecol.* **2017**, *129*, 439–447. [CrossRef] [PubMed]
7. Elzakkers, J.C.J.; van der Aa, M.A.; van Altena, A.M.; de Hullu, J.A.; Harmsen, M.G. Further Insights into the Role of Tumour Characteristics in Survival of Young Women with Epithelial Ovarian Cancer. *Gynecol. Oncol.* **2019**, *155*, 213–219. [CrossRef] [PubMed]
8. Yoshikawa, N.; Kajiyama, H.; Mizuno, M.; Shibata, K.; Kawai, M.; Nagasaka, T.; Kikkawa, F. Clinicopathologic Features of Epithelial Ovarian Carcinoma in Younger vs. Older Patients: Analysis in Japanese Women. *J. Gynecol. Oncol.* **2014**, *25*, 118–123. [CrossRef]
9. Duska, L.R.; Chang, Y.C.; Flynn, C.E.; Chen, A.H.; Goodman, A.; Fuller, A.F.; Nikrui, N. Epithelial Ovarian Carcinoma in the Reproductive Age Group. *Cancer* **1999**, *85*, 2623–2629. [CrossRef]
10. Lalrinpuii, E.; Bhageerathy, P.S.; Sebastian, A.; Jeyaseelan, L.; Thomas, A.; Chandy, R.; Peedicayil, A. Ovarian Cancer in Young Women. *Indian J. Surg. Oncol.* **2017**, *8*, 540–547. [CrossRef]
11. Chan, J.K.; Urban, R.; Cheung, M.K.; Osann, K.; Shin, J.Y.; Husain, A.; Teng, N.N.; Kapp, D.S.; Berek, J.S.; Leiserowitz, G.S. Ovarian Cancer in Younger vs Older Women: A Population-Based Analysis. *Br. J. Cancer* **2006**, *95*, 1314–1320. [CrossRef]
12. Copson, E.R.; Maishman, T.C.; Tapper, W.J.; Cutress, R.I.; Greville-Heygate, S.; Altman, D.G.; Eccles, B.; Gerty, S.; Durcan, L.T.; Jones, L.; et al. Germline BRCA Mutation and Outcome in Young-Onset Breast Cancer (POSH): A Prospective Cohort Study. *Lancet Oncol.* **2018**, *19*, 169–180. [CrossRef]
13. Carraro, D.M.; Koike Folgueira, M.A.A.; Garcia Lisboa, B.C.; Ribeiro Olivieri, E.H.; Vitorino Krepischi, A.C.; de Carvalho, A.F.; de Carvalho Mota, L.D.; Puga, R.D.; do Socorro Maciel, M.; Michelli, R.A.D.; et al. Comprehensive Analysis of BRCA1, BRCA2 and TP53 Germline Mutation and Tumor Characterization: A Portrait of Early-Onset Breast Cancer in Brazil. *PLoS ONE* **2013**, *8*, e57581. [CrossRef] [PubMed]
14. Hahnen, E.; Hauke, J.; Engel, C.; Neidhardt, G.; Rhiem, K.; Schmutzler, R.K. Germline Mutations in Triple-Negative Breast Cancer. *Breast Care* **2017**, *12*, 15–19. [CrossRef] [PubMed]

15. Brianese, R.C.; de Mello Nakamura, K.D.; Ramalho, R.F.; de Figueiredo Barros, B.D.; e Ferreira, E.N.; da Cruz Formiga, M.N.; de Andrade, V.P.; de Lima, V.C.C.; Carraro, D.M. BRCA1 Deficiency Is a Recurrent Event in Early-Onset Triple-Negative Breast Cancer: A Comprehensive Analysis of Germline Mutations and Somatic Promoter Methylation. *Breast Cancer Res. Treat.* **2018**, *167*, 803–814. [CrossRef] [PubMed]
16. Gonzalez-Angulo, A.M.; Timms, K.M.; Liu, S.; Chen, H.; Litton, J.K.; Potter, J.; Lanchbury, J.S.; Stemke-Hale, K.; Hennessy, B.T.; Arun, B.K.; et al. Incidence and Outcome of BRCA Mutations in Unselected Patients with Triple Receptor-Negative Breast Cancer. *Clin. Cancer Res.* **2011**, *17*, 1082–1089. [CrossRef] [PubMed]
17. Li, M.; Zhang, J.; Ouyang, T.; Li, J.; Wang, T.; Fan, Z.; Fan, T.; Lin, B.; Xie, Y. Incidence of BRCA1 Somatic Mutations and Response to Neoadjuvant Chemotherapy in Chinese Women with Triple-Negative Breast Cancer. *Gene* **2016**, *584*, 26–30. [CrossRef]
18. Xie, Y.; Gou, Q.; Wang, Q.; Zhong, X.; Zheng, H. The Role of BRCA Status on Prognosis in Patients with Triple-Negative Breast Cancer. *Oncotarget* **2017**, *8*, 87151–87162. [CrossRef] [PubMed]
19. Sugino, K.; Tamura, R.; Nakaoka, H.; Yachida, N.; Yamaguchi, M.; Mori, Y.; Yamawaki, K.; Suda, K.; Ishiguro, T.; Adachi, S.; et al. Germline and Somatic Mutations of Homologous Recombination-Associated Genes in Japanese Ovarian Cancer Patients. *Sci. Rep.* **2019**, *9*, 17808. [CrossRef] [PubMed]
20. Krivokuca, A.; Boljevic, I.; Jovandic, S.; Magic, Z.; Mandic, A.; Tomasevic, Z.; Brankovic-Magic, M. Germline Mutations in Cancer Susceptibility Genes in High Grade Serous Ovarian Cancer in Serbia. *J. Hum. Genet.* **2019**, *64*, 281–290. [CrossRef] [PubMed]
21. Maistro, S.; Teixeira, N.; Encinas, G.; Katayama, M.L.H.; Niewiadonski, V.D.T.; Cabral, L.G.; Ribeiro, R.M.; Gaburo Junior, N.; de Gouvêa, A.C.R.C.; Carraro, D.M.; et al. Germline Mutations in BRCA1 and BRCA2 in Epithelial Ovarian Cancer Patients in Brazil. *BMC Cancer* **2016**, *16*, 934. [CrossRef]
22. Bernards, S.S.; Norquist, B.M.; Harrell, M.I.; Agnew, K.J.; Lee, M.K.; Walsh, T.; Swisher, E.M. Genetic Characterization of Early Onset Ovarian Carcinoma. *Gynecol. Oncol.* **2016**, *140*, 221–225. [CrossRef]
23. Eoh, K.J.; Kim, H.M.; Lee, J.-Y.; Kim, S.; Kim, S.W.; Kim, Y.T.; Nam, E.J. Mutation Landscape of Germline and Somatic BRCA1/2 in Patients with High-Grade Serous Ovarian Cancer. *BMC Cancer* **2020**, *20*, 204. [CrossRef]
24. Lord, C.J.; Ashworth, A. BRCAness Revisited. *Nat. Rev. Cancer* **2016**, *16*, 110–120. [CrossRef] [PubMed]
25. The Cancer Genome Atlas Network. Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* **2012**, *490*, 61–70. [CrossRef] [PubMed]
26. Bowtell, D.D.; Böhm, S.; Ahmed, A.A.; Aspuria, P.-J.; Bast, R.C.; Beral, V.; Berek, J.S.; Birrer, M.J.; Blagden, S.; Bookman, M.A.; et al. Rethinking Ovarian Cancer II: Reducing Mortality from High-Grade Serous Ovarian Cancer. *Nat. Rev. Cancer* **2015**, *15*, 668–679. [CrossRef] [PubMed]
27. Sæther, N.H.; Skuja, E.; Irmejs, A.; Maksimenko, J.; Miklasevics, E.; Purkalne, G.; Gardovskis, J. Platinum-Based Neoadjuvant Chemotherapy in BRCA1-Positive Breast Cancer: A Retrospective Cohort Analysis and Literature Review. *Hered. Cancer Clin. Pract.* **2018**, *16*, 9. [CrossRef] [PubMed]
28. Muggia, F.; Safra, T.; Dubeau, L. BRCA Genes: Lessons Learned from Experimental and Clinical Cancer. *Ann. Oncol.* **2011**, *22* (Suppl. 1), i7–i10. [CrossRef]
29. O'Shaughnessy, J.; Osborne, C.; Pippen, J.; Yoffe, M.; Patt, D.; Monaghan, G.; Rocha, C.; Ossovskaya, V.; Sherman, B.; Bradley, C. Efficacy of BSI-201, a Poly (ADP-Ribose) Polymerase-1 (PARP1) Inhibitor, in Combination with Gemcitabine/Carboplatin (G/C) in Patients with Metastatic Triple-Negative Breast Cancer (TNBC): Results of a Randomized Phase II Trial. *J. Clin. Oncol.* **2009**, *27*, 3. [CrossRef]
30. Ledermann, J.; Harter, P.; Gourley, C.; Friedlander, M.; Vergote, I.; Rustin, G.; Scott, C.; Meier, W.; Shapira-Frommer, R.; Safra, T.; et al. Olaparib Maintenance Therapy in Platinum-Sensitive Relapsed Ovarian Cancer. *N. Engl. J. Med.* **2012**, *366*, 1382–1392. [CrossRef] [PubMed]
31. Dent, R.A.; Lindeman, G.J.; Clemons, M.; Wildiers, H.; Chan, A.; McCarthy, N.J.; Singer, C.F.; Lowe, E.S.; Watkins, C.L.; Carmichael, J. Phase I Trial of the Oral PARP Inhibitor Olaparib in Combination with Paclitaxel for First- or Second-Line Treatment of Patients with Metastatic Triple-Negative Breast Cancer. *Breast Cancer Res.* **2013**, *15*, R88. [CrossRef] [PubMed]
32. Shah, S.P.; Roth, A.; Goya, R.; Oloumi, A.; Ha, G.; Zhao, Y.; Turashvili, G.; Ding, J.; Tse, K.; Haffari, G.; et al. The Clonal and Mutational Evolution Spectrum of Primary Triple-Negative Breast Cancers. *Nature* **2012**, *486*, 395–399. [CrossRef] [PubMed]
33. Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C.; et al. Landscape of Somatic Mutations in 560 Breast Cancer Whole-Genome Sequences. *Nature* **2016**, *534*, 47–54. [CrossRef] [PubMed]
34. Kan, Z.; Ding, Y.; Kim, J.; Jung, H.H.; Chung, W.; Lal, S.; Cho, S.; Fernandez-Banet, J.; Lee, S.K.; Kim, S.W.; et al. Multi-Omics Profiling of Younger Asian Breast Cancers Reveals Distinctive Molecular Signatures. *Nat. Commun.* **2018**, *9*, 1725. [CrossRef]
35. Banerji, S.; Cibulskis, K.; Rangel-Escareno, C.; Brown, K.K.; Carter, S.L.; Frederick, A.M.; Lawrence, M.S.; Sivachenko, A.Y.; Sougnez, C.; Zou, L.; et al. Sequence Analysis of Mutations and Translocations across Breast Cancer Subtypes. *Nature* **2012**, *486*, 405–409. [CrossRef]
36. Cancer Genome Atlas Research Network. Integrated Genomic Analyses of Ovarian Carcinoma. *Nature* **2011**, *474*, 609–615. [CrossRef]
37. Sukhai, M.A.; Craddock, K.J.; Thomas, M.; Hansen, A.R.; Zhang, T.; Siu, L.; Bedard, P.; Stockley, T.L.; Kamel-Reid, S. A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. *Genet. Med.* **2016**, *18*, 128–136. [CrossRef] [PubMed]

38. Li, M.M.; Datto, M.; Duncavage, E.J.; Kulkarni, S.; Lindeman, N.I.; Roy, S.; Tsimberidou, A.M.; Vnencak-Jones, C.L.; Wolff, D.J.; Younes, A.; et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the association for molecular pathology, american society of clinical oncology, and college of american pathologists. *J. Mol. Diagn.* **2017**, *19*, 4–23. [CrossRef]
39. Abbott, K.L.; Nyre, E.T.; Abrahante, J.; Ho, Y.-Y.; Isaksson Vogel, R.; Starr, T.K. The Candidate Cancer Gene Database: A Database of Cancer Driver Genes from Forward Genetic Screens in Mice. *Nucleic Acids Res.* **2015**, *43*, D844–D848. [CrossRef]
40. Chakravarty, D.; Gao, J.; Phillips, S.M.; Kundra, R.; Zhang, H.; Wang, J.; Rudolph, J.E.; Yaeger, R.; Soumerai, T.; Nissan, M.H.; et al. Oncokb: A precision oncology knowledge base. *JCO Precis. Oncol.* **2017**, *1*, 1–16. [CrossRef] [PubMed]
41. Lange, S.S.; Takata, K.; Wood, R.D. DNA Polymerases and Cancer. *Nat. Rev. Cancer* **2011**, *11*, 96–110. [CrossRef]
42. Wood, R.D.; Mitchell, M.; Sgouros, J.; Lindahl, T. Human DNA Repair Genes. *Science* **2001**, *291*, 1284–1289. [CrossRef] [PubMed]
43. Chae, Y.K.; Anker, J.F.; Carneiro, B.A.; Chandra, S.; Kaplan, J.; Kalyan, A.; Santa-Maria, C.A.; Platanias, L.C.; Giles, F.J. Genomic Landscape of DNA Repair Genes in Cancer. *Oncotarget* **2016**, *7*, 23312–23321. [CrossRef] [PubMed]
44. Mayakonda, A.; Lin, D.-C.; Assenov, Y.; Plass, C.; Koeffler, H.P. Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* **2018**, *28*, 1747–1756. [CrossRef]
45. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Aparicio, S.A.J.R.; Behjati, S.; Biankin, A.V.; Bignell, G.R.; Bolli, N.; Borg, A.; Børresen-Dale, A.-L.; et al. Signatures of Mutational Processes in Human Cancer. *Nature* **2013**, *500*, 415–421. [CrossRef] [PubMed]
46. Reimand, J.; Kull, M.; Peterson, H.; Hansen, J.; Vilo, J. g: Profiler—A Web-Based Toolset for Functional Profiling of Gene Lists from Large-Scale Experiments. *Nucleic Acids Res.* **2007**, *35*, W193–W200. [CrossRef]
47. Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **2011**, *6*, e21800. [CrossRef] [PubMed]
48. Shyr, C.; Tarailo-Graovac, M.; Gottlieb, M.; Lee, J.J.Y.; van Karnebeek, C.; Wasserman, W.W. FLAGS, Frequently Mutated Genes in Public Exomes. *BMC Med. Genom.* **2014**, *7*, 64. [CrossRef] [PubMed]
49. Stratton, M.R.; Campbell, P.J.; Futreal, P.A. The Cancer Genome. *Nature* **2009**, *458*, 719–724. [CrossRef] [PubMed]
50. Giannakouros, P.; Matte, I.; Rancourt, C.; Piché, A. Transformation of NIH3T3 Mouse Fibroblast Cells by MUC16 Mucin (CA125) Is Driven by Its Cytoplasmic Tail. *Int. J. Oncol.* **2015**, *46*, 91–98. [CrossRef]
51. Yamamoto, Y.; Abe, A.; Emi, N. Clarifying the Impact of Polycomb Complex Component Disruption in Human Cancers. *Mol. Cancer Res.* **2014**, *12*, 479–484. [CrossRef] [PubMed]
52. Chen, X.; Zhang, G.; Chen, B.; Wang, Y.; Guo, L.; Cao, L.; Ren, C.; Wen, L.; Liao, N. Association between Histone Lysine Methyltransferase KMT2C Mutation and Clinicopathological Factors in Breast Cancer. *Biomed. Pharmacother.* **2019**, *116*, 108997. [CrossRef] [PubMed]
53. Gala, K.; Li, Q.; Sinha, A.; Razavi, P.; Dorso, M.; Sanchez-Vega, F.; Chung, Y.R.; Hendrickson, R.; Hsieh, J.J.; Berger, M.; et al. KMT2C Mediates the Estrogen Dependence of Breast Cancer through Regulation of ERα Enhancer Function. *Oncogene* **2018**, *37*, 4692–4710. [CrossRef] [PubMed]
54. Bertucci, F.; Ng, C.K.Y.; Patsouris, A.; Droin, N.; Piscuoglio, S.; Carbuccia, N.; Soria, J.C.; Dien, A.T.; Adnani, Y.; Kamal, M.; et al. Genomic Characterization of Metastatic Breast Cancers. *Nature* **2019**, *569*, 560–564. [CrossRef] [PubMed]
55. Niyomnaitham, S.; Parinyanitikul, N.; Roothumnong, E.; Jinda, W.; Samarnthai, N.; Atikankul, T.; Suktitipat, B.; Thongnoppakhun, W.; Limwongse, C.; Pithukpakorn, M. Tumor Mutational Profile of Triple Negative Breast Cancer Patients in Thailand Revealed Distinctive Genetic Alteration in Chromatin Remodeling Gene. *PeerJ* **2019**, *7*, e6501. [CrossRef] [PubMed]
56. Sato, K.; Akimoto, K. Expression Levels of KMT2C and SLC20A1 Identified by Information-Theoretical Analysis Are Powerful Prognostic Biomarkers in Estrogen Receptor-Positive Breast Cancer. *Clin. Breast Cancer* **2017**, *17*, e135–e142. [CrossRef]
57. Rummel, S.K.; Lovejoy, L.; Shriver, C.D.; Ellsworth, R.E. Contribution of Germline Mutations in Cancer Predisposition Genes to Tumor Etiology in Young Women Diagnosed with Invasive Breast Cancer. *Breast Cancer Res. Treat.* **2017**, *164*, 593–601. [CrossRef] [PubMed]
58. Bakhuizen, J.J.; Hogervorst, F.B.; Velthuizen, M.E.; Ruijs, M.W.; van Engelen, K.; van Os, T.A.; Gille, J.J.; Collée, M.; van den Ouweland, A.M.; van Asperen, C.J.; et al. TP53 Germline Mutation Testing in Early-Onset Breast Cancer: Findings from a Nationwide Cohort. *Fam. Cancer* **2019**, *18*, 273–280. [CrossRef] [PubMed]
59. Olivier, M.; Langerød, A.; Carrieri, P.; Bergh, J.; Klaar, S.; Eyfjord, J.; Theillet, C.; Rodriguez, C.; Lidereau, R.; Bièche, I.; et al. The Clinical Value of Somatic TP53 Gene Mutations in 1794 Patients with Breast Cancer. *Clin. Cancer Res.* **2006**, *12*, 1157–1167. [CrossRef] [PubMed]
60. Soussi, T.; Wiman, K.G. TP53: An Oncogene in Disguise. *Cell Death Differ.* **2015**, *22*, 1239–1249. [CrossRef]
61. Li, W.; Shao, D.; Li, L.; Wu, M.; Ma, S.; Tan, X.; Zhong, S.; Guo, F.; Wang, Z.; Ye, M. Germline and Somatic Mutations of Multi-Gene Panel in Chinese Patients with Epithelial Ovarian Cancer: A Prospective Cohort Study. *J. Ovarian Res.* **2019**, *12*, 80. [CrossRef]
62. Zhang, Y.; Shi, X.; Zhang, J.; Chen, X.; Zhang, P.; Liu, A.; Zhu, T. A Comprehensive Analysis of Somatic Alterations in Chinese Ovarian Cancer Patients. *Sci. Rep.* **2021**, *11*, 387. [CrossRef]
63. Chen, H.; Chong, W.; Yang, X.; Zhang, Y.; Sang, S.; Li, X.; Lu, M. Age-Related Mutational Signature Negatively Associated with Immune Activity and Survival Outcome in Triple-Negative Breast Cancer. *Oncoimmunology* **2020**, *9*, 1788252. [CrossRef]
64. Mealey, N.E.; O'Sullivan, D.E.; Pader, J.; Ruan, Y.; Wang, E.; Quan, M.L.; Brenner, D.R. Mutational Landscape Differences between Young-Onset and Older-Onset Breast Cancer Patients. *BMC Cancer* **2020**, *20*, 212. [CrossRef] [PubMed]

65. Azim, H.A.; Nguyen, B.; Brohée, S.; Zoppoli, G.; Sotiriou, C. Genomic Aberrations in Young and Elderly Breast Cancer Patients. *BMC Med.* **2015**, *13*, 266. [CrossRef] [PubMed]
66. Patch, A.-M.; Christie, E.L.; Etemadmoghadam, D.; Garsed, D.W.; George, J.; Fereday, S.; Nones, K.; Cowin, P.; Alsop, K.; Bailey, P.J.; et al. Whole-Genome Characterization of Chemoresistant Ovarian Cancer. *Nature* **2015**, *521*, 489–494. [CrossRef]
67. Ashley, C.W.; Da Cruz Paula, A.; Kumar, R.; Mandelker, D.; Pei, X.; Riaz, N.; Reis-Filho, J.S.; Weigelt, B. Analysis of Mutational Signatures in Primary and Metastatic Endometrial Cancer Reveals Distinct Patterns of DNA Repair Defects and Shifts during Tumor Progression. *Gynecol. Oncol.* **2019**, *152*, 11–19. [CrossRef] [PubMed]
68. Masoodi, T.; Siraj, S.; Siraj, A.K.; Azam, S.; Qadri, Z.; Parvathareddy, S.K.; Tulbah, A.; Al-Dayel, F.; AlHusaini, H.; AlOmar, O.; et al. Genetic Heterogeneity and Evolutionary History of High-Grade Ovarian Carcinoma and Matched Distant Metastases. *Br. J. Cancer* **2020**, *122*, 1219–1230. [CrossRef]
69. Yang, S.Y.C.; Lheureux, S.; Karakasis, K.; Burnier, J.V.; Bruce, J.P.; Clouthier, D.L.; Danesh, A.; Quevedo, R.; Dowar, M.; Hanna, Y.; et al. Landscape of Genomic Alterations in High-Grade Serous Ovarian Cancer from Exceptional Long- and Short-Term Survivors. *Genome Med.* **2018**, *10*, 81. [CrossRef] [PubMed]
70. Gupta, G.K.; Collier, A.L.; Lee, D.; Hoefer, R.A.; Zheleva, V.; Siewertsz van Reesema, L.L.; Tang-Tan, A.M.; Guye, M.L.; Chang, D.Z.; Winston, J.S.; et al. Perspectives on Triple-Negative Breast Cancer: Current Treatment Strategies, Unmet Needs, and Potential Targets for Future Therapies. *Cancers (Basel)* **2020**, *12*, 2392. [CrossRef] [PubMed]
71. Zhu, S.; Xu, Y.; Wang, L.; Liao, S.; Wang, Y.; Shi, M.; Tu, Y.; Zhou, Y.; Wei, W. Ceramide Kinase Mediates Intrinsic Resistance and Inferior Response to Chemotherapy in Triple-Negative Breast Cancer by Upregulating Ras/ERK and PI3K/Akt Pathways. *Cancer Cell Int.* **2021**, *21*, 42. [CrossRef]
72. Tsang, Y.T.; Deavers, M.T.; Sun, C.C.; Kwan, S.-Y.; Kuo, E.; Malpica, A.; Mok, S.C.; Gershenson, D.M.; Wong, K.-K. KRAS (but Not BRAF) Mutations in Ovarian Serous Borderline Tumour Are Associated with Recurrent Low-Grade Serous Carcinoma. *J. Pathol.* **2013**, *231*, 449–456. [CrossRef] [PubMed]
73. Meinhold-Heerlein, I.; Fotopoulou, C.; Harter, P.; Kurzeder, C.; Mustea, A.; Wimberger, P.; Hauptmann, S.; Sehouli, J. The New WHO Classification of Ovarian, Fallopian Tube, and Primary Peritoneal Cancer and Its Clinical Implications. *Arch. Gynecol. Obstet.* **2016**, *293*, 695–700. [CrossRef] [PubMed]
74. Rojas, V.; Hirshfield, K.M.; Ganesan, S.; Rodriguez-Rodriguez, L. Molecular Characterization of Epithelial Ovarian Cancer: Implications for Diagnosis and Treatment. *Int. J. Mol. Sci.* **2016**, *17*. [CrossRef] [PubMed]
75. Zhong, F.; Zhu, T.; Pan, X.; Zhang, Y.; Yang, H.; Wang, X.; Hu, J.; Han, H.; Mei, L.; Chen, D.; et al. Comprehensive Genomic Profiling of High-Grade Serous Ovarian Carcinoma from Chinese Patients Identifies Co-Occurring Mutations in the Ras/Raf Pathway with TP53. *Cancer Med.* **2019**, *8*, 3928–3935. [CrossRef]
76. Ma, D.; Jiang, Y.-Z.; Xiao, Y.; Xie, M.-D.; Zhao, S.; Jin, X.; Xu, X.-E.; Shao, Z.-M. Integrated Molecular Profiling of Young and Elderly Patients with Triple-Negative Breast Cancer Indicates Different Biological Bases and Clinical Management Strategies. *Cancer* **2020**, *126*, 3209–3218. [CrossRef] [PubMed]
77. Lee, S.; Zhao, L.; Rojas, C.; Bateman, N.W.; Yao, H.; Lara, O.D.; Celestino, J.; Morgan, M.B.; Nguyen, T.V.; Conrads, K.A.; et al. Molecular Analysis of Clinically Defined Subsets of High-Grade Serous Ovarian Cancer. *Cell Rep.* **2020**, *31*, 107502. [CrossRef] [PubMed]
78. Yamulla, R.J.; Nalubola, S.; Flesken-Nikitin, A.; Nikitin, A.Y.; Schimenti, J.C. Most Commonly Mutated Genes in High-Grade Serous Ovarian Carcinoma Are Nonessential for Ovarian Surface Epithelial Stem Cell Transformation. *Cell Rep.* **2020**, *32*, 108086. [CrossRef]
79. Aine, M.; Boyaci, C.; Hartman, J.; Häkkinen, J.; Mitra, S.; Campos, A.B.; Nimeus, E.; Ehinger, A.; Vallon-Christersson, J.; Borg, Å.; et al. Molecular Analyses of Triple-Negative Breast Cancer in the Young and Elderly. *Breast Cancer Res.* **2021**, *23*, 20. [CrossRef]