

**A.C. CAMARGO CANCER CENTER
FUNDAÇÃO ANTÔNIO PRUDENTE**

LARISSA DIAS DE SOUZA

**ABORDAGENS GENÔMICAS PARA RECLASSIFICAR VARIANTES DE
SIGNIFICADO CLÍNICO INCERTO NÃO CODIFICANTES E DE EFEITO EM
SPLICE EM GENES DE PREDISPOSIÇÃO HEREDITÁRIA AO CÂNCER**

**Tese de Doutorado, apresentada à Fundação Antônio
Prudente como requisito para a obtenção do título de Doutor
em Ciências da Saúde.**

Área de Concentração: Oncologia

Orientador: Dra. Giovana Tardin Torrezan

Co-Orientador: Dr. Alexandre Defelicibus

São Paulo

2025

FICHA CATALOGRÁFICA

Souza, Larissa Dias de.

Abordagens genômicas para reclassificar variantes de significado clínico incerto não codificantes e de efeito em splice em genes de predisposição hereditária ao câncer. / Larissa Dias de Souza. São Paulo, 2025.

182f.

Tese de Doutorado - Fundação Antônio Prudente. Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Giovana Tardin Torrezan.

1. Câncer hereditário, 2. Splicing, 3. Análise de RNA

CDU 616

Nome: Larissa Dias de Souza

Título: Abordagens genômicas para reclassificar variantes de significado clínico incerto não codificantes e de efeito em *splice* em genes de predisposição hereditária ao câncer.

Aprovado em: ____/____/____

Banca examinadora

Orientador: Giovana Tardin Torrezan

Instituição: Fundação Antônio Prudente

Membro da banca: Jose Cláudio Casali da Rocha

Instituição: Fundação Antônio Prudente

Membro da banca: Tiago Góss dos Santos

Instituição: Fundação Antônio Prudente

Membro da banca: Maria Lucia Hirata Katayama

Instituição: Faculdade de Medicina da Universidade de São Paulo

Membro da banca: Guilherme Lopes Yamamoto

Instituição: Faculdade de Medicina da Universidade de São Paulo

SUPORTE À PESQUISA POR AGÊNCIA DE FOMENTO

Este trabalho recebeu apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, através de auxílio à Pesquisa de 01/07/2017 a 28/02/2025. n° do processo 465682/2014-6. E da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, através de auxílio à Pesquisa de 01/04/2023 a 31/03/2026. n° do processo 2022/05162-8.

“O que sabemos é uma gota, o que ignoramos é um oceano.”

(Isaac Newton)

Dedico esse trabalho aos meus pais (Sandra e José Francisco) e à minha irmã (Letícia) pelo suporte e apoio em todas as etapas da minha vida pessoal e profissional.

AGRADECIMENTOS

Inicialmente eu gostaria de agradecer a todos os pacientes oncológicos que aceitaram participar deste estudo com suas amostras biológicas. Desejo muita força e apoio nessa batalha que enfrentam todos os dias. Acredito que meu estudo vai proporcionar um laudo definitivo para alguns e um tratamento mais adequado para outros.

À minha orientadora Dra. Giovana Tardin Torrezan, que acreditou em mim para realização desse projeto, a qual me deu a oportunidade de entrar no mestrado e logo em seguida para o Doutorado Direto no AC Camargo. Sou grata pelo seu apoio em todos os momentos, desde pequenas apresentações de grupo até apresentações em congressos nacionais/internacionais. E principalmente pelo seu lado humano de ser. Saiba que você é uma inspiração para mim como cientista e mulher.

Ao meu coorientador Dr. Alexandre Defelicibus, que desde o início me apoiou nas análises de Bioinformática, e foi a chave central para a superar as barreiras da programação Bash/R/Python e nas ferramentas de *splice*. Sou grata pela sua assistência, paciência e motivação nessa jornada, sempre com boas ideias e conselhos. Te admiro como profissional e educador, sempre estimulando a autonomia e resolução de problemas.

À pesquisadora principal do grupo, Dra. Dirce Maria Carraro e a todos os integrantes do grupo de Genômica Clínica e Funcional (GCF) pela oportunidade de fazer parte do grupo de pesquisa, por estarem sempre dispostos a ajudar, sendo atenciosos e abertos para conversas, solucionar dúvidas e ao aprendizado conjunto.

Aos membros da banca Dra. Maria Lúcia e Dr. Thiago Góss, que me acompanham desde o começo do Doutorado e pela contribuição com sugestões de grande valor para o refinamento desse estudo. Com destaque a Dra. Maria Lúcia que me acompanha como supervisora desde a minha especialização no ICESP. Sou grata pelos estímulos e contribuições para meu desenvolvimento profissional desde meus primeiros passos na pós-graduação.

Ao médico oncologista Leandro Jonata de Carvalho Oliveira da Oncoclínicas que contribuiu nesse estudo encaminhando o caso dos pacientes aparentados RNAP-09 e RNAP-09.2. Obrigada pela confiança e colaboração entre instituições no compartilhamento dos dados e amostras biológicas.

Às Enfermeiras de Pesquisa Silvana e Laís, que disponibilizaram seu tempo ajustando seus horários para conversar com os pacientes elegíveis do meu estudo e coletar suas amostras biológicas. Juntamente às funcionárias do Biobanco Louise e Aninha, as quais extraíram o RNA, testando diferentes protocolos e maneiras para obter um bom rendimento das amostras preciosas, obrigada por todo cuidado e paciência.

À Deus, o qual esteve sempre a frente no comando, me dando forças extras, quando a jornada era cansativa e desgastante, onde encontrei conforto para seguir a diante.

Aos meus pais e minha irmã, Sandra, José e Letícia, os quais estiveram comigo desde sempre, me oferecendo todo tipo de suporte nos dias bons e ruins, sempre estimulando meu crescimento

e jornada pessoal. Assim como toda minha família que sempre apoiou meus passos e minhas decisões na área da pesquisa científica.

Aos meus amigos de infância próximos, Íris e Ricardo, que desde o ensino fundamental me ensinaram o que é a amizade verdadeira. Sou grata pelo apoio, carinho e empatia em cada etapa da minha vida desde a adolescência à vida adulta. E ao meu amigo Ayrton que está comigo desde a faculdade, numa relação de vivência, estudos, carinho e admiração pessoal, a qual compartilhou muitos momentos comigo, sendo um porto seguro para mim, por isso gostaria de agradecer todo seu apoio em todas as facetas da minha vida pessoal e profissional como bióloga e cientista.

A todos meus colegas de laboratório, Ana Beatriz, Mariana, Anthony e Lucas, que fizeram meus dias mais felizes, engraçados, divertidos e mais leves, além de ótimas companhias nos congressos onde aprendemos muitas coisas e compartilhamos momentos únicos. À Nathalia de Angelis que esteve comigo também sempre de coração aberto dando conselhos no final do expediente, com altas conversas sobre vida, família e estudos. Para quem chegou mais na metade para o final do meu doutorado, Thainá, Nayara, Darle e Jaque, que complementaram o grupo com muita alegria e carisma, obrigada pela amizade e trocas de ideias.

Por fim, às agências de fomento CAPES e FAPESP, pelo suporte financeiro tornando possível a realização desse projeto de pesquisa, as quais tornam também possível a realização de diversos estudos brasileiros em diferentes áreas de conhecimento.

RESUMO

Souza, LD. **Abordagens genômicas para reclassificar variantes de significado clínico incerto não codificantes e de efeito em *splice* em genes de predisposição hereditária ao câncer.** [Tese]. São Paulo: Fundação Antônio Prudente; 2025.

Cerca de 50 a 80% dos indivíduos em risco para síndromes de predisposição hereditária ao câncer apresenta variantes de relevância clínica desconhecida (VUS) em regiões codificantes e/ou não codificantes intragênicas. Avanços recentes no desenvolvimento de ferramentas computacionais e moleculares (como o sequenciamento alvo de RNA) tem possibilitado a priorização e reclassificação de VUS em regiões não codificantes ou previstas de alterarem o *splicing*. Nesse sentido, o objetivo desse estudo foi estabelecer e validar um conjunto de ferramentas genômicas, computacionais e experimentais que permitam reclassificar VUS não codificantes ou codificantes com possível efeito em *splice*, e validar um painel de captura de RNA de 27 genes de predisposição ao câncer. A seleção de pacientes foi feita a partir da reavaliação de arquivos VCFs prévios provenientes de testes genéticos de DNA que avaliaram de 26 a 126 genes de predisposição hereditária ao câncer realizados no Laboratório de Diagnóstico Genômico e em projetos de pesquisa do A.C Camargo Cancer Center utilizando o software Varseq. Variantes raras exônicas (sinônimas, *missenses*) e não codificantes (intrônicas até 20 pb do limite do éxon e nas regiões 5' e 3' UTR) foram filtradas e avaliadas por ferramentas de predição de *splice* (MaxEntScan, dbSNV e *SpliceAI*) e de impacto na 5' UTR (UTRannotator). O sequenciamento alvo de RNA (RNA-cap) foi realizado em pacientes com variantes previstas de alterar *splicing*, utilizando RNA de leucócitos ou tecido normal na plataforma Illumina NextSeq 500. Os padrões de *splicing* foram analisados pelos softwares rMATS, LeafCutterMD e FRASER2. Realizamos a revisão e avaliação *in silico* dos VCFs de 2.076 pacientes, nos quais identificamos 4.518 variantes raras. Destas, 182 variantes são previstas de afetar *splice* em pelo menos um preditor, 32 de impactar a função da região 5'UTR e 898 estão na região 3'UTR em todos os genes. Para a validação do painel RNA-cap de 27 genes, foram sequenciados 50 indivíduos (14 pacientes com 14 variantes patogênicas/provavelmente patogênicas [P/PP] distintas, 18 pacientes com 14 VUS distintas e 19 controles). Na avaliação das variantes P/PP, a ferramenta rMATS foi a que identificou mais eventos de *splice* aberrantes (em 11/14 variantes), depois a LeafCutterMD e FRASER2 (ambas em 7 variantes), as quais concordaram em apenas 4 variantes. Das 14 VUS, três variantes *ATM*:c.7273G>A, *BRCA2*:c.7618-10T>G e *MSH2*:del 5'UTR + del parcial do E1 foram

reclassificadas para P/PP, cinco para benigna/provavelmente benigna e seis permaneceram VUS, alcançando uma taxa de reclassificação de 57% das variantes. Duas VUS apresentaram alteração de *splice* parcial e potencialmente prejudicial (*NBN*:c.2071-4A>G e *BRCA1*:c.4096+1G>A), sendo necessário estudos complementares. As análises direcionadas ao transcrito e desbalanço alélico complementaram as análises do RNA-cap com mais evidências. Estabelecemos critérios e um pipeline automatizado para a seleção de variantes raras codificantes e não codificantes previstas de alterar *splice*, padronização para análise de *splicing* diferencial e aberrante, assim como processamento/refinamento dos dados. A aplicação do conjunto de ferramentas genômicas, computacionais e experimentais forneceu evidências no nível de RNA suficientes para aprimorar a interpretação clínica das VUS relacionadas ao câncer hereditário.

Palavras-chave: Câncer hereditário. *Splicing*. Análise de RNA. Reclassificação de VUS.

ABSTRACT

Souza, LD. **Genomic approaches for reclassifying noncoding and splice altering VUS in cancer predisposing genes.** [Tese]. São Paulo: Fundação Antônio Prudente; 2025.

Approximately 50 to 80% of individuals at risk for hereditary cancer predisposition syndromes presents variants of unknown clinical significance (VUS) in coding and/or non-coding intragenic regions. Recent advances in the development of computational and molecular tools (such as targeted RNA sequencing) have enabled the prioritization and reclassification of VUS in non-coding regions or predicted to alter splicing. In this sense, the objective of this study was to establish and validate a set of genomic, computational, and experimental tools that allow for the reclassification of non-coding or coding VUS with a possible effect on splicing, and to validate a capture panel of RNA from 27 cancer predisposition genes. Patient selection was made from the re-evaluation of previous VCF files from genetic DNA tests that assessed 26 to 126 hereditary cancer predisposition genes conducted at the Genomic Diagnosis Laboratory and in research projects at the A.C. Camargo Cancer Center using the Varseq software. Rare exonic variants (synonymous, missense) and non-coding variants (intronic up to 20 bp from the exon boundary and in the 5' and 3' UTR regions) were filtered and evaluated by splice prediction tools (MaxEntScan, dbSNV, and SpliceAI) and impact on the 5' UTR (UTRannotator). Targeted RNA sequencing (RNA-cap) was performed in patients with variants predicted to alter splicing, using RNA from leukocytes or normal tissue on the Illumina NextSeq 500 platform. Splicing patterns were analyzed using the rMATS, LeafCutterMD, and FRASER2 software. We performed in silico review and evaluation of the VCFs from 2,076 patients, in which we identified 4,518 rare variants. Of these, 182 variants are predicted to affect splicing by at least one predictor, 32 are expected to impact the function of the 5'UTR region, and 898 are located in the 3'UTR region across all genes. For the validation of the RNA-cap panel targeting 27 genes, 50 individuals were sequenced (14 patients with 14 pathogenic/likely pathogenic [P/LP] variants, 18 patients with 14 VUS, and 19 controls). In the evaluation of P/LP variants, the rMATS tool identified the most aberrant splicing events (in 11/14 variants), followed by LeafCutterMD and FRASER2 (both detecting events in 7 variants), with concordance in only 4 variants. Among the 14 VUS, three variants—*ATM*:c.7273G>A, *BRCA2*:c.7618-10T>G, and *MSH2*: 5'UTR deletion + partial exon 1 deletion—were reclassified as P/LP, five as benign/likely benign, and six remained VUS, resulting in a reclassification rate of 57%. Two VUS showed partial and potentially deleterious splicing alterations (*NBN*:c.2071-

4A>G and *BRCA1*:c.4096+1G>A), requiring further studies. Transcript-level and allelic imbalance analyses complemented the RNA-cap analysis by providing additional evidence. We established criteria and an automated pipeline for the selection of rare coding and non-coding variants predicted to alter splicing, standardization for differential and aberrant splicing analysis, as well as data processing and refinement. The application of a comprehensive set of genomic, computational, and experimental tools provided sufficient RNA-level evidence to improve the clinical interpretation of VUS related to hereditary cancer.

Key-words: Hereditary cancer. Splicing. RNA analyses. VUS reclassification.

LISTA DE ILUSTRAÇÕES

Figura 1 - Representação do sítio canônico de <i>splicing</i> ou região consenso de <i>splicing</i> e o grau de conservação das bases.....	25
Figura 2 - Representação dos elementos <i>cis</i> regulatórios de <i>splicing</i> presentes no mRNA.....	25
Figura 3 - Processo de remoção da região intrônica destacando a região de <i>Branchpoint</i>	26
Figura 4 - Intervalo de análise da ferramenta MaxEntScan.....	27
Figura 5 - Intervalo de análise da ferramenta dbSCSNV.....	27
Figura 6 - Intervalo de análise da ferramenta SpliceAI.....	28
Figura 7 - Representação das regiões que se concentram as análises de RNA pelas ferramentas de bioinformática.....	31
Figura 8 - Esquema simplificado para o cálculo do PSI entre diferentes isoformas de um mesmo gene.....	33
Figura 9 - Equação geral das métricas ψ_5 e ψ_3	34
Figura 10 - Equação geral das métricas θ_5 e θ_3	34
Figura 11 - Representação de uma uORF e oORF <i>in-frame</i> e <i>out-of-frame</i> na região 5'UTR.....	36
Figura 12 - Representação do trecho correspondente a transcrição do mRNA no DNA.....	39
Figura 13 - Processamento geral das variantes.....	45
Figura 14 - Delineamento do estudo.....	46
Figura 15 - Heatmap do GTEx Portal demonstrando a expressão dos 27 genes de predisposição ao câncer em sangue total contra os principais tecidos analisados.....	49
Figura 16 - Frequência geral das variantes por tipo.....	57
Figura 17 - Fluxograma representando os resultados obtidos em cada etapa do processamento das variantes.....	59
Figura 18 - Resumo do processamento das variantes preditas de impactar <i>splice</i> nos 27 genes.....	60
Figura 19 - Posição das variantes preditas de serem spliceogênicas.....	75
Figura 20 - Quantidade de variantes P/PP por preditores <i>in silico</i> de <i>splice</i>	76
Figura 21 - Quantidade de variantes VUS por preditores <i>in silico</i> de <i>splice</i>	77
Figura 22 - Descrição da quantidade de pacientes e variantes que entraram via pipeline, extra-pipeline e os que foram sequenciados.....	83

Figura 23 - Perfil das bibliotecas de transcriptoma total com amostras representativas de cada um dos lotes de experimentos realizados.....	90
Figura 24 - Perfil das três capturas de cDNA realizadas.....	91
Figura 25 - Heatmap da expressão dos 27 genes de estudo no sequenciamento de RNA-cap entre amostras de tecido normal adjacente do tumor e leucócitos.....	94
Figura 26 - Resumo das variantes P/PP que foram identificados eventos de <i>splice</i> significativos entre as três ferramentas.....	97
Figura 27 - Boxplots com a distribuição do PSI dos controles contra os valores do PSI dos pacientes com variantes P/PP.....	103
Figura 28 - Sashimi plots dos eventos SE significativos encontrados.....	104
Figura 29 - Imagem do gel do Bioanalyzer do paciente RNAP-02 para análise da variante <i>PALB2</i> dup E11.....	107
Figura 30 - Análise de desbalanço alélico do paciente RNA-02 a partir da variante <i>missense</i> em <i>PALB2</i> :c.3350G>A (chr 16:23607864)	108
Figura 31 - Imagem do gel do Bioanalyzer do paciente RNAP-101 para análise da variante <i>MLH1</i> :c.588+5G>C.....	109
Figura 32 - Imagem do gel do Bioanalyzer do paciente RNAP-03 para análise da variante <i>BRCA2</i> :c.156_157insAlu.....	111
Figura 33 - Visualização da retenção de 59 pb do íntron 4, devido a variante <i>BRCA1</i> :c.213-11T>G.....	112
Figura 34 - Visualização do gráfico de sashimi gerado através do IGV.....	112
Figura 35 - Imagem do gel do Bioanalyzer do paciente RNAP-118 para análise da variante <i>BRCA1</i> :c.213-11T>G.....	113
Figura 36 - Boxplot com a distribuição do PSI dos controles x PSI dos pacientes com VUS (sem variantes em <i>CHEK2</i>)	120
Figura 37 - Boxplot com a distribuição do PSI dos controles x PSI dos pacientes com VUS com variantes em <i>CHEK2</i> :c.320-5T>A.....	121
Figura 38 - Reclassificação final de variantes após o estudo combinando as técnicas de RNA-cap, análise direcionada ao transcrito e desbalanço alélico.....	124
Figura 39 - Heredograma da paciente RNAP-05.....	125
Figura 40 - Visualização da análise direcionada ao transcrito para a variante <i>BRCA2</i> : c.7618-10T>G (chr 13:32357732) no IGV.....	100
Figura 41 - Imagem do gel do Bioanalyzer do paciente RNAP-05 para análise da variante <i>BRCA2</i> :c.7618-10T>G.....	127
Figura 42 - Heredograma do paciente RNAP-06.....	128

Figura 43 - Sashimi plot mostrando a utilização dos sítios alternativos e canônicos de <i>splice</i> no paciente RNAP-06 e três controles.....	129
Figura 44 - Análise dos BAMs do paciente RNAP-06 oriundos da análise direcionada ao transcrito e do RNA-cap.....	130
Figura 45 - Imagem do gel do Bioanalyzer do paciente RNAP-06 para análise da variante <i>ATM</i> :c.7273G>A.....	131
Figura 46 - Heredograma do probando com destaque para o histórico familiar de câncer na família materna, com maior frequência de incidência de câncer colorretal e uterino.....	132
Figura 47 - Análise de desbalanço alélico no paciente RNAP-09 com a variante <i>MSH2</i> :c.1168C>T (chr 2:47.429.833 – hg38 ou chr2:47.656.972 – hg19)	133
Figura 48 - Visualização da deleção de aproximadamente 1876 pb da região promotora e parcial do E1 do gene <i>MSH2</i>	134
Figura 49 - Imagem do gel do Bioanalyzer do paciente RNAP-103 para análise da variante <i>BRCA1</i> : c.547+5G>A.....	135
Figura 50 - Imagem do gel do Bioanalyzer do paciente RNAP-104 para análise da variante <i>MUTYH</i> : c.1434+2C>T.....	134
Figura 51 - Imagem do Sashimi plot gerada no IGV das amostras RNAP-105 e RNAP-106, as quais possuem a variante <i>MUTYH</i> :c.934-2A>G.....	137
Figura 52 - Imagem do gel do Bioanalyzer das pacientes RNAP-105 e RNAP-106 para análise da variante <i>MUTYH</i> :c.934-2A>G.....	138
Figura 53 - Imagem do gel do Bioanalyzer das pacientes RNAP-08, RNAP-09 e RNAP-123 para análise da variante <i>CHEK2</i> :c.320-5T>A.....	140
Figura 54 - Imagem do gel do Bioanalyzer da paciente RNAP-114 para análise da variante <i>SMARCA4</i> :c.859+5G>A.....	141
Figura 55 - Imagem do gel do Bioanalyzer da paciente RNAP-03 para análise da variante <i>RAD51C</i> :c.571+4A>G.....	143
Figura 56 - Gráfico de sashimi apresentando evento de salto do éxon 13 presente no paciente RNAP-110 e em um controle externo.....	145
Figura 57 - Análise de desbalanço alélico a partir da variante <i>missense</i> em <i>MLH1</i> :c.790C>G.....	146
Figura 58 - Imagem do gel do Bioanalyzer da paciente RNAP-115 para análise da variante <i>MLH1</i> :c.790C>G.....	147
Figura 59 - Visão geral dos 4 SNPs marcadores encontrados no E10 da paciente RNAP-124.....	149

LISTA DE TABELAS

Tabela 1 - Filtros personalizados aplicados no VarSeq para os VCFs.	42
Tabela 2 - Resumo da análise de dados pelas três ferramentas de análise de <i>splice</i>	52
Tabela 3 - Variantes P/PP previstas de alterarem o <i>splicing</i> nos 27 genes de estudo.....	62
Tabela 4 - Variantes de significado incerto previstas de alterarem o <i>splicing</i> nos 27 genes de estudo.....	66
Tabela 5 - Previsão de impacto em variantes únicas na região 5'UTR em todos os genes.....	79
Tabela 6 - Variantes adicionadas extra-pipeline.....	84
Tabela 7 - Pacientes que foram sequenciados com suas respectivas variantes e informações pessoais.....	87
Tabela 8 - Informações e métricas gerais do sequenciamento.....	91
Tabela 9 - Resultados preditos <i>versus</i> obtidos a partir das variantes P/PP.....	98
Tabela 10 - Resumo dos resultados gerados no rMATS para pacientes com variantes que apresentaram eventos de <i>splice</i> anormais evidentes.....	102
Tabela 11 - Resultados preditos <i>versus</i> obtidos a partir das variantes VUS.....	115
Tabela 12 - Resumo dos resultados gerados pelas três ferramentas para pacientes com VUS que apresentaram eventos de <i>splice</i> significativos.....	119
Tabela 13 - Resumo da reclassificação de VUS com os critérios ACMG aplicados antes e depois da análise de RNA-cap.....	123

LISTA DE ABREVIATURAS

A3SS	Uso alternativo do sítio 3' de <i>splice</i>
A5SS	Uso alternativo do sítio 5' de <i>splice</i>
ACMG	Colégio Americano de Genética e Genômica Médica (do inglês, <i>American College of Medical Genetics and Genomics</i>)
B/PB	Benigna/ Provavelmente Benigna
BAI	Arquivo índice do BAM
BAM	Arquivo com leituras mapeadas ao genoma referência do Homo sapiens (do inglês, <i>Binary Alignment Map</i>)
cDNA	DNA complementar
CTRLs	Indivíduos controles – pessoas saudáveis sem doença
DNA-seq	Sequenciamento de DNA
E(n)	Éxon (número do éxon)
ESE	Intensificador de splicing exônico (do inglês, <i>Exonic splicing enhancer</i>)
ESS	Silenciador de splicing exônico (do inglês, <i>Exonic splicing silencer</i>)
GPC	Genes de predisposição ao câncer
GTF	Arquivo de anotação de genes e transcritos (do inglês, <i>Gene Transfer Format</i>)
HBOC	Câncer hereditário de mama e ovário (do inglês, <i>Hereditary Breast and Ovarian Cancer</i>)
IGV	Integrative Genomics Viewer
ISE	Intensificador de splicing intrônico (do inglês, <i>Intronic splicing enhancer</i>)
ISS	Silenciador de splicing intrônico (do inglês, <i>Intronic splicing silencer</i>)
LDG	Laboratório de Diagnóstico Genômico do A.C.Camargo Cancer Center
miRNA	Micro-RNA
mRNA	RNA mensageiro
MXE	Uso de éxons mutuamente exclusivos
NGS	Sequenciamento de nova geração
NMD	Nonsense mediated decay
P/PP	Patogênica / Provavelmente Patogênica
PSI	do inglês <i>Percent Spliced in</i>
PTC	Códon de parada prematuro (do inglês, <i>premature termination codon</i>)
RBP	Proteínas de ligação ao RNA (do inglês, <i>RNA binding proteins</i>)
RI	Retenção de íntron
RNA-cap	Sequenciamento alvo de RNA (captura de RNA)
RNA-seq	Sequenciamento de RNA
rRNA	RNA ribossômico
SE	Salto de éxon
SNV	Variante (s) de nucleotídeo único (do inglês, <i>Single nucleotide variant</i>)
SPHC	Síndromes de predisposição hereditária ao câncer
TCLE	Termo de Consentimento Livre e Esclarecido
uAUG	AUG a montante (do inglês, <i>upstream AUG</i>)
uORF	Quadro de leitura aberta a montante (do inglês, <i>Upstream Open reading frame</i>)

UTR	Região não traduzida (do inglês, <i>Untranslated Region</i>)
VAF	Frequência alélica da variante (do inglês, <i>Variant Allele Frequency</i>)
VCF	Arquivo (s) de chamada de variantes (do inglês, <i>Variant Call File</i>)
VEP	<i>Variant Effect Predictor</i>
VUS	Variante (s) de significado incerto (do inglês, <i>Variant of Uncertain Significance</i>)
WES	Sequenciamento de exoma completo (do inglês, <i>Whole Exome Sequencing</i>)
WGS	Sequenciamento de genoma completo (do inglês, <i>Whole Genome Sequencing</i>)

SUMÁRIO

1. INTRODUÇÃO.....	21
1.1. Síndromes de Predisposição Hereditária ao Câncer.....	21
1.2. Testes genéticos para SPHC	22
1.3. VUS encontradas nos testes genéticos	23
1.4. Ferramentas de predição <i>in silico</i> de <i>splice</i>	26
1.5. Análise pareada de RNA e DNA	28
1.6. Ferramentas de bioinformática para análise de <i>splicing</i>	30
1.6.1. Métricas utilizadas para detectar <i>splicing</i> alternativo	32
1.7. Regiões 5' e 3' UTR do mRNA.....	34
2. JUSTIFICATIVA	38
3. OBJETIVOS	40
3.1. Objetivo geral	40
3.2. Objetivos específicos.....	40
4. MATERIAIS E MÉTODOS	41
4.1. Casuística	41
4.2. Revisão de testes genéticos prévios para identificação das VUS de interesse....	41
4.3. Delineamento da validação do RNA-cap.....	45
4.3.1. Critérios de inclusão de pacientes	46
4.4. Sequenciamento alvo de RNA (RNA-cap).....	47
4.4.1. Coleta de material biológico e extração de RNA.....	47
4.4.2. Construção da biblioteca por captura de RNA.....	48
4.5. Análises de Bioinformática	50
4.5.1. Alinhamento de RNA e análise de expressão gênica.....	50
4.5.2. Ferramentas de análise de eventos de <i>splice</i>	50
4.5.3. Gráficos de Sashimi e Boxplot.....	53
4.6. Avaliações complementares do efeito de VUS em transcritos.....	53
4.6.1. Análise direcionada de transcritos.....	53
4.6.2. Análise de desbalanço alélico	55
4.7. Reclassificação de variantes de <i>splice</i> e em regiões não codificantes.....	55
4.8. Análises estatísticas	56
5. RESULTADOS	57
5.1. Levantamento de dados e filtragem de VCFs.....	57

5.2. Seleção de variantes previstas de impactar o <i>splice</i> ou com impacto na região 5'UTR e as presentes na região 3'UTR	58
5.2.1. Variantes previstas de impactar <i>splicing</i>	59
5.2.2. Variantes com impacto na região 5'UTR e as presentes na região 3'UTR.....	77
5.3. Seleção de pacientes em seus respectivos grupos para validação do RNA-cap .	82
5.3.1. Seleção de pacientes para o grupo controle positivo.....	85
5.3.2. Seleção de pacientes para o grupo com VUS possíveis modificadoras de <i>splice</i>	85
5.4. Sequenciamento alvo de RNA (RNA-cap).....	86
5.4.1. Construção da Biblioteca de cDNA por captura	90
5.4.2. Expressão gênica entre casos e controles	93
5.5. Análises de dados.....	94
5.5.1. Comparação das ferramentas de análise de <i>splice</i> em controles positivos	95
5.5.2. Resultados – grupo controle positivo	97
5.5.3. Resultados – grupo de pacientes com VUS preditas de impactar <i>splice</i>	113
5.6. Reclassificação das VUS	121
5.6.1. VUS reclassificadas para P/PP	124
5.6.2. VUS reclassificadas para B/PB	134
5.6.3. VUS que permaneceram como VUS.....	142
6. DISCUSSÃO	150
7. CONCLUSÃO.....	164
8. REFERÊNCIAS.....	166
APÊNDICE A - Painéis de genes utilizados para teste genético	176
APÊNDICE B - Tabela com a relação Gene-Doença	180

1. INTRODUÇÃO

1.1. Síndromes de Predisposição Hereditária ao Câncer

Na última década, devido aos avanços tecnológicos nas plataformas de sequenciamento foi possível melhorar a compreensão da suscetibilidade ao câncer, na qual as síndromes de predisposição hereditária ao câncer (SPHC) representam aproximadamente de 5-10% de todos os casos de câncer. As SPHC ocorrem devido à presença de variantes genéticas germinativas patogênicas raras que elevam o risco para o desenvolvimento de algumas neoplasias, dependendo do gene afetado. Hoje em dia são conhecidas aproximadamente mais de uma centena de SPHC, as quais apresentam, em sua maioria, herança autossômica dominante, como é o caso da síndrome de câncer de mama e ovário hereditários e da síndrome de Lynch (Murff, Spigel *et al.*, 2004; Nagy, Sweet *et al.*, 2004, Wang *et al.*, 2023).

Sabe-se que o câncer é essencialmente uma doença genética, por isso o entendimento da sua genômica é fundamental para a prática clínica. Entender as bases moleculares das SPHC possibilita um diagnóstico e prognóstico adequados para os pacientes, assim como, acesso ao aconselhamento genético familiar, a adoção de medidas de rastreamento/prevenção direcionadas e personalizadas. Recentemente, diversas alterações genéticas germinativas passaram a ter papel preditivo para a escolha de terapias alvo e eficiência do tratamento oncológico, como por exemplo, a indicação de tratamento com inibidores de PARP para indivíduos com alterações nos genes *BRCA1* e *BRCA2* e indicação de imunoterapias naqueles com alterações nos genes da síndrome de Lynch (Iyevleva, Imyanitov *et al.*, 2016).

Já foram identificados mais de centena de genes que conferem riscos moderados a altos para o desenvolvimento de câncer, como por exemplo, *BRCA1*, *BRCA2*, *TP53*, *MLH1*, *MSH2*, *ATM*, *BRIP1*, *CHEK2* e *PALB2*, conhecidos como genes de predisposição ao câncer (GPC) (Rahman, 2014). O estudo destes genes fornece informações importantes para o processo de formação tumoral, podendo contribuir na busca de terapias mais efetivas. Grande parte dos GPCs atuam regulando o funcionamento do ciclo celular de duas principais formas, regulando negativamente a progressão de ciclo celular ou em vias de reparo de DNA, função desempenhada pelos genes supressores de tumor, ou estimulando a entrada do ciclo celular, função desempenhada pelos proto-oncogenes. Alterações genéticas que modificam ou anulam a função destes genes podem promover a oncogênese, devido à desregulação dos mecanismos

de controle do ciclo celular (Ward, 2002). A maioria dos GPCs já descritos apresentam funções de genes supressores de tumor.

O padrão de herança de predisposição ao câncer é diversificado, sendo que a maioria dos GPCs apresentam herança autossômica dominante, o que significa que, a presença de apenas uma única cópia alterada do gene já é suficiente para o aumento de risco de câncer. Porém, em outros casos, o mesmo gene ou grupos de genes relacionados à mesma síndrome podem apresentar padrão dominante e padrão recessivo, de forma que a condição recessiva se comporta de maneira mais agressiva, como por exemplo, o caso da síndrome de deficiência constitucional de reparo de mal-pareamento (CMMRD) e da síndrome de Lynch (Rahman, 2014). Considera-se a penetrância dos GPCs altamente heterogênea, na qual existem genes associados a alelos raros de alta penetrância e genes com alelos mais comuns de baixa penetrância ou penetrância moderada, na qual varia de forma ampla os riscos de seus portadores virem a desenvolver neoplasias. Existe ainda uma expressividade muito ampla da manifestação dos fenótipos associados a alterações nos GPCs, sendo que uma única variante patogênica pode causar fenótipos heterogêneos, enquanto o mesmo fenótipo pode acontecer pela identificação de diferentes variantes patogênicas. Ou então uma mesma variante pode ocasionar um fenótipo agressivo em um indivíduo, porém um fenótipo mais brando em outro. Isso revela a provável existência de variantes genéticas modificadoras de risco somado também a influência de fatores ambientais, reforçando a necessidade de desenvolvimento de ferramentas preditivas validadas por meio de aprendizado de máquina, por exemplo (Garutti *et al.*, 2023).

1.2. Testes genéticos para SPHC

Em razão ao avanço de novas tecnologias, o diagnóstico genético tornou-se mais efetivo nos últimos anos, como o sequenciamento de nova geração (NGS), o qual proporciona um entendimento melhor da base genética de doenças mendelianas, como as SPHC. O sequenciamento por NGS tornou-se a metodologia padrão para o diagnóstico de doenças genéticas, pois permite sequenciamento de muitos pacientes em um mesmo experimento e análise de várias regiões do genoma em uma única reação, sendo considerada uma metodologia rápida e custo-efetiva (Jamuar, Ene-Choo Tan, 2015). Logo, a análise por NGS com painéis multigenes se tornou o método mais adequado para investigação genética de pacientes com suspeita de SPHC, uma vez que existe uma ampla sobreposição de fenótipos entre as SPHC e,

muitas vezes, pode acontecer da história clínica pessoal e familiar do paciente ser compatível com mais de uma síndrome.

Devido ao amplo desenvolvimento da tecnologia de NGS na década passada, e um aumento de testes genéticos como genotipagem, sequenciamento de genes únicos, painéis de genes, exomas, genomas, transcriptomas e ensaios epigenéticos para doenças genéticas, foi necessário a criação de normas e diretrizes para a interpretação da complexidade das variantes genéticas. Criada em 2015 pelo *American College of Medical Genetics and Genomics* (ACMG), juntamente com *Association for Molecular Pathology* (AMP) e o *College of American Pathologists* (CAP), a diretriz mais utilizada atualmente para a classificação de variantes relacionadas a doenças mendelianas monogênicas de alta penetrância, recomenda o uso de uma terminologia padrão para classificação de variantes em 5 classes: 'Patogênica', 'Provavelmente patogênica', 'Significado incerto', 'Provavelmente benigna' e 'Benigna'. A classificação nessas classes tem como base critérios que usam evidências disponíveis para cada variante, como por exemplo, frequência da variante em bancos de dados populacionais, dados computacionais de preditores *in silico* de patogenicidade, dados funcionais obtidos por experimentos *in vitro* e *in vivo*, estudos de casos e controles, dados de segregação, etc. (Richards *et al.*, 2015).

Desde a publicação dessa diretriz em 2015, grupos de experts de consórcios internacionais, como por exemplo, o ClinGen (<https://www.clinicalgenome.org/>) e Canvar-UK (<https://canvaruk.org/>) foram elaborando ao longo do tempo recomendações específicas para aplicação mais detalhada dos critérios da ACMG, para genes específicos. E em 2023 foram publicadas recomendações adaptadas para classificação de variantes de *splicing* pelo grupo de especialistas em *splice* da ClinGen (Walker *et al.*, 2023), a qual possui novas orientações para ponderar pontuações para variantes de *splice*, como por exemplo: utilização de uma árvore de decisão para o critério PVS1 (do inglês, *Pathogenic Very Strong*) para os nucleotídeos dentro e fora da região doadora e aceptora de *splice*; Calibrar ferramentas de predição *in silico* de *splice*; Avaliar a qualidade do ensaio de RNA entre diferentes estudos para conseguir capturar mais evidências relacionadas ao *splice*; entre outras.

1.3. VUS encontradas nos testes genéticos

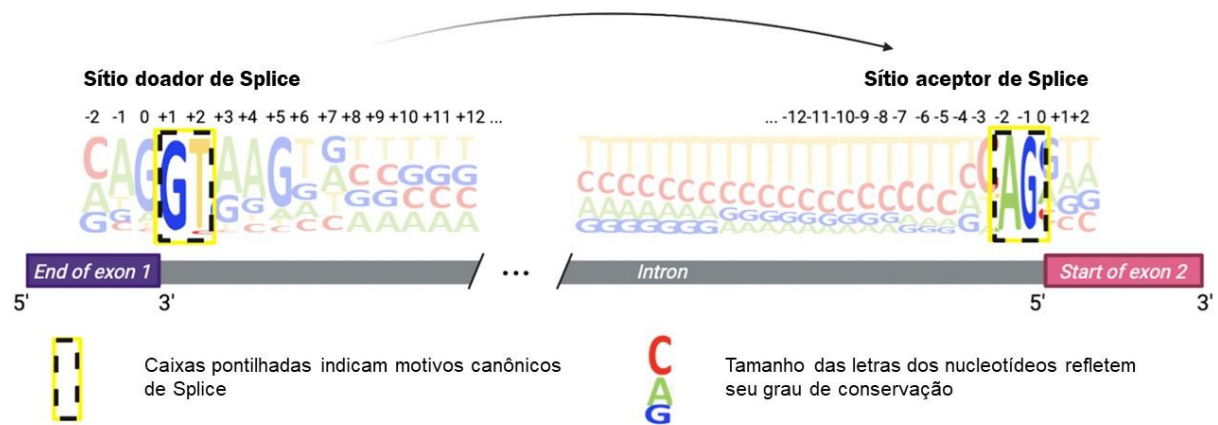
O uso de painéis multigênicos apesar de altamente vantajosos pelo seu custo-benefício e maior resolatividade, também leva a descobertas complexas, como a identificação de variantes patogênicas sem clara associação ao fenótipo e de muitas variantes de significado

clínico incerto (VUS). VUS é uma variante genética que não apresenta implicações claras para a função do gene, devido à falta de evidências científicas suficientes para determinar o seu significado biológico e clínico. Em geral, as VUS não devem orientar decisões clínicas, por isso são necessárias pesquisas adicionais para orientar na classificação dessas variantes (Clift *et al.*, 2020).

As VUS são principalmente variantes raras de substituições de nucleotídeo em regiões codificantes do tipo *missenses*, sinônimas ou inserções/deleções *in frame*, as quais podem estar em resíduos menos conservados ou em domínios funcionalmente menos relevantes para a proteína. Estas também podem estar em regiões envolvidas no *splicing* (variantes intrônicas ou exônicas) ou estar em regiões não codificantes como nas regiões 5' e 3' UTR. Dentre todas as VUS encontradas em sequenciamento NGS, as de *splice* são o segundo tipo de variante mais encontrada nos pacientes, ficando atrás apenas das variantes VUS *missenses* (Truty *et al.*, 2020; Horton *et al.*, 2022). As VUS de *splice* geralmente são aquelas que se encontram fora do sítio doador ou aceptor de *splice* (motivos canônicos de *splice*), em regiões menos conservadas, por isso é mais difícil de prever seu real impacto no *splice*, conforme demonstrado na Figura 1 (Truty *et al.*, 2020; Walker *et al.*, 2023). Diferentemente dessas variantes, as que se encontram nos dinucleotídeos conservados GT-AG, possuem uma probabilidade muito alta de impactar o *splice*, e por isso são, na maioria das vezes, classificadas como P/PP (Walker *et al.*, 2023).

Existem ainda múltiplas regiões regulatórias de *splicing* curtas (*motifs*) em *cis* de 6 a 8 pb não conservadas, no éxon chamadas de Intensificadores de *splicing* exônicos (*Exonic Splicing Enhancers* - ESE) e Silenciadores de *splicing* exônico (*Exonic Splicing Silencers* - ESS), e no íntron são chamadas de Intensificadores de *splicing* intrônico (*Intronic Splicing Enhancer* - ISE) e Silenciadores de *splicing* intrônico (*Intronic Splicing Silencer* - ISS), as quais aumentam ou diminuem a inclusão do éxon/íntron no mRNA maduro através da ligação de fatores de *splicing* proteicos (Figura 2) (Ramalho, 2012, Wang *et al.*, 2023). Variantes que alteram esses sítios são muito mais desafiadoras de serem interpretadas, uma vez que são ignoradas pelos *pipelines* utilizados em laboratórios de diagnóstico clínico, aparecendo como variantes *missenses*, sinônimas ou variantes de íntron profundo. Alterações nessas regiões são a base para ocorrência de muitas doenças genéticas raras (Wang *et al.*, 2023).

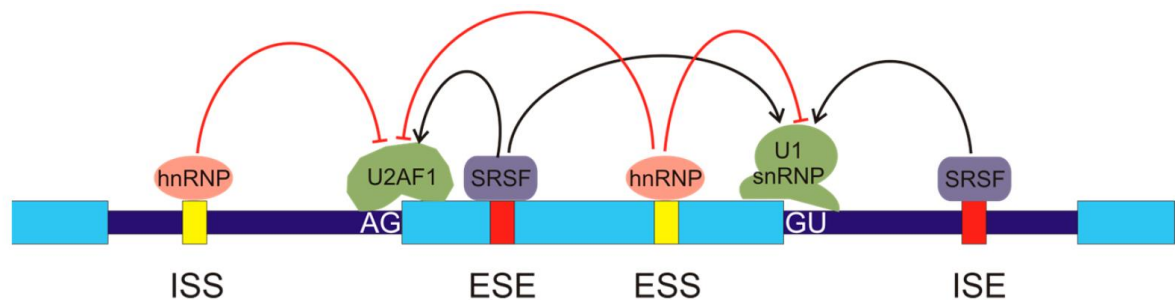
Figura 1 - Representação do sítio canônico de *splicing* ou região consenso de *splicing* e o grau de conservação das bases



Fonte: adaptado de Postel *et al.* (2022).

Legenda: Demonstra-se que o sítio doador e aceitador são regiões com alto grau de conservação dos dinucleotídeos GT-AG (variantes $\pm 1,2$ *splice*), porém as regiões mais distantes do limite éxon/intron são menos conservadas e por isso variantes que se encontram nessas regiões são muitas vezes classificadas como VUS.

Figura 2 - Representação dos elementos *cis* regulatórios de *splicing* presentes no mRNA



Fonte: Gimeno-Valiente *et al.* (2024).

Legenda: Regiões regulatórias silenciadoras de *splice* no éxon (*exonic splicing silencers* (ESSs)) e no íntron (*intronic splicing silencers* (ISSs)) estão representados pelas caixas amarelas, e regiões intensificadoras de *splice* no éxon (*exonic splicing enhancers* (ESEs)), e no íntron (*intronic splicing enhancers* (ISEs)) estão sendo representados pelas caixas vermelhas. As caixas azuis claro estão representando os éxons e as caixas roxas escuras os íntrons. Diferentes fatores regulatórios se ligam nessas sequências, os quais pertencem as duas famílias principais: as ribonucleoproteínas nucleares heterogêneas (hnRNPs) que se ligam nas regiões silenciadoras (em amarelo), e fatores ricos em serina/argina (SRSFs) que se ligam nas regiões intensificadoras (em vermelho). A ligação desses fatores pode ativar (arcos pretos) ou inibir (arcos vermelhos) o recrutamento dos componentes do spliceossomo para as sequências AG ou GU que estão nas bordas dos éxons, dessa forma ocorre a definição do éxon.

Ponto de ramificação (ou *branchpoint* em inglês) é outra região regulatória, que geralmente é uma adenina que se localiza entre 20 pb a 40 pb a montante de um sítio aceitador canônico ou críptico de *splice*, sendo um dos primeiros elementos reconhecidos pelo spliceossomo (Zhang *et al.*, 2023). Variantes que interrompem essas regiões podem levar a

pseudoexonização, retenção total/parcial de um íntron e salto de éxon resultando em um transcrito com perda de função (Barbosa *et al.*, 2023). A taxa de detecção de variantes patogênicas em regiões de *branchpoint* é baixa (48 variantes foram identificadas nos últimos 30 anos), isso ocorre pois não existe uma abordagem computacional eficiente para detectar e anotar tais variantes em sequenciamentos de NGS (Zhang *et al.*, 2022). Na Figura 3 está o esquema representando o papel do *branchpoint* no *splicing*.

Figura 3 - Representação da região de *Branchpoint* e processamento de retirada do íntron



Fonte: Adaptado de Zhang *et al.* (2022).

Legenda: Processo de remoção da região intrônica destacando a região de *Branchpoint*, representada pela letra 'A', na qual o spliceossomo utiliza como base para formação do laço circular, ocorrendo a remoção do íntron e união dos éxon adjacentes.

1.4. Ferramentas de predição *in silico* de *splice*

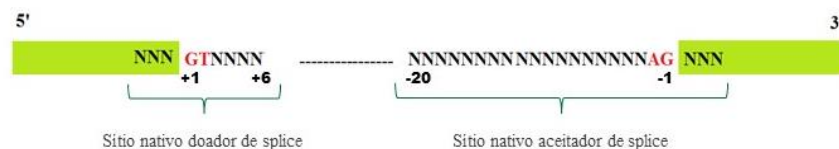
Algumas variantes no DNA podem afetar o processo normal de *splicing* resultando no salto de sequências codificantes (denominado *exon skipping*) ou então a inserção de sequências não codificantes no RNA mensageiro (mRNA) (denominado íntron *retention*), o que pode levar a perda de função do alelo. Identificar essas variantes é de extrema importância para a realização de análises posteriores complementares e para reportar no laudo clínico final, especialmente se o paciente apresentar teste genético negativo para variantes P/PP e identificação de variantes VUS preditas de causar erro em *splice* em genes relacionados ao seu fenótipo.

Para triar tais variantes das demais encontradas no sequenciamento de DNA são utilizadas ferramentas de predição *in silico* de *splice*, como por exemplo, MaxEntScan (Yeo *et al.*, 2004), dbSCNV (Jian *et al.*, 2014) e *SpliceAI* (Jaganathan *et al.*, 2019), as quais são as mais utilizadas atualmente no cenário clínico e que serão descritas a seguir. Essas três ferramentas estão disponíveis de forma gratuita no *Variant Effect Predictor* (VEP) do Ensembl (<https://www.ensembl.org/Tools/VEP>). A maior parte dos preditores foca suas análises em regiões canônicas, em que existe algum tipo de conservação de bases (conforme descrito anteriormente no tópico 1.3). Existem poucos preditores para análise das regiões regulatórias do mRNA, os quais ainda não são tão robustos, uma vez que essas regiões são altamente degeneradas podendo estar sobrepostas. Ainda, a previsão de *splice* pode ser mais desafiadora

porque as atividades que ocorrem nesses elementos regulatórios podem se diferir dependendo de sua localização ao longo do mRNA prematuro e estão relacionadas a um determinado contexto celular onde os níveis de expressão das suas RBPs parceiras também podem ser diferentes (Wang *et al.*, 2023).

O preditor MaxEntScan se baseia no princípio de entropia máxima utilizando sequências curtas (medida de desordem ou aleatoriedade - MEP), o dbscSNV utiliza dois métodos de montagem em conjunto para aperfeiçoar a predição, o Adaboost (Ada Score) e Random Forest (RF Score), ambos focam no possível impacto dentro das regiões consenso de *splice* levando em conta a conservação de bases e elementos *cis* de *splice*. Nas Figuras 4 e 5 estão detalhadas as regiões de análise de cada ferramenta, assim como o escore considerado ponto de corte para a variante ser considerada com potencial risco de causar alteração no *splice*.

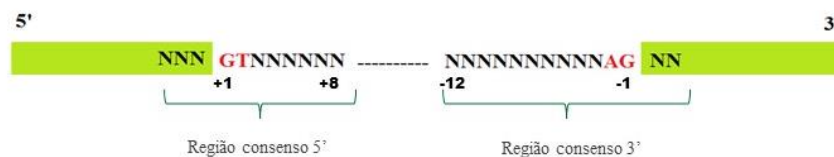
Figura 4 - Intervalo de análise da ferramenta MaxEntScan



Fonte: autoria própria.

Legenda: Esse preditor foca nas regiões do sítio nativo doador de *splice* (compreendendo os três últimos nt do éxon até o sexto nt do íntron), e do sítio aceptor de *splice* (compreendendo os últimos 20 nt do íntron até os três primeiros nt do éxon). O ponto de corte estabelecido para considerar uma variante com potencial risco de causar alteração no *splice* é a diferença $\geq 3,0$, do escores de entropia máxima da base referência e da base alternativa.

Figura 5 - Intervalo de análise da ferramenta dbscSNV



Fonte: autoria própria.

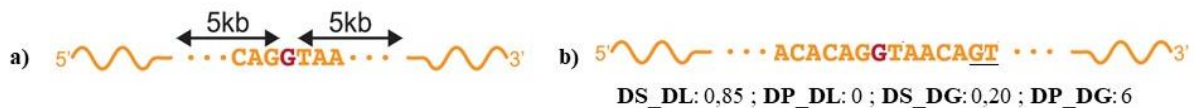
Legenda: Esse preditor foca nas regiões consenso 5' (compreendendo os três últimos nt do éxon até o oitavo nt do íntron), e na região consenso 3' (compreendendo os últimos 12 nt do íntron até os dois primeiros nt do éxon). O ponto de corte estabelecido para considerar uma variante com potencial risco de causar alteração no *splice* é ter o escore Ada ou RF $\geq 0,60$.

Já o *SpliceAI* se baseia em aprendizado de máquina, na qual uma rede neural profunda consegue prever alterações de *splice* a partir de uma sequência de pré-mRNA avaliando uma janela de 5 mil nucleotídeos antes e depois do nucleotídeo afetado, ou seja, não foca sua análise apenas nas regiões consenso de *splice*, como as duas ferramentas anteriores. E como vantagem ainda consegue prever a perda/ganho do sítio doador ou aceptor de *splice*, assim como a posição

de formação de sítios crípticos de *splice* (sítios de *splicing* que estão próximos dos sítios canônicos, os quais podem ser utilizados de forma alternativa após alterações nesses sítios). A Figura 6 exemplifica o intervalo de análise dessa ferramenta e a explicação dos escores gerados como resultados. Para mais informações o estudo de Garcia e colaboradores (2022) explica de forma simplificada como funciona cada uma dessas ferramentas.

Existe ainda uma ferramenta paga que compila vários preditores *in silico* de *splicing*, como: SpliceSiteFinder-like, MaxEntScan, GeneSplicer, NNSPLICE, ESEFinder, RESCUE-ESSE e EX-SKIP. Os três últimos fazem detecção de sítios de ligação de intensificadores de *splicing* exônico (ESE), que se baseiam em identificar ESEs e prever se as mutações exônicas podem inativar esses elementos podendo vir a causar salto de éxons (Cartegni *et al.*, 2003). Esse agregador de ferramentas se chama Alamut® Visual e é um módulo de predição de *splicing* dentro do *software* ‘Sophia Genetics’ que faz chamada e anotação de variantes geralmente utilizado em laboratórios de diagnóstico clínico.

Figura 6 - Intervalo de análise da ferramenta SpliceAI



Fonte: Jaganathan *et al.* (2019).

Legenda: a) Foco da análise nos 5 mil nt antes e depois da variante em questão, gerando como resultados escores de pontuação (*delta score* - DS) e escores de posição (*delta position* - DP), para a perda ou ganho dos sítios doadores e aceptores: perda do sítio doador (*delta score donor loss* - DS_DL) e a posição da perda (*delta position donor loss* - DP_DL); ganho de um sítio doador (*delta score donor gain* - DS_DG) e a posição do ganho (*delta position donor gain* - DP_DG); perda do sítio aceptor (*delta score acceptor loss* - DS_AL) e a posição da perda (*delta position acceptor loss* - DP_AL); ganho de um sítio aceptor (*delta score acceptor gain* - DS_AG) e a posição do ganho (*delta position acceptor gain* - DP_AG). A posição de perda/ganho é contabilizada a partir da posição da variante em avaliação (para valores positivos a posição estará à direita e para valores negativos à esquerda da variante). O ponto de corte estabelecido para considerar uma variante com risco moderado de causar alteração no *splice* é $\geq 0,20$ dos escores DS_DL ou DS_DG ou DS_AL ou DS_AG. b) Representação dos resultados gerados pela ferramenta para uma variante hipotética na posição chr 2:179415980, demonstrando que houve a predição de 85% de chance de perda do sítio doador de *splice* na posição da variante em análise e 20% de chance de ganho de um sítio doador seis nucleotídeos depois da variante, ou seja, na posição 2:179415986.

1.5. Análise pareada de RNA e DNA

Geralmente, os testes genéticos convencionais de DNA capturam com boa qualidade de cobertura as regiões codificantes (éxons) e curtos trechos dos íntrons flanqueadores adjacentes (± 20 pb do limite do éxon) considerada faixa reportável de variantes, característica considerada limitante, visto que variantes com potencial risco de serem patogênicas podem ficar fora dessa área de cobertura, e consequentemente podem não ser identificadas. No entanto, a interpretação

clínica dessas variantes ainda é muito difícil, assim como entender sua relevância na contribuição do impacto para a doença, uma vez que essas variantes podem não impactar diretamente a proteína, mas podem ter implicação em mecanismos de regulação pré ou pós transcricionais, assim como na própria regulação de *splicing* (Whiffin *et al.*, 2020).

Adicionar o sequenciamento de RNA (RNA-seq) ao sequenciamento de DNA (DNA-seq) pode capturar eventos de *splicing* incorretos causados por variantes intrônicas ou variantes exônicas sinônimas, e tornar possível a identificação de variantes patogênicas que poderiam ser ignoradas ou classificadas como VUS. Eventos de *splice* podem levar saltos de éxon e/ou inserções de íntron totais e/ou parciais de forma a perturbar a janela de leitura da tradução, sendo considerados eventos *out-of-frame*. Esse tipo de evento pode introduzir um códon de parada prematuro (PTC - do inglês, *premature termination codon*), que dependendo da sua localização no transcrito pode ativar a via de *nonsense-mediated-decay* (NMD), levando a sua degradação e eliminação da etapa de tradução, não ocorrendo a formação de uma proteína truncada. Também podem ter eventos de *splice* que não perturbam a janela de leitura da tradução, sendo considerados eventos *in-frame*. Nestes eventos é necessário observar se a alteração afeta um domínio clinicamente relevante para a doença, e quantos aminoácidos (aa) são perdidos em relação ao comprimento total da proteína, sendo considerado potencialmente prejudicial se essa perda for maior que 10%.

O RNA-seq mostrou ser efetivo em melhorar o rendimento diagnóstico e resolução de testes genéticos de DNA. Muitos estudos já identificaram variantes patogênicas em regiões intrônicas profundas em uma série de SPHC, como: câncer de mama e ovário hereditários (HBOC) (Zemankova *et al.*, 2024), síndrome de Lynch, polipose adenomatosa familiar, neurofibromatose, e síndrome de Li-Fraumeni (Landrith *et al.*, 2020, Horton *et al.*, 2022). Porém, pouco se sabe sobre a prevalência dessas variantes intrônicas profundas, e uma das causas é a escalabilidade limitada dos métodos de teste de RNA (Landrith *et al.*, 2020).

O estudo de Landrith e colaboradores (2020) juntamente com outros dois estudos mais recentes (Horton *et al.*, 2022; Kamps-Hughes *et al.*, 2023) realizaram sequenciamento de RNA de captura pareado com o sequenciamento de DNA em grandes coortes de 1.000, 43.524 e 20.317 pacientes, respectivamente, com painéis de 18 a 63 genes de predisposição hereditária ao câncer para o diagnóstico clínico. No geral, estes pacientes tinham suspeita de câncer hereditário, os quais já tinham sido submetidos ao sequenciamento de DNA para teste genético, e a análise de RNA foi feita de forma complementar sem seleção e/ou selecionando aqueles que

possuíam variantes previstas de causar erro de *splicing* e/ou histórico familiar forte visando aumentar da taxa de diagnóstico com a possível reclassificação de variantes. O perfil de *splicing* destes pacientes foram comparados utilizando um grupo de indivíduos controles saudáveis e um grupo de pacientes com variantes já conhecidas de alterar o *splice*, dessa forma alcançaram um aumento na taxa de diagnóstico positivo, ou seja, conseguiram reclassificar as VUS exclusivamente através da adição da análise de RNA (Landrith *et al.*, 2020; Horton *et al.*, 2022; Kamps-Hughes *et al.*, 2023). No trabalho de Landrith e colaboradores (2020) houve uma taxa de 9,1% de aumento relativo na detecção de variantes patogênicas, já no trabalho de Kamps-Hughes e colaboradores (2023) alcançaram uma taxa de reclassificação do total de VUS de 6,3%, em que 0,4% foram reclassificadas para P/PP e 5,9% para PB/B. Por último, o trabalho de Horton e colaboradores (2024) relataram reclassificação de 71,1% da VUS de *splice* para P/PB, devido a adição da análise de RNA, e aumento de 1,9% da taxa de diagnóstico positivo total. O destaque entre estes três estudos de forma resumida é que os ensaios de RNA diminuíram a taxa de VUS, ao mesmo tempo que aumentaram a taxa de positividade.

1.6. Ferramentas de bioinformática para análise de *splicing*

Nos últimos anos, ocorreu uma alta de análises de *splicing* através do RNA-seq como complementação para o diagnóstico, principalmente de indivíduos com doenças raras, quando o teste genético de DNA era inconclusivo (Wang *et al.*, 2023). Juntamente com essa alta houve o desenvolvimento e utilização de ferramentas de análise de RNA pós sequenciamento para detecção quantitativa de eventos de *splice* alternativos ou aberrantes associados a doenças genéticas, as quais podem ser divididos em três categorias principais: (i) análise de *splicing* diferencial; (ii) detecção de *splicing* aberrante e (iii) análise de rede relacionada ao *splicing* (Liu *et al.*, 2023). As análises de *splicing* avaliam as leituras (*reads*) que ficam principalmente nas junções éxon-éxon, pois através dessa informação é possível saber como está a ligação entre os éxons, ou seja, se ocorreu o processo de retirada dos íntrons de forma correta, pois as bordas dos éxons que serão afetadas e não o seu interior. A Figura 7 demonstra as *reads* que são priorizadas nesse tipo de análise.

Figura 7 - Representação das regiões que se concentram as análises de RNA pelas ferramentas de bioinformática



Fonte: Autoria própria.

Legenda: As ferramentas focam nas leituras do sequenciamento (*reads*) que ficam nas junções éxon-éxon (traços em vermelho), pois são informativas em relação a alguma alteração de *splice* como a representada acima que ocorreu salto do éxon 7, mas que também podem capturar eventos de saltos parciais de éxon, como também inserções totais e parciais de íntrons.

A **análise de *splicing* diferencial** tem como principal objetivo avaliar o perfil de *splicing* entre dois grupos de amostras de diferentes condições com múltiplas réplicas biológicas, como por exemplo, amostras casos e controles. Algumas ferramentas que se destacam nesse tipo de análise são: rMATS v.4.3.0 (Shen *et al.*, 2014), MISO (Katz *et al.*, 2010), DEX-seq (Anders *et al.*, 2012), MAJIQ (Norton *et al.*, 2018), Splice-q (de Melo Costa *et al.*, 2021) e DARTS (Zhang *et al.*, 2019) (Liu *et al.*, 2023). A ferramenta rMATS utilizada nesse estudo, faz detecção de eventos de *splicing* que são diferentes entre dois grupos de amostras. Para isso utiliza um modelo hierárquico para contabilizar a incerteza de amostragem em réplicas individuais, assim como também a variabilidade entre réplicas e realiza a detecção e quantificação de eventos de *splicing* alternativos diferenciais pela métrica de PSI (*Percent splice in*), como por exemplo, uso alternativo do sítio 3' e 5' (A3SS e A5SS), retenção de íntron (RI), uso mutuamente exclusivo de éxons (MXE) e salto de éxon (SE) (Shen *et al.*, 2014; Liu *et al.*, 2023).

A **análise de detecção de *Splice* aberrante ou *Outlier***, tem como principal objetivo encontrar um evento de *splicing* anormal/discrepante de um conjunto único com várias amostras de controles e pacientes a partir do sequenciamento de transcriptoma. Algumas ferramentas que se destacam nesse tipo de análise são: FRASER (Mertes *et al.*, 2021), LeafCutterMD (Jenkinson *et al.*, 2020), SPOT (Ferraro *et al.*, 2020) e Bisbee (Halperin *et al.*, 2021). As duas primeiras ferramentas foram utilizadas nesse estudo, e serão detalhadas a seguir. Na ferramenta LeafCutterMD, o *splicing* é medido como a excisão de íntrons (ao invés da inclusão de éxons) utilizando o modelo de distribuição Dirichlet-Multinomial, a qual faz a contagem de leituras do RNA-seq que abrangem as junções exon-exon alternativamente unidos (*spliced*) utilizando todas as amostras (pacientes e controles), dessa forma identificam as amostras individuais em que essas leituras se desviam significativamente da expectativa com base na distribuição ajustada (Wang *et al.*, 2023). A ferramenta FRASER (Find **RAre Splicing** Events in **RNA**-seq)

utiliza uma distribuição beta-binomial para avaliar a significância estatística de eventos *outliers* candidatos identificados com base em diferentes métricas de *splicing* (ψ_5 , ψ_3 e θ_5 , θ_3), dados que se desviarem do modelo ajustado são caracterizados como *outliers*. Essa ferramenta oferece como um diferencial a detecção de eventos de retenção de íntrons *outlier*, característica indisponível nos métodos das ferramentas anteriores (Liu *et al.*, 2023; Wang *et al.*, 2023). Essa ferramenta pode ser utilizada como um submódulo dentro do pipeline DROP, recentemente utilizado para diagnóstico de doenças raras, por realizar três análises principais com dados do RNA-seq: expressão gênica aberrante, *splicing* aberrante e expressão monoalélica específica (Yépez *et al.*, 2021).

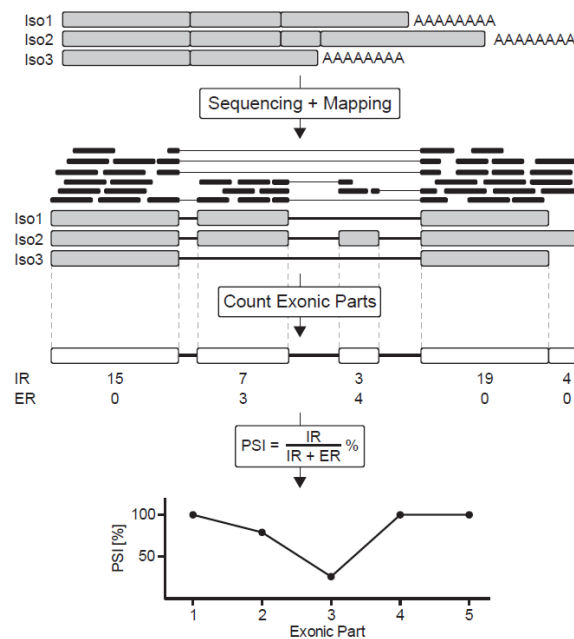
Por fim, a **análise de rede relacionada ao *splicing*** que diferente das anteriores focam na relação entre os diversos eventos e vias de *splice* no nível de rede, alguns exemplos são: SpliceNet (Yalamanchili *et al.*, 2014), CoSpliceNet (Aghamirzaie *et al.*, 2016), RWMG (Wang *et al.*, 2020) e SPIRON (Li *et al.*, 2021) (Liu *et al.*, 2023).

Em nosso estudo, utilizamos ferramentas dos dois primeiros tipos de análise, que apresentam aplicações para a detecção de eventos de *splicing* aberrante ocasionados por variantes genéticas raras.

1.6.1. Métricas utilizadas para detectar *splicing* alternativo

Existem muitas métricas para caracterizar eventos de *splice* alternativos ou aberrantes possuindo suas próprias vantagens e um foco específico de análise (Liu *et al.*, 2023). A quantificação mais comum utilizada para detecção de eventos de *splice* diferencial entre grupos, como no caso do rMATS, é o cálculo através do valor de **PSI** (*Percent Spliced in*), que é definido como número de leituras que incluem o evento de *splicing* alternativo dividido pela soma das leituras que incluem e excluem tal evento (número total de leituras no entorno da região do evento de *splicing*) (Schafer *et al.*, 2015). O valor de PSI varia de 0,0 a 1,0, no qual o valor 0 significa que nenhuma leitura incluiu determinado evento e o valor 1 significa que todas as leituras incluíram determinado evento, demonstrando uma relação de porcentagem entre 0% e 100%, respectivamente. Um esquema desse cálculo pode ser observado na Figura 8.

Figura 8 - Esquema simplificado para o cálculo do PSI entre diferentes isoformas de um mesmo gene



Fonte: Schafer *et al.* (2015).

Legenda: Nesse exemplo são mostradas três isoformas para um determinado gene, nas quais primeiro e último éxon são sempre inclusos, já os éxons centrais sofrem alteração de inclusão, dessa forma conta-se quantas reads incluem e excluem determinado éxon. Logo quanto maior o PSI, mais esse éxon é incluído (presente em muitas isoformas), e quanto menor o PSI, menos esse éxon é incluído (presente em poucas isoformas). **IR** – inclusion reads; **ER** – exclusion reads.

Outras medidas, como o **Índice de Splice** (ψ_5 e ψ_3) e **Eficiência de Splice** (θ_5 e θ_3), as quais são utilizadas pela ferramenta FRASER. A primeira métrica mede o índice de *splicing* nas extremidades 5' e 3' dos íntrons, de forma geral ψ_5 indica a proporção (razão) do número de leituras cortadas que abrangem o íntron entre o doador (D) e o receptor (A) por todos os receptores que sofreram *splice* junto com o doador de interesse, e o ψ_3 indica a proporção (razão) do número de leituras cortadas abrangendo o íntron entre o doador e o receptor por todos os doadores que sofreram *splice* junto com receptor de interesse (Figura 9). Já a segunda métrica indica o índice de completude do *splicing*, em que θ_5 é a proporção (razão) do número de todos os receptores que sofreram *splice* com o doador de interesse por leituras não-processadas e processadas que abrangem o limite exon-íntron do doador, e θ_3 é a proporção do número de todos os doadores que sofreram *splice* com o receptor de interesse para leituras não-processadas e processadas abrangendo o limite exon-íntron do receptor (Figura 10).

Figura 9 - Equação geral das métricas ψ_5 e ψ_3

$$\Psi_5(D, A) = \frac{n(D, A)}{\sum_{A'} n(D, A')}, \quad \Psi_3(D, A) = \frac{n(D, A)}{\sum_{D'} n(D', A)},$$

Fonte: Liu *et al.* (2023).

Legenda: Equação geral, em que $n(D, A)$ refere-se ao número de leituras processadas (*split reads*) abrangendo o íntron entre o doador e aceitador, e $n(D, A')$ são todas as leituras processadas do doador e aceitador, em que A' (aceptor de interesse) correu por todos os sítios aceptores; e $n(D', A)$ são as leituras processadas do doador e aceitador, em que D' (doador de interesse) correu por todos os sítios doadores.

Figura 10: Equação geral das métricas θ_5 e θ_3

$$\theta_5(D, A) = \frac{\sum_{A'} n(D, A')}{n(D) + \sum_{A'} n(D, A')}, \quad \theta_3(D, A) = \frac{\sum_{D'} n(D', A)}{n(A) + \sum_{D'} n(D', A)},$$

Fonte: Liu *et al.* (2023).

Legenda: Equação geral, em que $n(D)$ refere-se ao número de leituras não-processadas (*non-split reads*) abrangendo o limite éxon-íntron do doador e $n(A)$ refere-se ao número de leituras não-processadas (*non-split reads*) abrangendo o limite éxon-íntron do aceitador.

Em relação às ferramentas pós sequenciamento de RNA-seq para análise de eventos de *splicing* alternativos e aberrantes, as mais utilizadas atualmente são: rMATS, MAJIQ, LeafCutterMD e FRASER2. No entanto, essas ferramentas apresentam muitas vezes resultados discordantes entre si, assim como diferentes graus de sensibilidade e especificidade. Como não existe uma diretriz que indique quais os melhores parâmetros para otimizar a captura de eventos significativos, cada estudo adota suas próprias métricas (Jaramillo *et al.*, 2024).

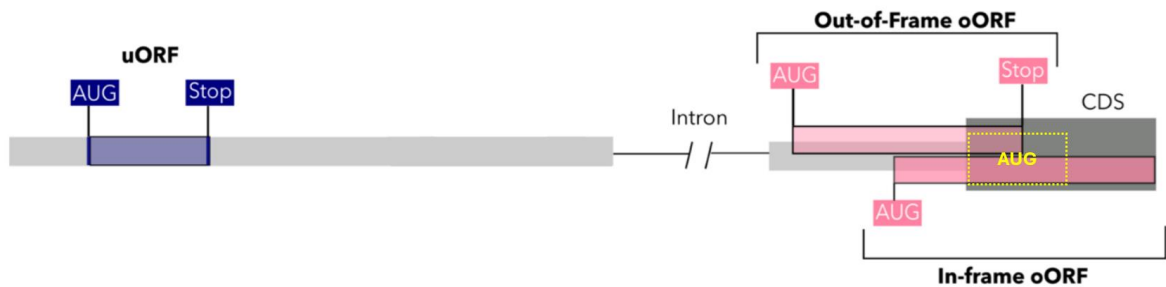
1.7. Regiões 5' e 3' UTR do mRNA

As regiões não traduzidas 5' e 3' (5' e 3' UTR, do inglês *untranslated region*) são trechos que se localizam no início e no final da sequência codificante do mRNA, ou seja, antes do primeiro éxon e depois do último éxon, respectivamente. Essas regiões são transcritas, porém raramente traduzem proteínas, pois atuam como reguladores pós-transcricionais essenciais na regulação da expressão gênica envolvidos no processamento e estabilização de mRNA e iniciação da tradução. Em pelo menos 13% dos transcritos de genes de mamíferos acontece *splicing* alternativo dentro das regiões UTR, que pode gerar um nível de controle pós-

transcricional diferente (Wieder *et al.*, 2024). Em estudos de Associação Genômica Ampla (GWAS) cerca de 3,7% das variantes encontradas estão localizadas em UTRs e através dos Sequenciamentos de Genoma Completo (WGS), cada vez mais são encontradas variantes relacionadas ao câncer nessas regiões, embora pouco se sabe como essas alterações genéticas podem contribuir funcionalmente para o surgimento da doença (Steri *et al.*, 2018; Schuster e Hsieh, 2019).

Nas regiões 5'UTR podem-se encontrar Quadros de Leitura Aberta a montante (uORFs – *Upstream Open Reading Frame*) que ocorrem em aproximadamente metade dos transcritos de mRNA humanos, sendo considerados importantes cis-reguladores de tradução tecido específico. A uORF é uma sequência no DNA que começa em um códon de iniciação AUG a montante (uAUG – *Upstream* AUG), o qual antecede o AUG da sequência codificante (AUG canônico) de determinado gene, e acaba num códon de término (UGA, UAG ou UAA) que pode estar localizado antes ou depois do códon de término canônico. Essa região regulatória, que pode ser traduzida em um pequeno peptídeo, funciona reduzindo em até 80% os níveis de expressão de uma proteína, pelo fato de atuar reduzindo a eficiência do início da tradução no ribossomo através de diferentes mecanismos (Barbosa *et al.*, 2013; Schulz *et al.*, 2018; Whiffin *et al.*, 2020). Por isso, mutações raras ou polimorfismos que criam uAUGs modificam ou interrompem o quadro de leitura de uORFs (por indels), ou que removem o códon de parada de uORFs, podem afetar diretamente na expressão proteica de determinado gene e estar relacionadas ao aparecimento de várias doenças humanas como distúrbios metabólicos ou neurológicos, síndromes hereditárias e câncer (Barbosa *et al.*, 2013; Whiffin *et al.*, 2020). Ao remover o códon de parada a montante (uStop-Codon) pode se formar uma ORF sobreposta (*overlapping* ORF - oORF) ao AUG canônico, em que códon de parada correspondente vai se estender além do início da região sequência codificante (CDS). Dessa forma, pode acontecer uma oORF dentro do quadro de leitura (*in-frame*) com a CDS formando um transcrito alongado (vai usar o mesmo códon de parada), ou ser fora do quadro de leitura (*out-of-frame*) terminando dentro da CDS (Wieder *et al.*, 2024). Na Figura 11 é possível observar esses tipos diferentes de ORFs: as ORFs a montante (uORF) e ORFs sobrepostas (oORFs).

Figura 11 - Representação de uma uORF e oORF *in-frame* e *out-of-frame* na região 5'UTR



Fonte: Adaptado de Wieder *et al.* (2024).

Legenda: A uORF tem uAUG com um códon de parada a montante (uStop-Codon) localizado antes do AUG canônico (em amarelo com linha pontilhada) da região codificante (CDS), representada pela cor roxo escura e a CDS com uma cor cinza transparente escura. A oORF tem um uAUG com um códon de parada que ultrapassa/sobrepõe o AUG canônico podendo estar na mesma janela de leitura da CDS (*in-frame*) formando um transcrito mais longo, ou fora do quadro de leitura (*out-of-frame*) terminando dentro da CDS antes do códon de parada canônico e a formação de um transcrito mais curto, ambas representadas pela cor rosa claro.

Para variantes detectadas na região 5'UTR, recentemente foi desenvolvido e validado um preditor *in silico* de impacto de variantes genéticas, chamado 'UTRannotator' (Zhang *et al.*, 2021). O *plugin* 'UTRannotator' está disponível para utilização via web do VEP ou via linha de comando através do repositório online (<https://github.com/ImperialCardioGenetics/UTRannotator>). Essa ferramenta anota variantes de alto impacto encontradas na região 5'UTR de eucariotos que podem criar ou interromper uORFs. Através desse *plugin* é possível anotar pequenas variações que incluem SNVs (1-5pb), variantes multinucleotídicas (MNV) e inserções/deleções de pequena escala (indels). Além de avaliar se a uORF já possui evidência experimental de tradução ou se determinada variante já é esperada para determinado transcrito através de uma lista curada de uORFs previamente identificados num repositório online de pequenos ORFs ([www. Sorfs.org](http://www.Sorfs.org)).

Essa ferramenta possui como saída possíveis cinco consequências que podem acontecer dentro de uma uORF: criação de um novo códon AUG (uAUG_gained), perda de um códon de início AUG existente (uAUG_lost), perda de um códon de parada de uma ORF *upstream* existente (uSTOP_lost), criação de um novo códon de parada dentro de uma ORF *upstream* existente (uSTOP_gained) e criação de uma mutação *frameshift* em um ORF *upstream* existente (uFrameshift) (Zhang *et al.*, 2021).

Em relação às regiões 3' UTR, estas regulam processos como: estabilidade, tradução e localização do mRNA, além de conter em sua sequência elementos responsáveis pela ligação de micro-RNAs (miRNAs) e proteínas (RBPs – *RNA binding proteins*). Modificações em sua sequência e em seu comprimento podem afetar a sinalização de poliadenilação, a ligação de miRNAs e RBPs, e consequentemente a expressão gênica (Steri *et al.*, 2018). Alguns exemplos mostram que elementos ricos em AU presentes na 3' UTR do mRNA estão super-representados no câncer podendo ter papel na progressão mitótica, e outro exemplo mostra que elementos ricos em G estão enriquecidos na região 3' UTR de 10 genes relacionados com o processo de transição epitélio mesênquima (Schuster e Hsieh, 2019). Os efeitos de alterações da região 3' UTR foram mais explorados no contexto de variantes comuns em estudos de GWAS e biomarcadores somáticos, e poucos trabalhos exploraram variantes nessa região como causas de doenças genéticas de herança Mendeliana, como nas SPHC.

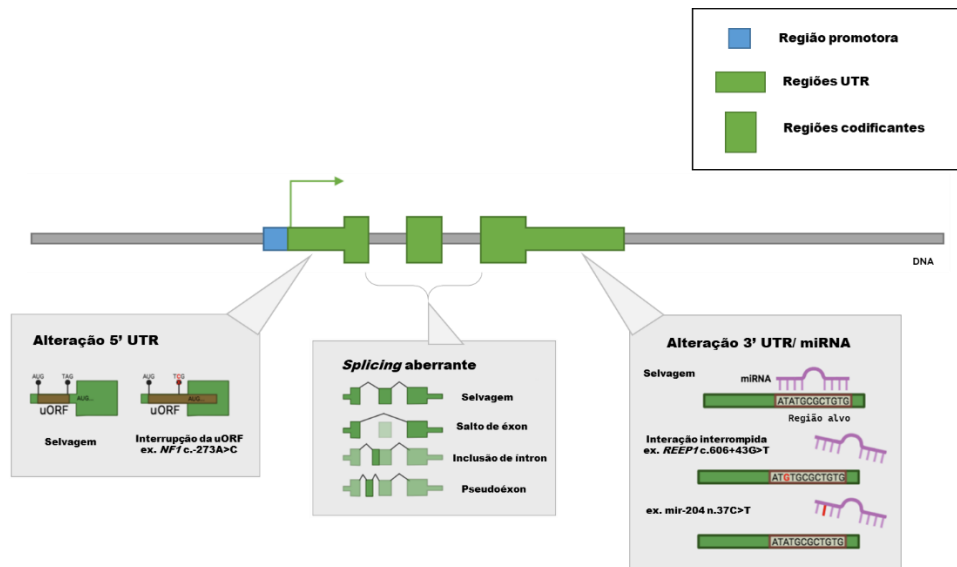
2. JUSTIFICATIVA

Os avanços recentes na genômica levaram ao reconhecimento de quase uma centena de novos genes de predisposição ao câncer. Estes avanços levaram também a um aumento dos estudos relacionados às SPHC e da utilização clínica do sequenciamento genético no contexto de câncer familiar. No entanto, uma significativa parcela (50-80%) dos pacientes investigados para SPHC apresentam variantes de significado clínico incerto. A identificação de VUS representa um desafio para a prática clínica e aconselhamento genético, uma vez que esse resultado é inconclusivo em relação à presença ou ausência de uma determinada SPHC. Algumas VUS, como variantes intrônicas em posições não-canônicas dos sítios de *splice* e variantes sinônimas raras, podem apresentar efeitos danosos às proteínas por alterar o *splice* normal do gene. A estratégia de realização do sequenciamento alvo pareado de DNA e RNA tem sido eficaz em auxiliar na reclassificação de VUS, pois o RNA pode fornecer mais evidências do efeito e da expressão das alterações detectadas no DNA.

Adicionalmente, as regiões 5' e 3' UTRs são pouco exploradas no contexto de testes genéticos clínicos, e variantes detectadas nessas regiões são de difícil interpretação (Schulz *et al.*, 2018; Whiffin *et al.*, 2020). No entanto, estudos prévios já demonstraram que variantes em regiões 5' UTR contribuíram para doenças humanas de alta penetrância quando nenhuma variante candidata foi encontrada na sequência codificante (Liu *et al.*, 1999; Matthes *et al.*, 2004; Calvo *et al.*, 2009; Wen *et al.*, 2009; Barbosa *et al.*, 2013; Occhi *et al.*, 2013; von Bohlen *et al.*, 2017; Schulz *et al.*, 2018).

Nesse sentido, a proposta desse estudo é analisar variantes VUS encontradas em regiões não codificantes 5' e 3' UTRs, além das encontradas em regiões exônicas e intrônicas (intragênicas) que são previstas de alterarem o *splicing*, para aprimorar a reclassificação dessas variantes nos diagnósticos clínicos moleculares, já que essas regiões podem abrigar variantes germinativas associadas a SPHC. Na Figura 12 é possível ter uma visão ampla dos elementos regulatórios no mRNA, com enfoque nas regiões que esse estudo irá investigar, as quais podem alterar as uORFs nas regiões de 5' UTR, causar *splicing* aberrante na região intragênica a partir de variantes intrônicas ou exônicas, e modificar a regulação do mRNA por miRNA através da alteração da região 3' UTR.

Figura 12 - Representação do trecho correspondente a transcrição do mRNA no DNA



Fonte: Modificada de Ellingford e colaboradores (2022).

Legenda: Destaque para as possíveis modificações genéticas em regiões codificantes e não codificantes que podem modificar a estrutura de um transcrito (por *splicing* aberrante) ou influenciar na regulação de sua tradução (alterações 5' e 3' UTR).

3. OBJETIVOS

3.1. Objetivo geral

Estabelecer e validar um conjunto de ferramentas genômicas, computacionais e experimentais que permitam reclassificar VUS não codificantes ou codificantes com possível efeito em *splice*, que possam ser implementadas na clínica para auxiliar no diagnóstico das síndromes de predisposição hereditária ao câncer.

3.2. Objetivos específicos

- Identificar variantes não codificantes (intrônicas e nas regiões 5' e 3' UTRs) ou codificantes com possível efeito em *splice* em testes genéticos prévios;
- Desenvolver e validar uma metodologia para sequenciamento alvo de RNA (RNA-cap) de 27 genes de predisposição hereditária ao câncer;
- Avaliar por sequenciamento de RNA direcionado VUS preditas de afetar *splice*;
- Avaliar a relevância das ferramentas *in silico* para análise de eventos de *splicing* alternativos/aberrantes e da avaliação de RNA para a reclassificação de VUS.

4. MATERIAIS E MÉTODOS

4.1. Casuística

Este projeto de pesquisa foi aprovado no Comitê de Ética em Pesquisa (CEP) sob o nº 3051/21 (CAAE: 43174921.3.0000.5432). Os pacientes que foram avaliados nesse estudo na etapa de revisão de variantes genéticas realizaram sequenciamento de DNA germinativo para painéis de 26 a 126 genes entre 2016 e 2022 para fins de projetos de pesquisa prévios ou de diagnóstico clínico (testes do laboratório de diagnóstico genômico (LDG) do A.C Camargo ou de laboratórios externos para pacientes do Departamento de Oncogenética). Em relação aos projetos de pesquisa prévios, foram revisados dados de 4 projetos com as seguintes casuísticas: pacientes com câncer colorretal abaixo de 50 anos (97 pacientes); pacientes com câncer gástrico e suspeita de predisposição hereditária (idade jovem, múltiplos casos na família ou gástrico difuso – 109 pacientes); pacientes com sarcoma abaixo de 40 anos (110 pacientes); pacientes com tumores neuroendócrinos abaixo de 40 anos (82 pacientes). Em relação aos testes genéticos clínicos realizados no LDG ou em laboratórios externos, os pacientes adicionados ao estudo realizaram os testes apresentando suspeita de predisposição hereditária ao câncer e em geral preencheram critérios clínicos recomendados pela Agência Nacional de Saúde ou *guidelines* específicos seguidos na prática clínica, como os do *National Comprehensive Cancer Network* (NCCN) e outras recomendações específicas de cada síndrome. Os pacientes dessa coorte em sua maioria são portadores de câncer de mama e ovário.

4.2. Revisão de testes genéticos prévios para identificação das VUS de interesse

Para identificação dessas variantes, foi levantado os dados de projetos anteriores do grupo, e adicionalmente os resultados de laudos LDG, chegando ao total de 2.076 pacientes.

A seleção das variantes foi feita através dos arquivos de chamada de variantes brutos (VCFs) gerados a partir do sequenciamento de painéis de 26, 62, 112, 113 ou 126 genes. Para esse estudo, foi utilizado um painel de captura de RNA 27 genes de predisposição hereditária ao câncer, dos quais 20 estão contidos em todos os painéis utilizados nos testes de DNA prévios (*APC*, *ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CHEK2*, *EPCAM*, *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *NBN*, *PALB2*, *PMS2*, *PTEN*, *RAD51C*, *RAD51D*, *STK11* e *TP53*), mais informações sobre cada painel com seus respectivos genes estão no APÊNDICE A. A escolha dos 27 genes

para esse estudo deve-se a sua relação definitiva e/ou moderada com as principais SPHC, as informações sobre a relação de gene-doença para os 27 genes, tipo de classificação (definitiva/moderada) e tipo de herança estão no APÊNDICE B.

Os VCFs de cada paciente foram avaliados no programa VarSeq (GoldenHelix) utilizando 3 filtros específicos relacionados a ‘**Qualidade da chamada de variante**’, ‘**Frequência populacional**’ e em relação ao seu ‘**Efeito/impacto no gene**’, com intuito de selecionar variantes germinativas raras não benignas ou provavelmente benignas. Na Tabela 1 resume os detalhes dos filtros aplicados.

Nesse estudo foram avaliadas variantes germinativas em regiões codificantes e não codificantes, como:

- **Exônicas:** sinônimas ou *missenses*; ou
- **Intrônicas:** em até 20 pares de bases do limite do éxon ou nas regiões 5’ e 3’ UTR.

Para refinar a análise foram retiradas as variantes mitocondriais e possíveis variantes consideradas artefatos do sequenciamento, ou seja, falsas variantes que aparecem em regiões repetitivas em muitos pacientes.

Apesar dos painéis de genes utilizados não apresentarem sondas desenhadas para a totalidade das regiões 5’ e 3’ UTRs, devido ao procedimento de fragmentação enzimática e captura do DNA, estas regiões podem ser capturadas indiretamente por sondas direcionadas ao primeiro e último éxon (denominadas regiões fora de alvo ou *off-target*).

Tabela 1 - Filtros personalizados aplicados no VarSeq para os VCFs

(continua)

Qualidade da chamada de variante	Todas as variantes	Q >20
		Cobertura > 15X leituras
		VAF > 35%
Frequência populacional	Todas as variantes	< 5% nas amostras da coorte
		≤ 1% frequência no banco gnomAD

Fonte: Elaboração dos autores.

Legenda: Em parêntesis o nome original do filtro utilizado na ferramenta. **VAF:** Frequência alélica; **B:** Benigna; **PB:** Provavelmente Benigna; **nt:** Nucleotídeo (s).

Tabela 1 - Filtros personalizados aplicados no VarSeq para os VCFs

(conclusão)

Efeito/impacto do gene	Todas as variantes	Exclui B e PB do Clinvar, inclui VUS (<i>conflicting</i> , <i>Uncertain Significance</i>), P/PP (<i>Likely Pathogenic</i> , <i>Pathogenic</i>) e não descritas.
	Variantes exônicas e intrônicas (não-UTR)	Inclui: <i>missense</i> (<i>missense variant</i>), sinônima (<i>synonymous variant</i>), doador e aceitador de <i>splice</i> (<i>splice donor variant</i> , <i>splice acceptor variant</i>) [± 1 nt a ± 2 nt], região de <i>splice</i> (<i>splice region variant</i>) [± 3 nt a ± 8 nt], variantes intrônicas (<i>intron variant</i>) [± 8 nt a ± 20 nt].
	Variantes em 5'UTR	Inclui: localizadas na região 5'UTR (<i>5 prime UTR variant</i> , <i>5 prime UTR premature start codon gain variant</i> , <i>Initiator codon variant</i>).
	Variantes em 3' UTR	Inclui: localizadas na região 3'UTR (<i>3 prime UTR variant</i>).

Fonte: Elaboração dos autores.**Legenda:** Em parêntesis o nome original do filtro utilizado na ferramenta. **VAF:** Frequência alélica; **B:** Benigna; **PB:** Provavelmente Benigna; **nt:** Nucleotídeo (s).

As variantes que foram selecionadas por esses filtros do VarSeq foram posteriormente avaliadas através da ferramenta *Variant Effect Predictor* (VEP) do Ensembl (<https://www.ensembl.org/Tools/VEP>), utilizando programas de predição *in silico* quanto à probabilidade de resultarem em alteração do *splice* (MaxEntScan, dbSNV e *SpliceAI*) e impacto na região 5'UTR (UTRannotator) nos transcritos clinicamente relevantes, ou transcritos MANE (do inglês “*The Matched Annotation from the NCBI and EMBL-EBI*”). O projeto MANE é colaborativo e tem como principal objetivo concentrar-se na anotação de genes e transcritos humanos e definir um conjunto amplo de transcritos representativos e proteínas correspondentes (quando aplicável) para genes humanos (Morales *et al.*, 2022). Ainda existem os transcritos MANE Plus Clinical, que são transcritos adicionais, em que o transcrito MANE selecionado não é suficiente para relatar as variantes P/PP. Os transcritos MANE/ MANE Plus Clinical selecionados para cada gene são: *APC* (NM_000038.6), *MUTYH* (NM_001048174.2 -

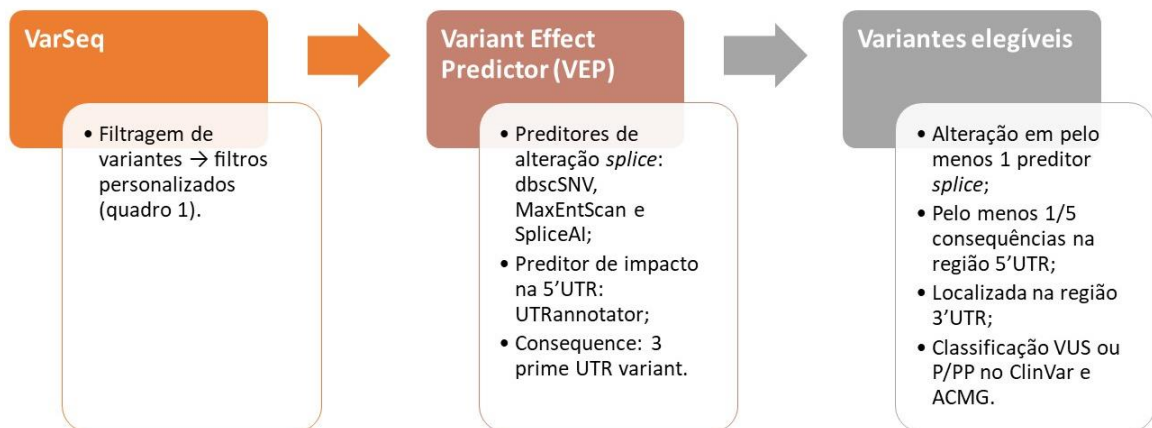
MANE; NM_001128425.2 – MANE Plus Clinical); *ATM* (NM_000051.4), *BARD1* (NM_000465.4), *BMPR1A* (NM_004329.3), *BRCA1* (NM_007294.4), *BRCA2* (NM_000059.4), *BRIP1* (NM_032043.3), *CDH1* (NM_004360.5), *CDK4* (NM_000075.4), *CDKN2A* (NM_000077.5 – MANE; NM_058195.4 - MANE Plus Clinical), *CHEK2* (NM_007194.4), *EPCAM* (NM_002354.3), *MLH1* (NM_000249.4), *MSH2* (NM_000251.3), *MSH6* (NM_000179.3), *NBN* (NM_002485.5), *PALB2* (NM_024675.4), *PMS2* (NM_000535.7), *PTEN* (NM_000314.8), *RAD51C* (NM_058216.3), *RAD51D* (NM_002878.4), *RET* (NM_020975.6), *SMARCA4* (NM_003072.5), *STK11* (NM_000455.5), *TP53* (NM_000546.6), *NF1* (NM_001042492.3).

Foram selecionadas variantes que apresentaram escores indicativos de alteração no *splicing* em pelo menos 1 dos 3 programas (diferença >3 para MaxEntScan, ada e rf scores >=0,6 para dbSCSNV e delta score >=0,2 *SpliceAI*); Para impacto na região 5'UTR, foram selecionadas as variantes que apresentaram pelo menos uma das cinco consequências possíveis de alteração: ganho de códon de início a montante (uAUG_gained), perda de códon de início a montante (uAUG_lost), perda do códon de parada a montante (uSTOP_lost), ganho de códon de parada a montante (uSTOP_gained) e mutação de mudança de quadro de leitura a montante (uFrameshift); Para região 3'UTR, como ainda não existe uma ferramenta de predição de impacto no VEP ou outra opção clinicamente validada, foram selecionadas todas as variantes que na coluna “Consequence” da tabela gerada pelo VEP correspondiam a “3 prime UTR variant”.

Antes da publicação da recomendação da *Clingen* (Walker *et al.*, 2023) para variantes de *splice* estávamos adicionando ao estudo apenas variantes com *cut-off* do delta score >=0,50, após essa publicação foi estabelecido reduzir o *cut-off* para 0,20, pois foi visto que esse escore abrange variantes com risco moderado de alterar o *splice*.

Na Figura 13 está o resumo do processamento geral das variantes desde sua filtragem personalizada no VarSeq, predição *in silico* com ferramentas específicas no VEP, e os critérios principais para sua elegibilidade e inclusão nesse estudo.

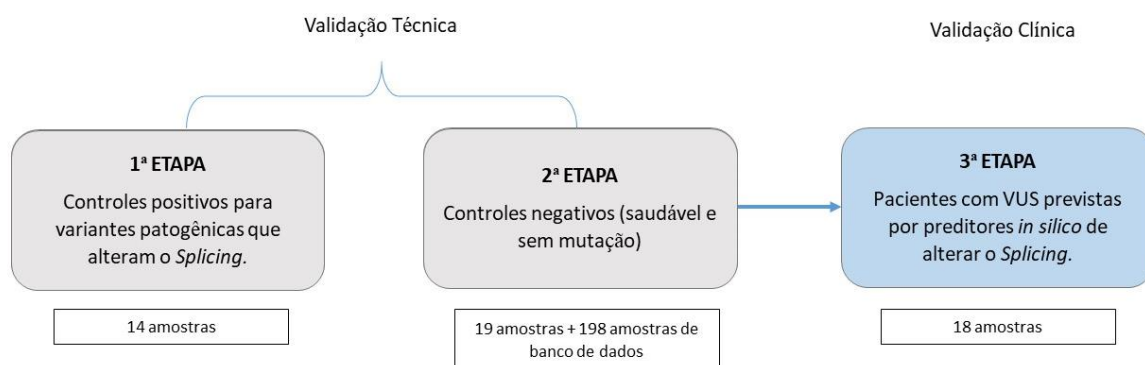
Figura 13 - Processamento geral das variantes



Fonte: autoria própria.

4.3. Delineamento da validação do RNA-cap

A validação do RNA-cap foi realizada em três etapas distintas, descritas na Figura 14. Na primeira etapa foi feita a validação de 14 amostras positivas para alterações de *splicing* patogênicas ou provavelmente patogênicas sabidamente já conhecidas e caracterizadas na literatura, compondo o grupo controle positivo. A segunda etapa consistiu na formação do grupo controle negativo, ou seja, pessoas saudáveis sem câncer, composto por 19 amostras controles do A.C Camargo (controle interno) mais 198 amostras do banco de dados disponível a partir do artigo de Landrith *et al.* 2020. Essas duas etapas compreendem o que denominamos de Validação Técnica, que tem como objetivo a padronização das técnicas experimentais, como também as de análise de dados, que envolvem: a construção de biblioteca de cDNA, captura por hibridização, sequenciamento de nova geração e análise de dados através das ferramentas rMATS, LeafCutterMD e FRASER2. A terceira etapa é denominada Validação Clínica, na qual selecionamos 18 pacientes com VUS previstas por softwares de predição de afetar o padrão de *splicing* para realizar o teste genético de RNA. Essa etapa tem como objetivo aplicar a técnica padronizada para avaliar seu poder de reclassificação de VUS através do RNA-cap. Os pacientes do controle positivo e com VUS foram referenciados e anonimizados nesse estudo como ‘RNAP’.

Figura 14 - Delineamento do estudo

Fonte: autoria própria.

4.3.1. Critérios de inclusão de pacientes

Os critérios de inclusão de cada grupo do estudo descrito acima estão detalhados nos itens 4.3.1.1, 4.3.1.2 e 4.3.1.3.

Para os pacientes selecionados, foi avaliada a disponibilidade de material biológico (leucócito congelado ou tecido normal) para extração de RNA no Biobanco institucional do A.C. Camargo Cancer Center. Os pacientes que possuíam material disponível e que tinham concordado com o uso do material em pesquisas futuras sem necessidade de re-consentimento foram incluídos diretamente no estudo. Aqueles que não possuíam material no Biobanco ou desejavam ser re-consentidos a cada estudo foram monitorados através do sistema Tasy para possível inclusão prospectiva após alguma consulta médica. Caso aceitassem participar do estudo, estes assinaram o Termo de Consentimento Livre e Esclarecido (TCLE), e dessa forma uma amostra de sangue em tubo EDTA foi coletada por uma enfermeira de pesquisa do hospital.

4.3.1.1. Seleção de indivíduos saudáveis do grupo controle negativo

O grupo controle negativo foi composto por 2 subgrupos: 19 indivíduos controle sem câncer e sem critérios clínicos para síndromes de predisposição hereditária ao câncer, com RNA de leucócito congelado no biobanco do A.C. Camargo Cancer Center (Controle negativo interno) e 198 indivíduos controles com dados de sequenciamento de RNA-cap realizado a partir de amostras de leucócitos, com um painel de 18 genes supressores de tumor (*APC*, *ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CHEK2*, *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *NF1*, *PALB2*, *PMS2*, *PTEN*, *RAD51C*, *RAD51D*, e *TP53*) do estudo realizado por Landrith e colaboradores (2020), os quais estão disponíveis no *Sequence Read Archive* sob o número PRJNA603342 (Controle negativo externo). Desse grupo, foram baixados os dados de 345 controles, que foram avaliados

quanto a métricas de qualidade de sequenciamento e selecionados apenas aqueles com mais de 85% da região alvo cobertas acima de 10X com intuito de dar mais confiança para as análises posteriores, atingindo 198 indivíduos controles desse banco de dados.

4.3.1.2. Seleção de pacientes do grupo controle positivo

Após as etapas de revisão dos VCFs prévios para seleção de variantes e predição do seu efeito em *splice*, como descrito no tópico 4.2, selecionou-se 14 pacientes portadores de variantes que ocorreram em um dos 27 genes do painel e classificadas como patogênicas (P) ou provavelmente patogênicas (PP) pelos critérios da ACMG. Para essas variantes, o efeito de modificação do *splice* já foi caracterizado em projetos anteriores do grupo ou foram publicadas por outros pesquisadores no ClinVar, ou seja, selecionamos variantes sabidamente spliceogênicas. Já aquelas variantes preditas de afetar *splice*, mas que foram classificadas como P ou PP devido a outras circunstâncias, como por exemplo, devido à troca do aminoácido e não devido ao seu efeito em *splice* foram excluídas do estudo.

4.3.1.3. Seleção de pacientes com VUS possivelmente modificadoras de *splice*

Para identificação dos pacientes para este grupo foram utilizadas as mesmas abordagens de revisão e avaliação de variantes descritas no tópico 4.2. Dessa forma, selecionamos 18 pacientes portadores de VUS em um dos 27 genes do painel e que foram previstas de alterarem o *splice* em pelo menos 1 dos 3 preditores *in silico* avaliados ou por possuírem grandes deleções/duplicações em que um efeito de *splice* é esperado, mas desconhecido, e não pode ser previsto com as ferramentas.

Demos prioridade para variantes nas quais os respectivos pacientes têm o fenótipo relacionado com o gene, e que não possuam quaisquer variantes P/PP no resultado do sequenciamento de painel. Exemplos: pacientes com suspeita de Síndrome de Câncer de Mama e Ovário Hereditários com a identificação de uma VUS em genes relacionados, mas sem apresentar nenhuma variante P ou PP em *BRCA1* e/ou *BRCA2* em seu resultado diagnóstico, ou pacientes com suspeita de Síndrome de Lynch que apresentaram a identificação de uma VUS, mas sem quaisquer variantes P ou PP nos genes de reparo do mal pareamento.

4.4. Sequenciamento alvo de RNA (RNA-cap)

4.4.1. Coleta de material biológico e extração de RNA

Para os pacientes elegíveis para inclusão no grupo controle positivo ou no grupo com

variantes VUS previstas de alterar *splice* foi verificado se estava presente amostras de leucócitos ou tecido congelado não neoplásico no biobanco. Esses pacientes foram considerados retrospectivos, e em todos eles o RNA foi extraído a partir do tecido normal. Durante esse estudo, pacientes diagnosticados com variantes P/PP ou VUS previstas de alterar o *splice*, ou VUS em regiões 5' e 3' UTR também foram convidados para participarem do estudo pelos médicos oncogeneticistas durante consultas de aconselhamento genético, e caso a resposta fosse positiva, estes eram recrutados de forma prospectiva, sendo coletada amostra de sangue periférico em um tubo de EDTA. De forma geral, pacientes retrospectivos tiveram RNA extraído a partir de tecido normal, e pacientes prospectivos tiveram RNA extraído a partir de leucócitos. As amostras dos indivíduos do grupo controle negativo interno foram todas de sangue periférico de indivíduos sem câncer provenientes do biobanco.

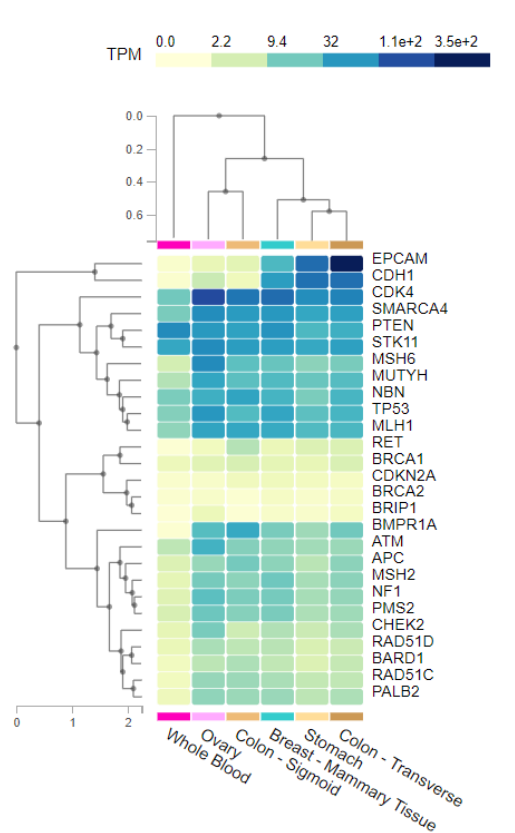
As extrações de RNA total de leucócito e tecido não neoplásico foram realizadas no Biobanco dessa instituição, de acordo com os protocolos operacionais padrão estabelecidos para cada tipo de amostra com o Kit QIASymphony RNA (Qiagen). A integridade do RNA foi avaliada pelo equipamento TapeStation (Agilent) e a concentração pelo equipamento Qubit® com o uso do reagente QUANT-IT RNA assay kit (Life Technologies).

4.4.2. Construção da biblioteca por captura de RNA

O sequenciamento de captura de RNA (RNA-cap) envolveu primeiro a construção da Biblioteca de transcriptoma total a partir de amostras de RNA total (sem depleção para rRNA), e depois a etapa de captura por hibridização para enriquecimento das regiões alvo dos 27 genes.

Considerando foram utilizadas tanto amostras de RNA de leucócitos quanto de tecido normal adjacente, para avaliarmos a expressão esperada dos 27 genes em leucócitos e nos principais tecidos utilizados no estudo utilizamos o banco de dados GTEx Portal (<https://www.gtexportal.org/home/multiGeneQueryPage>). A expressão de todos os genes pode ser observada na Figura 15. Pode-se observar que a expressão dos 27 genes no sangue é semelhante em relação aos tecidos avaliados (ovário, cólon, mama e estômago), embora para os genes *EPCAM* e *CDH1* foi observada uma baixa expressão em sangue total, e avaliamos experimentalmente se a quantidade de sequências obtidas com o RNA-cap seria suficiente para realização das análises posteriores.

Figura 15 - Heatmap do GTEx Portal demonstrando a expressão dos 27 genes de predisposição ao câncer em sangue total contra os principais tecidos analisados



Fonte: Gerado no GTEx Portal com os genes e tecidos de interesse.

As bibliotecas de DNA complementar (cDNA) foram geradas usando o kit xGen™ RNA Library (Integrated DNA Technologies) para a construção da biblioteca cDNA (transcriptoma total). Estabeleceu-se entrar com 300 ng de RNA total sem a necessidade de realizar a depleção de rRNA, apenas as amostras CTRL_100, CTRL_103, RNAP-105, RNAP-109, RNAP-111 e RNAP-112 que entraram com 500 ng iniciais, por questão de teste. Após avaliar a viabilidade com entrada de 300 ng, todas as outras amostras entraram com essa quantidade de massa.

Resumidamente, utilizamos 300 ng de RNA total, o qual é fragmentado por calor a 94° C por 10 min para obtenção do tamanho entre 300-350 pb e logo em seguida a primeira fita de cDNA é sintetizada através de primers aleatórios (random primers). Após a síntese da segunda fita e adição da cauda A, é realizada a ligação dos adaptadores e índices, uma rodada de amplificação por PCR com 12 ciclos seguida de purificação com *beads* magnéticas (Ampure beads). As bibliotecas amplificadas são diluídas em 22 µL de Low EDTA TE, e avaliadas quanto ao rendimento e perfil de fragmentos usando o Bioanalyzer (Agilent) com kit DNA HS.

As bibliotecas que passaram na avaliação de qualidade foram agrupadas em *pools* com 12 a 16 amostras, sendo a entrada de 200 ng totais de cDNA de cada amostra. Os *pools* de amostras foram submetidos à captura com sondas de DNA direcionadas com o IDT gene panel - xGEN lockdown probe, no período de aproximadamente 12h. As bibliotecas pós-captura foram submetidas a uma segunda rodada de amplificação por PCR de 10 ciclos, purificadas com *beads* magnéticas e então avaliadas novamente no Bioanalyzer. As bibliotecas foram sequenciadas no equipamento NextSeq 500 (Illumina), com kits de leitura bilateral (2 X 150 ou 2 X 75 pares de base). Pacientes e controles sequenciados tiveram seus arquivos FASTQs brutos depositados no SRA (do inglês, *Sequence Read Archive*) do NCBI sob o código de acesso e link: PRJNA1255641 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA1255641?reviewer=bnm7bnok9k9n1k1vgb7ip5qun1&archive=sra>).

4.5. Análises de Bioinformática

4.5.1. Alinhamento de RNA e análise de expressão gênica

As análises de bioinformática foram realizadas em colaboração com a *facility* de Bioinformática do A.C. Camargo Cancer Center. Os dados gerados do sequenciamento de RNA foram processados inicialmente na plataforma Illumina para identificação dos indexadores, trimagem de indexes e geração dos arquivos FASTQ. As leituras foram alinhadas ao genoma de referência humano hg38 usando o alinhador STAR v2.7.10b (Dobin *et al.*, 2013). Foi realizada análise de controle de qualidade para avaliar a eficiência de captura, % rRNA e taxa de cobertura dos 27 genes avaliados. As amostras foram consideradas adequadas quando $\geq 85\%$ das regiões alvo apresentaram cobertura média $\geq 50X$.

Para análise de expressão gênica, após o alinhamento de RNA, a contagem de leituras por gene foi feita utilizando a ferramenta HTSeq (Anders *et al.*, 2015). O nível de expressão gênica foi normalizado em transcritos por milhão (TPM) e os resultados apresentados através de gráficos de heatmap utilizando a métrica do z-score.

4.5.2. Ferramentas de análise de eventos de *splice*

Foram utilizadas três ferramentas para análise de eventos de *splicing* alternativos e aberrantes: rMATS v.4.3.0 (<https://github.com/Xinglab/rMATS>), LeafCutterMD v.0.2 (<http://davidaknowles.github.io/leafcutter/>) e FRASER2 v.1.99.4 como um submódulo através

do pipeline DROP v.1.4.0 (<https://github.com/gagneurlab/drop>). Para cada ferramenta foi feito um tipo de alinhamento no STAR recomendado pela documentação individual de cada método: no rMATS foi utilizado com os parâmetros padrões de alinhamento da ferramenta quando o arquivo de entrada é um FASTQ e com o GTF do Ensembl (GRCh38 versão 104); no LeafCutterMD foi utilizado com o parâmetro *--twopassMode Basic* e com o mesmo GTF do Ensembl; e no FRASER2 foi utilizado com o *--twopassMode Basic*, e com o GTF do Gencode (versão 46). O modo Two Pass do STAR permite o aumento da sensibilidade para o descobrimento de novas junções, conseguindo detectar mais leituras de *splice* mapeadas a novas junções.

Utilizamos essas 3 ferramentas de análise de *splicing* em sua maior parte com os parâmetros padrões (standard), com algumas modificações específicas. O rMATS foi executado com dois grupos de amostras, que correspondem ao grupo controle negativo (indivíduos saudáveis sem câncer) contra o paciente que possui uma variante P/PP (grupo controle positivo) ou contra o paciente com VUS predita de causar *splice*. Utilizou-se parâmetros padrão adicionando *--NovelSS*, o qual permite a detecção de novos sítios de *splice* levando em consideração sítios de *splice* não anotados. Ao executar a ferramenta, notamos que caso pelo menos uma amostra dos controles (CTRLs) estivesse sem cobertura em alguma junção de *splice*, gerava o valor 0 para essa junção, dessa forma o programa não conseguia calcular de forma correta os valores de pvalue e FDR (do inglês, *False Discovery Rate*), anotando ambos os valores como 1,0. Por isso, realizamos 2 análises distintas com o rMATS: uma com todos os CTRLs internos e externos (n = 217) contra a amostra do paciente, e uma segunda análise apenas os controles internos (n = 19) contra o paciente (uma vez que a cobertura dos controles internos foi maior e mais homogênea do que dos externos). Para o LeafCutterMD, cada paciente foi avaliado contra todos os controles negativos de forma individual e parâmetros padrão. O FRASER2 foi executado dentro da ferramenta DROP, com todas as amostras em conjunto num grupo único para os 27 genes de estudo através da criação de um arquivo yaml, o qual foi adicionado no arquivo de configuração, e o único parâmetro alterado foi o *quantileForFiltering* para 0.01.

Para a filtragem dos eventos significativos, cada ferramenta foi processada de maneira diferente. No rMATS, foram analisados os arquivos tabulares com extensão ‘JC.txt’ (do inglês, *junction count*) que continham as análises de *splicing* apenas das leituras que estavam alinhadas nas junções éxon-éxon, referente aos eventos A3SS, A5SS, MXE, RI e SE. De maneira geral foi feita a união de todos os dataframes referente aos 5 eventos e aplicado os filtros com as

seguintes condições: pvalue e FDR $\leq 0,05$ e “IncLevelDifference” $\geq \pm 0,10$ (diferença de PSI), obtendo uma tabela final e única contemplando todos os 5 eventos de *splicing* estatisticamente significativos para todos os genes sequenciados para cada paciente. No LeafCutterMD utilizou-se o arquivo final “leafcutter_outlier_pVals.txt” filtrando a coluna do paciente para $\leq 0,05$, indicando qual evento de excisão de íntron era significativa para o paciente em relação aos controles. No Drop/FRASER2 já existe um filtro padrão para detecção de eventos aberrantes no arquivo de configuração e o resultado com os eventos significativos ficam no arquivo “results_per_junction.tsv”, o qual foi analisado neste estudo. A Tabela 2 apresenta um resumo das três ferramentas com os destaques para cada parte da análise de dados. Para cada ferramenta os comandos foram automatizados e os códigos estão disponíveis no repositório online (https://github.com/LarissaDS21/RNAcap_splice).

Tabela 2 - Resumo da análise de dados pelas três ferramentas de análise de *splice*

	rMATS	LeafCutterMD	FRASER2
Alinhamento STAR	Parâmetros padrão de alinhamento da ferramenta, não utilizado --twopassMode Basic	Utilizado --twopassMode Basic	Utilizado --twopassMode Basic
Parâmetros modificados ou acrescentados	Adicionado --NovelSS	Nenhum, todos parâmetros padrão	Modificado ‘quantileForFiltering’ para 0.01 e adição do arquivo yaml com 27 genes
Como foi rodado (casos e CTRLs)	CTRLs internos e externos X cada paciente; CTRLs internos X cada paciente.	Cada paciente com todos os CTRLs de forma individual.	Pacientes P/PP com todos os CTRLs e Pacientes VUS com todos os CTRLs.
Processamento e/ou arquivos com eventos significativos	União dos arquivos JC.txt dos 5 eventos de <i>splice</i> , com filtros: pvalue e FDR $\leq 0,05$ e “IncLevelDifference” $\geq \pm 0,10$	Arquivo “leafcutter_outlier_pVals.txt” com filtro $\leq 0,05$ na coluna do paciente.	Arquivo final já filtrado: “results_per_junction.tsv”. Sem nenhum processamento posterior.

Fonte: Elaboração dos autores.

4.5.3. Gráficos de Sashimi e Boxplot

Os eventos de *splicing* obtidos pela metodologia de RNA-cap e pela validação por sequenciamento de *amplicons* foram representados visualmente através de gráficos do tipo *sashimi plot*, gerados usando o programa *rmats2sashimipLOT* (<https://github.com/Xinglab/rmats2sashimipLOT>). E, para observar a diferença e distribuição dos valores de PSI entre o caso e controles, foram gerados gráficos de boxplot em R para o evento significativo de *splice* encontrado, uma vez que esse tipo de gráfico simplifica a visualização da dispersão dos valores de PSI dos controles em relação ao paciente estudado.

A ferramenta *rmats2sashimipLOT* foi utilizada com o evento de *splice* significativo, o qual foi filtrado e individualizado do arquivo JC.txt original gerado na ferramenta rMATS com dois controles negativos internos e dois externos para visualização e comparação dos eventos entre caso e controles. Para visualização dos valores de dispersão do PSI dos CTRLs contra o PSI do paciente foi gerado um gráfico de Boxplot com o pacote *ggplot2*, no qual foi gerado um ponto de destaque vermelho que representa o PSI do paciente através da função ‘*geom_point*’ e a uma legenda com descrição HGVS da sua respectiva variante através da função ‘*geom_label*’. As análises foram realizadas utilizando o software R versão 4.3.

4.6. Avaliações complementares do efeito de VUS em transcritos

A avaliação do efeito nos transcritos das variantes codificantes e não-codificantes selecionadas foi feita por duas abordagens: além da realização o sequenciamento de RNA-cap, como descrito em 4.4, para todas as variantes VUS também realizamos uma análise direcionada ao transcrito do gene afetado para validar o evento de *splice*, através de RT-PCR e sequenciamento NGS, e quando possível, foi realizada a avaliação de desbalanço alélico tanto nos dados de RNA-cap quanto de RT-PCR.

4.6.1. Análise direcionada de transcritos

As variantes genéticas VUS foram avaliadas através do sequenciamento NGS de *amplicon*. Essa análise também foi realizada para pacientes com variantes patogênicas complexas, RNAP-03 (*BRCA2*: c.156_157insAlu) e RNAP-02 (*PALB2* dup E11), e outros dois pacientes com variantes P/PP RNAP-101 (*MLH1*:c.588+5G>C) e RNAP-118 (*BRCA1*:c.213-11T>G) para validar seu efeito aberrante.

Para isso, o RNA extraído de amostras de tecido normal adjacente ou sangue periférico foi convertido em cDNA através de reação de transcriptase reversa (RT). Resumidamente, de 100-250 ng de RNA total foi convertido em cDNA utilizando oligoDT e a enzima SuperScriptIII (Invitrogen), conforme protocolo padrão do fabricante. Este cDNA foi utilizado para a reação de PCR (RT-PCR) utilizando iniciadores desenhados para amplificar os possíveis transcritos.

Para variantes intrônicas ou exônicas previstas de afetar *splice*, foram desenhados primers nos éxons flangeadores da alteração (2 éxons à montante e 2 éxons à jusante do éxon predito de ser alterado), a fim de gerar os fragmentos que possam conter o erro de *splicing*. Para desenhar os *primers*, foram utilizadas as ferramentas Frankiln by Genoox (<https://franklin.genoox.com/clinical-db/home>), UCSC Genome Browser (Santa Cruz, CA), Primer3Plus (versão 3.3.0, HHMI, EUA) e UCSC In Silico PCR (Santa Cruz, CA). Por essa razão, para pacientes com alterações no éxon 1 ou íntron 1, não foi possível realizar a análise de alteração de *splicing* por transcrito direcionado, uma vez que se ocorre algum evento de salto do éxon 1 de forma total ou parcial não haverá uma região existente a jusante para o primer se ancorar e amplificar a região de interesse. Para esses pacientes realizamos apenas a análise do desbalanço alélico, o qual é explicado com mais detalhes no tópico 4.6.2, quando possível.

Juntamente com as amostras de cada paciente foram realizadas reações de RT-PCR com RNAs de 5 indivíduos controle. Os produtos das RT-PCR de pacientes e controles foram visualizados em gel de agarose 2% corado com gel *sybr safe* (Invitrogen) em luz UV com transiluminador (Alpha Innotec) e avaliados no equipamento de eletroforese automatizada Agilent 2100 Bioanalyzer, para comparação do padrão de produtos amplificados.

Os produtos amplificados foram preparados para sequenciamento NGS de *amplicon*, utilizando o kit Ion Plus Fragment Library (Thermo Fisher, SP, Brasil) na construção de biblioteca e o sequenciamento foi realizado através da plataforma IonS5 (Thermo Fisher Scientific, MA, EUA), seguindo as instruções do fabricante. Os arquivos do sequenciamento foram alinhados no STAR conforme descrito no tópico 4.5.1, e os BAMs foram analisados utilizando as ferramentas IGV e/ou CLCbioMain Workbench (CLC Bio, Denmark). As junções éxon-éxon foram avaliadas com a função “sashimi plot” do IGV (Robinson *et al.*, 2011).

É importante destacar que os resultados referentes ao tamanho real dos *amplicons* observados num ensaio do Bioanalyzer podem variar um pouco de ensaio para ensaio, apresentando tamanhos aproximados. Nos resultados, o transcrito completo ou *amplicon*

esperado foi identificado pela sigla FL (do inglês, *full length*) com uma seta da cor azul e o transcrito anormal esperado pelo erro de *splicing* foi identificado com uma seta de cor vermelha. A nomenclatura do impacto de *splice* nos transcritos de RNA foi padronizada nas figuras, para que: I. o símbolo ‘ Δ ’ seguido de um número represente o ‘salto total’ do respectivo éxon; II. Para salto parcial, o ‘ Δ ’ é seguido pelo número do éxon, depois pela letra ‘p’ se a deleção for no sentido 5’ do éxon, ou seguido da letra ‘q’ se o sentido for 3’ e entre parênteses a quantidade de pares de base deletados; III. Para eventos de inserção/retenção de íntron utilizamos o símbolo ‘ ∇ ’ seguido do número do éxon adjacente e entre parênteses a quantidade de pares de base inseridos (Morak *et al.*, 2022).

4.6.2. Análise de desbalanço alélico

Aqueles pacientes com VUS que possuíam uma variante exônica heterozigota no mesmo gene de ocorrência da variante foram avaliados em relação ao desbalanço alélico. A análise de desbalanço alélico consiste em observar se houve diferença de expressão, através da frequência alélica observada no RNA entre os dois alelos do gene, utilizando como marcador algum polimorfismo em heterozigose encontrado em região codificante no mesmo gene que se encontra a VUS (SNP marcador). Para isso, nos pacientes foram avaliados os VCFs brutos do sequenciamento prévio de DNA no software VarSeq, filtrando o gene em questão e utilizando apenas os filtros relacionados a ‘qualidade da chamada de variantes’, como descrito em 4.2 (quadro 1) para a identificação de polimorfismos em heterozigose na região codificante. A visualização da região dos polimorfismos foi feita através do software IGV a partir dos BAMs gerados no próprio sequenciamento de RNA-cap e/ou com sequenciamento de RNA direcionado com *amplicon* para investigar se houve uma expressão diferencial entre os alelos do gene.

4.7. Reclassificação de variantes de *splice* e em regiões não codificantes

As classificações das variantes elegíveis para o estudo antes da análise de RNA foram feitas utilizando os critérios gerais da ACMG (Richards *et al.*, 2015) e atualizações feitas até 2021 (data de inclusão dos participantes no estudo) ou baseadas nos critérios informados nos laudos clínicos. Após a análise de RNA, além dos critérios anteriores, incorporamos na classificação das variantes as recomendações do artigo de Walker e colaboradores (2023) - criada pelo grupo de especialistas da *ClinGen SVI Splicing Subgroup* e que fornece recomendações para a alteração da aplicação de códigos ACMG/AMP existentes relacionados

ao *splicing* e a redefinição de outros códigos para capturar evidências relacionadas ao *splicing*. Ainda, utilizamos as diretrizes dos Registros de Especificação de Critérios da ClinGen (*Criteria Specification Registry - CSPEC*) para os genes que apresentam CSPECs publicados: *ATM*, *BRCA1*, *BRCA2*, *MLH1* e *TP53*, e a classificação ACMG baseada em pontos (Tavtigian *et al.*, 2020).

Já para as variantes nas regiões 5' e 3' UTR, nesse estudo não foi possível reclassificarmos essas variantes, pois outros experimentos funcionais seriam necessários para validar seu efeito no gene. Nesse estudo, a inclusão dessas variantes deu-se de forma exploratória apenas na etapa de filtragem de variantes.

4.8. Análises estatísticas

Para identificar as variantes de *splicing*, foi avaliada a diferença de PSI entre cada paciente (controles positivos e grupo validação clínica) e os controles saudáveis, utilizando um teste *t* bilateral de uma amostra. Na ferramenta rMATS um evento de *splicing* de um paciente foi considerado como significativo quando a diferença do seu PSI em relação à média do PSI dos doadores saudáveis for $\geq 10\%$, com valor de $p \leq 0,05$ e FDR $\leq 0,05$, e o PSI do evento no paciente caso for $>10\%$ (para eventos RI e A3SS/A5SS) ou $<90\%$ para eventos de SE. Para a ferramenta LeafCutterMD o evento de *splice* tinha que ter o valor significativo de $p \leq 0,05$, sendo também significativo em menos de 2 controles. Na ferramenta FRASER2 dentro do pipeline DROP a variante precisava ter passado pelo filtro de qualidade padrão da ferramenta: variação mínima do delta PSI de 0,05 e valor do delta PSI maior que 0,1 foram considerados *outliers* e cut-off de 0,1 para indicar o FDR máximo que um evento pode ter para ser considerado um outlier.

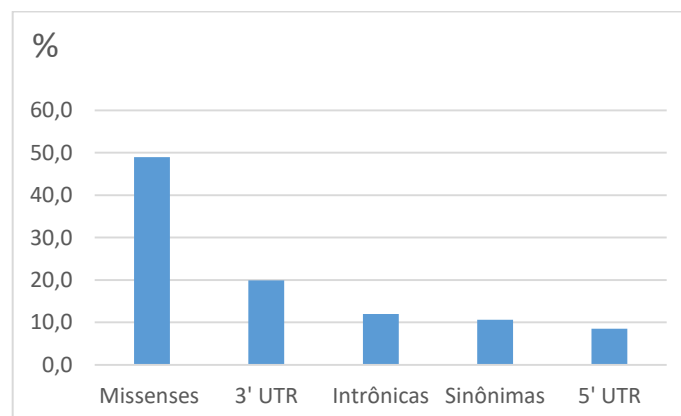
5. RESULTADOS

5.1. Levantamento de dados e filtragem de VCFs

Realizamos o levantamento dos dados de variantes germinativas de 1.026 pacientes dos projetos anteriores do grupo de pesquisa e 1.050 pacientes avaliados pelo Laboratório de Diagnóstico Genômico do A.C.Camargo Cancer Center (LDG), correspondendo a um total de 2.076 pacientes, que haviam sido submetidos a testes genéticos de DNA de 26 a 126 genes de predisposição hereditária ao câncer e possuíam o arquivo VCF bruto disponível.

A partir dos arquivos VCFs, foram utilizados filtros personalizados no software VarSeq para selecionar variantes germinativas raras **exônicas** (*missenses* ou sinônimas) ou **intrônicas** (± 20 pb limite éxon-intron, regiões 5' e 3' UTR). Obtivemos 1.950 pacientes (93,9%) que apresentaram pelo menos uma variante selecionada pelos filtros aplicados no VarSeq, correspondendo a 9.174 variantes totais e 4.518 variantes únicas, após retirar possíveis 200 variantes artefatuais. As variantes do tipo *missenses* foram as mais frequentes em relação ao total de variantes únicas com 2.212 (49,0%), seguida das variantes na 3'UTR que representam 898 (19,9%), depois as variantes intrônicas com 543 (12%), variantes sinônimas com 481 (10,6%) e por último as variantes na região 5'UTR com 384 (8,5%). Na Figura 16 pode-se observar um balanço geral das variantes selecionadas por tipo na nossa coorte de estudo.

Figura 16 - Frequência geral das variantes por tipo



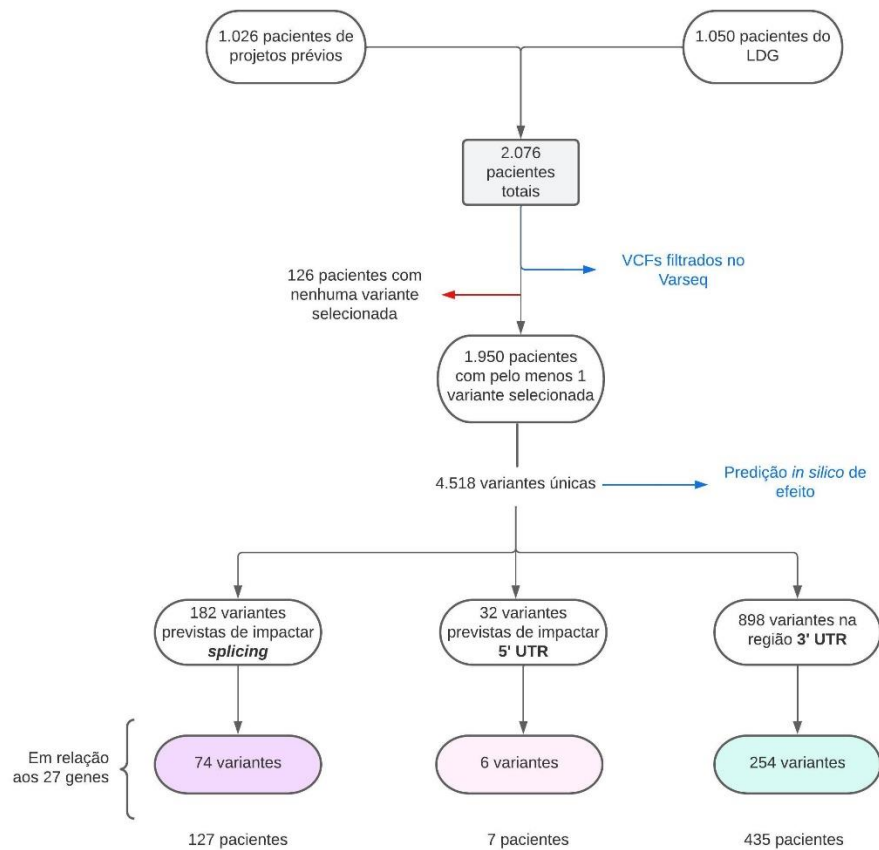
Fonte: autoria própria.

5.2. Seleção de variantes previstas de impactar o *splice* ou com impacto na região 5'UTR e as presentes na região 3'UTR

As variantes selecionadas na etapa anterior foram avaliadas quanto ao seu impacto no *splicing* utilizando os três preditores *in silico* disponíveis no VEP: MaxEntScan, dbSCSNV e SpliceAI. Já as variantes consideradas variantes de códon de início, variante de códon de início prematuro e na região 5'UTR foram previstas quanto ao seu impacto na região 5'UTR utilizando o preditor UTRannotator. Como não existe nenhum preditor estabelecido na clínica para avaliação das variantes na região 3'UTR para doenças genéticas mendelianas, todas as variantes selecionadas pelos filtros do VarSeq “3 prime UTR variant” foram incluídas no estudo sem nenhum tipo de priorização. Todas as variantes após seleção foram classificadas através da sua revisão no ClinVar e seguindo os critérios da ACMG.

Na Figura 17 está o resumo dos resultados obtidos a partir do processamento das variantes desde a filtragem dos VCFs brutos no VarSeq até a obtenção das variantes de *splice* e das regiões 5' e 3'UTR para o estudo dos 27 genes.

Figura 17 - Fluxograma representando os resultados obtidos em cada etapa do processamento das variantes



Fonte: autoria própria.

Legenda: LDG: Laboratório de Diagnóstico Genômico do A.C.Camargo Cancer Center.

5.2.1. Variantes previstas de impactar *splicing*

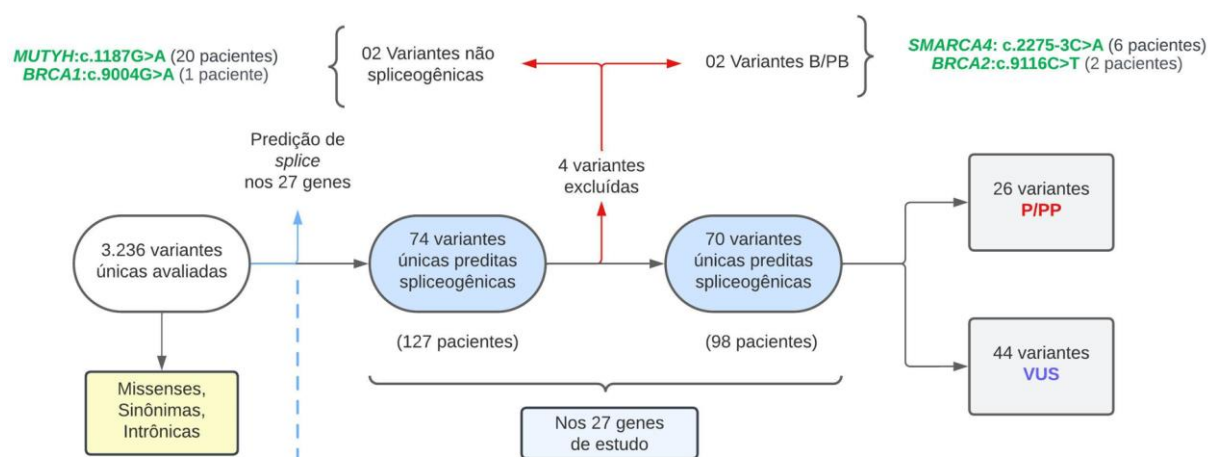
Das 3.236 variantes únicas *missenses*, sinônimas e intrônicas avaliadas *in silico* foram encontradas 182 variantes únicas preditas de afetar o *splice* (5,7%) em 297 pacientes, as quais apresentaram impacto em pelo menos um preditor e em todos os genes presentes nos VCFs brutos. Destas, 93 (51,6%) eram variantes intrônicas, 72 (39,5%) *missenses* e 17 (8,9%) sinônimas. Para os 27 genes de predisposição do painel do estudo foram encontradas 74 variantes previstas de impactar o *splice* em 127 pacientes, em pelo menos 1 preditor. A partir dessas variantes foi feita a classificação ACMG e notou-se que as variantes *missenses* classificadas como patogênicas *MUTYH*:c.1187G>A e *BRCA2*:c.9004G>A, apesar de apresentarem pontuação que indique impacto de *splicing*, são patogênicas devido a troca do aminoácido e não devido a sua spliceogenicidade (efeito em *splice*), por essa razão 20 e 1 pacientes que possuíam exclusivamente essas variantes, respectivamente, foram excluídos do

estudo. Além disso, variantes VUS que apesar de serem conflitantes segundo o ClinVar, a maior parte dos laboratórios já as classificam como benignas (B) ou provavelmente benignas (PB), também foram excluídas do estudo, *SMARCA4*: c.2275-3C>A e *BRCA2*:c.9116C>T, com 6 e 2 pacientes, respectivamente.

Dessa forma, para o estudo dos 27 genes restaram 70 variantes em 98 pacientes (um paciente possui tanto uma variante P como uma VUS). Destas 70 variantes, 26 variantes foram classificadas como P/PP em 32 pacientes (4 exônicas e 22 intrônicas), as quais foram incluídas no grupo Controle Positivo e serão discutidas detalhadamente no tópico 5.3.1. E 44 variantes foram classificadas como VUS em 67 pacientes (24 exônicas e 20 intrônicas), que foram incluídas no grupo com variantes VUS com possível efeito em *splice* que serão discutidas detalhadamente no tópico 5.3.2.

O esquema do processamento e refinamento das variantes preditas de impactar *splice* estão resumidas na Figura 18.

Figura 18 - Resumo do processamento das variantes preditas de impactar *splice* nos 27 genes



Fonte: autoria própria.

Legenda: Processamento envolveu a exclusão de variantes não spliceogênicas, e as classificadas como B/PB. Foram selecionadas aquelas previstas de serem spliceogênicas e classificadas como P/PP ou VUS. **B:** Benigna; **PB:** Provavelmente benigna; **P:** Patogênica; **PP:** Provavelmente patogênica; **VUS:** variante de significado incerto.

Das 26 variantes P/PP selecionadas, 24 tinham dados de literatura com estudos funcionais ou estudos de RNA (RNA alvo direcionado [RT-PCR], ou RNA de captura ou transcriptoma total, ensaio alélico específico com mRNA) confirmando efeito em *splice*, e apenas 2 variantes não tinham estudos publicados até o momento (*ATM*:c.5763-2A>T e *PMS2*: c.250+1G>C). Em relação às 44 VUS, 18 já tinham estudos funcionais ou de RNA depositado

no ClinVar, sendo que para 11 foram descritos algum efeito em *splice* e 7 não foi identificada nenhuma anormalidade, enquanto 26 variantes não possuem nenhum tipo de estudo descrito.

Na Tabela 3 está o resumo geral de todas as variantes classificadas como P/PP e na Tabela 4 está o resumo daquelas classificadas como VUS, as quais apresentaram o impacto em *splice* em pelo menos um dos três preditores. Nessas tabelas estão presentes algumas características como: gene, aminoácido alterado (p.), distância do éxon, número de pacientes com aquela variante e os respectivos tumores primários. Valores em vermelho indicam que aquele valor atingiu o ponto de corte para o respectivo preditor (diferença ≥ 3 para MaxEntScan, Ada e RF scores $>0,6$ para dbSCSNV e delta score $\geq 0,2$ no *SpliceAI*). As variantes com o fundo cinza claro na tabela foram sequenciadas com RNA-cap nesse estudo.

Tabela 3 - Variantes P/PP previstas de alterarem o *splicing* nos 27 genes de estudo

(continua)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred _DS _AG	Pred _DS _AL	Pred _DS _DG	Pre d_D S_ DL
01	ATM	c.2921+1G>A	na	1 nt	P	1	Câncer de mama	0,99	0,93	-1,9	6,29	8,18	0,00	0,00	0,00	0,91
02	ATM	c.8010+1delG	na	1 nt	P	1	Câncer de próstata	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,05	1,00
03	ATM	c.5763-2A>T	na	2 nt	P	1	Câncer de mama metastático	0,99	0,93	0,44	8,80	8,37	0,99	1,00	0,00	0,00
04	MSH2	c.942+3A>T	na	3 nt	P	1	Câncer colorretal	0,99	0,98	2,54	8,59	6,06	0,00	0,00	0,00	0,49
05	RAD51C	c.905-2_905-1delAG	na	2 nt	P	1	Câncer colorretal	(-)	(-)	(-)	(-)	(-)	0,00	0,91	0,00	0,00
06	MLH1	c.677G>A	R226Q	0 nt	P	1	Câncer gástrico	0,99	1,0	5,00	9,22	4,22	0,00	0,00	0,29	0,44
07	MLH1	c.116+5G>C	na	5 nt	P	1	Câncer colorretal	0,99	0,98	4,81	8,60	3,79	0,00	0,00	0,00	0,74

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **P:** Patogênica; **PP:** Provavelmente patogênica.

Tabela 3 - Variantes P/PP previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred _DS _AG	Pred _DS _AL	Pred _DS _DG	Pre d_D S_ DL
08	<i>MLH1</i>	c.1667+1G>A	na	1 nt	P	1	Câncer de mama e Câncer de colon	0,99	0,85	-1,56	6,62	8,18	0,00	0,00	0,02	0,98
09	<i>MUTYH</i>	c.933+3A>C	na	3 nt	P	1	Câncer colorretal	0,99	0,87	-4,51	6,31	10,82	0,00	0,00	0,00	0,96
10	<i>CHEK2</i>	c.846+4_846+7 del	na	7 nt	PP	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,00	0,98
11	<i>CHEK2</i>	c.793-1G>A	na	1 nt	P	1	Câncer de próstata	0,99	0,93	-2,59	6,16	8,75	0,92	0,99	0,00	0,0
12	<i>CHEK2</i>	c.1008+2T>G	na	2 nt	P	1	Câncer de mama	0,99	0,88	0,96	8,59	7,64	0,00	0,00	0,00	1,0
13	<i>BRCA1</i>	c.5074+2T>C	na	2 nt	P	4	Câncer de mama e Câncer de Ovário	0,99	0,6	0,28	7,48	7,76	0,00	0,03	0,00	0,74

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **P:** Patogênica; **PP:** Provavelmente patogênica.

Tabela 3 - Variantes P/PP previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred _DS _AG	Pred _DS _AL	Pred _DS _DG	Pre d_D S_ DL
14	<i>BRCA1</i>	c.211A>G	R71G	1 nt	P	4	Câncer de mama	0,99	0,99	3,89	7,84	3,96	0,00	0,00	0,00	0,64
15	<i>BRCA1</i>	c.213-11T>G	na	11 nt	PP	1	Câncer de mama	0,99	0,96	-1,50	4,84	6,34	0,20	0,71	0,00	0,00
16	<i>BRCA1</i>	c.441+2T>A	na	2 nt	P	1	Câncer de mama	0,99	0,47	-4,95	3,23	8,18	0,00	0,00	0,03	1,00
17	<i>BRCA1</i>	c.4484G>T	R1495M	0 nt	P	1	Câncer de ovário	0,99	0,99	5,55	10,57	5,01	0,00	0,00	0,01	0,60
18	<i>BRCA1</i>	c.5152+2T>C	na	2 nt	P	1	Câncer de mama	0,99	0,79	0,21	7,96	7,75	0,00	0,01	0,00	0,90
19	<i>BRCA1</i>	c.5153-2A>C	na	2 nt	P	1	Câncer de mama	0,99	0,93	0,73	8,78	8,04	0,21	0,99	0,02	0,00
20	<i>BRCA1</i>	c.547+2T>A	na	2 nt	P	1	Câncer de mama	0,99	0,90	0,90	9,08	8,18	0,00	0,00	0,00	0,99

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **P:** Patogênica; **PP:** Provavelmente patogênica.

Tabela 3 - Variantes P/PP previstas de alterarem o *splicing* nos 27 genes de estudo

(conclusão)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pre d_DS_DL
21	<i>CDH1</i>	c.1565+1G>T	na	1 nt	P	1	Câncer gástrico	0,99	0,93	2,23	10,34	8,50	0,00	0,00	0,73	1,00
22	<i>EPCAM</i>	c.556-14A>G	na	14 nt	PP	1	Câncer de mama	(-)	(-)	5,98	7,66	1,67	0,96	0,16	0,00	0,00
23	<i>BRCA2</i>	c.7976G>A	R2659 K	0 nt	P	2	Câncer de próstata e Câncer de mama	(-)	(-)	2,32	3,10	0,78	0,00	0,00	0,00	0,82
24	<i>BRCA2</i>	c.517-1G>A	na	1 nt	P	1	Câncer de mama	0,99	0,95	1,25	10,00	8,75	0,57	0,78	0,00	0,00
25	<i>BRCA2</i>	c.7806-2A>C	na	2 nt	P	1	Paraganglioma de retroperitônio	0,99	0,93	0,29	8,33	8,04	0,73	0,97	0,00	0,01
26	<i>PMS2</i>	c.250+1G>C	na	1 nt	P	1	Câncer colorretal	0,99	0,92	-0,20	8,07	8,27	0,00	0,00	0,28	1,00

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **P:** Patogênica; **PP:** Provavelmente patogênica.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(continua)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pred_DS_DL
01	<i>ATM</i>	c.6975G>A	A2325=	0	VUS	3	Câncer de pulmão (neuroendócrino) e Câncer de mama	0,99	0,98	-2,89	4,02	6,90	0,00	0,00	0,01	0,85
02	<i>ATM</i>	c.7273G>A	G2425S	34 nt	VUS	1	Câncer de próstata	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,14	0,69
03	<i>ATM</i>	c.2944C>T	R982C	22 nt	VUS	2	Câncer de mama e tireóide	(-)	(-)	(-)	(-)	(-)	0,30	0,05	0,00	0,00
04	<i>ATM</i>	c.4871A>G	H1624R	38 nt	VUS	4	Câncer de mama, ovário, e Câncer de ductos coletores	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,27	0,05
05	<i>ATM</i>	c.5005+19C>G	na	19 nt	VUS	1	Tumor neuroendócrino de intestino	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,45	0,03

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pred_DS_DL
06	<i>ATM</i>	c.7174C>T	R2392W	84 nt	VUS	2	Câncer colorretal e Leiomiossarcoma uterino	(-)	(-)	(-)	(-)	(-)	0,21	0,00	0,00	0,00
07	<i>CHEK2</i>	c.320-5T>A	na	5 nt	VUS/L B♦	5	Câncer colorretal, Câncer de tireoide, Câncer de mama	0,80	0,72	5,95	7,66	1,71	0,02	0,01	0,00	0,00
08	<i>CHEK2</i>	c.1376-13A>G	na	13 nt	VUS	1	Lipossarcoma	(-)	(-)	3,68	5,95	2,27	0,99	0,85	0,00	0,00
09	<i>CHEK2</i>	c.846+5G>C	na	5 nt	VUS	1	Câncer de mama	0,98	0,88	5,36	8,31	2,95	0,00	0,00	0,00	0,98
10	<i>CDH1</i>	c.833-3C>G	na	3 nt	VUS	1	Câncer gástrico	0,99	0,89	1,37	8,10	6,73	0,00	0,55	0,00	0,00
11	<i>MUTYH</i>	c.934-2A>G	na	2 nt	VUS	9	Câncer gástrico, colorretal, mama	0,99	0,78	-0,48	7,48	7,96	0,32	0,95	0,00	0,00

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pred_DS_DL
12	<i>MUTYH</i>	c.1518+3A>T	na	3 nt	VUS	1	Câncer colorretal	0,22	0,45	7,07	11,01	3,95	0,00	0,00	0,00	0,85
13	<i>MLH1</i>	c.790C>G	H264D	0 nt	VUS	1	Câncer de mama	0,73	0,42	11,81	10,43	-1,38	0,00	0,00	0,12	0,00
14	<i>MLH1</i>	c.359C>G	A120G	21 nt	VUS	2	Câncer de mama feminino e Câncer de mama masculino	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,21	0,01
15	<i>MLH1</i>	c.92C>G	A31G	24 nt	VUS	1	Câncer de mama masculino	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,40	0,03
16	<i>MLH1</i>	c.116+242A>G	na	242 nt	VUS	1	Informação indisponível	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,46	0,00
17	<i>RAD51D</i>	c.80C>A	T27K	2 nt	VUS	1	Câncer de mama	0,97	0,75	6.78	6,54	-0,24	0,00	0,00	0,00	0,00
18	<i>RAD51C</i>	c.571+4A>G	na	4 nt	VUS	1	Câncer de mama	0,19	0,49	8,099	10,45 ₁	2,352	0,00	0,00	0,24	0,38

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pred_DS_DL
19	<i>NBN</i>	c.995-16C>G	na	16 nt	VUS	1	Sem câncer, mas com HF	(-)	(-)	6,68	8,29	2,24	0,90	0,63	0,00	0,00
20	<i>NBN</i>	c.2071-4A>G	na	4 nt	VUS	1	Câncer colorretal	0,99	0,94	5,17	8,86	3,68	0,00	0,00	0,00	0,00
21	<i>NBN</i>	c.1454C>T	T485M	56 nt	VUS	1	Câncer de ovário	(-)	(-)	(-)	(-)	(-)	0,23	0,00	0,00	0,00
22	<i>BRCA1</i>	c.4986+13A>G	na	13 nt	VUS	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,87	0,03
23	<i>BRCA2</i>	c.8755-11A>G	na	11 nt	VUS	1	Câncer de mama + Câncer de papila duodenal	0,99	0,92	-0,88	3,78	4,66	0,99	0,92	0,00	0,00
24	<i>BRCA2</i>	c.9006A>T	E3002D	52 nt	VUS	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,35	0,00	0,00	0,00
25	<i>BRCA2</i>	c.9101A>G	Q3034R	16 nt	VUS	1	Câncer de ovário	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,24	0,00

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pred_DS_DL
26	<i>BRCA2</i>	c.9086C>T	A3029V	31 nt	VUS	2	Câncer de ovário e Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,02	0,30
27	<i>BARD1</i>	c.1569-3C>G	na	3 nt	VUS	1	Câncer de mama + Câncer de papila duodenal	0,99	0,75	-8,89	5,42	14,30	(-)	(-)	(-)	(-)
28	<i>BARD1</i>	c.1835A>T	D612V	24 nt	VUS	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,65	0,27	0,00	0,00
29	<i>BARD1</i>	c.556A>G	S186G	191 nt	VUS	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,93	0,08	0,00	0,00
30	<i>BARD1</i>	c.1808C>T	T603I	2 nt	VUS	1	Câncer de mama	0,97	0,77	7,65	10,07	2,42	0,00	0,00	0,00	0,18
31	<i>BRIP1</i>	c.93+4_93+7delAGTA	na	4 nt	VUS	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,00	0,94

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_ DS_ AG	Pred_ DS_ AL	Pred_ DS_ DG	Pred_ DS_ DL
32	<i>BRIP1</i>	c.2487T>C	L829=	5 nt	VUS	1	Sem câncer, mas tem histórico familiar	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,31	0,00
33	<i>CDK4</i>	c.218G>A	R73Q	0 nt	VUS	1	Tumor neuroendócrino de estômago	0,99	1,0	3,28	8,72	4,99	0,00	0,00	0,14	0,77
34	<i>PTEN</i>	c.1026+4A>T	na	4 nt	VUS	1	Câncer de ovário	0,97	0,71	6,42	8,67	2,53	0,00	0,00	0,00	0,00
35	<i>PTEN</i>	c.114T>G	P38=	34 nt	VUS	1	Câncer de ovário	(-)	(-)	(-)	(-)	(-)	0,32	0,00	0,08	0,00
36	<i>RET</i>	c.1759+10G>T	na	10 nt	VUS	2	Câncer gástrico, Câncer de mama	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,60	0,06

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(continuação)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pred_DS_DL
37	<i>RET</i>	c.2166G>T	K722N	12 nt	VUS	2	Tumor neuroendócrino de pâncreas e de estômago	(-)	(-)	(-)	(-)	(-)	0,50	0,03	0,00	0,00
38	<i>RET</i>	c.3183C>G	L1061=	4 nt	VUS	1	Câncer de ovário	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,57	0,27
39	<i>SMARCA4</i>	c.859+5G>A	na	5 nt	VUS	2	Câncer colorretal, Mioma e cistos ovarianos	0,99	0,66	5,54	9,60	4,06	0,00	0,00	0,01	0,19
40	<i>SMARCA4</i>	c.5007+5G>A	na	5 nt	VUS	1	Câncer colorretal	0,99	0,85	4,41	8,41	4,00	0,00	0,00	0,10	0,28
41	<i>STK11</i>	c.1108G>C	G370R	0 nt	VUS	1	Câncer de mama	0,99	0,98	8,40	10,90	2,49	0,00	0,00	0,01	0,01
42	<i>STK11</i>	c.1296G>A	Q432=	6 nt	VUS	1	Tumor neuroendócrino hepático	(-)	(-)	(-)	(-)	(-)	0,00	0,00	0,43	0,03

Fonte: Elaboração dos autores.

Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

Tabela 4 - Variantes de significado incerto previstas de alterarem o *splicing* nos 27 genes de estudo

(conclusão)

								Preditores de <i>Splicing</i>								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	N de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_ DS_ AG	Pred_ DS_ AL	Pred_ DS_ DG	Pred_ DS_ DL
43	TP53	c.114A>G	Q38=	17 nt	VUS	1	Câncer de endométrio	(-)	(-)	(-)	(-)	(-)	0,61	0,15	0,00	0,00
44	TP53	c.376-2dupA	na	0 nt	VUS	1	Sem câncer, mas com HF	(-)	(-)	(-)	(-)	(-)	0,29	0,02	0,00	0,00

Fonte: Elaboração dos autores.

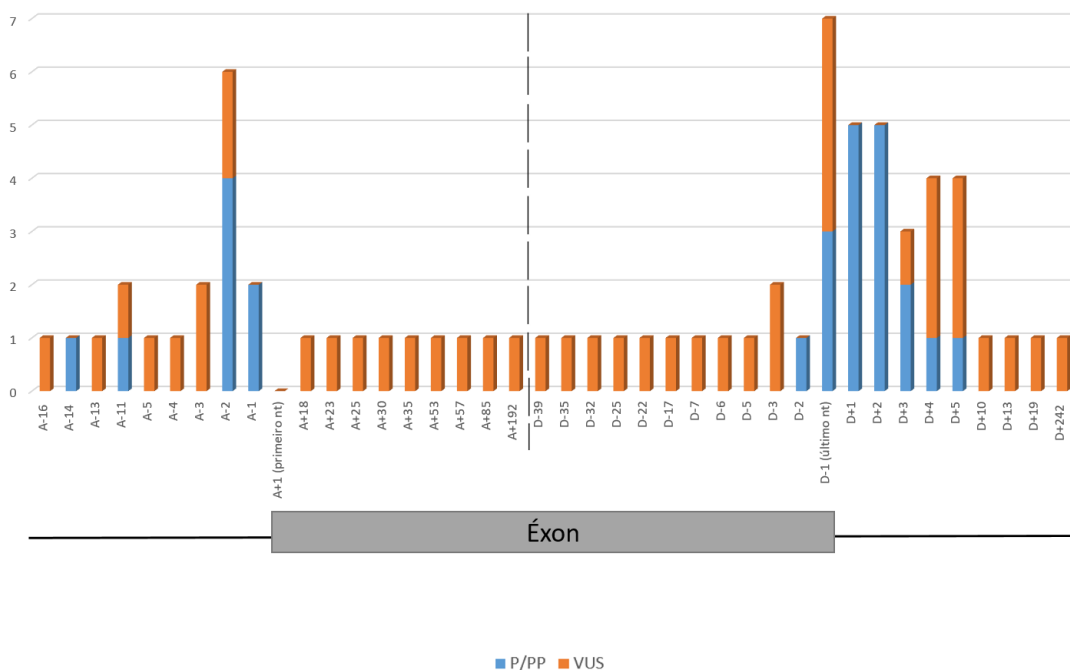
Legenda: Variantes em cinza foram sequenciadas. **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto; **HF:** histórico familiar; ♦ Variante em *CHEK2* entrou no estudo com conflitos de interpretações no ClinVar (VUS/LB), devido aos resultados incertos de análises de RNA.

5.2.1.1. Posição exônica ou intrônica e a frequência entre variantes P/PP versus VUS

Analisando de forma geral as posições das 26 variantes P/PP preditas de serem spliceogênicas nos 27 genes (4 eram exônicas e 22 intrônicas), ocorreu um enriquecimento de cerca de 16/26 variantes na região dos sítios doadores e aceptores canônicos de splice, $\pm 1,2$ pb do íntron (A-2, A-1 e D+1 e D+2), e todas as 4 variantes exônicas estavam entre o penúltimo (D-2) e último nucleotídeo do éxon (D-1). Das 22 variantes intrônicas, 4 estavam mais próximas do doador canônico: 2 variantes estavam na posição D+3, 1 na posição D+4 e 1 na posição D+5. E apenas 2/22 variantes intrônicas estavam mais próximas do sítio acceptor canônico de *splicing* nas posições A-11 e A-14.

Em relação às 44 VUS que geraram pontuação indicativa de erro de *splicing* nos 27 genes (24 eram exônicas e 20 intrônicas), das variantes exônicas ocorreu um enriquecimento de 4/24 variantes no último nucleotídeo do éxon (D-1) e de 2/24 variantes no antepenúltimo nucleotídeo do éxon (D-3), já as demais variantes 18/24 tinham apenas 1 exemplar de cada com distância de 4 até 191 pb dentro do éxon. Das variantes intrônicas, apenas 2/20 estão na região canônica de *splice* ambas na posição A-2 e 18/20 intrônicas com distância de ± 3 a ± 19 pb do éxon (foi capturada uma profunda de 242 pb de distância do íntron). Das 18 variantes intrônicas fora da região canônica de *splice*, 11 estavam próximas do sítio doador de *splice*: 3 variantes na posição D+4, 3 na posição D+5, uma na posição D+3, D+10, D+13, D+19 e D+242. Apenas 7/18 estavam próximas da região acceptora: 2 variantes na posição A-3, e apenas uma nas posições A-4, A-5, A-11, A-13 e A-16. A variante *MLH1*:c.116+242A>G está na posição +242 pb do éxon, a qual foi anotada com base no transcrito MANE, porém é considerada uma variante exônica no transcrito alternativo não MANE (ENST00000441265.6) à 79 pb dentro do éxon 1, por essa questão ela foi selecionada pelos filtros iniciais aplicados no VarSeq (selecionada por ser exônica *missense*). Todos esses resultados podem ser visualizados no Figura 19 abaixo.

Em resumo, a maior parte das variantes P/PP selecionadas para esse estudo estão na região canônica de *splice* (sítio acceptor e doador), já a maioria das variantes VUS estão fora da região canônica de *splice*, na região intrônica a partir da posição ± 3 pb ou mais para dentro do éxon de A+18 (18 nt de distância da posição acceptora) a D-3 (3 nt de distância da posição doadora).

Figura 19 - Posição das variantes preditas de serem spliceogênicas

Fonte: autoria própria

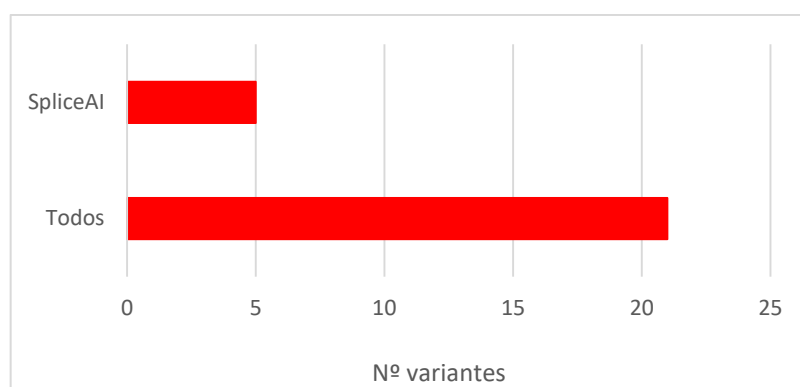
5.2.1.2. Comparação entre os três preditores *in silico*

Em relação aos resultados obtidos dos preditores *in silico* foi possível notar que os preditores dbScSNP e MaxEntScan conseguiram gerar escores para as regiões canônicas de *splicing*, considerando como sítio doador as posições (-3 nt do éxon e +8 nt do íntron), e (-3 nt do éxon e +6 nt do íntron), respectivamente, e como sítio aceptor (-12 nt do íntron e +2 do éxon), e (-20 nt do íntron e +3 do éxon), respectivamente. Pequenas deleções e inserções não geraram pontuações em ambos preditores, apenas no SpliceAI fazendo com que entrasse para o estudo dos 27 genes três variantes P/PP (*RAD51C*:c.905-2_905-1delAG, *ATM*:c.8010+1delG, *CHEK2*:c.846+4_846+7delAGTA) e duas VUS (*BRIP1*:c.93+4_93+7delAGTA; *TP53*:c.376-2dupA). Como avaliamos as variantes com distância máxima de 20 nt do éxon, os preditores dbScSNV e MaxEntScan teriam a capacidade de avaliar o impacto no *splice* na maior parte das variantes intrônicas, porém as variantes exônicas mais profundas poderiam não gerar pontuações nesses preditores.

Das 26 variantes únicas P/PP, 3/4 variantes que eram exônicas e estavam no último ou penúltimo nucleotídeo do éxon geraram pontuação nos três preditores e uma variante (*BRCA2*:c.7976G>A) gerou pontuação apenas no SpliceAI (DS_DL 0,82). Em relação às

variantes intrônicas, as 16 que estavam na região canônica de *splice* ($\pm 1,2$ pb do éxon) todas tiveram predição indicativa de erro de *splicing* nas 3 ferramentas, exceto 2 que eram *indels*, *ATM:c.8010+1delG* e *RAD51C:c.905-2_905-1delAG*, tiveram predição apenas pelo SpliceAI. Das seis variantes intrônicas fora da região canônica de *splice*, 4/6 foram preditas pelas 3 ferramentas, enquanto apenas duas tiveram escore de impacto apenas pela ferramenta SpliceAI (*CHEK2:c.846+4_846+7delAGTA* e *EPCAM:c.556-14A>G*). De forma geral, cerca de 21/26 variantes tinham escores indicativo de erro de *splice* nas 3 ferramentas e apenas 5/26 variantes com predição em apenas 1 preditor, que foi no SpliceAI (4 intrônicas e 1 exônica). Na Figura 20 está um resumo da quantidade de variantes que foram preditas de impacto no *splicing* e seus respectivos preditores.

Figura 20 - Quantidade de variantes P/PP por preditores *in silico* de *splice*

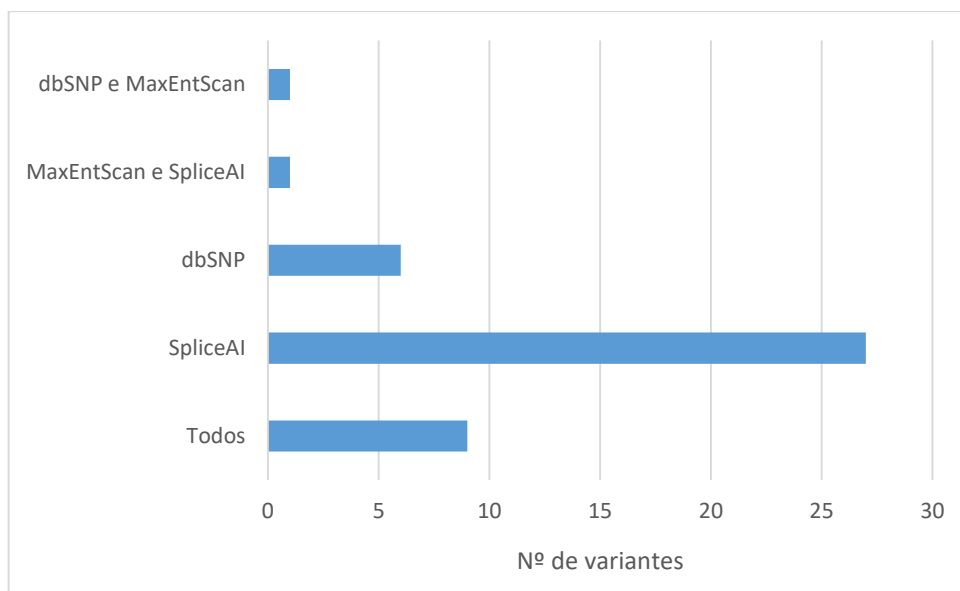


Fonte: autoria própria.

Das 44 variantes VUS, 2/24 variantes exônicas têm predição indicativa de erro de *splice* nas 3 ferramentas (*ATM:c.6975G>A* e *CDK4:c.218G>A*) ambas localizadas no último nucleotídeo do éxon, e todas as outras 22/24 exônicas têm predição em apenas uma ferramenta: 4 variantes apenas no dbSNV que estavam localizadas entre os três últimos nucleotídeos do éxon, e 18 variantes apenas no SpliceAI com quatro nucleotídeos de distância ou mais para dentro do éxon. Em relação as variantes intrônicas, apenas duas estavam na região canônica de *splice* A-2 pb, *MUTYH:c.934-2A>G* foi previsto *splice* nas três ferramentas, e a *TP53:c.376-2dupA* por ser uma *indel* foi prevista apenas no SpliceAI. Cerca de 18/44 variantes intrônicas que estão fora da região canônica, 6/18 foram previstas de impacto no *splice* pelas 3 ferramentas, 2/18 em 2 ferramentas, em que *MUTYH:c.1518+3A>T* foi prevista pelos preditores MaxEntScan e SpliceAI e *SMARCA4:c.859+5G>A* no dbSNV e MaxEntScan, e 10/18 variantes previstas em apenas uma ferramenta, em que 8/10 variantes foram pelo

SpliceAI e 2/10 foram pelo dbSNV. Na Figura 21 está um resumo da quantidade de variantes que foram preditas de impacto no *splicing* e seus respectivos preditores.

Figura 21 - Quantidade de variantes VUS por preditores *in silico* de *splice*



Fonte: autoria própria.

De modo geral foi possível perceber que SpliceAI foi o único preditor que conseguiu gerar pontuações para VUS intrônicas com ± 10 pb (exceto *BRCA2*:c.8755-11A>G), exônicas com distância de 5 até 191 pb para dentro do éxon e que sozinho, no geral, detectou mais variantes com potencial de *splice* em variantes tanto P/PP quanto VUS. Outro ponto a se destacar é que a maior parte das variantes P/PP é vista de impactar o *splice* nos 3 preditores (73,1%), enquanto a maior parte das variantes VUS (75%) é prevista de alterar *splice* em apenas 1 preditor.

5.2.2. Variantes com impacto na região 5'UTR e as presentes na região 3'UTR

Das 450 variantes presentes na região 5'UTR encontrou-se 32 variantes (7,1%) em 46 pacientes que foram previstas de impactar a região 5'UTR pelo preditor UTRannotator em todos os genes. As variantes *PTCH1* (NM_000264.5):c.202-568G>A e *BRF1* (NM_001519.4):c.544+3917_544+3918insGG presentes em 5 e 8 pacientes, respectivamente, foram excluídas por serem muito frequentes em nossa coorte (alta frequência populacional é indicativo de que a variante tem baixo risco de causar a doença), restando então 33 pacientes com 26 variantes únicas. Após a classificação ACMG e revisão de seus dados no ClinVar notou-

se que 9 variantes únicas foram classificadas como B/PB, 16 como VUS e apenas 1 foi classificada como P por causar a perda no códon de início no gene *BRCA2* que tem alta penetrância. Das 26 variantes apenas 7 já foram descritas e 19 ainda não estavam descritas no ClinVar. Para os 27 genes de estudo foram encontrados apenas 7 pacientes com 6 variantes. Na Tabela 5 estão todas as variantes em que o impacto foi predito na região 5'UTR em todos os genes, e os transcritos MANE (cl clinicamente relevante) estão em negrito. Nessa tabela ainda estão apresentados os resultados das consequências de impacto gerados pelo preditor UTRannotator, sendo identificado o ‘ganho de um códon de início a montante (uAUG *gained*)’ como a consequência mais frequente.

Para a região 3'UTR foram encontradas muitas variantes, cerca de 898 variantes em todos os genes em 999 pacientes, para os 27 genes do estudo foram encontradas 254 variantes em 435 pacientes.

Tabela 5 - Previsão de impacto em variantes na região 5'UTR em todos os genes

(continua)

	Gene	HGVS c.	p.	Dist. AUG♦	ACMG	ClinVar	Pac.	Tipo de tumor	Consequência 5'UTR	Transcrito de referência
1	<i>BAP1</i>	NM_004656.4:c.-28T>G	(-)	27 nt	VUS	(-)	1	Osteossarcoma osteoblástico	uAUG_lost	NM_004656.3
2	<i>BRCA2*</i>	NM_000059.4:c.2T>G	Met1?	0 nt	P	P	2	Câncer de mama e ovário; Mama masculina	uAUG_lost	ENST00000530893.2
3	<i>BRF1</i>	NM_001519.4:c.544+3930_544+3931insTCGC	(-)	Íntron 5	B	(-)	1	Tumor de reto alto	uFrameshift	ENST00000550208.1
4	<i>BRF1</i>	NM_001519.4:c.544+3920 dupC	(-)	Íntron 5	VUS	(-)	1	Adenocarcinoma de colon	uFrameshift	ENST00000550208.1
5	<i>CEP57</i>	NM_014679.5:c.-162G>A	(-)	162 nt	VUS	(-)	1	Câncer de ovário	uAUG_gained	NM_001243777.1; NM_014679.4
6	<i>CEP57</i>	NM_014679.5:c.45G>A	Ser15=	44nt (éxon 1)	VUS	(-)	1	Câncer de mama	uAUG_gained	ENST00000544522.1
7	<i>CHEK2*</i>	NM_007194.4:c.538C>T	Arg180 Cys	Éxon 4	VUS	VUS	1	Câncer de ovário	uAUG_gained	ENST00000425190.2 e ENST00000544772.1 (NM_001257387.1)
8	<i>CHEK2*</i>	NM_007194.4:c.-10C>T	(-)	7.053 nt	VUS	VUS	1	Informação indisponível	uAUG_gained	NM_007194.4; NM_145862.2; NM_001257387.1
9	<i>CTNNA1</i>	NM_001903.5:c.1063-10497T>C	(-)	Íntron 7	VUS	(-)	1	Tumor neuroendócrino grau 1	uSTOP_lost	ENST00000520260.1, ENST00000522013.1, ENST00000540387.1

Fonte: Elaboração dos autores.

Legenda: **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. AUG:** Distância do códon de início; **Pac.:** pacientes; **uAUG_lost:** perda do códon de início a montante; **uFrameshift:** a variante causa uma alteração *frameshift* na janela de leitura da região regulatória a montante; **uAUG_gained:** ganho de um códon de início a montante; **uSTOP_lost:** perda do códon de parada a montante; **uSTOP_gained&uAUG_gained:** dois eventos concomitantes, em que ocorreu um ganho do códon de parada e do códon de início; Em negrito na coluna “Efeito no transcrito de referência” estão os transcritos MANE; * variantes presentes nos 27 genes de estudo; ♦ Distância do AUG canônico do transcrito MANE.

Tabela 5 - Previsão de impacto em variantes na região 5'UTR em todos os genes

(continuação)

	Gene	HGVS c.	p.	Dist. AUG♦	ACMG	ClinVar	Pac.	Tipo de tumor	Consequência 5'UTR	Transcrito de referência
10	<i>MSH3</i>	NM_002439.5:c.-12C>G	(-)	11 nt	PB	(-)	2	Câncer de mama; Neoplasia maligna do tecido conjuntivo e de outros tecidos moles	uAUG_gained	NM_002439.5
11	<i>ERCC2</i>	NM_000400.4:c.6-146_6-145delCT	(-)	162 nt	VUS	(-)	1	Informação indisponível	uSTOP_lost	ENST00000391941.2, ENST00000586856.1
12	<i>EXT2</i>	NM_207122.2:c.-31+403G>A	Met1?	0 nt	VUS	(-)	1	Informação indisponível	uAUG_lost	ENST00000343631.3
13	<i>FANCL</i>	NM_001114636.1:c.-34T>G	(-)	33 nt	PB	(-)	2	Câncer de ovário e Câncer de mama	uSTOP_lost	ENST00000449070.1
14	<i>KLLN</i>	NM_001126049.2:c.-523delC	(-)	518 nt	VUS	(-)	1	Câncer renal e de parótida	uFrameshift	NM_001126049.2
15	<i>KLLN</i>	NM_001126049.2:c.-492_-491delCT	(-)	487 nt	VUS	(-)	1	Câncer de testículo	uFrameshift	NM_001126049.2
16	<i>LZTR1</i>	NM_006767.4:c.-38T>A	(-)	38 nt	VUS	Conflitling	1	Câncer de mama	uAUG_gained	NM_006767.4
17	<i>MSH6*</i>	NM_000179.3:c.533G>A	Arg178 His	Éxon 3	PB	VUS	1	Câncer renal do tipo células claras	uAUG_gained	NM_001281494.1
18	<i>PDGFRA</i>	NM_006206.6:c.-13+10626C>T	(-)	11 nt	VUS	(-)	2	Tumor de Wilms	uAUG_gained	ENST00000512143.1
19	<i>PTCH1</i>	NM_000264.5:c.201+155G>A	(-)	353 nt	PB	(-)	2	Leiomiossarcoma, Tumor neuroendócrino de baixo grau	uAUG_lost	NM_001083605.1

Fonte: Elaboração dos autores.

Legenda: **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. AUG:** Distância do códon de início; **Pac.:** pacientes; **uAUG_lost:** perda do códon de início a montante; **uFrameshift:** a variante causa uma alteração *frameshift* na janela de leitura da região regulatória a montante; **uAUG_gained:** ganho de um códon de início a montante; **uSTOP_lost:** perda do códon de parada a montante; **uSTOP_gained&uAUG_gained:** dois eventos concomitantes, em que ocorreu um ganho do códon de parada e do códon de início; Em negrito na coluna “Efeito no transcrito de referência” estão os transcritos MANE; * variantes presentes nos 27 genes de estudo; ♦ Distância do AUG canônico do transcrito MANE.

Tabela 5 - Previsão de impacto em variantes na região 5'UTR em todos os genes

(conclusão)

	Gene	HGVS c.	p.	Dist. AUG♦	ACMG	ClinVar	Pac.	Tipo de tumor	Consequência 5'UTR	Transcrito de referência
20	<i>PTCH1</i>	NM_000264.5:c.202-425T>G	(-)	1335 nt	PB	(-)	2	Sarcoma de células dendríticas foliculares de baixo grau, Câncer de mama	uSTOP_lost	NM_001083607.1
21	<i>PTCH1</i>	NM_000264.5:c.202-469T>A	(-)	1291 nt	B	(-)	1	Lipossarcoma mixóide de alto grau	uAUG_lost	NM_001083607.1
22	<i>RAD50</i>	NM_005732.4:c.-82C>T	(-)	80 nt	PB	(-)	2	Câncer de mama (TN), Câncer de mama	uAUG_gained	NM_005732.3
23	<i>RAD51C*</i>	NM_058216.3:c.3G>A	Met1?	0 nt	VUS	VUS	1	Adenocarcinoma de endométrio	uAUG_lost	ENST00000461271.1
24	<i>RUNX1</i>	NM_001754.5:c.98-1229G>A	(-)	Íntron 3	B	(-)	1	Carcinoma papilífero de tireoide	uAUG_gained	NM_001001890.2;NM_001122607.1
25	<i>SDHD</i>	NM_003002.4:c.-4C>T	(-)	3 nt	VUS	VUS	1	Rabdomiossarcoma embrionário de alto grau metastático	uSTOP_gained & uAUG_gained	NM_003002.3 ; NM_001276504.1; NM_001276506.1; NM_001276503.1
26	<i>STK11*</i>	NM_000455.5:c.-363G>C	(-)	360 nt	VUS	(-)	1	Câncer de mama	uSTOP_lost	NM_000455

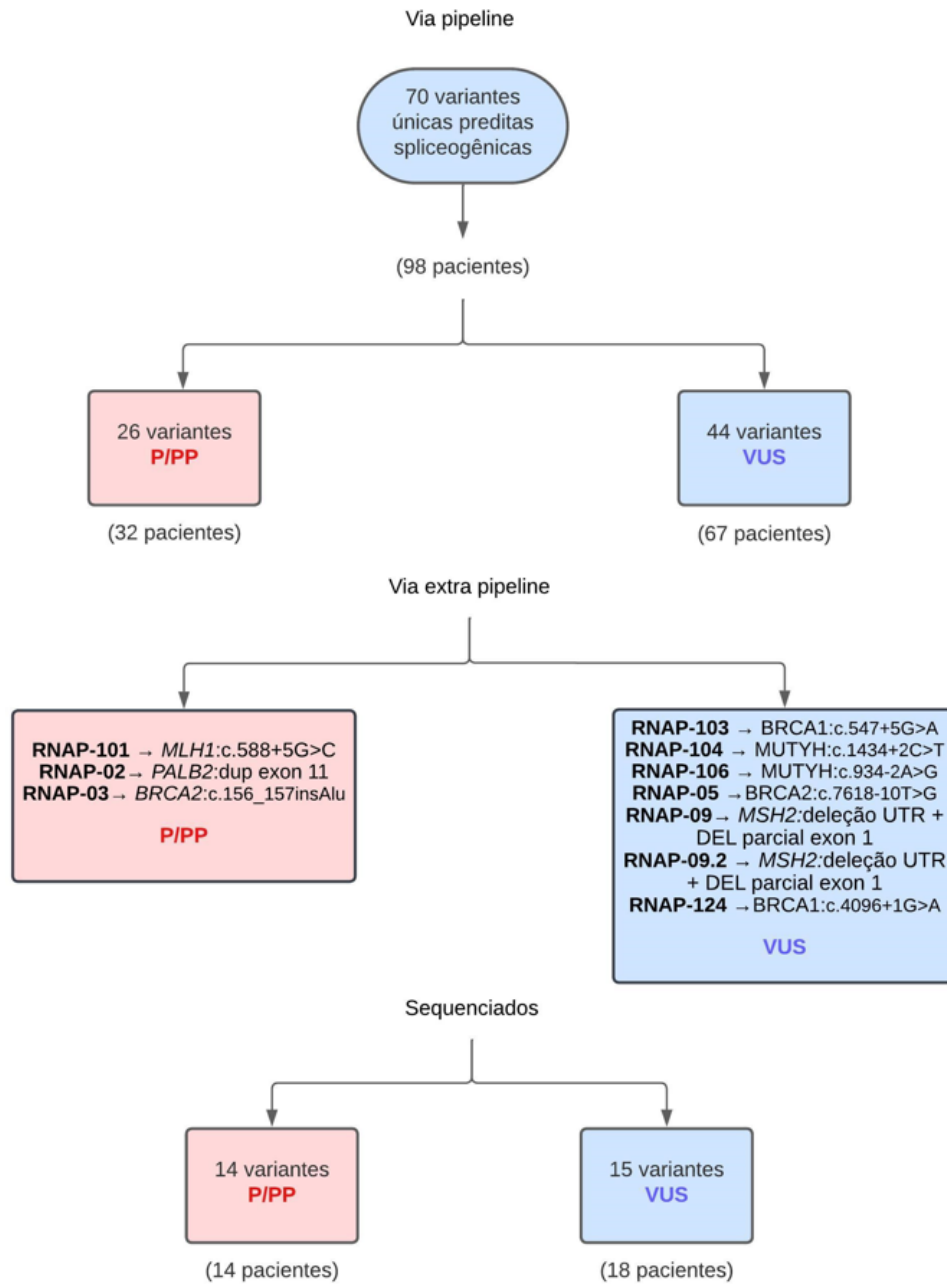
Fonte: Elaboração dos autores.

Legenda: **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. AUG:** Distância do códon de início; **Pac.:** pacientes; **uAUG_lost:** perda do códon de início a montante; **uFrameshift:** a variante causa uma alteração *frameshift* na janela de leitura da região regulatória a montante; **uAUG_gained:** ganho de um códon de início a montante; **uSTOP_lost:** perda do códon de parada a montante; **uSTOP_gained&uAUG_gained:** dois eventos concomitantes, em que ocorreu um ganho do códon de parada e do códon de início; Em negrito na coluna “Efeito no transcrito de referência” estão os transcritos MANE; * variantes presentes nos 27 genes de estudo; ♦ Distância do AUG canônico do transcrito MANE.

5.3. Seleção de pacientes em seus respectivos grupos para validação do RNA-cap

Um total de 98 pacientes foram considerados elegíveis para o estudo, pois apresentaram variantes classificadas como P/PP (já conhecidas de serem spliceogênicas) ou VUS preditas de causar dano em *splice* por pelo menos 1 preditor dentro dos 27 genes do estudo, os quais compuseram os grupos controle positivo e grupo com VUS. Além disso, incluímos mais 9 pacientes extra-pipeline (3 controles positivos e 6 com VUS). Três pacientes controles positivos já tinham sido identificados em estudos anteriores do grupo/testes de rotina, logo suas variantes já apresentavam resultados prévios indicando serem spliceogênicas (RNAP-101, RNAP-02, RNAP-03). Os seis pacientes (e um familiar) com VUS incluídos extra-pipeline foram identificados, em testes de rotina do LDG e laboratório externo, como sendo interessantes para o estudo de *splice* (RNAP-103 RNAP-104, RNAP-106, RNAP-05, RNAP-09, RNAP-09.2 e RNAP-124), os quais serão discutidos com mais detalhes a seguir e estão resumidos na Figura 22 e Tabela 6.

Figura 22 - Descrição da quantidade de pacientes e variantes que entraram via pipeline, extra-pipeline e os que foram sequenciados



Fonte: autoria própria.

Tabela 6 - Variantes adicionadas extra-pipeline

								Preditores de Splicing								
								dbscSNV		MaxEntScan			SpliceAI			
	Gene	HGVS c.	p.	Dist. éxon	ACMG	Nº de Pac.	Tipos de tumor	Ada Score	Rf Score	Alt.	Ref.	Dif.	Pred_DS_AG	Pred_DS_AL	Pred_DS_DG	Pre d_DS_DL
1	MLH1	c.588+5G>C	na	5 nt	PP	1	Câncer colorretal	0,99	0,89	4,33	10	5,39	0,00	0,00	0,00	0,89
2	PALB2	NC_000016.9:g.(?_23625315)_(23625422_?)dup, referenciado como dup E11	na	na	P	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
3	BRCA2	c.156_157insAlu	na	87 nt	P	1	Câncer de mama	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
4	BRCA1	c.547+5G>A	na	5 nt	VUS	1	Câncer de mama	0,99	0,84	4,20	9,08	4,89	0,00	0,00	0,10	0,25
5	MUTYH	c.1434+2C>T	na	2 nt	VUS	1	Adenocarcinoma gástrico	(-)	(-)	7,52	0,23	-	0,00	0,00	0,00	0,00
6	MUTYH	c.934-2A>G	na	2 nt	VUS	10	Adenocarcinoma de Sigmóide	0,99	0,78	-	7,48	7,95	0,49	0,95	0,00	0,00
7	BRCA2	c.7618-10T>G	na	10 nt	VUS	1	Câncer de mama	0,99	0,98	0,88	7,11	6,23	0,96	0,93	0,00	0,00
8	MSH2	c.Del UTR + Del parcial E1	na	na	VUS	2	Câncer colorretal	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
9	BRCA1	c.4096+1G>A	na	1 nt	VUS	1	Câncer de mama	0,99	0,98	-	5,64	8,18	0,00	0,00	0,01	0,24

Fonte: Elaboração dos autores.

Legenda: **HGVS c.:** nomenclatura de variantes segundo o *Human Genome Variation Society*; **p.:** proteína; **ACMG:** classificação de variantes segundo os critérios do *American College of Medical Genetics and Genomics* (ACMG); **nt:** nucleotídeos; **na:** Não se aplica; **Dist. éxon:** Distância do éxon; **Pac.:** pacientes; **Alt.:** alteração; **Ref.:** referência; **Dif.:** diferença; **Pred_DS_AG:** delta score predito de ganho de sítio aceptor; **Pred_DS_AL:** score predito de perda de sítio aceptor; **Pred_DS_DG:** score predito de ganho de sítio doador; **Pred_DS_DL:** score predito de perda de sítio doador; **VUS:** Variante de significado incerto.

5.3.1. Seleção de pacientes para o grupo controle positivo

Foram identificados 32 pacientes que possuíam 26 variantes P/PP spliceogênicas únicas (Figura 18). Além desses pacientes, foram adicionados mais dois que possuíam variantes P não passíveis de serem avaliadas pelos preditores de *splice* (RNAP-02 e RNAP-03), por serem alterações de grandes alterações/inserções. Os preditores utilizados têm como limitação avaliar com maior rigor variantes de substituição de base única que estão próximas da região canônica de *splicing*, logo variantes com grandes inserções/deleções não geram nenhum escore, pois a ferramenta não é capaz de analisar seu devido impacto. A paciente RNAP-02 apresenta uma variante de duplicação do éxon 11 no gene *PALB2* e uma variante *missense* no mesmo gene (c.3350G>A, p.Arg117Lys) localizada no último nucleotídeo do éxon 12, classificada como P pelo ClinGen expert group por seu efeito em *splicing*, que leva a um salto do E12 *out-of-frame* (apesar de não ter sido predita de causar danos no *splicing* em nenhum preditor utilizado nesse estudo). A paciente RNAP-03 apresenta uma variante de inserção de um elemento transponível Alu no gene *BRCA2*, e tinha entrado para o estudo por conter a variante VUS *RAD51C*:c.571+4A>G.

Ainda foi adicionado o paciente RNAP-101, com a variante patogênica *MLH1* c.588+5G>C, a qual já foi validada por técnicas de NGS e Sanger como tendo impacto em *splicing* causando o salto total do éxon 7, e por isso essa amostra entrou diretamente em nosso controle positivo, dessa forma alcançou-se 35 pacientes ao total.

Dos 35 indivíduos elegíveis para essa etapa foi dada preferência para 11 que já possuíam material disponível em biobanco (RNAP-101, RNAP-107, RNAP-111, RNAP-112, RNAP-113, RNAP-116, RNAP-117, RNAP-118, RNAP-120, RNAP-121 e RNAP-122). Três pacientes que não possuíam tecido disponível no biobanco foram recrutados prospectivamente (RNAP-01, RNAP-02 e RNAP-03). Todos estes 14 pacientes entraram para etapa de validação técnica no sequenciamento de RNA alvo, os quais estão descritos na Tabela 7.

5.3.2. Seleção de pacientes para o grupo com VUS possíveis modificadoras de *splice*

Foram identificados 67 pacientes que possuíam 44 variantes VUS previstas de impactar *splice* (Figura 18). A esse grupo adicionou-se mais sete pacientes, com 6 VUS únicas: os pacientes RNAP-103, RNAP-104 e RNAP-106, os quais já tinham sido identificados com variantes preditas spliceogênicas em estudos anteriores, e os pacientes RNAP-05, RNAP-124, RNAP-09 e RNAP-09.2 entraram de forma prospectiva através dos exames de rotina do LDG

ou de laboratório externo. O paciente RNAP-09 é o probando e sua mãe é a paciente RNAP-09.2, os quais possuem a mesma variante complexa que causa a deleção da região 5' UTR e uma deleção parcial do éxon 1 no gene *MSH2*, porém com impacto no gene ainda indeterminado. Ao final, alcançamos um total de 74 pacientes elegíveis com 49 VUS únicas.

Desse grupo foram recrutados um total de 18 pacientes para validação clínica no sequenciamento de RNA alvo, em que 11 já tinham material disponível em biobanco (RNAP-103, RNAP-104, RNAP-105, RNAP-106, RNAP-108, RNAP-109, RNAP-110, RNAP-114, RNAP-115, RNAP-123 e RNAP-124) e 7 foram recrutados prospectivamente (RNAP-03, RNAP-04, RNAP-05, RNAP-06, RNAP-07, RNAP-09 e RNAP-09.2). Estes pacientes estão descritos na Tabela 7.

5.4. Sequenciamento alvo de RNA (RNA-cap)

Realizou-se o sequenciamento de RNA-cap com 50 pacientes (14 com variantes P/PP, 18 com VUS – sendo que um paciente apresentou uma variante P e uma VUS - e 19 controles). A Tabela 7 contém a relação dos pacientes e seus dados clínicos e moleculares (sexo, idade ao diagnóstico, suas respectivas variantes e qual tecido foi avaliado).

Tabela 7 - Pacientes que foram sequenciados com suas respectivas variantes e informações pessoais

(continua)

Paciente	Tumor	Sexo	Idade no dx	Outros tumores primários (idade)	HF?	Gene	Variante	ACMG	Tecido avaliado
RNAP-01	Câncer de mama	F	47	Não	Sim	CHEK2	c.846+4_846+7delAGTA	LP	Leucócitos
RNAP-02	Câncer de mama	F	40	Não	Sim	PALB2	NC_000016.9:g.(?_23625315)_(23625422_?)dup → dup E11	P	Leucócitos
							c.3350G>A	P	
RNAP-03	Câncer de mama	F	61	Não	Sim	BRCA2	c.156_157insAlu	P	Leucócitos
						RAD51C	c.571+4A>G	VUS	
RNAP-04	Câncer de mama	F	57	Adenocarcinoma ductal pancreático (61)	Sim	RAD51D	c.80C>A	VUS	Leucócitos
RNAP-05	Câncer de mama	F	45	Não	(-)	BRCA2	c.7618-10T>G	VUS	Leucócitos
RNAP-06	Câncer de Próstata	M	73	Ca pele: Carcinoma de células escamosas bem diferenciado e Carcinoma basocelular (73)	Sim	ATM	c.7273G>A	VUS	Leucócitos
RNAP-07	Câncer de Endométrio	F	76	Adenocarcinoma reto distal (61)	Sim	TP53	c.114A>G	VUS	Leucócitos
RNAP-09	Câncer colorretal	M	31	Não	Sim	MSH2	Deleção UTR + DEL parcial exon 1	VUS	Leucócitos
RNAP-09.2	Câncer colorretal	F	61	Não	Sim	MSH2	Deleção UTR + DEL parcial exon 1	VUS	Leucócitos
RNAP-101	Câncer colorretal	F	36	Câncer endométrio (47) Carcinoma basocelular (58)	(-)	MLH1	c.588+5G>C	PP	Retossigmóide normal
RNAP-103	Câncer de mama	F	69	Não	Sim	BRCA1	c.547+5G>A	VUS	Mama normal

Fonte: Elaboração dos autores.**Idade no dx** - idade no diagnóstico; **HF**- histórico familiar; **(-)**: não informado.

Tabela 7 - Pacientes que foram sequenciados com suas respectivas variantes e informações pessoais

(continuação)

Paciente	Tumor	Sexo	Idade no dx	Outros tumores primários (idade)	HF?	Gene	Variante	ACMG	Tecido avaliado
RNAP-104	Adenocarcinoma gástrico	M	42	Não	(-)	<i>MUTYH</i>	c.1434+2C>T	VUS	Estômago normal
RNAP-105	Câncer colorretal	F	48	Não descrito	(-)	<i>MUTYH</i>	c.934-2A>G	VUS	Reto normal
RNAP-106	Câncer colorretal	F	Não disponível	Não	Sim	<i>MUTYH</i>	c.934-2A>G	VUS	Metástase de CCR em fígado
RNAP-107	Câncer colorretal	M	37	Não	Sim	<i>MSH2</i>	c.942+3A>T	P	Tumor CCR
RNAP-108	Câncer de Tireóide	F	Não disponível	Não	(-)	<i>CHEK2</i>	c.320-5T>A	VUS	Tireóide normal
RNAP-109	Câncer colorretal	F	49	Não descrito	Sim	<i>CHEK2</i>	c.320-5T>A	VUS	Junção retossigmoid e
RNAP-110	Câncer colorretal	F	47	Não	(-)	<i>NBN</i>	c.2071-4A>G	VUS	Reto normal
RNAP-111	Câncer colorretal	M	28	Não	Sim	<i>RAD51C</i>	c.905-2_905-1delAG	P	Apêndice normal
RNAP-112	Câncer gástrico	M	46	Não descrito	Sim	<i>MLH1</i>	c.677G>A	P	Estômago normal
RNAP-113	Câncer colorretal	F	36	Não descrito	(-)	<i>MLH1</i>	c.116+5G>C	P	Cólon normal
RNAP-114	Câncer colorretal	F	33	Não	(-)	<i>SMARCA4</i>	c.859+5G>A	VUS	Cólon normal
RNAP-115	Câncer de mama	F	52	Não	Sim	<i>MLH1</i>	c.790C>G	VUS	Mama normal
RNAP-116	Câncer colorretal	M	50	Não	(-)	<i>MUTYH</i>	c.933+3A>C	LP	Retossigmoi de normal

Fonte: Elaboração dos autores.**Idade no dx** - idade no diagnóstico; **HF**- histórico familiar; **(-)**: não informado.

Tabela 7 - Pacientes que foram sequenciados com suas respectivas variantes e informações pessoais

(conclusão)

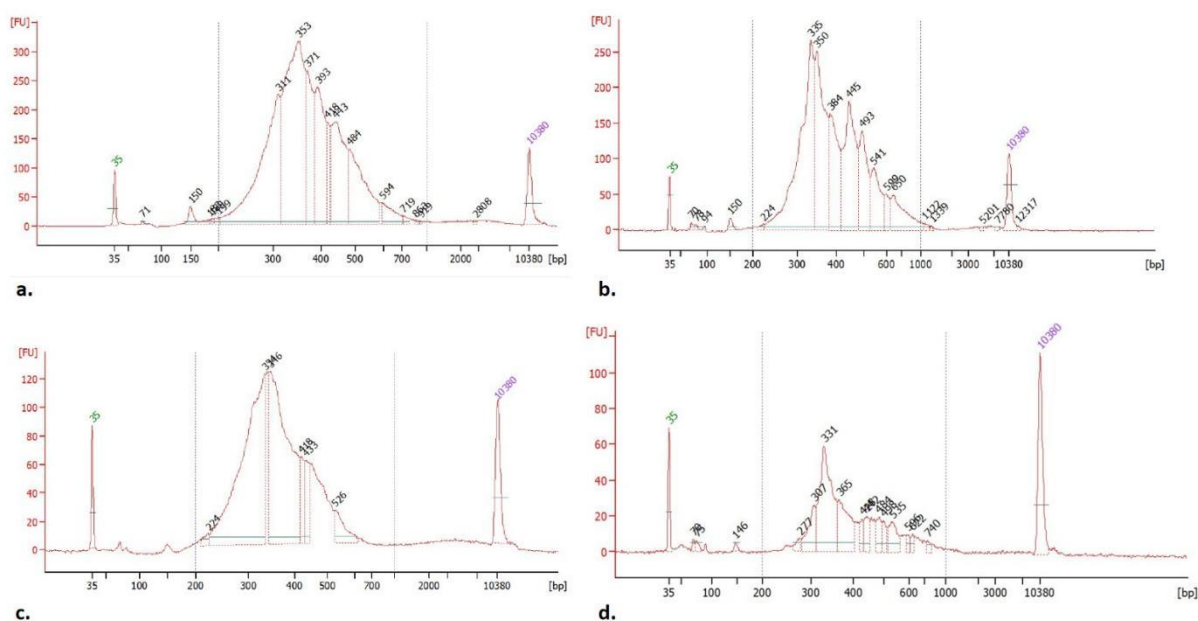
Paciente	Tumor	Sexo	Idade no dx	Outros tumores primários (idade)	HF?	Gene	Variante	ACMG	Tecido avaliado
RNAP-117	Câncer de mama	F	43	Não descrito	(-)	<i>BRCA1</i>	c.547+2T>A	P	Ovário normal
RNAP-118	Câncer de mama	F	41	Não	Sim	<i>BRCA1</i>	c.213-11T>G	PP	Mama normal
RNAP-120	Câncer de ovário	F	68	Não	Sim	<i>BRCA1</i>	c.5074+2T>C	P	Baço - tumor primário
RNAP-121	Câncer de mama	F	62	Não	Sim	<i>ATM</i>	c.2921+1G>A	P	Mama normal
RNAP-122	Câncer de mama	F	33	Não	Sim	<i>BRCA1</i>	c.5153-2A>C	P	Mama normal
RNAP-123	Câncer de mama	F	49	Não	Não	<i>CHEK2</i>	c.320-5T>A	VUS	Mama normal
RNAP-124	Câncer de mama	F	34	Adenoma tubular com displasia de baixo grau (46)	(-)	<i>BRCA1</i>	c.4096+1G>A	VUS	Mama normal

Fonte: Elaboração dos autores.**Idade no dx** - idade no diagnóstico; **HF**- histórico familiar; **(-)**: não informado.

5.4.1. Construção da Biblioteca de cDNA por captura

O rendimento médio das bibliotecas de cDNA foi de 68 ng/uL em 20 uL finais e gerou fragmentos com tamanhos aproximados de 300-400 pb estando de acordo com o esperado pelas especificações do kit (Figura 23).

Figura 23 - Perfil das bibliotecas de transcriptoma total com amostras representativas de cada um dos lotes de experimentos realizados



Fonte: Autoria própria.

Legenda: (a) RNAP-112 com tamanho de fragmento mais frequente de 353 pb, (b) RNAP-114 com 335 pb, e (c) RNAP-121 com 334 pb, (d) RNAP-06 com 331 pb.

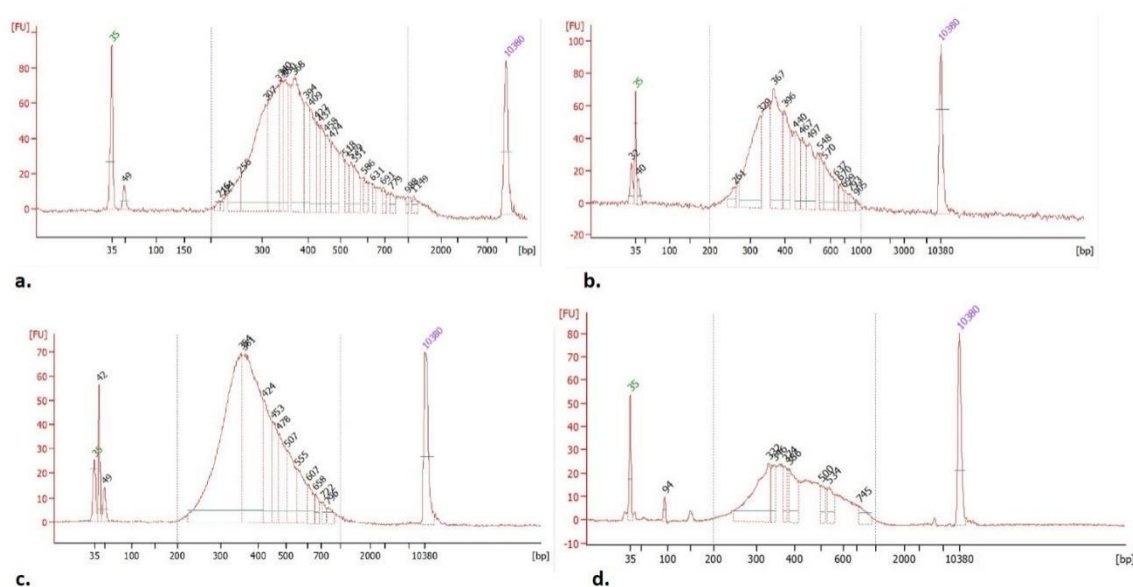
O rendimento final médio das capturas foi de 2,5 ng/μL. A Figura 24 descreve o perfil da biblioteca de captura obtido a partir do Bioanalyzer, apresentando o enriquecimento de fragmentos entre 300 e 400 pb.

Foram feitas 4 rodadas de experimentos, que compreendeu a construção de biblioteca por captura e sequenciamento. Foi obtido um enriquecimento médio de 33% para as regiões alvo, e a média de cobertura por base >50X foi de 96,25%. Informações gerais do sequenciamento como input, RIN, rendimento do sequenciamento, enriquecimento e % de rRNA nas amostras estão descritos na Tabela 8.

Em relação à presença de RNA ribossomal (rRNA) nas amostras, a média foi 10,59% de leituras totais mapeadas no rRNA, sendo que no último sequenciamento quase todas as amostras tiveram mapeamento maior que 9%. Nos sequenciamentos anteriores apenas quatro

amostras tiveram mais que 10% de leituras mapeadas no rRNA. Todas essas informações de métricas de qualidade por amostra e lote de experimentos podem ser visualizadas na Tabela 8.

Figura 24 - Perfil das três capturas de cDNA realizadas



Fonte: Autoria própria.

Legenda: (a) captura 1, (b) captura 2, (c) captura 3 e (d) captura 4.

Tabela 8 - Informações e métricas gerais do sequenciamento

(continua)

Amostra	Tipo	Input RNA	RIN	Nº total reads mapeadas	Enriquecimento (%)	Méd. cobertura	Cob. 50X (%)	rRNA (%)	Lote
CTRL_100	Leucócito	500	8,6	12665444	31,0	8292	95,2	2,8%	1
CTRL_102	Leucócito	300	8,2	19966565	33,0	13795	95,9	1,8%	1
CTRL_103	Leucócito	500	8,2	26571451	38,0	21013	97,2	1,1%	1
CTRL_104	Leucócito	300	(-)	21187382	34,0	15136	95,6	1,5%	1
RNAP-03	Leucócito	300	4,4	3155721	43,0	2865	99,9	0,7%	1
RNAP-105	Tecido	500	7,1	7254688	30,0	4497	97,1	4,8%	1
RNAP-108	Tecido	300	8	13225622	33,0	9049	97,1	1,8%	1
RNAP-109	Tecido	500	6	15308959	32,0	10305	97,5	1,4%	1
RNAP-110	Tecido	300	4,6	9347800	37,0	7199	96,5	2,6%	1
RNAP-111	Tecido	500	9,5	7519640	26,00	4123	95,3	3,7%	1

Fonte: Elaboração dos autores.

Legenda: Em vermelho amostras que tiveram RIN<4, cobertura 50X< 85%, ou mais que 10% de rRNA mapeadas no sequenciamento.

Tabela 8 - Informações e métricas gerais do sequenciamento

(continuação)

Amostra	Tipo	Input RNA	RIN	Nº total reads mapeadas	Enriquecimento (%)	Méd. cobertura	Cob. 50X (%)	rRNA (%)	Lote
RNAP-112	Tecido	500	3,7	1850247	30,0	1234	89,5	9,8%	1
RNAP-113	Tecido	300	3,2	5593648	38,0	4413	98,6	5,3%	1
CTRL_105	Leucócito	300	10	16762374	40,0	14132	96,3	1,7%	2
CTRL_106	Leucócito	300	10	16567277	42,0	14711	95,6	1,7%	2
RNAP-106	Tecido	300	8,8	26149010	36,0	20053	96,4	1,9%	2
RNAP-107	Tecido	300	9,5	8096094	25,0	4208	98,7	2,3%	2
RNAP-114	Tecido	300	3,5	10499287	36,0	7899	97	2,3%	2
RNAP-115	Tecido	300	6,4	12469085	35,0	9223	99,5	2,0%	2
RNAP-116	Tecido	300	3,1	5636018	40,0	4700	99,1	5,1%	2
RNAP-117	Tecido	300	7,6	12099583	24,0	6150	97,6	1,8%	2
RNAP-118	Tecido	300	5,7	6999898	39,0	5657	99,3	4,9%	2
CTRL_107	Leucócito	300	9,5	9607407	34,0	6871	93,6	6,8%	3
CTRL_108	Leucócito	300	9,5	12033317	36,0	9239	94,4	4,7%	3
CTRL_109	Leucócito	300	9,6	15784429	36,0	11829	96,6	4,6%	3
CTRL_110	Leucócito	300	8,6	28003711	37,0	21710	97,4	3,0%	3
CTRL_111	Leucócito	300	7,1	8439275	33,0	6159	91,7	12,2%	3
CTRL_112	Leucócito	300	8	13582424	35,0	10127	99,5	6,9%	3
RNAP-01	Leucócito	300	7,9	6346729	32,0	4278	94,8	11,2%	3
RNAP-02	Leucócito	300	7	5348297	33,0	3730	95,2	12,5%	3
RNAP-04	Leucócito	300	6,7	10601314	35,0	7881	98,4	7,6%	3
RNAP-101	Tecido	300	5,3	19984893	31,0	12835	99,2	2,6%	3
RNAP-103	Tecido	300	9	16741383	29,0	10277	99,2	5,2%	3
RNAP-104	Tecido	300	6,5	4700425	28,0	2780	94,6	18,6%	3
RNAP-120	Tecido	300	7,6	4462588	40,0	3701	99,7	4,1%	3
RNAP-121	Tecido	300	7,8	8542160	28,0	4943	97,4	9,5%	3
RNAP-122	Tecido	300	8,2	11142137	29,0	6714	98,8	6,6%	3
RNAP-123	Tecido	300	5,4	13782042	30,0	8620	99,7	6,5%	3
CTRL_113	Leucócito	300	6,5	16298577	34,0	11772	95,4	31,6%	4
CTRL_114	Leucócito	300	9,5	23847006	32,0	16303	96,9	22,2%	4
CTRL_115	Leucócito	300	9,3	24684687	32,0	16923	96,6	22,4%	4
CTRL_116	Leucócito	300	9,1	15801938	34,0	11308	94,8	23,5%	4
CTRL_117	Leucócito	300	8,8	51929617	37,0	40206	97,6	9,6%	4
CTRL_118	Leucócito	300	6,6	42688283	35,0	31267	98,5	11,5%	4
CTRL_119	Leucócito	300	8	31808644	33,0	22161	97,5	19,9%	4
RNAP-05	Leucócito	300	7,6	16325566	33,0	11398	97,4	24,0%	4
RNAP-06	Leucócito	300	5,6	10533611	31,0	7256	82,4	31,9%	4

Fonte: Elaboração dos autores.**Legenda:** Em vermelho amostras que tiveram RIN<4, cobertura 50X< 85%, ou mais que 10% de rRNA mapeadas no sequenciamento.

Tabela 8 - Informações e métricas gerais do sequenciamento

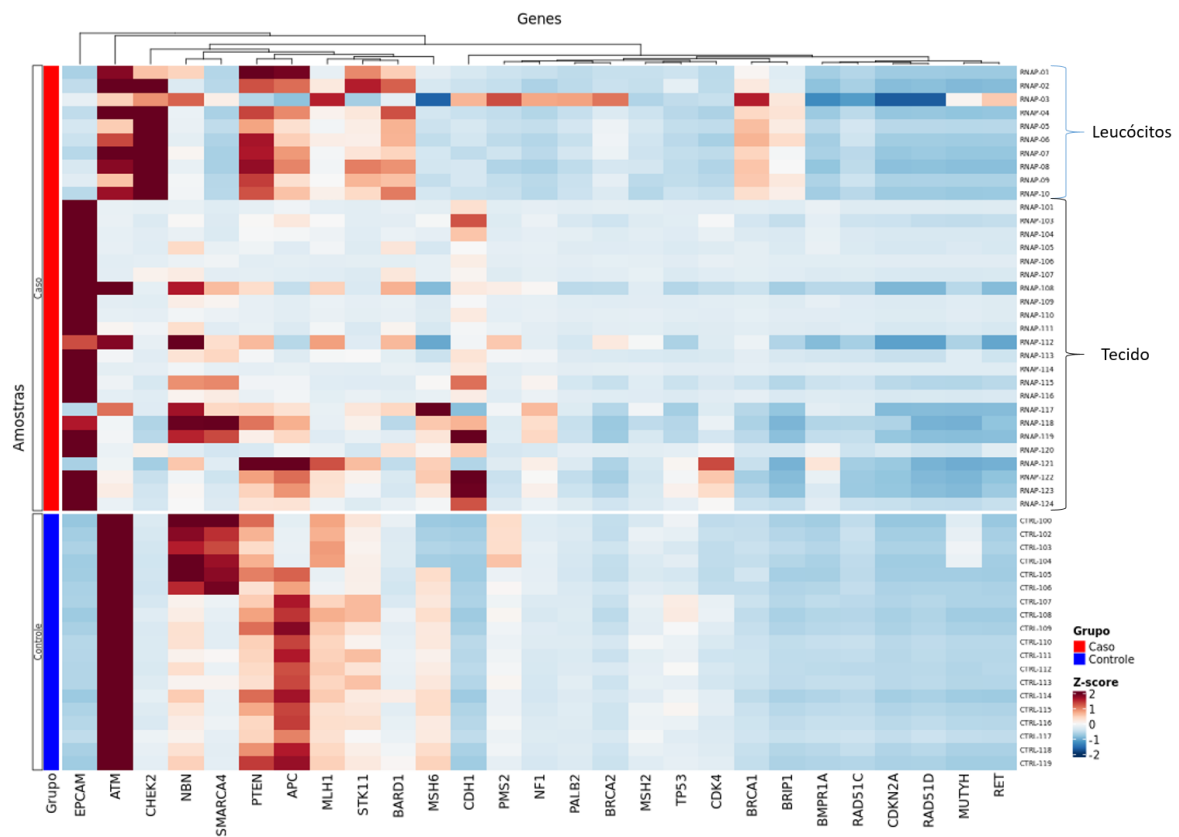
(conclusão)

Amostra	Tipo	Input RNA	RIN	Nº total reads mapeadas	Enriquecimento (%)	Méd. cobertura	Cob. 50X (%)	rRNA (%)	Lote
RNAP-07	Leucócito	300	3,4	7359821	29,0	4825	77,8	44,3%	4
RNAP-09	Leucócito	300	6,3	10639413	31,0	7112	94,3	37,9%	4
RNAP-09.2	Leucócito	300	5,4	12280825	31,0	8212	95,8	33,5%	4
RNAP-124	Tecido	300	8,9	25763928	27,0	14732	99,5	21,8%	4
MÉDIA				14705931,33	33,4	10447	96,3	10,6%	

5.4.2. Expressão gênica entre casos e controles

Os resultados obtidos em relação a expressão dos 27 genes nos leucócitos e tecido normal adjacente ao tumor são ilustrados na Figura 25. Nota-se o mesmo padrão de expressão encontrado pela plataforma GTEx, em que o tecido sanguíneo e o tecido normal adjacente apresentam algumas diferenças esperadas como na expressão dos genes *CDH1* e *EPCAM*, uma vez que ambos codificam moléculas de adesão/contato célula-célula e são mais expressos em células epiteliais, por isso são mais expressos no tecido se comparado com os leucócitos. Embora nos pacientes prospectivos, que foi feita a extração do RNA a partir de leucócitos (RNAP-01, RNAP-02, RNAP-03, RNAP-04, RNAP-05, RNAP-06, RNAP-07, RNAP-08, RNAP-09 e RNAP-09.2), a expressão desses dois genes estivesse um pouco mais alta em comparação com os controles negativos, a expressão estava mais baixa em relação as amostras de tecidos normais adjacentes, com exceção dos pacientes RNAP-121 e RNAP-117 (ambos tecidos mama). Por último, vale destacar que nenhum paciente sequenciado possuía variantes em nenhum desses dois genes, fato que não atrapalhou nas análises posteriores. Alguns outros genes também apresentam variações em níveis de expressão entre as amostras, porém, como avaliado pelo dado de cobertura, foi possível obter com sequenciamento RNA-cap uma cobertura adequada (pelo menos 50X em 85% da região alvo) para a maioria das amostras.

Figura 25 - Heatmap da expressão dos 27 genes de estudo no sequenciamento de RNA-cap entre amostras de tecido normal adjacente do tumor e leucócitos



Fonte: Autoria própria.

Legenda: Heatmap de indivíduos controles e pacientes. Os valores de expressão gênica representados estão em escala logarítmica de transcritos por milhão (TPM).

5.5. Análises de dados

Nessa etapa, as três ferramentas (rMATS, LeafCutterMD e FRASER2) foram testadas com os 14 pacientes controles positivos (com 14 variantes P/PP) e os 217 controles negativos (19 internos e 198 externos) para o melhor entendimento dos resultados de saída de cada ferramenta em si, assim como para estabelecermos pontos de corte de significância e identificarmos limitações (validação técnica). Após essa etapa analisamos as amostras dos pacientes com as VUS (validação clínica). Para cada paciente reuniu-se as informações referentes a cada variante, como: localização, erro de *splice* (qual o impacto no transcrito de RNA e na proteína), abrangência de algum domínio proteico relevante, desbalanço alélico e descrições no ClinVar por laboratórios clínicos de referência que realizaram teste de RNA (Invitae e Ambry Genetics).

5.5.1. Comparação das ferramentas de análise de *splice* em controles positivos

A ferramenta que mais detectou eventos de *splice* nas amostras com variantes sabidamente spliceogênicas foi o rMATS (11 eventos), seguido pelo LeafCutterMD (7 eventos) e FRASER2 (7 eventos). Observamos que as ferramentas apresentaram pouca concordância entre si, visto que apenas 4/14 variantes P/PP analisadas apresentaram o mesmo evento de *splice* nas três ferramentas (a variante *BRCA1*: c.547+2T>A apresentou o evento de salto do E7 no rMATS e FRASER2, porém diferente no LeafCutterMD identificando salto do E8 e E9), 6/14 variantes tiveram o mesmo evento capturado por duas ferramentas (3 eventos identificados pelo rMATS e LeafCutterMD e 3 eventos pelo rMATS e FRASER2), a variante *MLH1*:c.677G>A apresentou evento de *splice* discordante entre rMATS e FRASER2. Para 3 variantes não foi encontrado nenhum efeito de *splice* significativo em nenhuma ferramenta. Todas essas informações estão resumidas e podem ser visualizadas no diagrama de Venn (Figura 26). O detalhamento dos eventos esperados e identificados com cada ferramenta são apresentados no item 5.5.2 e 5.5.3.

A ferramenta rMATS foi a mais completa para a análise de *splice*, pois ela tem como vantagem a possibilidade de observar vários eventos de *splicing* que podem acontecer ao mesmo tempo próximo da variante, além de poder acrescentar como *input* um número grande de controles contra uma única amostra do paciente ou réplicas biológicas. Com ela é possível entender com exatidão quais os éxons envolvidos nos eventos, pois nas tabelas de resultados tem suas as coordenadas iniciais-finais, também pode-se visualizar a cobertura de quantas leituras incluíram e excluíram o evento de *splice* para o grupo de controles e o paciente, além do valor do PSI e a diferença entre os grupos. Em relação às limitações, podemos destacar a dificuldade de observar onde os eventos de *splice* estão acontecendo, pois nas tabelas de saída só aparecem as coordenadas genômicas de onde começa e termina cada éxon, por isso foi necessário a cada análise utilizar o Genome Browser UCSC (disponível online em <https://genome.ucsc.edu/cgi-bin/hgGateway>) para conseguir localizar os éxons e íntrons envolvidos nos eventos para melhor interpretação. Diferentes IDs de eventos serão criados se as posições iniciais-finais de cada éxon envolvido no evento se diferenciarem um pouco, devido as quebras das leituras do sequenciamento, gerando vários eventos muito semelhantes entre si.

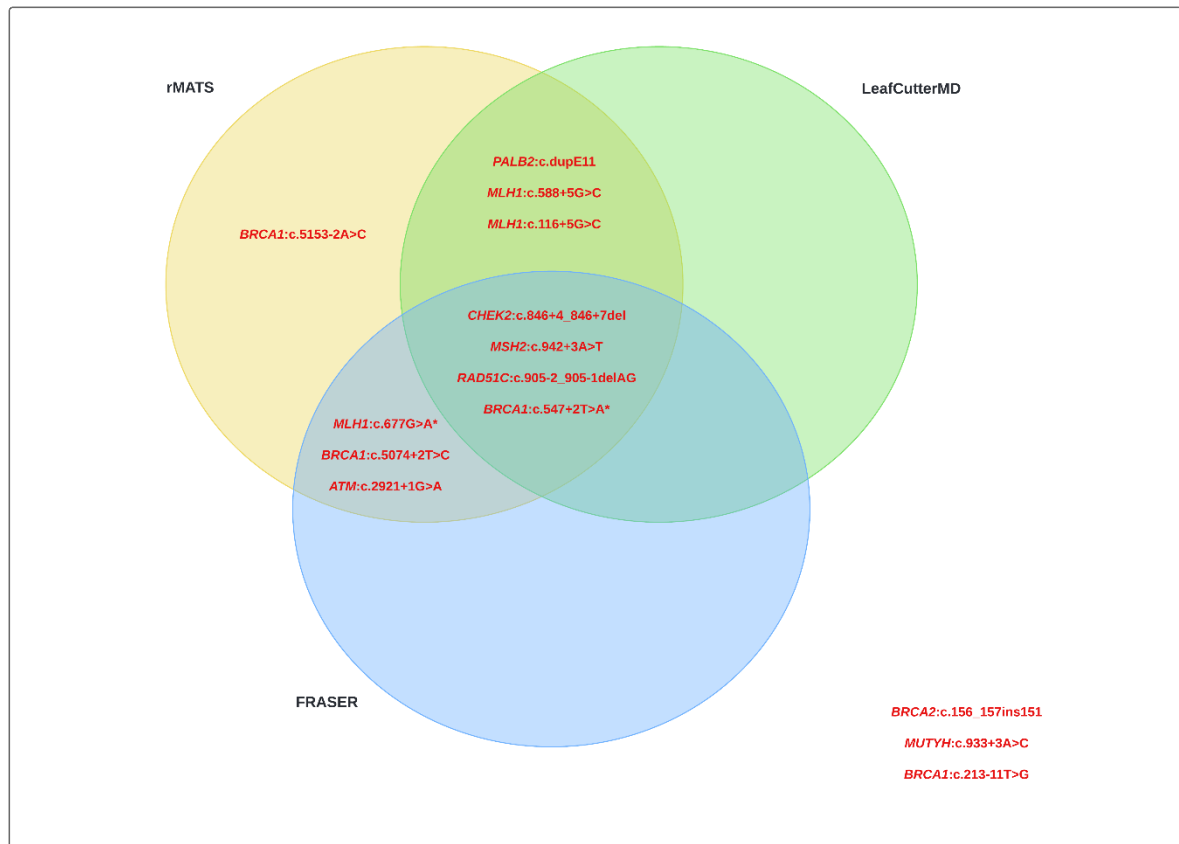
O LeafCutterMD gera como arquivo final uma tabela com a descrição de cada evento em uma linha contendo suas posições genômicas (o cromossomo e coordenadas da excisão de íntrons com as junções), e cada coluna representa um paciente e seu valor de p para aquele evento. Quando o p valor é $\leq 0,05$ aquela excisão é significativa para o paciente comparado

com todas outras amostras. Filtramos a coluna do paciente para esse valor, e se o evento referente a variante era significativo no paciente, também verificamos se algum controle também apresentava o valor $\leq 0,05$. A desvantagem é que a tabela só apresenta os valores de p, sem nenhuma outra informação sobre o evento de *splice*, junção, número de reads que incluíram ou não aquela junção.

O resultado gerado pelo DROP com a ferramenta FRASER2 é uma tabela já filtrada para os eventos significativos, com as métricas de PSI, pValue, colunas que descrevem o potencial impacto de *splice* e a probabilidade de causar uma alteração *Frameshift*. Para rodar a ferramenta através do pipeline DROP é possível ajustar diversos parâmetros para personalizar os filtros de variantes significativas, porém mesmo realizando tentativas de ajustar diversos parâmetros, não identificamos eventos adicionais.

Alguns pontos negativos que podemos destacar entre as três ferramentas é que eventos de retenção ou salto parcial de éxon ou íntron são difíceis de serem detectados nas três ferramentas, assim como também eventos que acontecem em regiões que tem pouca cobertura (poucas leituras).

Figura 26 - Resumo das variantes P/PP que foram identificados eventos de *splice* significativos entre as três ferramentas



Fonte: Autoria própria.

Legenda: Visão geral do desempenho de detecção de variantes spliceogênicas pelas três ferramentas. As três variantes para as quais não foi encontrado nenhum evento de *splice* significativo estão descritas fora dos círculos.

*Variantes em que o evento de *Splice* encontrado não foi exatamente o mesmo entre as ferramentas.

5.5.2. Resultados – grupo controle positivo

De forma geral, dos 14 pacientes com 14 variantes P/PP já conhecidas de serem spliceogênicas, foram encontrados eventos significativos de *splice* em 11 variantes, e apenas para 3 variantes não foi encontrado nenhum efeito de *splice* por nenhuma das ferramentas. O tipo de evento mais encontrado foi salto de éxon (SE) em 10 variantes, sendo 2 variantes com evento SE *in-frame* (quando a janela de leitura de tradução não é alterada perdendo alguns aminoácidos da cadeia de polipeptídeos) e 8 variantes com evento SE *out-of-frame* (quando a janela de leitura de tradução é alterada causando na maioria dos casos na formação de um códon de parada prematuro e uma proteína incompleta). Apenas foi encontrado um evento de uso alternativo do sítio 5' de *splice*. Os resultados detalhados estão descritos na Tabela 9 e serão discutidos com detalhes a seguir.

Tabela 9 - Resultados preditos *versus* obtidos a partir das variantes P/PP

(continua)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δ score posição no RNA)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante
RNAP-01	CHEK2	c.846+4_846+7del; (I7)	DL 0,98 5	Salto do E7 <i>in-frame</i> , perdendo 18 aa no domínio Serina-treonina/tirosina quinase. r.793_846del, p.(Asp265_His282del)	rMATS, LeafcutterMD e FRASER2
RNAP-02	PALB2	NC_000016.9:g.(?_23625315)_(23625422_?)dup. Referenciado como dup E11.	Sem predição	Salto do E11 e E12 <i>in-frame</i> perdendo 79 aa do domínio WD40 (Parceiro e localizador do domínio BRCA2), r.3115_3351del e p.(Asn1039_Arg1117del); Salto do E12 <i>out-of-frame</i> , não previsto de sofrer NMD, perdendo o domínio WD40. r.3203_3351del e p.(Gly1068Valfs*5).	rMATS; LeafCutterMD
	PALB2	c.3350G>A (E12)	DL 0,14 0		

Tabela 9 - Resultados preditos *versus* obtidos a partir das variantes P/PP

(continuação)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δ score posição no RNA)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante
RNAP-03	BRCA2	c.156_157insAlu (E3)	Sem predição	Salto do E3 <i>in-frame</i> perdendo 83 aa do domínio de ativação transcricional. r.68_316del e p.(Asp23_Leu105del).	Nenhuma. Identificado por inspeção visual do Bioanalyzer.
RNAP-101	MLH1	c.588+5G>C (I7)	DL 0,89 -5	Salto do E6 <i>out-of-frame</i> , previsto de sofrer NMD. r.456_547del, p.(Glu153Phefs*8)	rMATS, LeafcutterMD
RNAP-107	MSH2	c.942+3A>T (I5)	DL 0,49 -3	Salto do E5 <i>in-frame</i> , perdendo 49 aa do domínio MutS_II e MutS_III. r.794_943del, p.(Val265_Gln314del)	rMATS, LeafcutterMD e FRASER2
RNAP-111	RAD51C	c.905-2_905-1delAG (I6)	AL 0,91 3	Salto do E7 <i>out-of-frame</i> , previsto de sofrer NMD. r.906_966del, p.(Glu303Trpfs*41)	rMATS, LeafcutterMD e FRASER2
RNAP-112	MLH1	c.677G>A (E8)	DL 0,44 0; DG 0,29 -4	Salto do E8 <i>out-of-frame</i> , afetando o domínio DNA <i>mismatch repair</i> e resultando em PTC, previsto de sofrer NMD. r.589_677del, p.(Gln197Argfs*8)	rMATS
RNAP-113	MLH1	c.116+5G>C (I1)	DL 0,74 -5	Uso alternativo do sítio 5' <i>splice</i> a 231 pb depois do E1 utilizado menos pelo paciente.	rMATS, LeafcutterMD

Tabela 9 - Resultados preditos *versus* obtidos a partir das variantes P/PP

(continuação)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δ score posição no RNA)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante
RNAP-116	<i>MUTYH</i>	c.933+3A>C (I10)	DL 0,96 3	Sem eventos significativos identificados	Nenhuma
RNAP-117	<i>BRCA1</i>	c.547+2T>A (I7)	DL 0,99 2	Salto do E7 <i>out-of-frame</i> , previsto de sofrer NMD. r.442_547del, p.(Gln148Aspfs*51)	rMATS e FRASER2
RNAP-118	<i>BRCA1</i>	c.213-11T>G (I4)	AL 0,71 -11; AG 0,20 48	Retenção de 59 pb na extremidade 5' do éxon 4, resultando em PTC e previsto de sofrer NMD. r. 212_213ins 213-1_213-59. p.(Arg71Serfs*11).	Nenhuma. Identificado por inspeção visual do BAM no IGV.
RNAP-120	<i>BRCA1</i>	c.5074+2T>C (I16)	DL 0,74 2	Salto do E16 <i>out-of-frame</i> , afetando o domínio BRCT e resultando em PTC e previsto de sofrer NMD. r.4992_5079del, p.(Val1665Serfs*8).	rMATS, FRASER2
RNAP-121	<i>ATM</i>	c.2921+1G>A (I19)	DL 0,91 -1	Salto do E19 <i>out-of-frame</i> , afetando o domínio serina-treonina quinase, resultando em PTC e previsto de sofrer NMD. r.2839_2921del, p.(Tyr947Glnfs*9).	rMATS, FRASER2

Tabela 9 - Resultados preditos *versus* obtidos a partir das variantes P/PP

(conclusão)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δ score posição no RNA)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante
RNAP-122	BRCA1	c.5153-2A>C (I17)	AL 0,99 -2; AG 0,21 -13	Salto do E18 <i>out-of-frame</i> , afetando domínio BRCT, resultando em PTC e previsto de sofrer NMD. r.5153_5193del, p.(Trp1718Serfs*2).	rMATS

Fonte: Elaboração dos autores.**Legenda:** **DL** – perda do sítio doador (Donor Loss); **DG** – ganho de um sítio doador (Donor Gain); **AL** – perda do sítio aceptor (Acceptor Loss); **AG** – ganho de um sítio aceptor (Acceptor Gain); **PTC** – códon de parada prematuro; **NMD** – *nonsense mediated decay*.

O resumo dos resultados da análise do rMATS pode ser visualizado na Tabela 10, no gráfico de boxplots (Figura 27) e no gráfico de *sashimi plots* (Figura 28), referente a cada evento, os quais serão discutidos a seguir.

Tabela 10 - Resumo dos resultados gerados no rMATS para pacientes com variantes que apresentaram eventos de *splice* significativos

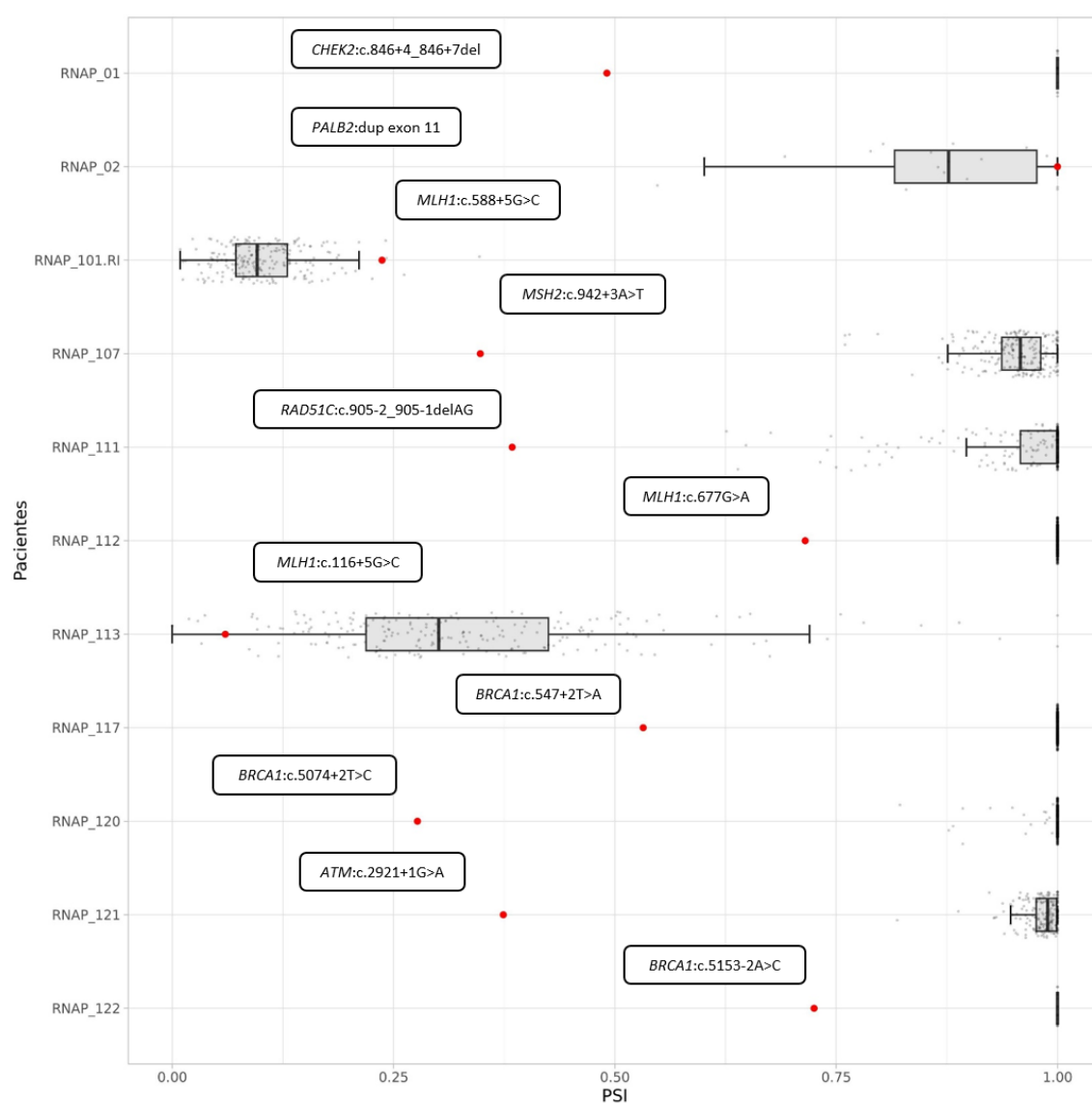
Paciente	Variante	Evento <i>splice</i>	Pvalue	FDR	PSI Ctrl (média)	PSI Caso	Dif. PSI	Predição <i>Splice</i>
RNAP-01	<i>CHEK2</i> : c.846+4_846+7del	SE (E7)	0	0	1,0-1,0 (1,0)	0,491	0,509	SpAI
RNAP-02	<i>PALB2</i> dup E11	SE (11 e 12)	0,01	0,047	0,548-1,0 (0,86)	1,0	-0,138	Sem predição
RNAP-101	<i>MLH1</i> : c.588+5G>C	RI (I7)	0	0	0,009-0,347 (0,10)	0,237	-0,134	AdRf, MES, SpAI
RNAP-107	<i>MSH2</i> : c.942+3A>T	SE (E5)	0	0	0,76 - 1,0 (0,95)	0,348	0,605	AdRf, MES, SpAI
RNAP-111	<i>RAD51C</i> : c.905-2_905-1delAG	SE (E7)	0	0	0,626 – 1,0 (0,96)	0,384	0,578	SpAI
RNAP-112	<i>MLH1</i> :c.677G>A	SE (E8)	0	0	1,0 – 1,0 (1,0)	0,715	0,285	AdRf, MES
RNAP-113	<i>MLH1</i> : c.116+5G>C	A5SS (E1)	0	0	0,004-1,0 (0,33)	0,056	0,278	AdRf, MES, SpAI
RNAP-117	<i>MUTYH</i> : c.933+3A>C	SE (E7)	0	0	1,0 – 1,0 (1,0)	0,532	0,468	AdRf, MES, SpAI
RNAP-120	<i>BRCA1</i> : c.5074+2T>C	SE (E16)	0	0	0,822 - 1,0 (1,0)	0,277	0,719	AdRf, MES, SpAI
RNAP-121	<i>ATM</i> : c.2921+1G>A	SE (E19)	0	0	0,819 - 1,0 (0,98)	0,374	0,609	AdRf, MES, SpAI
RNAP-122	<i>BRCA1</i> : c.5153-2A>C	SE (E18)	0	0	1,0 - 1,0 (1,0)	0,725	0,275	AdRf, MES, SpAI

Fonte: Elaboração dos autores.

Legenda: Preditores de *splice* - Ada e Rf score (**AdRf**), MaxEntScan (**MES**) e SpliceAI (**SpAI**). **E**: éxon; **SE**: salto de éxon; **A5SS**: uso alternativo do sítio 5' de *splice*; **RI**: retenção de íntron.

Foi visto que, nos eventos SE significativos de *splice* a diferença de PSI entre controles e pacientes teve como média ~0,4 (40%), e que eventos nos quais essa diferença é $\geq 10\%$, devem ser levados em consideração para uma análise mais profunda do evento de *splice*.

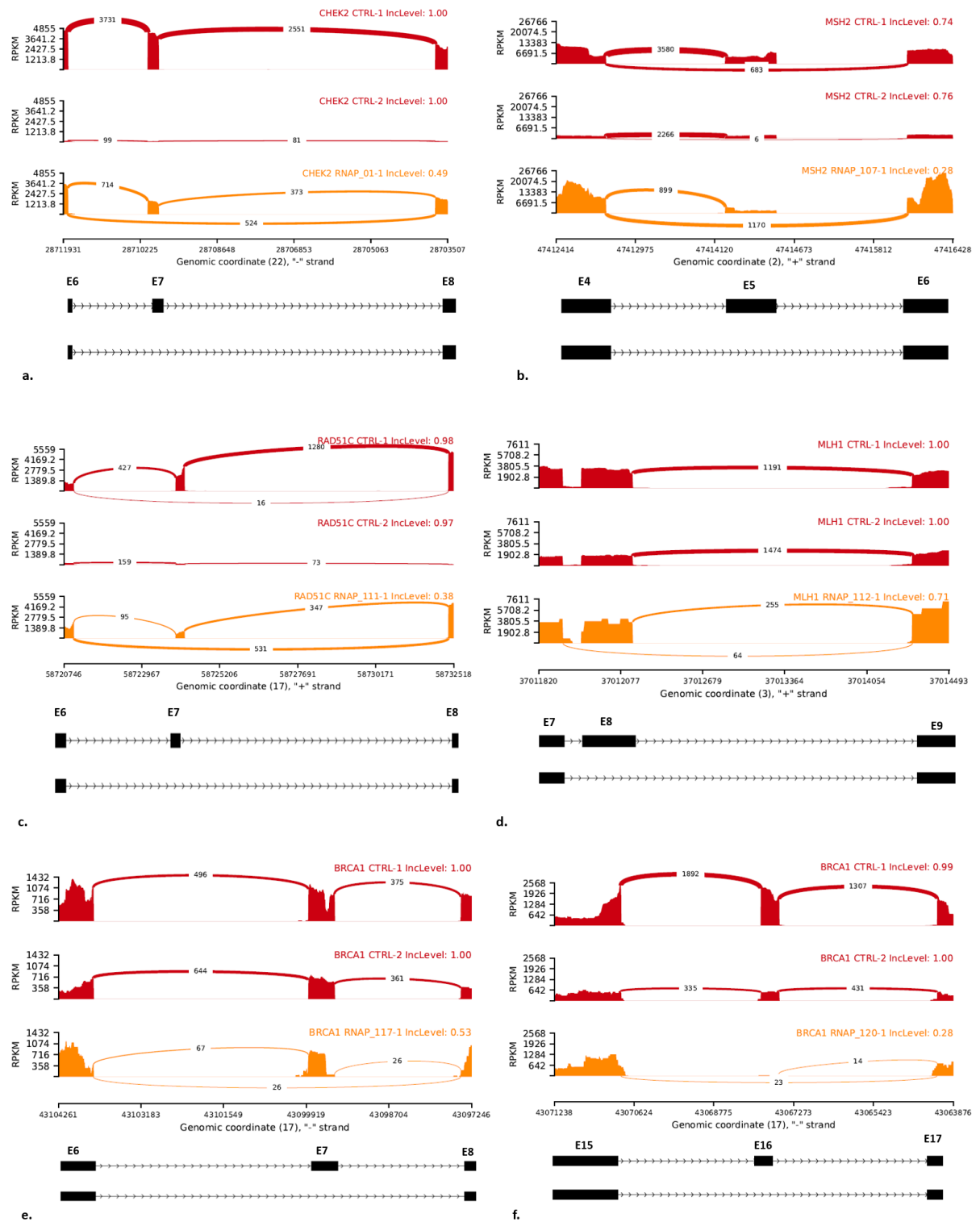
Figura 27 - Boxplots com a distribuição do PSI dos controles contra os valores do PSI dos pacientes com variantes P/PP

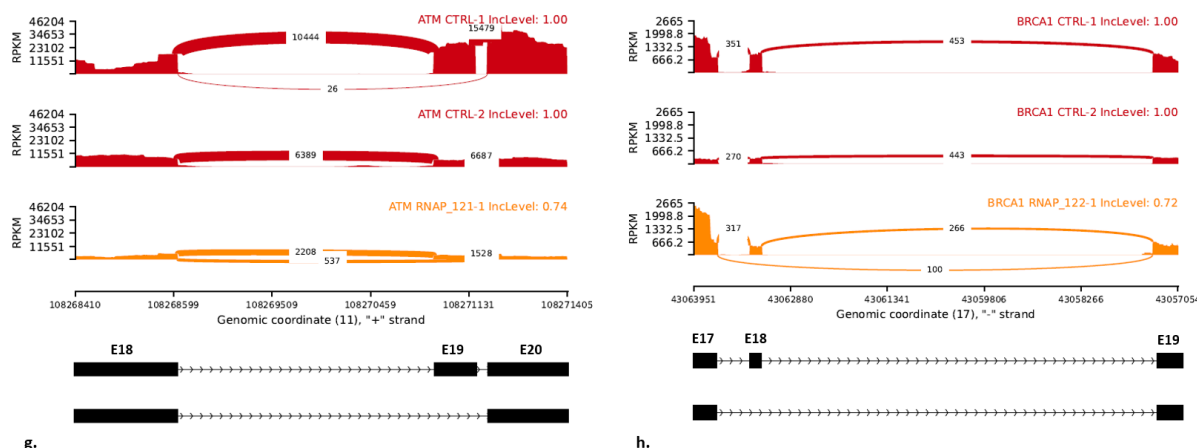


Fonte: Autoria própria.

Legenda: Os boxplots em cinza mostram o intervalo da distribuição dos valores de PSI dos controles negativos, e os pontos vermelhos em destaque mostram os valores do PSI dos pacientes com cada variante patogênica em questão.

Figura 28 - Sashimi plots dos eventos SE significativos encontrados





Fonte: Autoria própria.

Legenda: Os gráficos de sashimi mostram o salto de éxon que pode ser observado com mais frequência e/ou exclusivamente nos pacientes casos (na cor laranja claro) em comparação com indivíduos controles (na cor vermelha). No gráfico o CTRL-1 se refere a um indivíduo controle negativo interno e CTRL-2 a um indivíduo controle negativo externo. IncLevel refere-se ao PSI de inclusão do éxon nas amostras avaliadas. **a.** RNAP-01, **b.** RNAP-107, **c.** RNAP-111, **d.** RNAP-112, **e.** RNAP-117, **f.** RNAP-120, **g.** RNAP-121, **h.** RNAP-122.

Os 2 eventos de salto de éxon *in-frame* foram identificados pelas 3 ferramentas em dois pacientes. O primeiro com a variante em *CHEK2*:c.846+4_846+7del localizada no íntron 7 causou o salto *in-frame* do éxon 7 levando a perda de 18 aminoácidos do domínio tirosina-quinase. Nos registros do ClinVar a Invitae descreveu que a variante causou o salto apenas do éxon 7 que é *in-frame*, e descreveram o salto do éxon 7-8 sendo *out-of-frame*, no primeiro caso a janela de leitura é preservada e no segundo caso o transcrito sofre degradação via NMD. Essa paciente não possuía nenhuma variante exônica para análise de desbalanço alélico. O segundo paciente com a variante em *MSH2*:c.942+3A>T localizada no íntron 5 causou o salto *in-frame* do éxon 5 levando a perda de 49 aminoácidos que fazem parte de um trecho dos domínios MutS_II e MutS_III, no qual a Invitae e Ambry Genetics também já constataram o salto do éxon 5 e preservação da janela de leitura. Esse paciente também não possuía SNP marcador para a análise de desbalanço alélico.

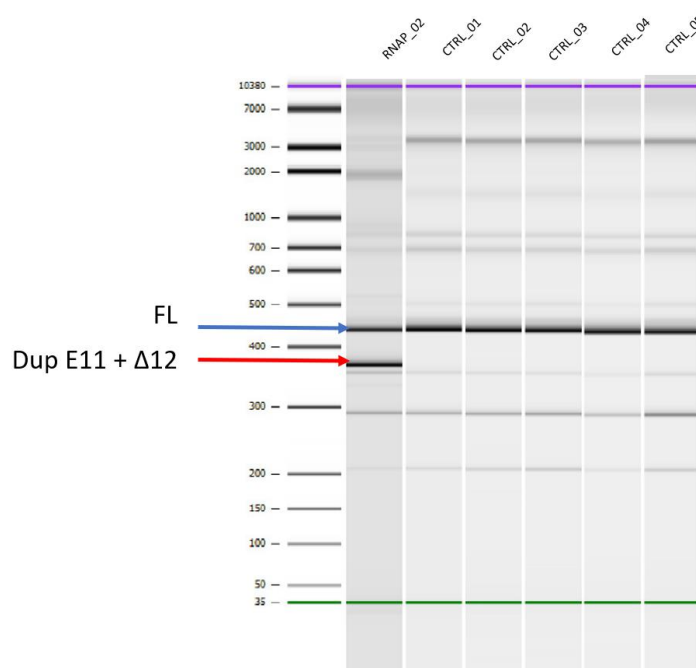
Todos os outros 8 saltos de éxons ocasionam transcritos *out-of-frame*, gerando códons de parada prematuros nos pacientes RNAP-111 (encontrado nas 3 ferramentas), RNAP-02, RNAP-101, RNAP-117, RNAP-120, RNAP-121 (encontrados em 2 ferramentas), RNAP-112 e RNAP-122 (encontrado em apenas 1 ferramenta). Sete dessas variantes já foram descritas pela Ambry e Invitae por análises de RNA apontando os mesmos achados de formação de um códon de parada prematuro e quase todos os transcritos sofrendo degradação via NMD.

Apenas a variante *BRCA1*:c.5153-2A>C do RNAP-122 tem 3 depósitos no ClinVar, porém nenhum deles tem descrição detalhada do efeito de *splice* da variante e não foram depositados por esses dois laboratórios de referência. Nessa paciente, o evento de SE 18 foi encontrado apenas pela ferramenta rMATS, quando avaliado contra os controles internos, em que 100% das leituras apresentaram a inclusão do E18 nos controles, em contraste com 73% das leituras do paciente, portanto, apresentando o salto do éxon em 27% das leituras. Essa alteração levará a formação de um códon de parada prematuro e é predita de levar à NMD.

Na paciente RNAP-02, com a duplicação do E11 no gene *PALB2* e a variante c.3350G>A (p.Arg1117Lys), foi encontrado o salto do E11 e E12 pela ferramenta rMATS (apenas contra com os CTRLs internos), um evento *in-frame* com a perda de 79 aminoácidos. Esse evento ocorre em isoformas normais, e a paciente não apresentou leituras com essa isoforma (PSI de inclusão dos éxons 11 e 12 foi de 1,0 na paciente, e variou de 0,55 a 1,0 nos controles). Já na ferramenta LeafCutterMD foi identificado salto apenas do E12 que é *out-of-frame*, com códon de parada após 5 aminoácidos. Nenhuma das ferramentas conseguiu identificar a presença de 2 éxons 11 (duplicação) no RNA, o que foi comprovado pela análise direcionada ao transcrito (Figura 29).

Através da análise direcionada ao transcrito na paciente RNAP-02 (*PALB2* dup E11) identificamos uma banda forte não presente nos CTRLs, que é indicativo de duplicação do E11 com o salto do E12, ou seja, ocorre uma ligação direta do E11 duplicado com o E13 (Figura 29). A confirmação da estrutura dessa banda foi realizada por sequenciamento de *amplicon* e análise no CLC Genomics. Relembrando que, para todas as imagens de eletroforese capilar, o transcrito completo ou *amplicon* esperado está representado pela sigla FL com uma seta da cor azul, e o transcrito anormal esperado pelo erro de *splicing* está representado com uma seta de cor vermelha. A nomenclatura do impacto de *splice* nos transcritos de RNA foi padronizada de acordo com o que foi descrito no tópico 4.6.1.

Figura 29 - Imagem do gel do Bioanalyzer do paciente RNAP-02 para análise da variante *PALB2* dup E11

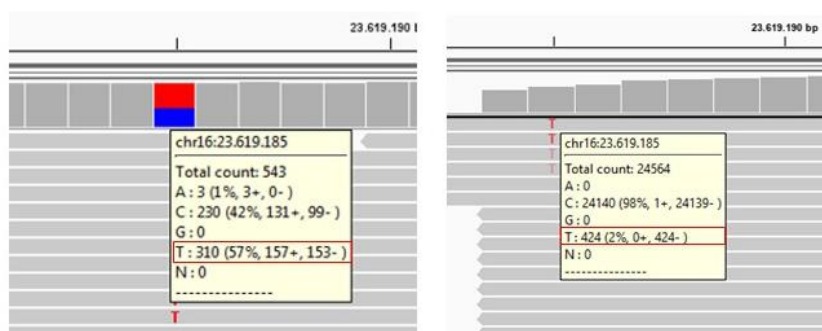


Fonte: Autoria própria.

Legenda: O paciente está na primeira canaleta, a seta azul indica o *amplicon* esperado de 440 pb (FL), e a seta vermelha indica o *amplicon* alterado esperado de 379 pb indicativo de duplicação do E11 juntamente com salto do E12, a qual está ausente nos controles.

Através da análise de desbalanço alélico foi possível identificar se a variante *missense* presente no último nucleotídeo do E12 (c.3350G>A) estava em *cis* (no mesmo cromossomo) com a duplicação do E11 ou se estava em *trans* (em cromossomos distintos). Ao avaliar no IGV, o alelo alternativo estava expresso em apenas 2% das leituras, enquanto o alelo referência está com expressão de 98% (Figura 30), indicando que a duplicação e o alelo alternativo da variante c.3350G>A estão no mesmo cromossomo, visto que pela análise direcionada ao transcrito o paciente apresenta exclusivamente uma banda de tamanho molecular equivalente a duplicação do E11 sem a presença do E12 (Figura 29). Dessa forma ocorre o salto do E12 fazendo com que essa variante não seja incluída no transcrito maduro. No ClinVar, essa variante já tinha sido vista em combinação com a duplicação do E11 em pelo menos 4 pacientes sem Anemia de Falconi (indicando estarem em *cis*), o que foi confirmado por nossos dados.

Figura 30 - Análise de desbalanço alélico do paciente RNA-02 a partir da variante *missense* em *PALB2*:c.3350G>A (chr16:23607864)



Fonte: Autoria própria.

Legenda: O mapeamento da esquerda demonstra a variante em heterozigose no sequenciamento de DNA. O mapeamento da direita demonstra o sequenciamento alvo de RNA, podendo-se observar claramente o desbalanço alélico, em que a expressão do alelo selvagem tem VAF de 98% das leituras e o alelo alternativo apenas 2%.

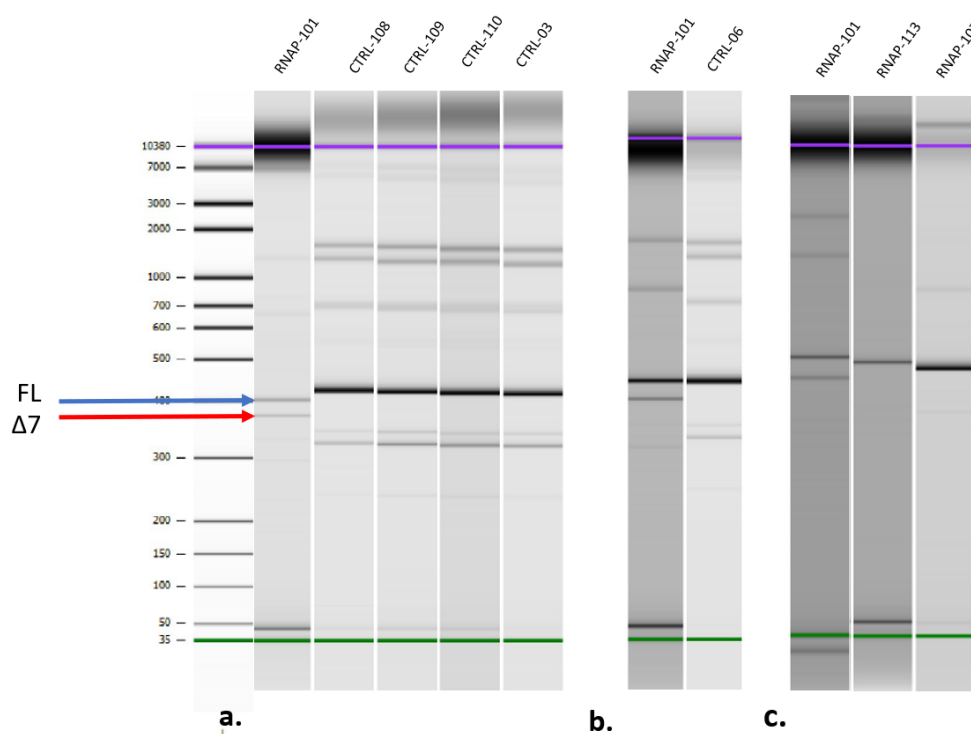
Na paciente RNAP-113 com a variante em *MLH1*: c.116+5G>C (localizada no íntron 1), foi detectado apenas um evento de uso alternativo do sítio 5' de *splice*, em que existe uma ampla utilização desse sítio nos indivíduos controles, porém esse sítio é menos utilizado na paciente. Possivelmente o evento principal esperado para esse paciente (não inclusão do éxon 1) não é passível de detecção pelas ferramentas utilizadas, uma vez que não seria gerada uma “nova junção” a partir desse evento. Essa paciente não possui SNP marcador. Não foi possível realizar a análise direcionada ao transcrito, pois devido a localização da variante os *primers* não conseguem se ancorar em uma região anterior ao E1 caso esse éxon seja saltado na paciente.

Na paciente RNAP-101 que possui a variante em *MLH1*:c.588+5G>C, localizada no íntron 7, foi detectado como significativo pelo rMATS e pelo LeafCutterMD um menor nível de salto do éxon 6. Nesse evento o E6 é mais incluído na paciente que nos controles com uma diferença muito pequena de PSI de 5,2%, que apesar de ser significativa, não entra no nosso estudo, pois utilizamos o *cut-off* de diferença de PSI de 10%. Observou-se também um evento de retenção do íntron 7 significativo, sendo essa isoforma mais incluída na paciente, com a diferença do PSI de 13,4%. O salto do E6 nesse gene é sabidamente uma isoforma frequente em indivíduos saudáveis.

Essa variante foi demonstrada, em outro estudo com participação de nosso grupo, de causar o salto do E7 e E6 + E7 (eventos *out-of-frame*), através da análise de RNA de sangue periférico por RT-PCR e sequenciamento NGS (Piñero *et al.*, 2020). Na análise direcionada ao

transcrito, a paciente com RNA extraído do retossigmoide apresentou uma banda exclusiva mais fraca (seta vermelha), que pode ser indicativo de salto do E7. No terceiro ensaio corremos a amostra da paciente contra amostras controles de RNAs extraídos do cólon. Observou-se que o perfil de bandas (*splice*) do paciente é mais semelhante às amostras de cólon do que das amostras de leucócitos, apresentando menos bandas (menos isoformas de transcritos). Ainda é possível reparar que apenas na paciente RNAP-101 apareceu essa segunda banda de tamanho menor ao esperado para o transcrito *full length* (Figura 31).

Figura 31 - Imagem do gel do Bioanalyzer do paciente RNAP-101 para análise da variante *MLH1*:c.588+5G>C



Fonte: Autoria própria.

Legenda: Os três ensaios estão representados pelas letras (a), (b) e (c), respectivamente. Apenas no terceiro ensaio correu o paciente contra os pacientes RNAP-107 e RNAP-113, que possuíam RNA extraído do cólon. A seta azul indica o *amplicon* esperado de 450 pb (FL), e a seta vermelha o *amplicon* alterado esperado de 407 pb, indicativo do salto do E7 e que está ausente nos CTRLs de leucócitos e de tecido.

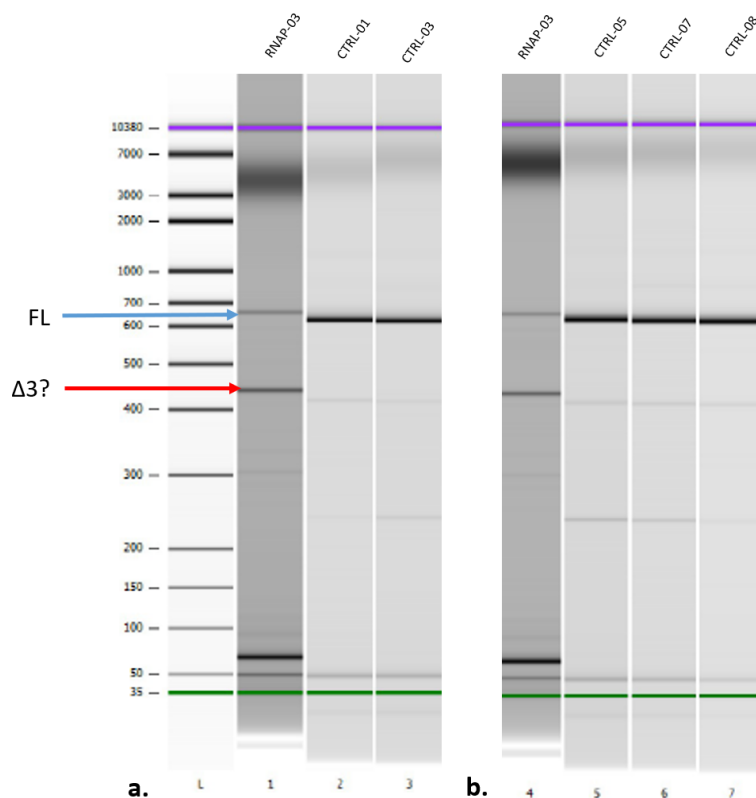
Em resumo, as ferramentas do RNA-cap não detectaram o evento esperado que era salto do E6-E7 e salto do E7, mas detectaram diferenças significativas na proporção das isoformas normais dessa região, indicando alterações relevantes de *splice* envolvendo esses éxons. Através da eletroforese capilar, foi possível perceber que as amostras de leucócitos têm uma

diversidade maior de isoformas (apresentam mais bandas), devido ao *splice* alternativo de *MLH1*, do que nas amostras dos tecidos retosigmóide/cólon, que apresentam menos bandas.

As variantes P/PP que não apresentaram eventos significativos detectáveis pelas ferramentas foram: a variante de inserção Alu em *BRCA2*:c.156_157insAlu (RNAP-03); a variante em *MUTYH*:c.933+3A>C (RNAP-116) e *BRCA1*: c.213-11T>G (RNAP-118). A variante que está no gene *MUTYH* no paciente que RNAP-116 pode não ter apresentado eventos significativos por conta da grande quantidade de isoformas transcritas para esse gene, que por apresentar íntrons muito pequenos, apresenta retenções de íntrons tanto nos casos como nos controles. As ferramentas de análise de *splicing* podem ter dificuldade em reconhecer múltiplos eventos de *splicing* que ocorrem ao mesmo tempo através do sequenciamento de leituras curtas, o que pode ter contribuído para a não identificação de eventos anormais para essa variante. Apesar de não estar numa posição conservada, essa variante foi prevista de causar o salto do éxon 10, e consequentemente junção E9-E11, por enfraquecer o sítio doador nativo de *splice*.

A análise direcionada ao transcrito foi realizada apenas nas pacientes RNAP-03 e RNAP-118. A variante *BRCA2*: c.156_157insAlu na paciente RNAP-03 é conhecida como a variante fundadora portuguesa, em que o elemento Alu do subtipo Ya5 é inserido no éxon 3 no mRNA do *BRCA2* (c.156_157insAluYa5). Estudos identificaram que essa inserção leva ao salto *in-frame* do E3 com a deleção dos resíduos de 23 a 105 na proteína, região que contém dois sítios de ativação de transcrição, a qual é necessária para a fosforilação de *BRCA2*. O salto do E3 já foi visto também a níveis baixos em células normais e com alteração de *splicing* parcial de 5% em amostras normais. Ao fazer a análise direcionada ao transcrito, o padrão de bandas da paciente se diferenciou um pouco dos controles, com uma diminuição da proporção do transcrito *full lenght* e um aumento da banda com salto do E3 (Figura 32).

Figura 32 - Imagem do gel do Bioanalyzer do paciente RNAP-03 para análise da variante *BRCA2*:c.156_157insAlu

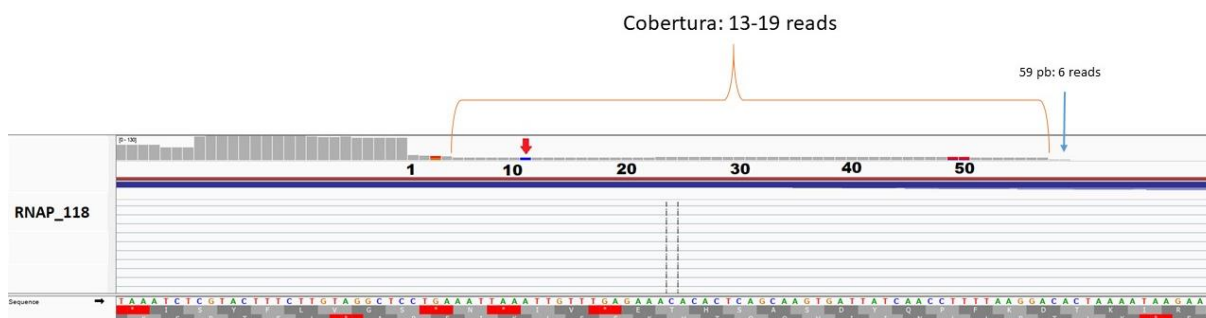


Fonte: Autoria própria.

Legenda: Na primeira canaleta do gel (a) e (b) está a amostra do paciente e nas demais canaletas as amostras controles. A seta azul indica o *amplicon* esperado com 636 pb (FL), e a seta vermelha indica a possível banda que representa o *amplicon* com o salto do E3, o qual também pode estar presente nos CTRLs, os quais apresentam bandas mais fracas de tamanhos aproximados. Observa-se também no paciente uma banda exclusiva forte com tamanho em torno de 70 pb, para a qual não foi possível prever a origem.

A paciente RNAP-118 possui a variante *BRCA1*: c.213-11T>G que está localizada no íntron 4, e ao fazer a inspeção visual do BAM do sequenciamento de RNA-cap no IGV conseguimos identificar uma retenção de 59 pb do íntron 4 ao início do E5 (Figura 33). Também conseguimos identificar a ligação do éxon 4 (éxon anterior) ao primeiro nucleotídeo dessa retenção que é suportado por 75 leituras, enquanto apenas 14 leituras suportam o evento de *splice* normal que pode ser visualizado na Figura 34. Essa retenção provavelmente aconteceu, pois a variante enfraqueceu o sítio aceptor canônico de *splice*, fortalecendo um sítio críptico de *splice* na posição 60 e 61 pb de distância do E5. Esse evento de retenção já tinha sido descrito por alguns laboratórios no ClinVar, sendo um evento que leva à criação de um códon de parada prematuro e NMD.

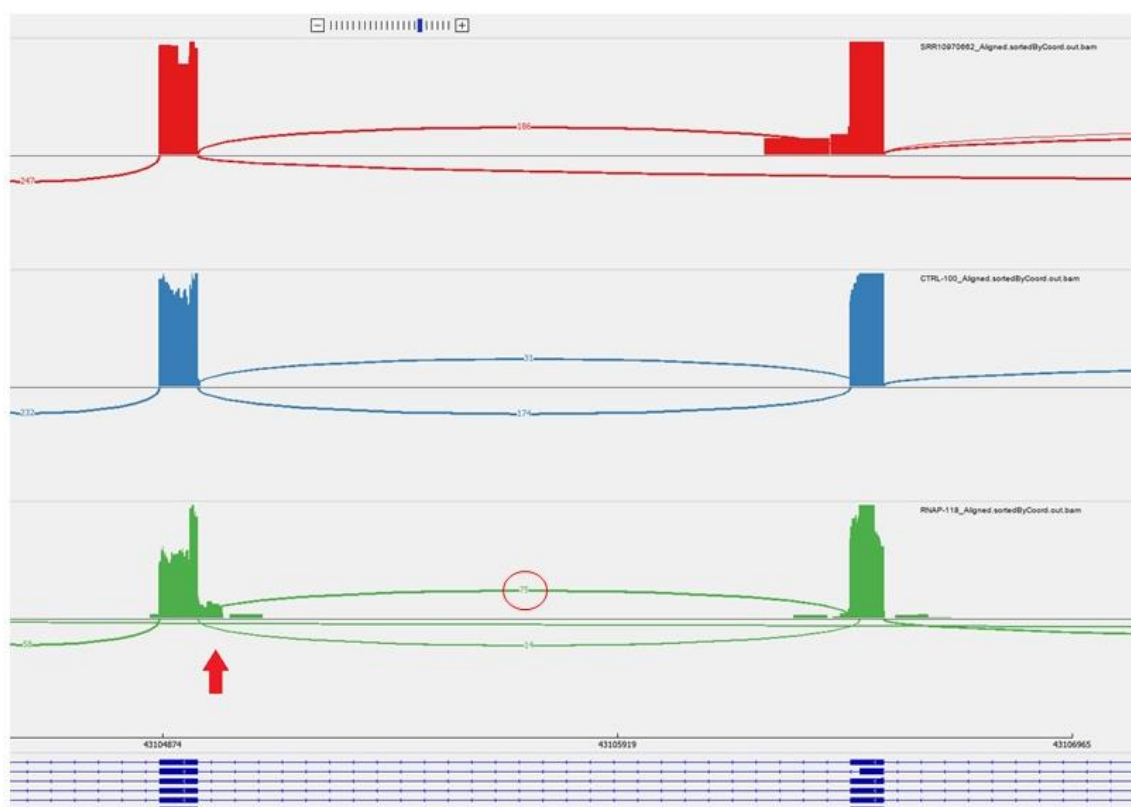
Figura 33 - Visualização da retenção 59 pb do íntron 4, devido a variante *BRCA1*:c.213-11T>G



Fonte: Autoria própria.

Legenda: Evento com cobertura de 13 a 19 leituras nos 59 nucleotídeos anteriores do éxon 5.

Figura 34 - Visualização do gráfico de sashimi gerado através do IGV



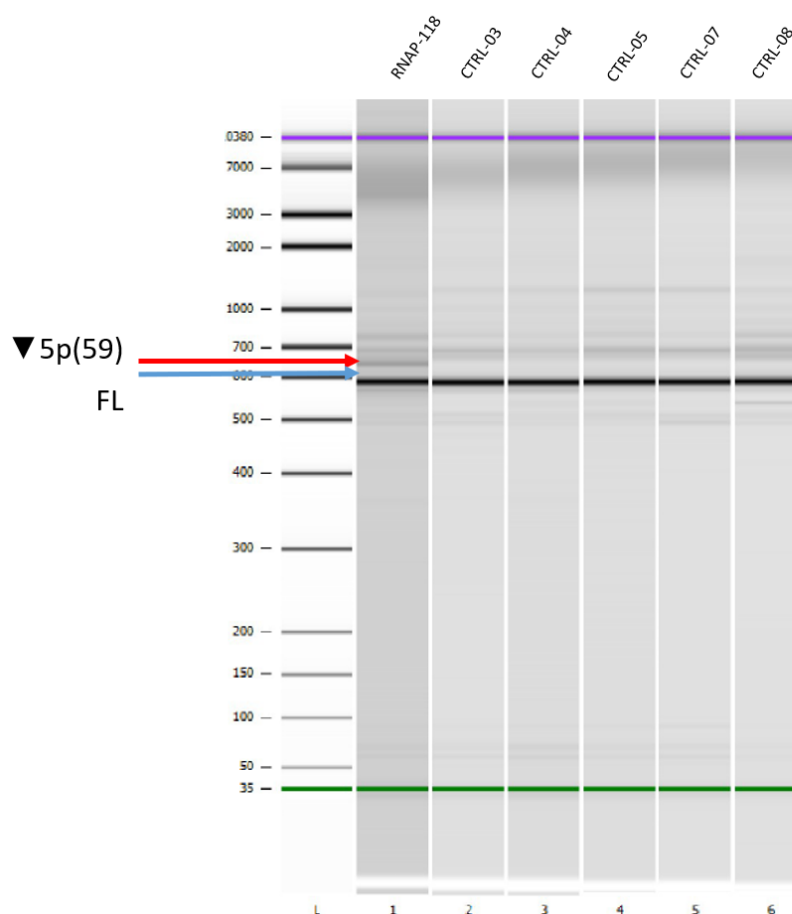
Fonte: Autoria própria.

Legenda: As duas primeiras tracks são o CTRL externo e o CTRL externo, respectivamente, e a última track é a do paciente. Pode-se observar que apenas no paciente é visto esse evento de *splice* em que ocorre a união do E4 aos nucleotídeos que foram retidos do final do íntron 4 (seta vermelha), não visto nos controles.

Através da análise direcionada ao transcrito identificou exclusivamente uma banda maior em comparação com o produto principal amplificado esperado, que pode ser indicativo dessa retenção (Figura 35). Essa paciente possui sete SNPs marcadores (cinco no éxon 10, um

no éxon 12 e outro no éxon 15), em que a maior parte das VAFs estão na proporção de 70% e 30%, apenas a variante c.3113A>G no éxon 10 e c.4308T>C no éxon 12 que estavam balanceadas.

Figura 35 - Imagem do gel do Bioanalyzer do paciente RNAP-118 para análise da variante *BRCA1*:c.213-11T>G



Fonte: Autoria própria.

Legenda: Na primeira canaleta está a amostra do paciente, a seta azul indica o *amplicon* principal que tem 590 pb (FL), a seta vermelha indica o *amplicon* alterado esperado de 649 pb, em que são retidos 59 pb do íntron 4 no início do E5. Enquanto as próximas bandas com peso molecular mais alto nos CTRLs têm tamanho entre 670 a 690 pb.

5.5.3. Resultados – grupo de pacientes com VUS preditas de impactar *splice*

Na etapa de validação clínica foram avaliadas 14 variantes únicas em 18 pacientes nas três ferramentas e apenas 5 variantes únicas apresentaram efeito em *splice* detectado (*CHEK2*: c.320-5T>A, *NBN*: c.2071-4A>G, *MLH1*: c.790C>G, *ATM*: c.7273G>A e *BRCA1*: c.4096+1G>A), sendo que o rMATS detectou todos os 5 eventos, o FRASER2 detectou 2 e o

LeafCutterMD detectou 1. Dessa forma, foi identificado um evento de uso alternativo do sítio 5' de *splice* para a variante em *ATM*, salto de éxon *in-frame* para a variante em *NBN* e saltos de éxon *in-frame* e *out-frame* para as variantes em *CHEK2*, *MLH1* e *BRCA1*.

Através da análise direcionada ao transcrito conseguimos identificar alteração de *splice* para 2 variantes com resultado negativo no RNA-cap: na paciente RNAP-05 (*BRCA2*: c.7618-10T>G), que ocorreu a inserção de 9 nt do íntron 15 ao éxon 16, e nos pacientes RNAP-105 e RNAP-106 (*MUTYH*: c.934-2A>G), que ocorreu a perda de 9 nt iniciais do éxon 11. Não conseguimos fazer esse tipo de análise nas amostras RNAP-07 (*TP53*:c.114A>G) e RNAP-110 (*NBN*: c.2071-4A>G), pois o RIN dessas amostras não estava adequado, assim como nos pacientes RNAP-04 (*RAD51D*:c.80C>A), RNAP-09 e RNAP-09.2 (*MSH2*:del 5'UTR + del parcial do E1), nas quais suas variantes estavam localizadas no E1 ou I1. Porém, conseguimos identificar perda do transcrito que abrigava o alelo mutante através da técnica de desbalanço alélico no paciente RNAP-09. Dessa forma, foram encontrados mais 3 eventos anormais de *splice* (nos genes *BRCA2*, *MUTYH* e *MSH2*), chegando num total de 8/14 VUS confirmadas como spliceogênicas combinando as técnicas RNA-cap, análise direcionada ao transcrito ou desbalanço alélico.

Cada paciente e suas variantes reclassificadas ou não serão descritos com mais detalhes a seguir no próximo tópico 5.6, onde foram reunidos todos resultados referentes às análises do RNA-cap, transcrito direcionado e desbalanço alélico. O resumo dos resultados preditos *versus* obtidos para cada variante estão descritos na Tabela 11, os resultados detalhados dos eventos detectados pelas três ferramentas na Tabela 12 e os Boxplots com as amostras que apresentaram efeito em *splice* na ferramenta rMATS são apresentados nas Figuras 36 e 37.

Tabela 11 - Resultados preditos *versus* obtidos a partir das variantes VUS

(continua)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δscore pos no RNA); (outros preditores)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante	Transcrito direcionado (inspeção visual na eletroforese capilar)	Desbalanço alélico (visualização no IGV dos Alelos alt/ref)
RNAP-03	<i>RAD51C</i>	c.571+4A>G (I3)	DL 0,38 -4; DG 0,24 0	Sem eventos significativos identificados	Nenhuma	Banda normal em relação aos CTRLs.	Não tem SNP marcador.
RNAP-04	<i>RAD51D</i>	c.80C>A (E1)	na; AdRf	Sem eventos significativos identificados	Nenhuma	Variante no E1, não é possível analisar.	SNP marcador é a própria variante. VAF alelo ref/alt (8% / 92%).
RNAP-05	<i>BRCA2</i>	c.7618-10T>G (I15)	AL 0,93 10; AG 0,96 1	Retenção <i>in-frame</i> de 9 nt do íntron 15 no início do E16, inserindo um PTC. r.7617_7618ins7618-9_7618- 1 e p.(Gln2539_Leu2540insPheVal*).	Nenhuma	Banda anormal em relação aos CTRLs.	VCF não disponível para análise de SNP marcador.
RNAP-06	<i>ATM</i>	c.7273G>A (E49)	DL 0,64 34; DG 0,14 -4	Uso alternativo do sítio 5' de <i>splice</i> , com a perda de 37 pb finais do E49, que afeta o domínio FAT, levando o salto parcial <i>out-of-frame</i> , resultando em PTC e previsto de sofrer NMD. r.7270_7307del e p.(Val2424Ilefs*13).	rMATS, FRASER2	Banda anormal em relação aos CTRLs.	SNP marcador é a própria variante. VAF alelo ref/alt (100% / 0%).
RNAP-07	<i>TP53</i>	c.114A>G (E4)	AG 0,61 -1	Não avaliável (baixa cobertura)	Não avaliável	Não foi possível analisar, RIN da amostra insatisfatório.	

Tabela 11 - Resultados preditos *versus* obtidos a partir das variantes VUS

(continuação)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δ score pos no RNA); (outros preditores)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante	Transcrito direcionado (inspeção visual na eletroforese capilar)	Desbalanço alélico (visualização no IGV dos Alelos alt/ref)
RNAP-09 e RNAP-09.2	<i>MSH2</i>	c.del 5'UTR + del parcial do E1 (E1)	na	Deleção da região promotora e do início do E1. Ausência de transcrição neste alelo. Chr2:g.47401398_47403273delinsA lu (del 1876bp).	Nenhuma	Variante de grande deleção na região 5'UTR e parcial do E1, não sendo possível a análise.	SNP marcador c.1168C>T (E7). VAF alelo ref/alt do probanda (100% / 0%).
RNAP-103	<i>BRCA1</i>	c.547+5G>A (I7)	DL 0,25 5; DG 0,10 -1	Sem eventos significativos identificados	Nenhuma	Banda normal em relação aos CTRLs.	Não tem SNP marcador.
RNAP-104	<i>MUTYH</i>	c.1434+2C>T (I14)	na; Mes	Sem eventos significativos identificados	Nenhuma	Banda normal em relação aos CTRLs.	Não tem SNP marcador.
RNAP-105 e RNAP-106	<i>MUTYH</i>	c.934-2A>G (I10)	AL 0,95 -2; AG 0,32 -11	Deleção <i>in-frame</i> dos 9 nt iniciais do E11, perdendo 3 aa do domínio NUDIX (não crítico para a proteína). r.934_942del e p.(Val312_Gln314del).	Nenhuma	Eletroforese capilar não informativa. Visualização do evento no gráfico de sashimi (IGV), a partir do BAM da análise direcionada ao transcrito.	RNAP-105: Não tem SNP marcador. RNAP-106: VCF não disponível para análise.

Fonte: Elaboração dos autores.**Legenda:** DL – perda do sítio doador (Donor Loss); DG – ganho de um sítio doador (Donor Gain); AL- perda do sítio aceptor (Acceptor Loss); AG – ganho de um sítio aceptor (Acceptor Gain); na: não se aplica; AdRf – Ada e Rf Score; Ada – Ad Score apenas; MES – MaxEntScan; PTC – códon de parada prematuro; NMD – *nonsense mediated decay*.

Tabela 11 - Resultados preditos *versus* obtidos a partir das variantes VUS

(continuação)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δscore pos no RNA); (outros preditores)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante	Transcrito direcionado (inspeção visual na eletroforese capilar)	Desbalanço alélico (visualização no IGV dos Alelos alt/ref)
RNAP-108, RNAP-109 e RNAP-123	<i>CHEK2</i>	c.320-5T>A (I2)	na; AdRf	Salto do E3, E4 e E5 <i>out-of-frame</i> perdendo o domínio FHA, r.320_683del e p.(Glu107Valfs*7), resultando em PTC e previsto de sofrer NMD. Salto do E3 e E4 <i>in-frame</i> perdendo 91 aa e afetando domínio FHA r.320_592del e p.(Glu107_Lys197del). Alongamento do E3 (FRASER2).	rMATS, FRASER2	Banda anormal em relação aos CTRLs.	Nenhuma amostra possui SNP marcador.
RNAP-110	<i>NBN</i>	c.2071-4A>G (I13)	AG 0,92 -4; AL 0,13 -16	Salto do E13 <i>in-frame</i> , perdendo 52 aa, afetando o domínio DNA damage repair protein Nbs1 domain. r.1915_2070del, p.(Asn639_Lys690del)	rMATS	Não foi possível analisar, RIN da amostra insatisfatório.	SNP marcador c.102G>A (E2) VAF alelo ref/alt (42% / 58%). c.1197T>C (E10) VAF alelo ref/alt (1% / 99%).
RNAP-114	<i>SMARCA4</i>	c.859+5G>A (I5)	na; AdRf, Mes	Sem eventos significativos identificados	Nenhuma	Duas bandas fracas diferentes dos CTRLs.	Não tem SNP marcador.

Tabela 11 - Resultados preditos *versus* obtidos a partir das variantes VUS

(conclusão)

Paciente	Gene	Variante (c.); éxon (E) ou íntron (I)	Predição do SpliceAI (tipo Δ score pos no RNA); (outros preditores)	Anormalidade de <i>splicing</i> observada. Nomenclatura HGVS r. e p.	Ferramentas que identificaram o <i>splicing</i> aberrante	Transcrito direcionado (inspeção visual na eletroforese capilar)	Desbalanço alélico (visualização no IGV dos Alelos alt/ref)
RNAP-115	<i>MLH1</i>	c.790C>G (E9)	na; Ad	Vários eventos de salto entre os éxons 9, 10 e 11, que está localizado o domínio DNA mismatch repair protein S5. SE (9) r.678_790del e p.(Glu227Serfs*42); SE (10) r.791_884del e p.(His264Leufs*2); SE (9-10) r.678_884del e p.(Glu227_Ser295del) salto <i>in-</i> <i>frame</i> de 69 aa; SE (10-11) r.791_1038del e p.(Arg265Phefs*14);	rMATS	Banda normal em relação aos CTRLs com RNA extraído de mama normal.	A própria variante é um SNP marcador VAF alelo ref/alt (58% / 42%). Outro: c.655A>G (E8) VAF alelo ref/alt (61% / 39%).
RNAP-124	<i>BRCA1</i>	c.4096+1G>A (I10)	DL 0,69 1; DG 0,10 -44	Salto do E10 <i>in-frame</i> , perdendo 1.142 aa e afetando o domínio BRCT. r.671_4096del. p.(Ala224_Leu1365del).	rMATS e LeafCutterMD	Não foi possível analisar.	Cinco SNPs marcadores. c.1067A>G (E10) VAF alelo ref/alt (90% / 10%); c.2612C>T (E10) (34% / 66%); c.3113A>G (E10) (34% / 66%); c.3548A>G (E10) (17% / 83%); c.4837A>G (E15) (82% / 18%).

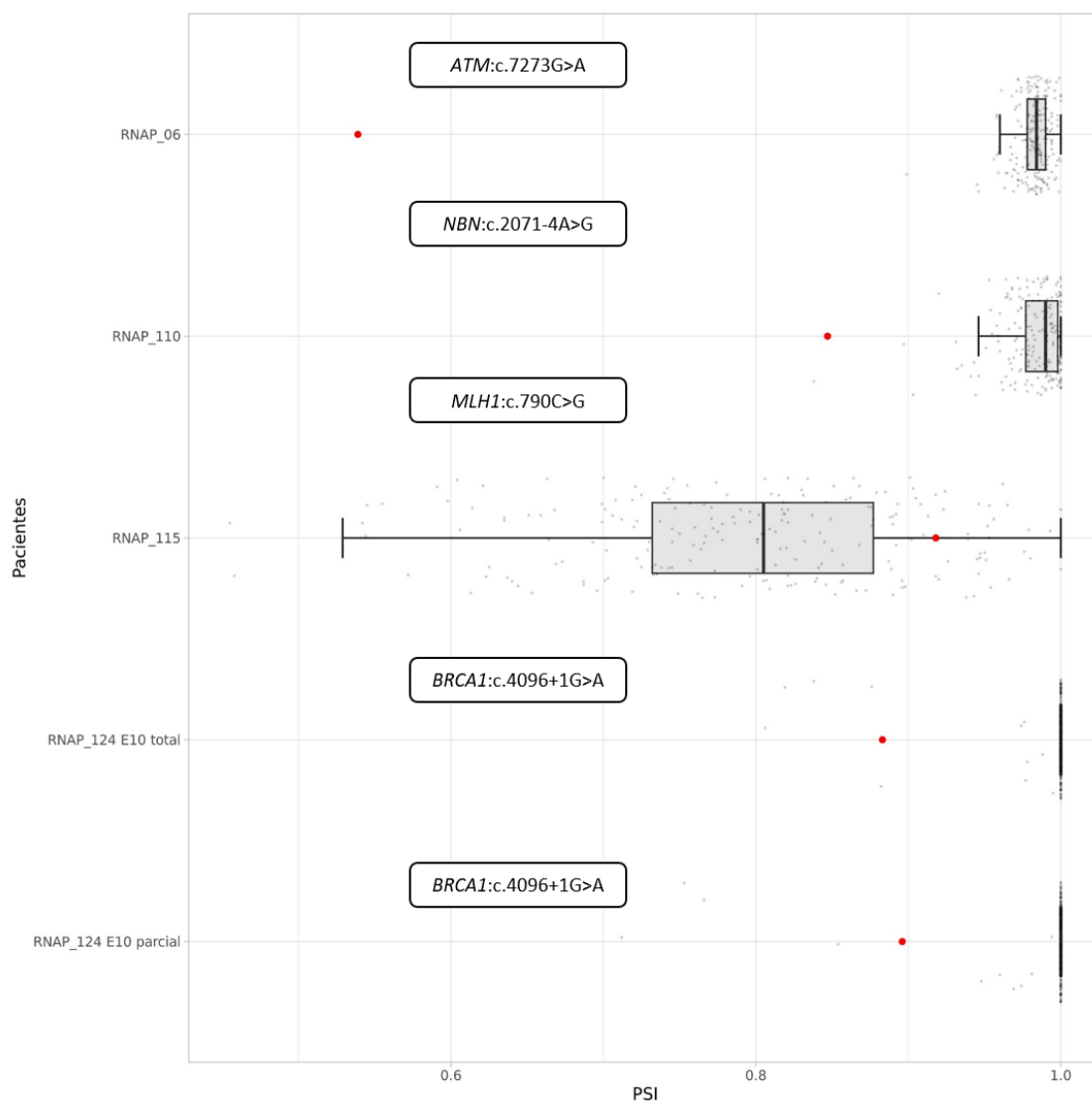
Tabela 12 - Resumo dos resultados gerados pelas três ferramentas para pacientes com VUS que apresentaram eventos de *splice* significativos

Paciente	Variante	Ferramenta	Evento <i>splice</i>	Pvalue	FDR	PSI Ctrl (média)	PSI Caso	Dif. PSI	Predição <i>Splice</i>
RNAP-06	ATM: c.7273G>A	rMATS	A5SS (E49)	0	0	0,899 - 1,0 (0,98)	0,539	0,444	SpAI
		FRASER2	Éxon truncado (49) e uso reduzido do íntron 49	9,945E-06	na	na	na	0,23	SpAI
RNAP-06									
RNAP-108	CHEK2: c.320-5T>A	rMATS	SE (3, 4 e 5)	0	0	1,0-1,0 (1,0)	0,873	0,127	AdRf
		FRASER2	Alongamento do E3	0,01	na	na	na	0,16	AdRf
RNAP-108									
RNAP-109	CHEK2: c.320-5T>A	rMATS	SE (3, 4 e 5)	0	0	1,0 - 1,0 (1,0)	0,852	0,148	AdRf
RNAP-123	CHEK2: c.320-5T>A	rMATS	SE (3, 4 e 5)	0	0	1,0 - 1,0 (1,0)	0,931	0,069	AdRf
		rMATS	SE (3 e 4)	0	0	0,989- 1,0 (1,0)	0,929	0,071	AdRf
RNAP-123									
RNAP-110	NBN: c.2071-4A>G	rMATS	SE (13)	0	0	0,838 - 1,0 (0,98)	0,847	0,136	AdRf, MES, SpAI
RNAP-115	MLH1: c.790C>G	rMATS	SE (10 e 11)	6,22E-06	1,50E-04	0,455 - 1,0 (0,80)	0,918	0,12	Ad, na
RNAP-124	BRCA1: c.4096+1G>A	rMATS	SE (10) parcial	0	0	0,662-1,0 (0,99)	0,896	0,11	AdRf, MES, SpAI
		LeafCutterMD	SE (10) parcial	0,02	na	na	na	na	
		rMATS	SE (10) total	0	0	0,806-1,0 (1,0)	0,883	0,1	
RNAP-124									

Fonte: Elaboração dos autores.

Legenda: SE- Salto de éxon; A5SS- Uso alternativo do sítio 5' *splice*; **SpAI**- SpliceAI; **AdRf** – Ada e Rf Score; **Ad, na** – apenas Ada Score; **MES** – MaxEntScan; **na** – não se aplica.

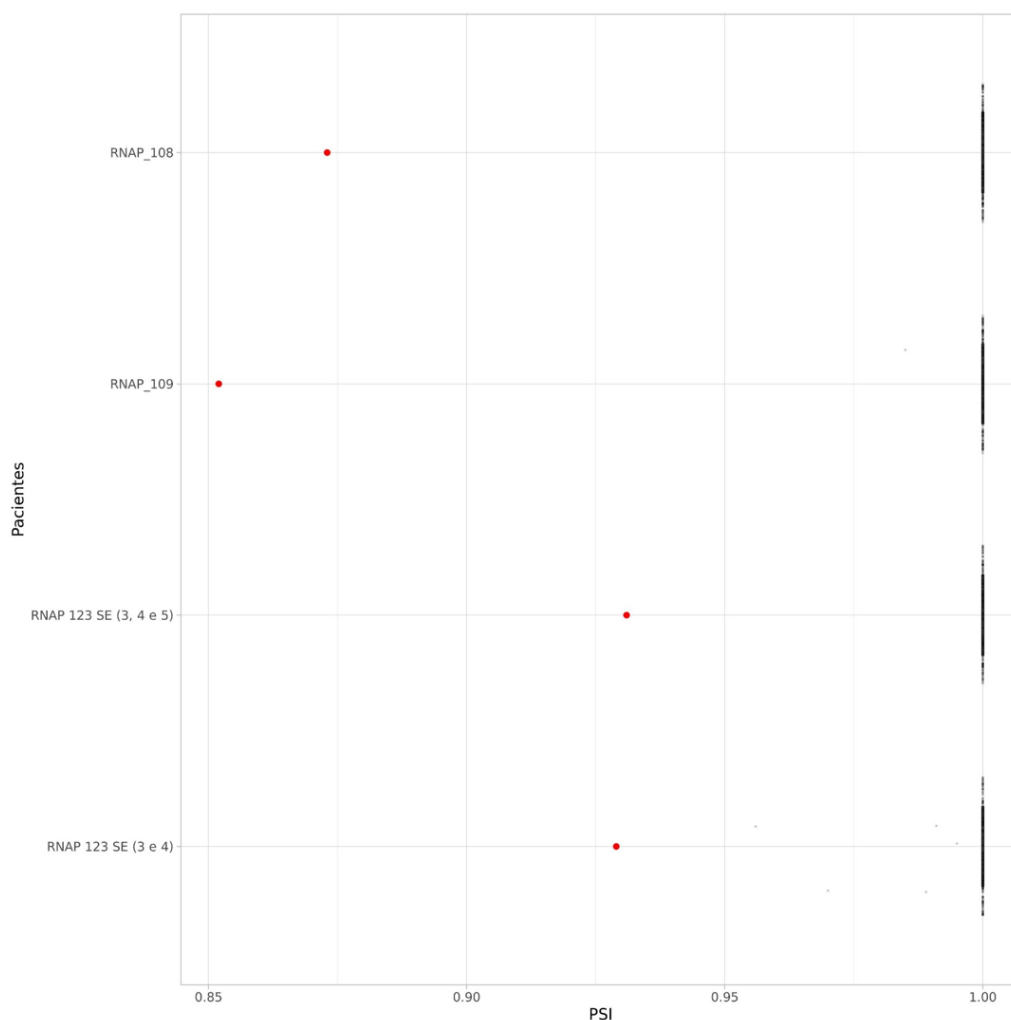
Figura 36 - Boxplot com a distribuição do PSI dos controles x PSI dos pacientes com VUS (sem variantes em *CHEK2*)



Fonte: Autoria própria.

Legenda: Os boxplots em cinza mostram o intervalo da distribuição dos valores de PSI dos controles negativos, e os pontos vermelhos em destaque os valores do PSI dos pacientes com as variantes VUS em questão, pontos pequenos em cinza são a dispersão dos PSI dos controles.

Figura 37 - Boxplot com a distribuição do PSI dos controles x PSI dos pacientes com VUS com variantes em *CHEK2:c.320-5T>A*



Fonte: Autoria própria.

Legenda: Os boxplots em cinza mostram o intervalo da distribuição dos valores de PSI dos controles negativos, e os pontos vermelhos em destaque os valores do PSI dos pacientes com as variantes VUS em questão, pontos pequenos em cinza são a dispersão dos PSI dos controles.

5.6. Reclassificação das VUS

A partir da reclassificação das VUS (realizada seguindo os critérios descritos no item 4.7) dividimos as variantes em três categorias principais: (i) VUS para as quais foram encontradas evidências suficientes para elevá-las para P ou PP ($VUS \rightarrow P/PP$); (ii) VUS para as quais as evidências não foram suficientes para sua reclassificação ($VUS = VUS$); e (iii) as que foram encontradas evidências suficientes para rebaixá-las para B ou PB ($VUS \rightarrow B/PB$).

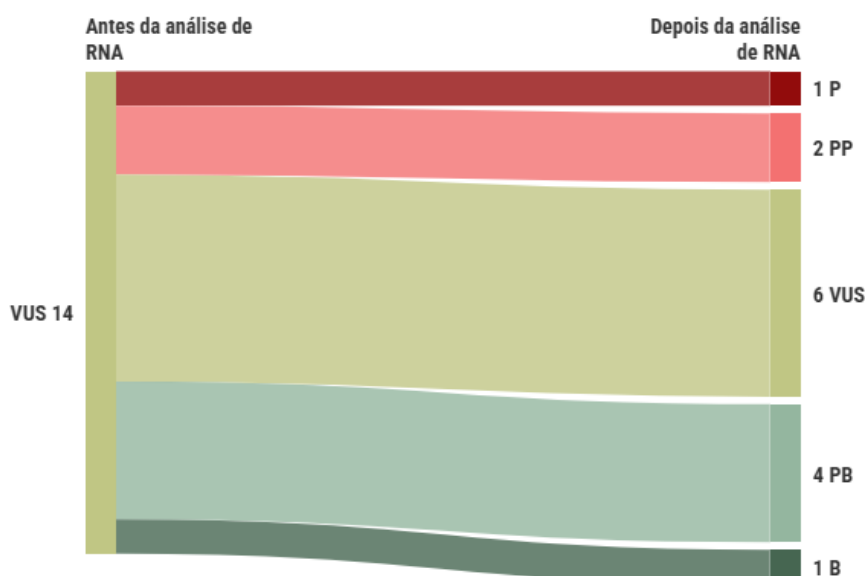
No geral, conseguimos reclassificar 3 variantes para P/PP, 5 para B/PB e 6 continuaram VUS, totalizando uma taxa de reclassificação de VUS de 57% (8/14). Na Tabela 13 a seguir está o resumo dos critérios ACMG aplicados antes e depois do RNA-cap e, na Figura 38, uma visão geral da reclassificação final de variantes. Os resultados das análises direcionadas ao transcrito obtidos no Bioanalyzer através da eletroforese capilar serão apresentados a seguir individualmente para cada paciente.

Tabela 13 - Resumo da reclassificação de VUS com os critérios ACMG aplicados antes e depois da análise de RNA-cap

Paciente	Variante (c.)	Antes RNA-cap	Depois do RNA-cap	Classificação final (points-based)
RNAP-03	<i>RAD51C</i> c.571+4A>G	PM2_sup, PP3	PM2_sup (1), PP3 (1)	VUS
RNAP-04	<i>RAD51D</i> :c.80C>A	PM2_sup	PM2_sup (1)	VUS
RNAP-05	<i>BRCA2</i> :c.7618-10T>G	PM2_sup, PP3	PVS1_strong (RNA) (4), PM5_strong (PTC) (4), PM2_sup (1)	Provavelmente patogênica
RNAP-06	<i>ATM</i> :c.7273G>A	PM2_sup, PP3	PVS1 (RNA) (8), PM2_sup (1)	Provavelmente Patogênica
RNAP-07	<i>TP53</i> :c.114A>G	PM2_sup, PP3	PM2_sup (1), PP3 (1)	VUS
RNAP-09 e RNAP-09.2	<i>MSH2</i> :del 5'UTR + del parcial do E1	Laboratório externo	PVS1 (8), PM2_sup (1) e PP4 (1)	Patogênica
RNAP-103	<i>BRCA1</i> :c.547+5G>A	PM2_sup, PP3	PM2_sup (1), PP3 (1), BP7_strong (RNA) (-4)	Provavelmente benigna
RNAP-104	<i>MUTYH</i> :c.1434+2C>T	PM2_sup, BP4	PM2_sup (1), BP4 (-1), BP7_strong (RNA) (-4)	Provavelmente benigna
RNAP-105 e RNAP-106	<i>MUTYH</i> :c.934-2A>G	PVS1, BS1, BS2	PVS1(RNA_NA) (0), BS1_strong (-4), BS2_strong (-4)	Benigna
RNAP-108, RNAP-109 e RNAP-123	<i>CHEK2</i> :c.320-5T>A	PM2_sup, BS2_sup, BP4*	BS2_sup (-1), BP4 (-1), PVS1 (RNA_NA)	Provavelmente benigna
RNAP-110	<i>NBN</i> :c.2071-4A>G	PM2_sup, PP3	PM2_sup (1), PP3 (1), PVS1_mod (RNA_NA) (0)	VUS
RNAP-114	<i>SMARCA4</i> :c.859+5G>A	PM2_sup	PM2_sup (1), BP7_strong (RNA) (-4)	Provavelmente benigna
RNAP-115	<i>MLH1</i> :c.790C>G	PM2_sup, PP3	PM2_sup (1), BP4 (-1), BP7_Strong (RNA_NA) (0)	VUS
RNAP-124	<i>BRCA1</i> :c.4096+1G>A	PVS1_strong, PM2_sup e PP3	PVS1_mod (RNA) (2), PM2_sup(1), PP3 (1)	VUS

Fonte: Elaboração dos autores.

Figura 38 - Reclassificação final de variantes após o estudo combinando as técnicas de RNA-cap, análise direcionada ao transcrito e desbalanço alélico

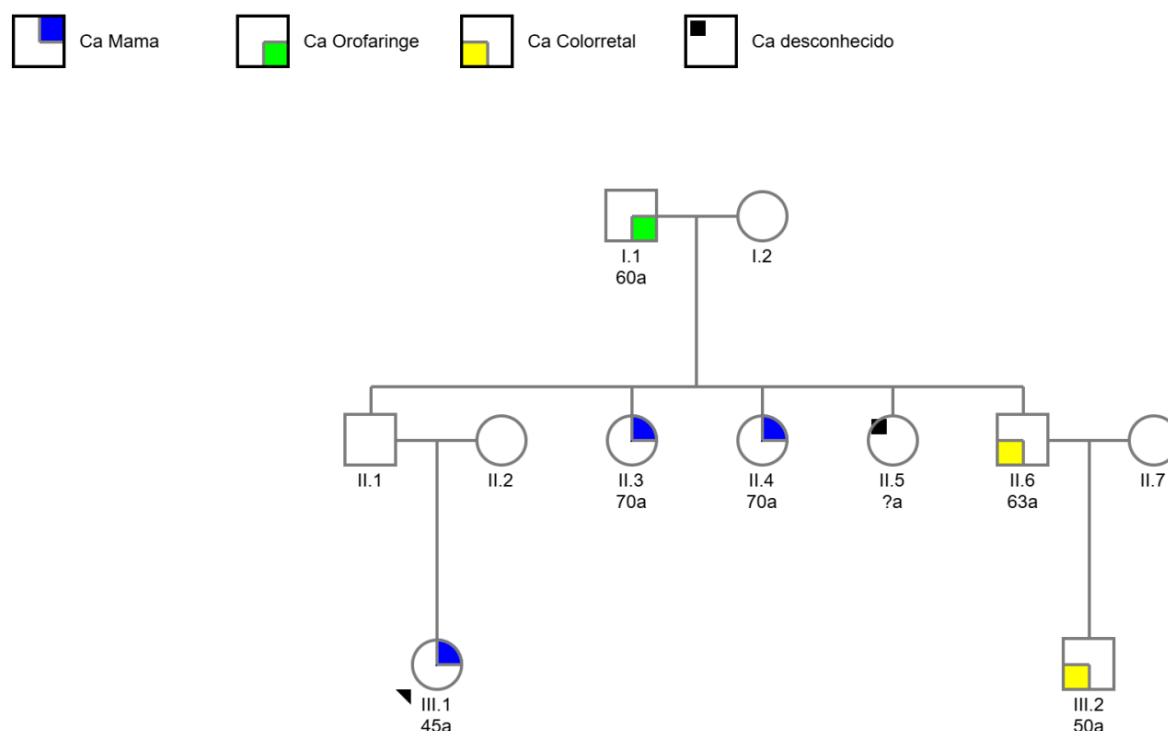


Fonte: Autoria própria (feito no Flourish).

Legenda: Após a análise combinada foi possível reduzir o número inicial de VUS encontradas no sequenciamento de painel de DNA.

5.6.1. VUS reclassificadas para P/PP

A paciente **RNAP-05** tem um histórico familiar relevante afetando todas as 3 gerações do lado paterno, conforme demonstrado no heredograma abaixo (Figura 39), a qual é portadora da variante em *BRCA2*: c.7618-10T>G que está localizada no íntron 15 a dez nucleotídeos de distância do E16. Esta variante não apresentou nenhum evento de *splice* significativo detectável pelas 3 ferramentas, porém foi reclassificada no ClinVar no final do ano de 2023 para provavelmente patogênica. Este depósito foi feito pela Ambry que detectou *splicing* anormal, na qual a variante enfraquece o sítio nativo aceptor de *splice* ao criar ou aumentar a força de um novo sítio aceptor de *splice* introduzindo um códon de parada prematuro, e gerando um transcrito que sofre NMD. O VCF do paciente não está disponível para análise de SNP marcador.

Figura 39 - Heredograma da paciente RNAP-05

Fonte: Autoria própria.

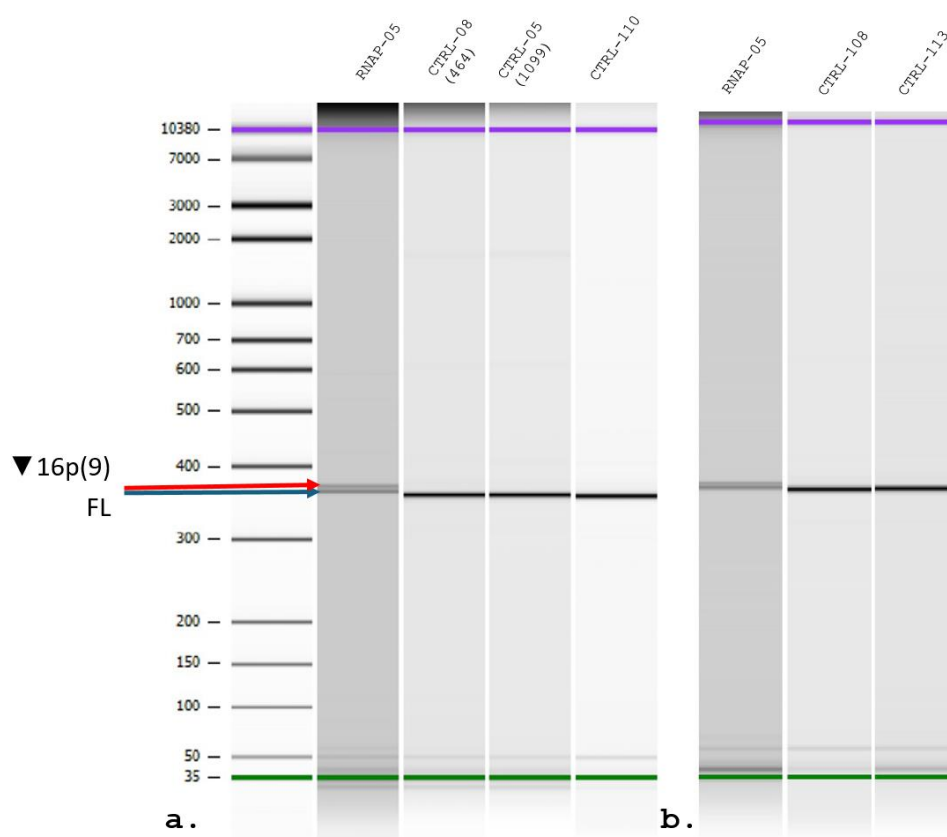
Legenda: O heredograma demonstra as três gerações afetadas predominantemente por câncer de mama e colorretal, na qual a legenda encontra-se na parte superior esquerda da imagem. A idade na imagem se refere a idade ao diagnóstico dos respectivos cânceres.

Através da análise direcionada ao transcrito foi possível observar a retenção no mRNA de 9 nucleotídeos antes do E16 (TTTGTGTAG; mRNA r.7617_7618insuuuguguag) com uma cobertura aproximada de 900X presente apenas na paciente, a qual insere um códon de parada prematuro 'TAG → UGA' (p.Gln2539_Leu2540insPheVal*) (Figura 40), consequentemente levando à perda de função desse alelo.



Ainda com os *amplicons* gerados através da técnica de análise direcionada ao transcrito e realizando a eletroforese capilar, foi possível observar uma segunda banda muito próxima e acima da banda do *amplicon* principal, evidenciando um fragmento maior que é resultado da retenção adicional desses 9 nt, conforme demonstrado na Figura 41. A partir dos resultados desse estudo e demais evidências, foi possível reclassificar a variante como provavelmente patogênica (PVS1_strong (RNA), PM5_strong (PTC), PM2_sup).

Figura 41 - Imagem do gel do Bioanalyzer do paciente RNAP-05 para análise da variante *BRCA2*:c.7618-10T>G



Fonte: Autoria própria.

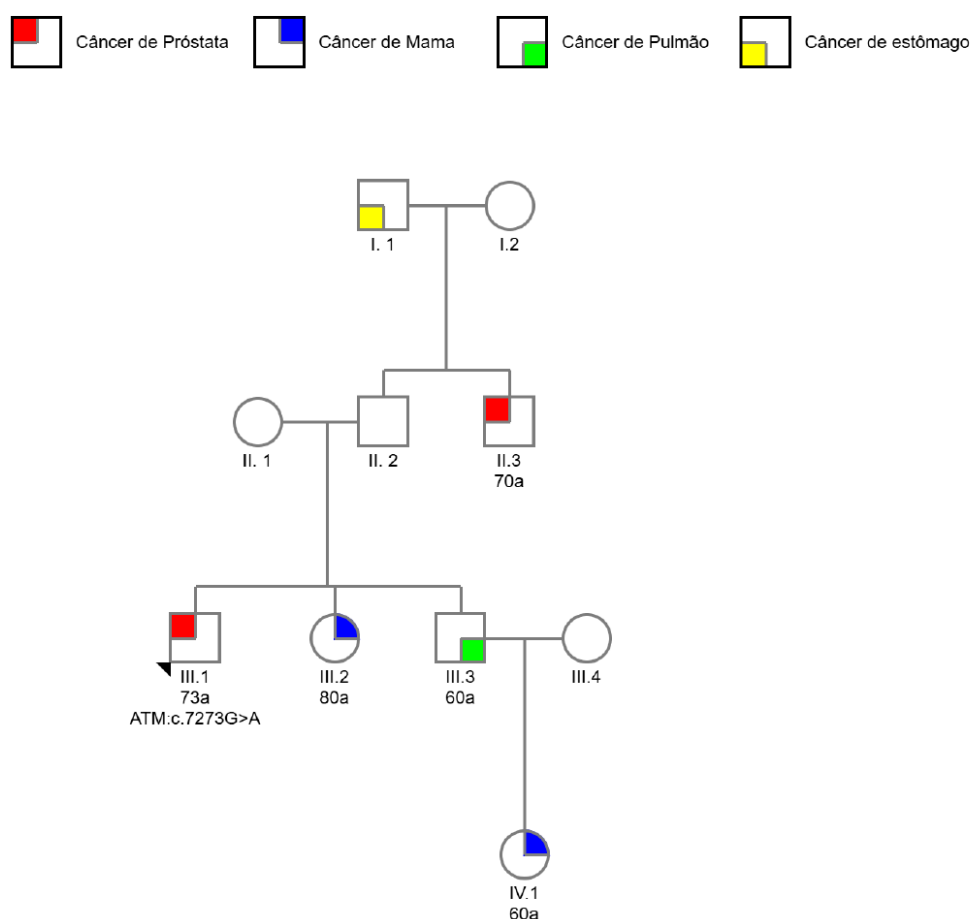
Legenda: No gel (a) e (b) o paciente está na primeira canaleta, possuindo duas bandas com tamanhos muito próximos. A seta azul indica o *amplicon* esperado de 342 pb (FL), enquanto a seta vermelha indica o transcrito aberrante esperado com 351 pb, devido a inserção de 9 nt no início do E16, o qual está ausente nos CTRLs.

O paciente **RNAP-06** tem um forte histórico de câncer familiar de herança paterna afetando todas as gerações, conforme mostrado no heredograma (Figura 42), o qual é portador da variante exônica em *ATM*:c.7273G>A, que está localizada a 34 nucleotídeos de distância do final do E49. Essa variante apresentou um evento significativo de uso alternativo do sítio 5' de *splice*, tanto pela ferramenta rMATS quanto pelo FRASER2. Nesse evento os controles possuíam o éxon 49 inteiro com inclusão de 90% a 100% das leituras, enquanto o paciente tem inclusão de 53% das leituras que cobrem o éxon completo, e 47% das leituras cobrem apenas os 181 pb iniciais do E49, ocorrendo uma deleção parcial dos 38 pb finais. No gráfico de sashimi (Figura 43) é possível ver o respectivo evento. Esse mesmo resultado foi visto nos dois tipos de análise, tanto na análise do BAM do RNA-cap como do BAM do transcrito direcionado, representado na Figura 44. Essa alteração enfraquece o sítio canônico de *splicing* presente na

borda, utilizando um sítio críptico de splice 4 nucleotídeos antes da variante, por isso a alteração na posição 34 pb utiliza um sítio na posição 38 pb antes do éxon causando um salto parcial do E49. Nesse evento o salto parcial é considerado *out-of-frame* formando um códon de parada prematuro depois de 13 aminoácidos, e o transcrito é previsto de sofrer NMD.

Essa variante já é um SNP marcador, e como ocorre o uso do sítio de *splice* 4 nucleotídeos antes da variante, o alelo alterado é deletado, apresentando o alelo referência em quase 100% das leituras e redução de cobertura de leituras até o final do E49 (Figura 44). No ClinVar essa variante tem ID 1686480, com apenas um depósito feito pela Mendelics como VUS sem descrição personalizada.

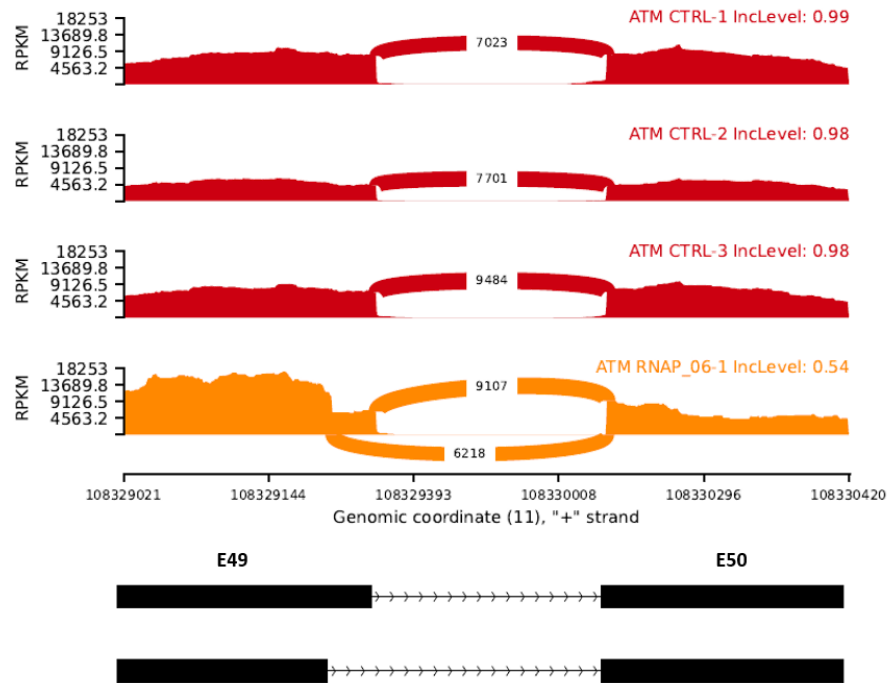
Figura 42 - Heredograma do paciente RNAP-06



Fonte: Autoria própria.

Legenda: O heredograma demonstra as quatro gerações afetadas por diferentes tipos de câncer descritos na parte superior da imagem. A idade na imagem se refere a idade ao diagnóstico dos respectivos cânceres.

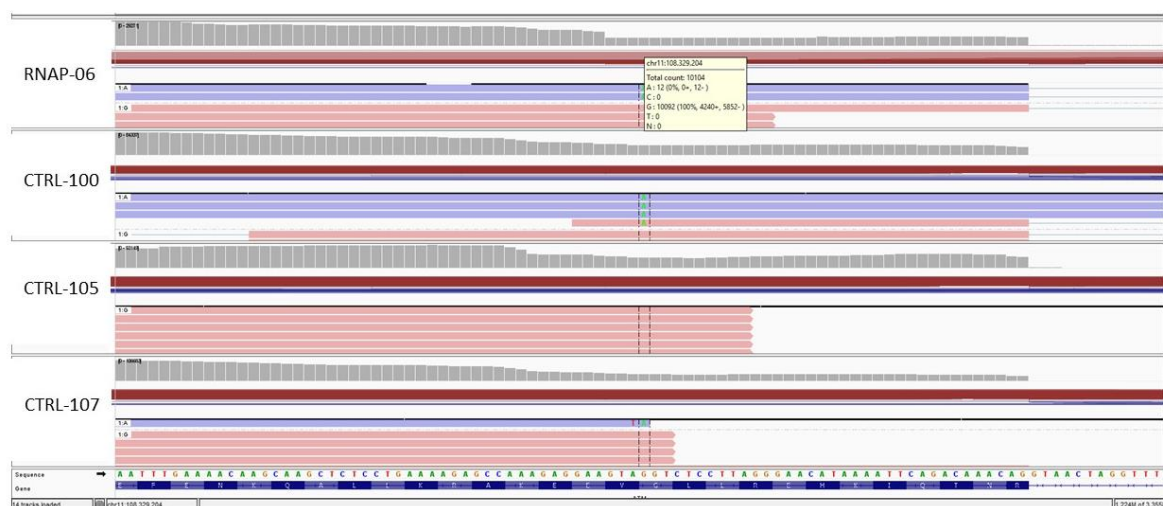
Figura 43 - Sashimi plot mostrando a utilização dos sítios alternativos e canônicos de *splice* no paciente RNAP-06 e três controles



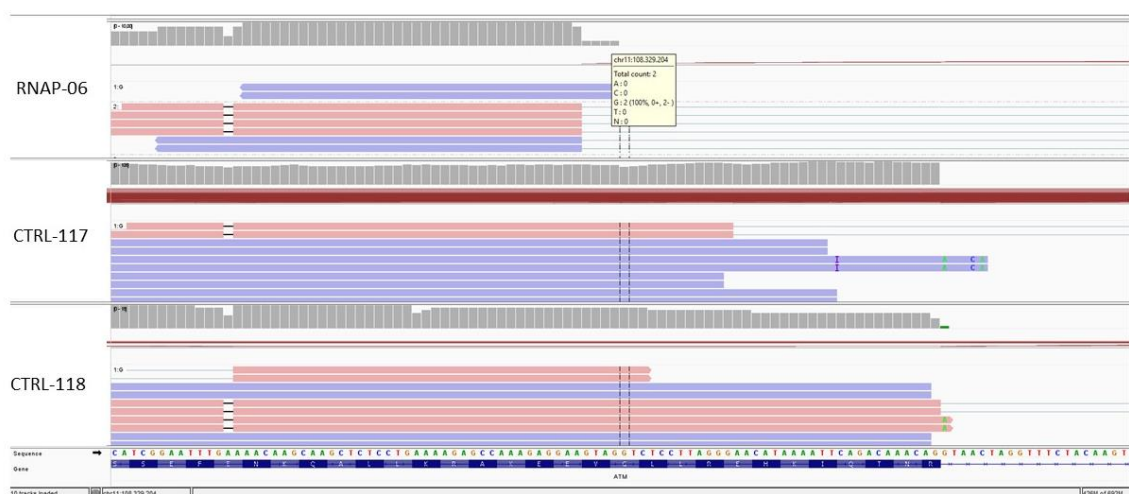
Fonte: Autoria própria.

Legenda: Visualização do evento com três controles externos (cor vermelha) e a amostra do paciente (cor laranja). O uso do sítio alternativo de *splicing* causada pela variante dentro do éxon tem suporte de 6.218 leituras, vista apenas no paciente.

Figura 44 - Análise dos BAMs do paciente RNAP-06 oriundos da análise direcionada ao transcrito e do RNA-cap



a)



b)

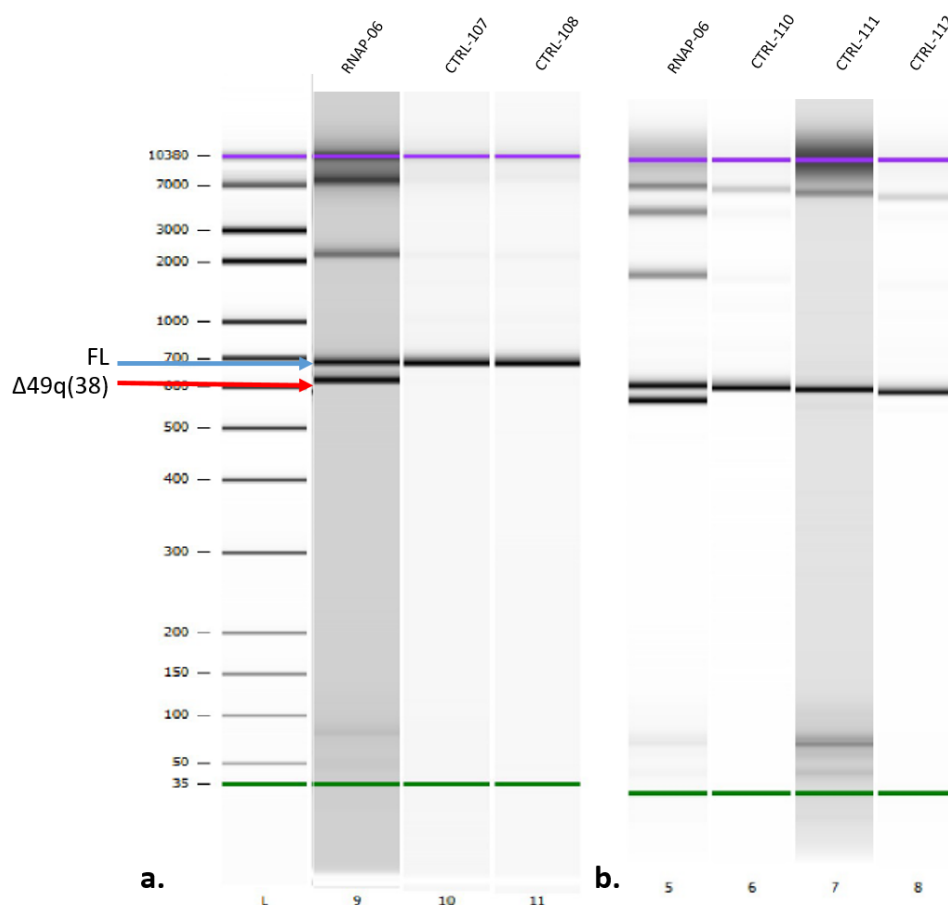
Fonte: Autoria própria.

Legenda: Análises (a) direcionada ao transcrito; (b) RNA-cap. É possível observar uma redução de leituras mais acentuada no paciente do que nos controles a partir da posição 11:108329200, conforme previsto pelo SpliceAI para a formação de um novo sítio doador (GT) a 4 pb antes da variante. A própria variante já é um SNP marcador, podendo observar que ocorre apenas a expressão do alelo referência, isso indica que o alelo alternativo fortalece o sítio de splicing alternativo 4 pb antes (11:108329200-108329201 – ‘GT’), o qual é mais utilizado que o sítio canônico no final do éxon salto, dessa forma esse alelo não chega a ser expresso, apenas o de referência com as leituras que se estendem até o final do éxon 49.

Resultados semelhantes foram obtidos fazendo a inspeção visual da análise direcionada ao transcrito, que na eletroforese capilar apresentou uma banda forte com peso molecular menor

e ausente nos controles. Esse fato corrobora com os resultados encontrados no rMATS, com o éxon 49 mais curto devido ao uso alternativo da região 5' de *splice* dentro do éxon, forma um fragmento de 623 pb visível apenas no paciente (Figura 45). Dessa forma, a partir dos resultados desse estudo e demais evidências, foi possível reclassificar a variante como provavelmente patogênica.

Figura 45 - Imagem do gel do Bioanalyzer da paciente RNAP-06 para análise da variante *ATM:c.7273G>A*



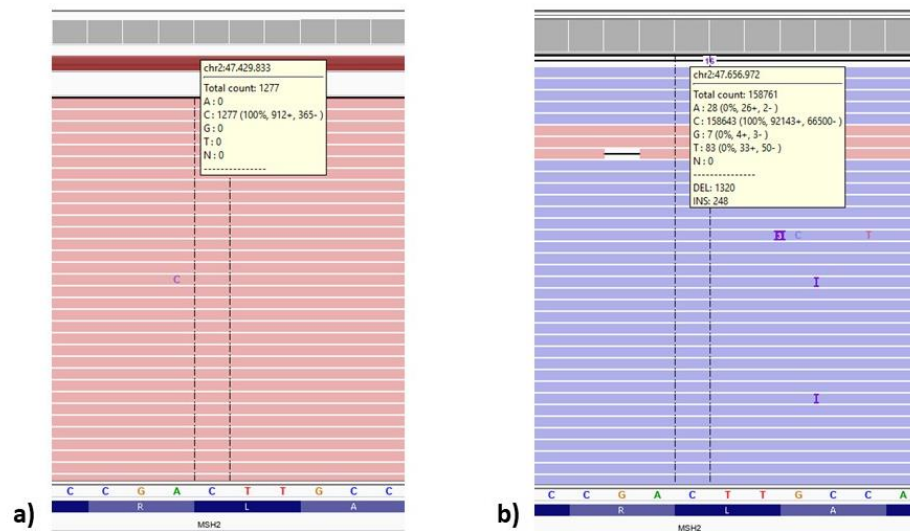
Fonte: Autoria própria.

Legenda: No gel (a) e (b) o paciente está na primeira canaleta, possuindo a banda referente ao *amplicon* esperado com 677 pb (seta azul - FL), e a banda referente ao transcrito anormal com aproximadamente 639 pb (seta vermelha), indicativa do salto parcial do final do E49, ausente nos CTRLs.

Os pacientes **RNAP-09** e **RNAP-09.2** são filho e mãe, respectivamente, de ancestralidade asiática, os quais apresentam forte histórico familiar de câncer e possuem câncer no reto. O heredograma pode ser visualizado através da Figura 46.

alinhamento no IGV e CLCGenomics, foi possível identificar o ponto de quebra da variante, demonstrando que ela leva à deleção da região promotora, UTR e início do éxon 1 (cerca de 1876 bp) (Figura 48), além de ocorrer uma inserção de sequência Alu impedindo a transcrição do alelo alterado. Dessa forma, essa variante foi reclassificada como uma variante patogênica recebendo os critérios PVS1, PM2_supp, PP4 e nomenclatura Chr2:47401398_47403273delinsAlu (del 1876bp), NM_000251.3:c.-1794_82delinsAlu.

Figura 47 - Análise de desbalanço alélico no paciente RNAP-09 com a variante *MSH2*:c.1168C>T (chr 2:47.429.833 – hg38 ou chr2:47.656.972 – hg19)



Fonte: Autoria própria.

Legenda: A expressão do alelo selvagem é observada em 100% das leituras, no sequenciamento de RNA-cap alinhado ao genoma referência hg38 (a), como no sequenciamento de *amplicon* alinhado ao genoma referência hg19 (b).

Figura 48 - Visualização da deleção de aproximadamente 1876 pb da região promotora e parcial do E1 do gene *MSH2*



Fonte: Autoria própria.

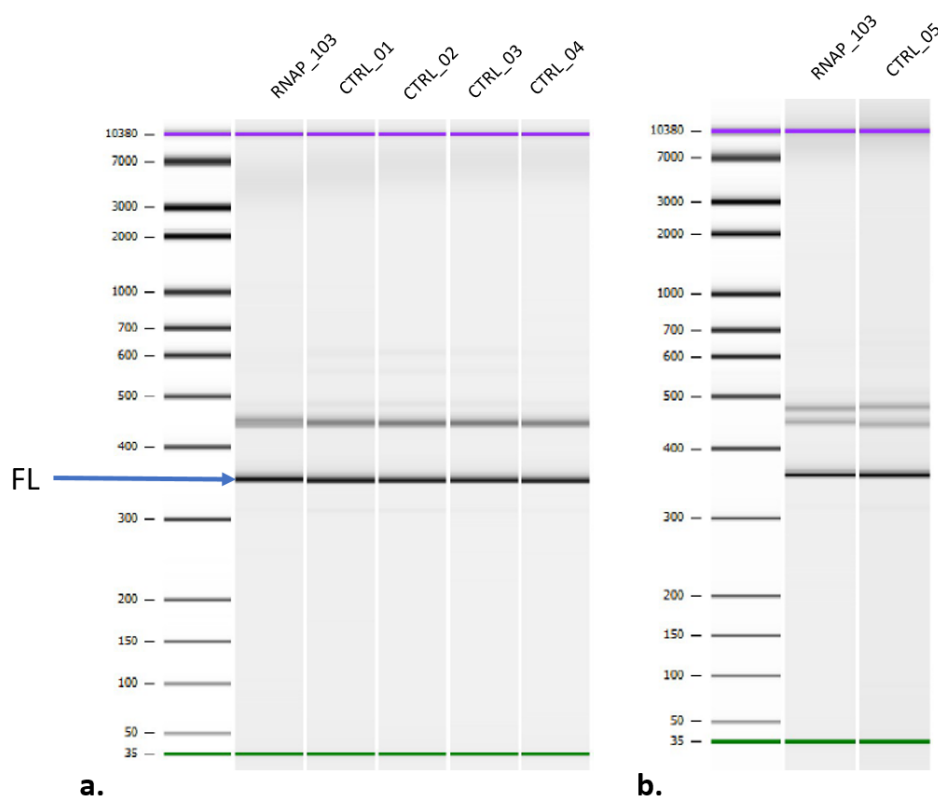
Legenda: Na imagem destaca-se o ponto de quebra (seta em vermelho), que possui queda da cobertura.

5.6.2. VUS reclassificadas para B/PB

Cinco VUS foram rebaixadas para B ou PB nos pacientes: RNAP-103, RNAP-104, RNAP-105 e RNAP-106, RNAP-108, RNAP-109, RNAP-114.

A paciente RNAP-103 com a variante no íntron 7, *BRCA1*:c.547+5G>A não apresentou nenhum evento significativo de *splice*. No ClinVar existe apenas um depósito para essa variante pelo laboratório da Invitae, em que diz que esta não está presente no banco de dados populacional do gnomAD, não foi relatada em estudos com indivíduos em condições relacionadas com alteração em *BRCA1* e que, apesar dos preditores demonstrarem que pode ocorrer dano no sítio consenso de *splice* ainda não foi confirmada a partir de estudos transcricionais. Esta paciente não possui SNP marcador, todas suas variantes exônicas estavam em homozigose. Na análise direcionada ao transcrito, o padrão de bandas do paciente ficou igual ao dos indivíduos controles, conforme pode ser visualizado na Figura 49, sendo indicativo de não alteração de *splice*.

Figura 49 - Imagem do gel do Bioanalyzer do paciente RNAP-103 para análise da variante *BRCA1*: c.547+5G>A



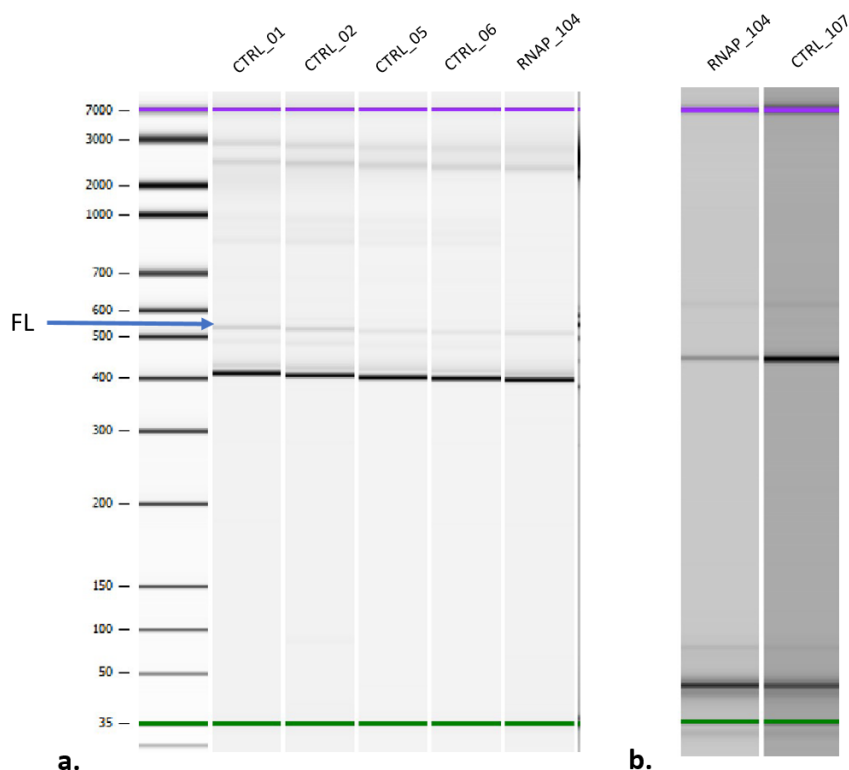
Fonte: Autoria própria.

Legenda: No gel (a) e (b) apresentam dois ensaios, em que o perfil de bandas do paciente está igual ao dos controles, sendo indicativo de não alteração de splice. A seta azul indica o *amplicon* esperado com 365 pb (FL).

O paciente RNAP-104 com a variante *MUTYH*: c.1434+2C>T no íntron 14 não apresentou nenhum evento significativo de *splice* próximo do éxon 14, porém no ClinVar os laboratórios Ambry Genetics e Invitae classificaram a variante como B e PB, respectivamente. É importante destacar que essa variante troca um sítio doador atípico de *splice* ‘GC’ para um sítio doador consenso ‘GT’, dessa forma o sítio fica mais forte que antes e por isso é mais utilizado, sem ocasionar erro de *splice*.

Este paciente também não possui SNP marcador para análise de desbalanço alélico. Na análise direcionada ao transcrito o padrão de bandas do paciente ficou igual ao dos indivíduos controles, conforme pode ser visualizado na Figura 50, sendo indicativo de não alteração de *splice*. Diante dessas evidências e demais estudos, foi possível reclassificar essa variante para provavelmente benigna (PM2_sup, PP3, BP7_strong (RNA)).

Figura 50 - Imagem do gel do Bioanalyzer do paciente RNAP-104 para análise da variante *MUTYH*: c.1434+2C>T



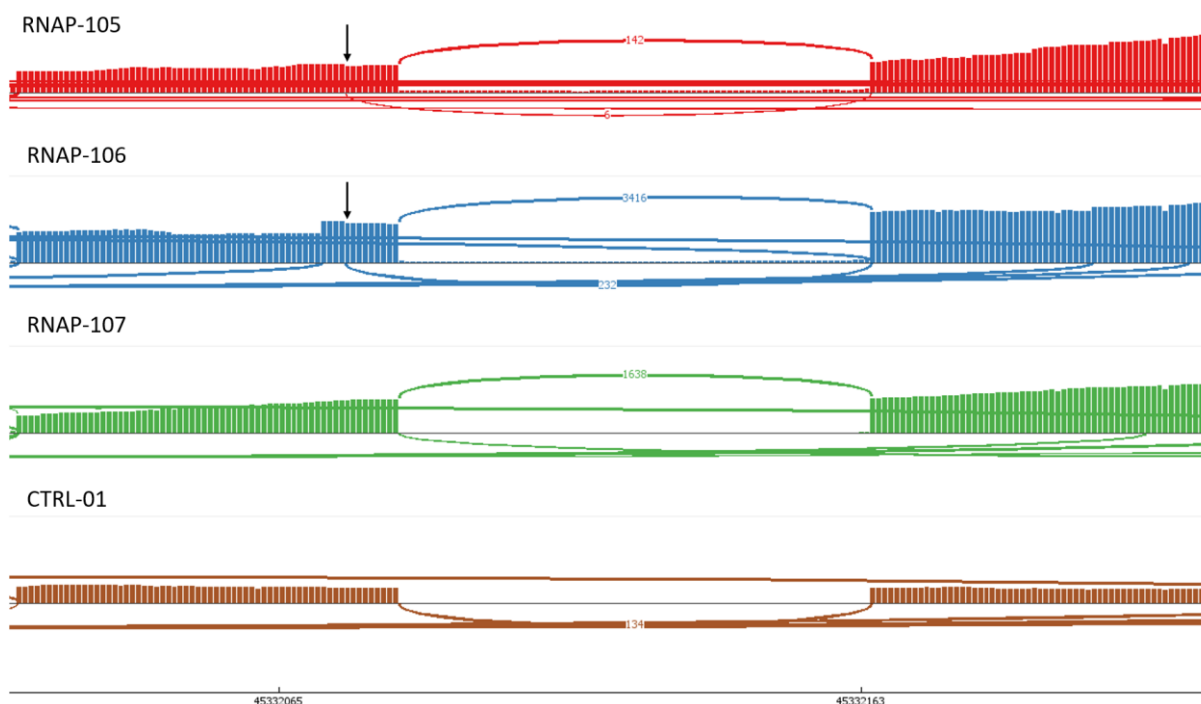
Fonte: Autoria própria.

Legenda: No gel (a) e (b) demonstram 2 ensaios realizados em que o perfil de bandas do paciente está igual ao dos controles, sendo indicativo de não alteração de *splice*. A seta azul indica o *amplicon* esperado com 532 pb (FL).

As pacientes RNAP-105 e RNAP-106 possuem a variante *MUTYH*:c.934-2A>G no íntron 10, e não apresentaram nenhum evento significativo detectado relacionado à variante em análise. A paciente RNAP-105 não possui SNP marcador, enquanto o VCF do paciente RNAP-106 está indisponível. A variante *MUTYH*:c.934-2A>G já foi descrita por Hernandez e colaboradores (2022) que classificaram como provavelmente benigna devido à diversos fatores, um deles é que essa variante foi identificada em mais de 800 indivíduos em homozigose ou heterozigose, nos quais os indivíduos em homozigose (com os dois alelos alterados com essa variante) não apresentavam o fenótipo de Polipose Associada ao *MUTYH* (MAP). A frequência alélica dessa variante é considerada relativamente alta no gnomAD com 1,54%, que leva em conta a frequência máxima subpopulacional (exoma + genoma) com pelo menos 2 mil alelos, e 2,1% no banco de dados de Referência Japonês Multi Omics “jMorp” (Tadaka *et al.*, 2024) baseado em 4.773 genomas de indivíduos japoneses sequenciados. Ao realizarem a análise de

RNA em indivíduos não aparentados observaram uma deleção de 9 nucleotídeos no começo do éxon 11, levando a um evento *in-frame* com perda de 3 aminoácidos num domínio não crítico para a proteína, evento que não foi observado em nenhum indivíduo controle. Conseguimos observar esse mesmo evento apenas com a análise dos BAMs gerados pela técnica da análise direcionada ao transcrito, porém com pouca cobertura, conforme mostrado na Figura 51.

Figura 51 - Imagem do Sashimi plot gerada no IGV das amostras RNAP-105 e RNAP-106



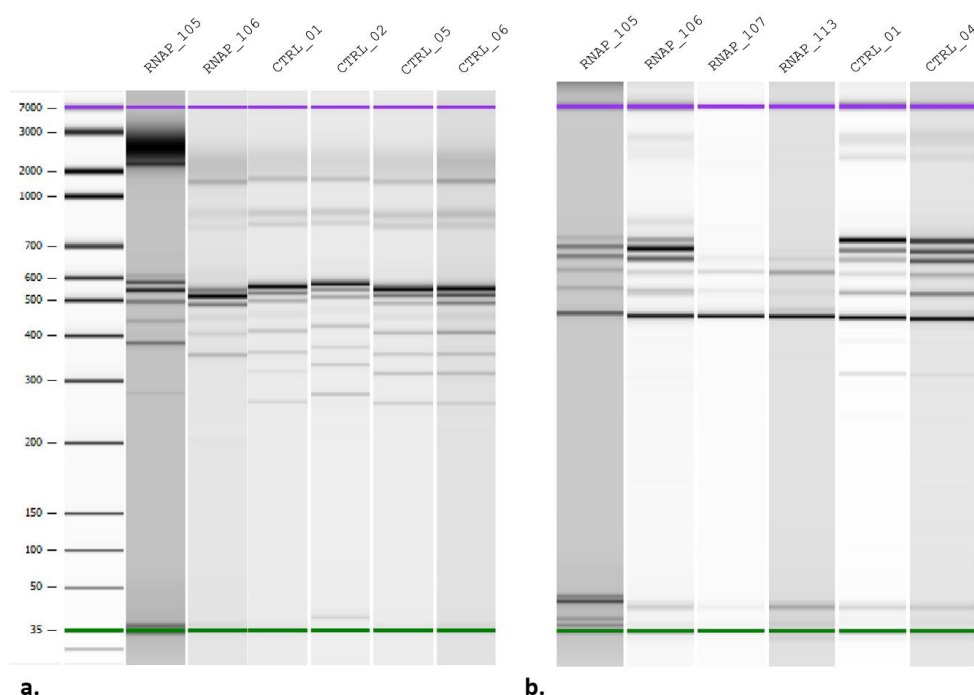
Fonte: Autoria própria.

Legenda: Ambos os pacientes possuem a variante *MUTYH*:c.934-2A>G e apresentam um salto dos primeiros 9 nt do E11 (setas pretas) com cobertura na primeira amostra de 6 leituras e na segunda de 232 leituras, evento não visualizado no RNAP-107 que é uma amostra de cólon normal sem a variante e CTRL-01 é uma amostra de leucócito.

Também através da análise direcionada ao transcrito foi possível observar muitos padrões de bandas, tanto nos pacientes como nos controles, demonstrando a grande quantidade de isoformas que esse gene possui. No primeiro ensaio observa-se que tanto as pacientes RNAP-105 e RNAP-106 têm menos bandas que os indivíduos controles, porém a amostra de RNA da primeira paciente é de reto e a amostra da segunda é de uma metástase de CCR no fígado, enquanto as amostras dos controles são de leucócitos. Por isso, foi realizado um segundo ensaio com amostras de cólon de outros pacientes, RNAP-107 e RNAP-113 (sem variantes patogênicas ou VUS em *MUTYH*), com dois indivíduos controles (leucócitos) para avaliar o padrão de bandas (*splicing*). Ainda assim apareceram muitas bandas, as pacientes RNAP-105

e RNAP-106 tiveram padrões de bandas similares, porém um pouco diferente dos pacientes com RNA do cólon, e semelhantes com as amostras de leucócitos dos controles, exceto pela ausência de duas bandas com peso molecular mais baixo presente apenas nos leucócitos. Isso demonstra que o *splicing* ocorre de forma diferenciada nesses três tecidos avaliados, ficando difícil a comparação de *splicing* normal e anormal, por isso a eletroforese capilar não é informativa nesses casos. Os dois ensaios podem ser visualizados na Figura 52.

Figura 52 - Imagem do gel do Bioanalyzer das pacientes RNAP-105 e RNAP-106 para análise da variante *MUTYH*:c.934-2A>G



Fonte: Autoria própria.

Legenda: No gel (a) e (b) demonstram dois ensaios para a avaliação de *splicing* no tecido de cólon (RNAP-107 e RNAP-113) e de leucócitos (CTRL_01 e CTRL_04) contra os pacientes RNAP-105 e RNAP-106.

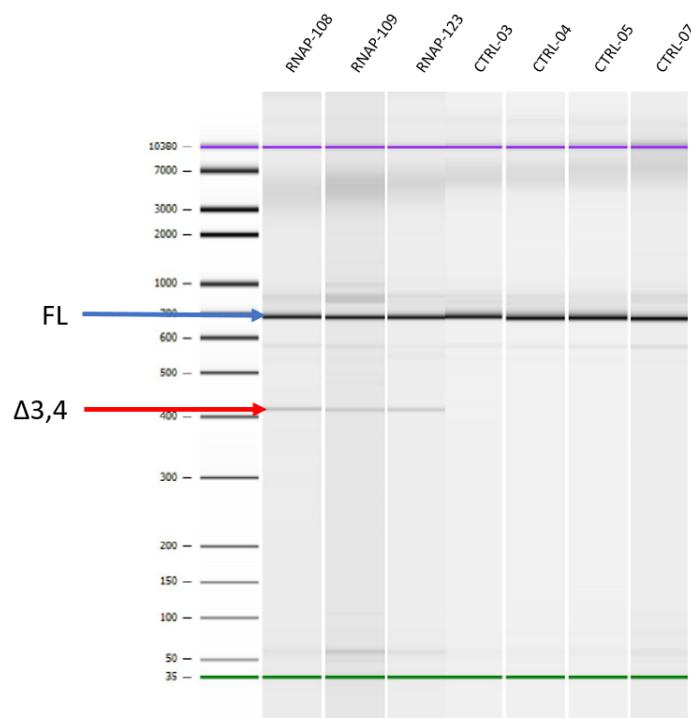
Dessa forma, com o nosso estudo e demais evidências essa variante foi reclassificada para benigna (PVS1(RNA_NA), BS1_strong, BS2_strong).

Os pacientes RNAP-108, RNAP-109 e RNAP-123 possuem a variante *CHEK2*:c.320-5T>A no íntron 2 e, segundo a ferramenta rMATS apresentaram o evento de *splice* significativo de salto do éxon 3 ocorrendo a junção do E2 com o E6, logo houve também o salto dos éxons 4 e 5, e outro evento que apareceu apenas na paciente RNAP-123, que foi o salto do éxon 3 e 4. O evento que salta os três éxons é *out-of-frame* levando a um códon de parada 7 aminoácidos depois, nesse evento que é comum nos três pacientes, o éxon 3 é sempre incluído pelos controles

(PSI= 1) e nas pacientes esse éxon é só um pouco menos incluso, com os PSI iguais a 0,873, 0,852 e 0,929. Enquanto o salto do E3 e E4 é *in-frame* com a perda de 91 aminoácidos, os controles sempre incluem o E3 e no paciente RNAP-123 a inclusão é de 93%, com uma diferença mínima de 6,9%. Os éxons 3 e 4 codificam domínio FHA da proteína *CHEK2*, porém com esse salto, esse domínio não será incluído na proteína. Esse mesmo evento já foi descrito pelo laboratório GeneDx que diz que o salto desses dois éxons ocorre de maneira *in-frame* presente em 20% dos transcritos, e classificaram como provavelmente benigna, juntamente com a Ambry Genetics. Os três pacientes não possuem SNPs marcadores para análise de desbalanço alélico.

Na análise direcionada ao transcrito, os pacientes RNAP-108, RNAP-109 e RNAP-123 tiveram seu RNA extraído do tecido normal da tireoide, junção retossigmóide e mama, respectivamente, e apresentaram em conjunto uma banda exclusiva que é indicativa do salto dos éxons 3 e 4, ausente nos controles, e demonstrado na Figura 53. As bandas são bem parecidas entre tecidos diferentes, podendo relacionar esse resultado ao fato de que o *splicing* do gene *CHEK2* parece ser semelhante em diferentes tecidos apresentando as mesmas isoformas. Com base nas evidências encontradas nesse e demais estudos, essa variante foi reclassificada para provavelmente benigna (BS2_sup, BP4, PVS1 (RNA_NA)).

Figura 53 - Imagem do gel do Bioanalyzer das pacientes RNAP-08, RNAP-09 e RNAP-123 para análise da variante *CHEK2:c.320-5T>A*



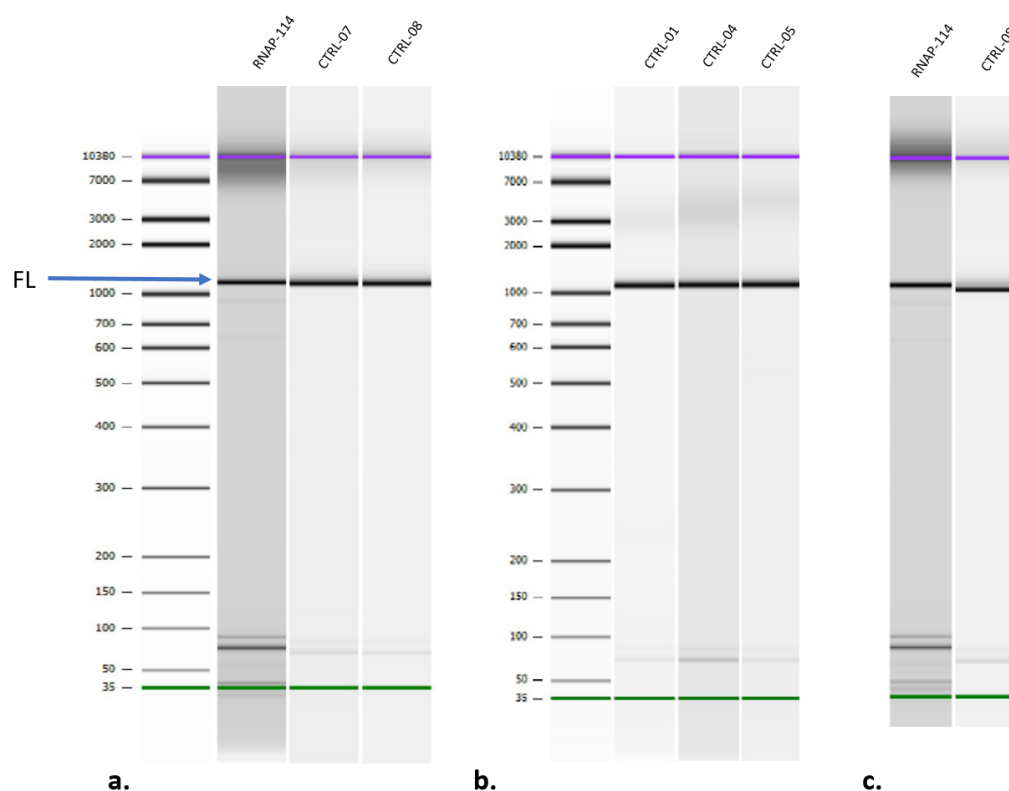
Fonte: Autoria própria.

Legenda: As três primeiras canaletas são das pacientes portadoras da variante. A seta azul indica o *amplicon* esperado com 688 pb (FL) e a seta vermelha indica o *amplicon anormal* esperado com 521 pb, que é indicativo do salto do E3 e E4, ausente nos CTRLs.

A paciente RNAP-114 com variante em *SMARCA4:c.859+5G>A* localizada no íntron 5 não apresentou eventos de *splicing* alternativos significativos em regiões próximas da variante. Segundo a Ambry, que realizou estudo de RNA não observou impacto no mRNA e somando outros parâmetros que não foram associados à doença, a variante foi classificada como PB. Por outro lado, a Invitae classificou a variante como VUS, pois estudos internos mostraram que a variante alterou o *splicing* resultando em uma proteína com impacto desconhecido. Esse paciente não possui nenhum SNP marcador para a análise de desbalanço alélico.

Na análise direcionada ao transcrito, o padrão de bandas do paciente ficou parecido com os indivíduos controles, conforme demonstrado na Figura 54.

Figura 54 - Imagem do gel do Bioanalyzer da paciente RNAP-114 para análise da variante *SMARCA4*:c.859+5G>A



Fonte: Autoria própria.

Legenda: No gel (a), (b) e (c) estão os três ensaios realizados. A seta azul indica o *amplicon* esperado com 974 pb (FL).

Utilizando a adaptação dos códigos ACMG para variantes de *splice* fora do sítio canônico de *splice* (Walker *et al.*, 2023), essa variante segundo o SpliceAI possui Delta Score 0,19 para perda do sítio doador, indicando que não é possível assinalar o código PP3 (predição *in silico* em que ferramentas de predição suportam efeito deletério no gene), pois o ponto de corte para essa ferramenta é acima ou igual a 0,20. Ainda com essa adaptação é possível aplicar o código BP7_strong para variantes fora dos sítios canônicos de *splice* $\pm 1,2$, as quais não se observou nenhum impacto através de dados de RNA-seq. Com modificação desses códigos e utilizando a classificação ACMG baseada em pontos, essa variante é reclassificada para provavelmente benigna (PM2_sup, BP7_strong (RNA)).

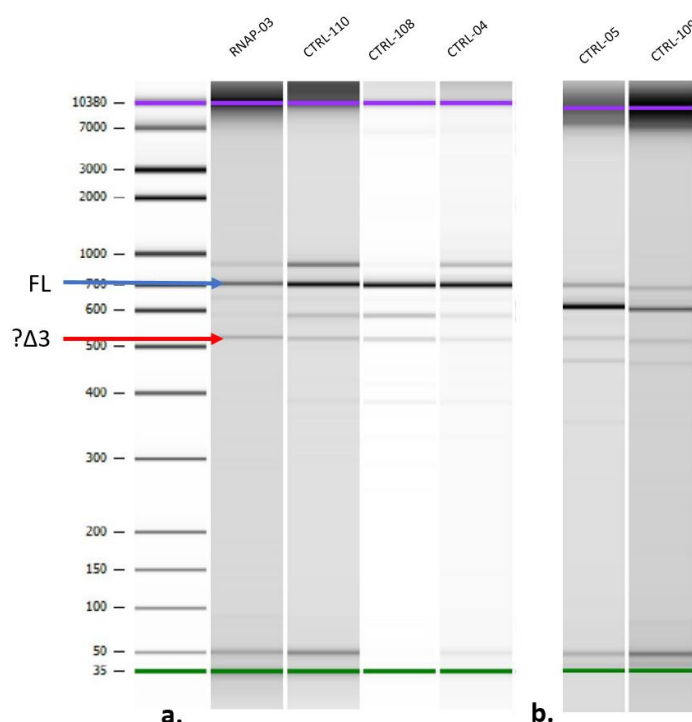
5.6.3. VUS que permaneceram como VUS

As 8 variantes VUS encontradas nos pacientes: RNAP-03, RNAP-04, RNAP-07, RNAP-110, RNAP-115 e RNAP-124 continuaram como VUS.

A paciente RNAP-03 possui uma variante P e a variante intrônica no *RAD51C*:c.571+4A>G localizada no íntron 3 predita de causar dano no *splice* devido a sua pontuação no SpliceAI (Donor Loss 38% e Donor Gain 24%). Porém nenhuma ferramenta detectou nenhum evento significativo e essa paciente não tem nenhuma variante exônica nesse gene para a análise do desbalanço alélico. Segundo os depósitos no ClinVar, Invitae detectou níveis inconclusivos de *splicing* alterado, e estudos internos da Ambry Genetics apontaram *splice* alterado também ocorrendo em indivíduos sem câncer, porém com impacto clínico desconhecido, ambos a classificando ainda como uma variante de significado incerto. O laboratório *Color* classifica a variante como PP, considerando os resultados de estudos que apontam o salto do E3, levando a *frameshift*.

Na análise direcionada ao transcrito, o perfil de bandas ficou parecido com os controles, lembrando que a amostra de RNA do paciente também foi extraída de leucócitos, exceto por alguma variação de intensidade das bandas entre o paciente e controles. Seus fragmentos estão representados por setas na Figura 55. Dessa forma, essa variante continua como VUS, com aplicação dos critérios PM2_sup, PP3.

Figura 55 - Imagem do gel do Bioanalyzer da paciente RNAP-03 para análise da variante *RAD51C:c.571+4A>G*



Fonte: Autoria própria.

Legenda: No gel (a) e (b) mostra os dois ensaios realizados no Bioanalyzer. O *amplicon* esperado em azul com 688 pb (FL) e a possível banda do transcrito sem o éxon 3 está indicado com a seta vermelha com 521 pb, e presente também nos CTRLs conforme descrito pela Ambry Genetics.

A paciente RNAP-04 possui a variante *missense* em *RAD51D:c.80C>A*, que se localiza 2 pb do final do primeiro éxon. Geralmente variantes que se localizam no primeiro éxon ou no primeiro íntron de um gene são desafiadoras para serem analisadas quanto ao seu efeito em *splicing*, pois por ser o primeiro éxon ele não possui éxons a montante (apenas a jusante), logo identificar um possível salto de éxon com o RNA-cap e rMATS não é possível. Por isso, como esperado, a análise do rMATS não evidenciou nenhum evento significativo próximo do éxon 1 (E1). Pelo fato da variante c.80C>A se localizar em região exônica, ela também é considerada um SNP marcador para avaliação de “Desbalanço alélico”, no qual foi visto que o alelo normal está sendo bem mais expresso, com frequência alélica (VAF) de 92%, enquanto o alelo variante foi detectado em apenas 8% das leituras. Esse resultado sugere que ocorra um salto do éxon 1 no alelo contendo a variante, onde está localizado códon de início ‘AUG’. Entretanto existe um códon de início *in-frame* alternativo no E5 e no E6 do transcrito MANE e em outras isoformas alternativas, por isso se ocorrer um salto do E1, devido à presença dessa variante, a proteína ainda será traduzida nos AUGs alternativos *in-frame*, que não alterarão o quadro de leitura da

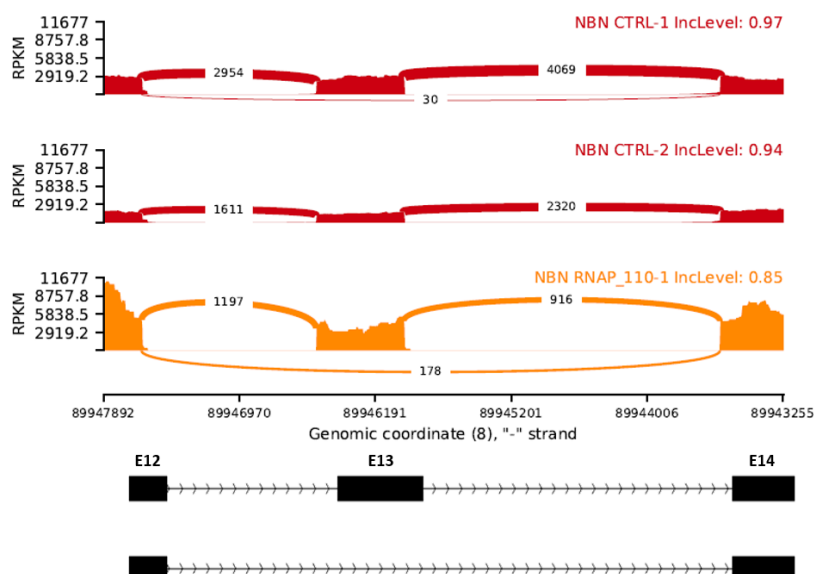
proteína. Essa variante ainda tem alta frequência no banco de dados populacional do Japão não esperada para a doença de 0,26%.

Yang e colaboradores (2022) detectaram uma variante PP na posição c.82C>G (p.Val28Met), a dois nucleotídeos de distância desta variante no próximo códon, a qual foi realizada a análise de *splicing* por RT-PCR, clonagem e sequenciamento de DNA. Identificaram alteração de *splicing* com a perda do final da região 5'UTR junto com o éxon 1 inteiro (c.-86_c.82) compreendendo um fragmento 168 pb. Pelo método de clonagem observaram a abolição completa do *splicing* normal pelo alelo mutante, não gerando nenhum transcrito normal. Devido a evidências insuficientes, essa variante continua sendo uma VUS, com aplicação do critério PM2_sup.

A paciente RNAP-07, com a variante sinônima em TP53:c.114A>G (p.Gln38=) localizada a 17 pb de distância do final do E4, não apresentou evento significativo em *splice* em nenhuma das ferramentas. Uma vez que essa amostra teve uma qualidade ruim de integridade do RNA (RIN 3,4) e baixa cobertura total (77,8% de cobertura em 50X) e no gene TP53 (81,4% em 50X), consideramos que a análise de RNA-cap foi inconclusiva, pois ao observar a cobertura dessa variante no IGV, a qual também é um SNP marcador, só tinha cobertura de 1 leitura. Essa variante também não tem depósito no ClinVar, ou seja, nenhuma evidência clínica. Devido a evidências insuficientes, essa variante continua sendo VUS, com aplicação dos critérios PM2_sup, PP3.

A paciente RNAP-110, com variante em NBN:c.2071-4A>G localizada no íntron 13 apresentou salto do éxon 13. Enquanto os controles apresentaram o intervalo de inclusão do éxon de 0,84 a 1,0 (84% a 100%), o paciente teve um PSI de ~0,85 (85%), apresentando uma diferença de PSI com a média dos controles de ~0,14 (14%). Isso significa que esse evento também ocorre de forma alternativa nos controles, o qual pode ser visualizado através da Figura 56. Esse salto do éxon ocorre de forma *in-frame*, não alterando a janela de leitura de tradução, mas esse evento causa a perda de 52 pb que correspondem a parte inicial do domínio “DNA damage repair protein Nbs1” e cerca de 6,9% da proteína. Essa variante foi descrita pela Invitae como produzindo uma proteína alterada devido a dano no *splicing*, mas com o impacto funcional ainda desconhecido. Além disso, a pela Ambry Genetics após a análise de RNA constataram que a variante resulta em um defeito de *splicing* incompleto e que o impacto clínico dessa variante ainda é desconhecido.

Figura 56 – Gráfico de sashimi apresentando evento de salto do éxon 13 presente no paciente RNAP-110 e em um controle externo



Fonte: Autoria própria.

Legenda: O evento de SE 13 foi visto no paciente e nos controles, porém no paciente esse salto ocorre numa proporção maior de leituras, sendo considerado um evento significativo em comparação com os controles.

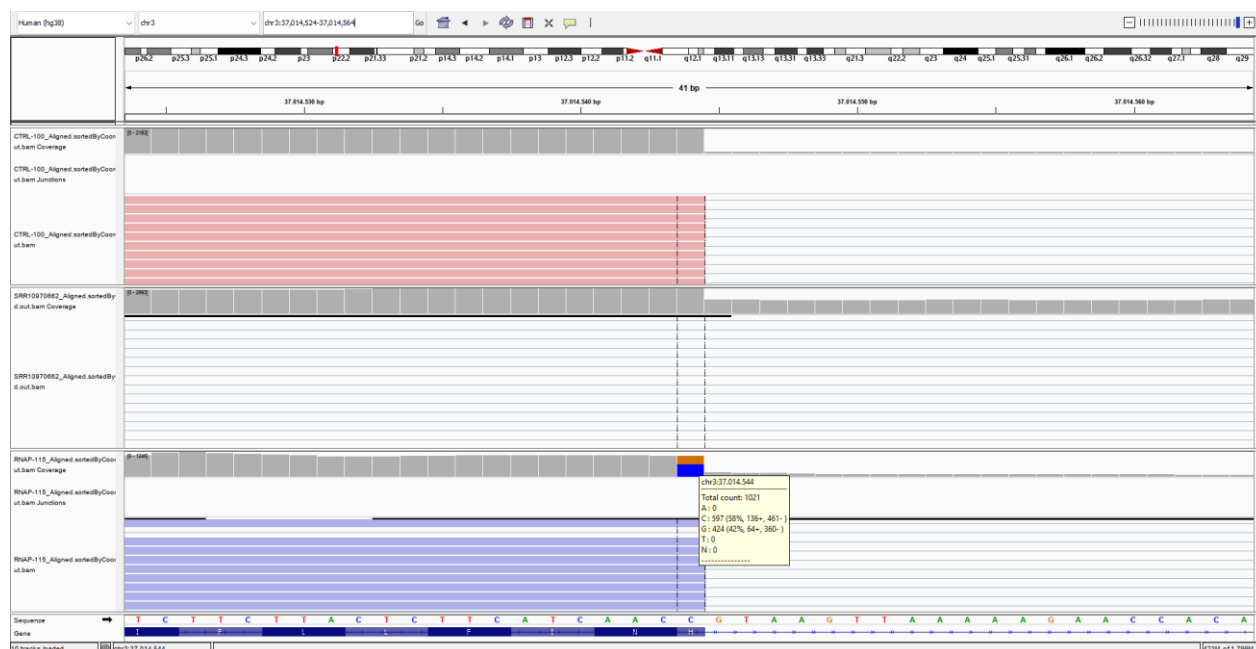
Na análise de desbalanço alélico, o alelo alterado da variante sinônima (NM_002485.5:c.1197T>C – no éxon 10) teve VAF de 99% do alelo alternativo, enquanto a VAF do alelo referência foi de 1%. Em outro SNP marcador na variante sinônima (NM_002485.5:c.102G>A - éxon 2), a VAF do alelo referência foi 42% e do alelo variante 58%. Não foi possível realizar a análise direcionada ao transcrito com esse paciente, pois devido a integridade do RNA não estar adequada (RIN = 4,6) e, mesmo testando de dois pares de *primers* diferentes, a amostra não amplificou, apenas nos indivíduos controles. Devido a evidências insuficientes, essa variante continua sendo VUS, com aplicação dos critérios PM2_sup, PP3, PVS1_mod(RNA_NA).

A paciente RNAP-115, com variante em *MLH1*:c.790C>G localizada no último nucleotídeo do éxon 9, apresentou pela ferramenta rMATS quatro eventos de *splice* significativos de salto de éxon *in-frame* e *out-of-frame* envolvendo os éxons 9, 10 e 11 que codificam para o domínio DNA mismatch repair protein S5 da proteína MLH1. Foi visto 3 saltos de éxons *out-of-frame* (salto do E9, salto do E10, salto do E10-E11) e um salto de éxon *in-frame* (E9-E10) perdendo 69 aminoácidos. Em todos os eventos, o PSI do paciente estava incluso no intervalo de PSI dos CTRLs, evidenciando que esses eventos ocorrem naturalmente

em indivíduos sem a variante e que esses éxons são altamente variáveis na inclusão no transcrito maduro através do *splicing* alternativo. Isso também reflete na diferença de PSI que é muito pequena entre o paciente e os CTRLs, apenas o evento do salto do E10-E11 apresentou uma diferença acima de 10%, por isso foi o único evento que foi depositado na tabela 12 e representado no gráfico boxplot (Figura 36).

Nos depósitos do ClinVar, a Invitae classificou como provavelmente benigna, porém sem descrição, e a Ambry não fez estudos de RNA e não descreveu casos na literatura de indivíduos afetados com essa variante, apenas a troca de aminoácido que possui propriedades químicas semelhantes, por isso a classificaram como VUS. Por ser uma variante exônica, também foi possível avaliar se ocorreu o desbalanço alélico a partir desta mesma variante. O resultado foi que a variante apresentou uma expressão balanceada, na qual a VAF do alelo normal foi de 58% e do alelo alterado de 42%. Isso indica que ambos os alelos estão sendo expressos em proporções equivalentes, não ocorrendo degradação do transcrito com o alelo alterado, conforme pode ser visualizado na Figura 57. A partir dessa análise também observou que não teve um enriquecimento de deleção do E9 no alelo alternativo.

Figura 57 - Análise de desbalanço alélico a partir da variante *missense* em *MLH1*:c.790C>G

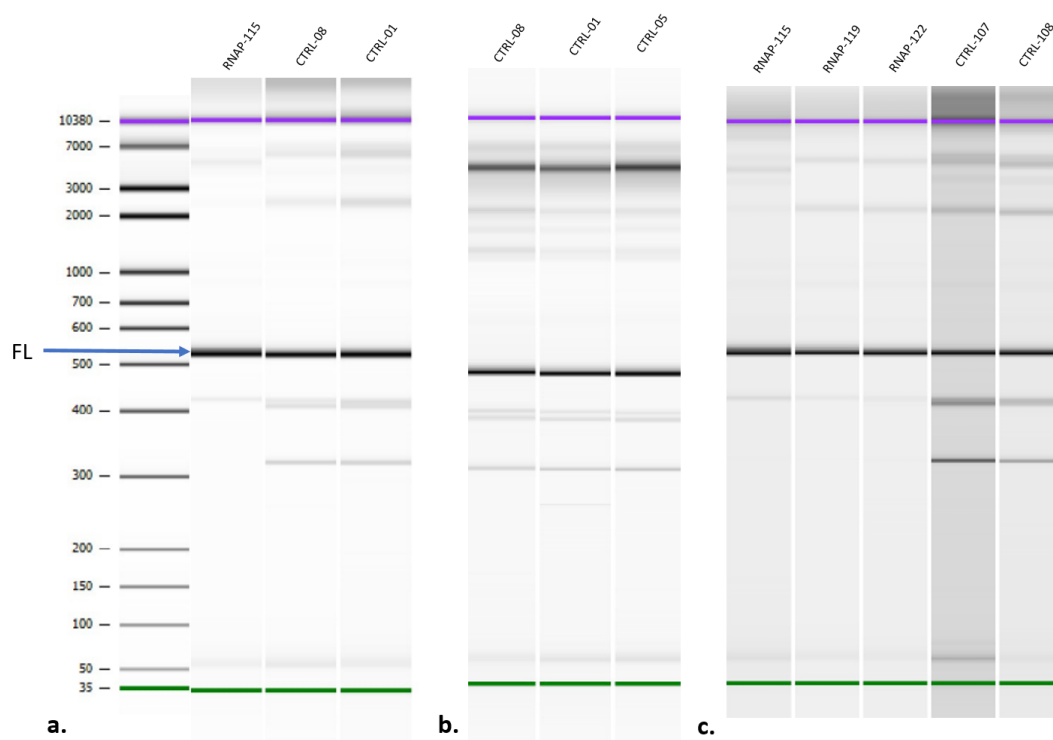


Fonte: Autoria própria.

Legenda: Na imagem é possível observar que expressão do alelo selvagem e o alterado estão balanceados.

Para análise direcionada ao transcrito foram realizados três ensaios com o Bioanalyzer, representados na Figura 58: o primeiro analisou a amostra do paciente oriunda do tecido da mama normal contra três CTRLs de tecido de leucócito, o segundo ensaio apenas com mais três indivíduos controles e o último com dois pacientes, RNAP-119 e RNAP-107 de tecido da mama normal (que possuíam variantes patogênicas apenas em *BRCA2* e *BRCA1*, respectivamente) com mais dois controles (leucócitos) contra o paciente. Através da última análise foi visto que as bandas dos pacientes RNAP-119 e RNAP-107 eram semelhantes com as bandas do paciente RNAP-115, porém um pouco diferentes dos controles de leucócitos. Dessa forma, pode-se explicar que o *splicing* alternativo de *MLH1* é diferente entre o tecido da mama e os leucócitos, em que o primeiro possui menos isoformas que o segundo, ou seja, menos diversidade de isoformas. Também foi possível observar que o padrão de bandas da paciente estava idêntico com os pacientes RNAP-119 e RNAP-107, sendo indicativo de que não houve uma alteração de *splicing* na mama.

Figura 58 - Imagem do gel do Bioanalyzer da paciente RNAP-115 para análise da variante *MLH1*:c.790C>G



Fonte: Autoria própria.

Legenda: No gel (a), (b) e (c) estão os três ensaios realizados. No último ensaio estão as amostras dos pacientes com RNA oriundo da mama normal (RNAP-115, RNAP-119 e RNAP-122) contra as amostras dos indivíduos controles de leucócitos, notando-se um padrão de banda diferente, com uma diversidade menor de bandas que as amostras de leucócitos, embora com padrão parecido entre si. A seta azul indica o *amplicon* esperado com 540 pb.

Considerando esse resultado de padrão de *splicing* distinto entre o tecido mamário e leucócitos, realizamos uma análise extra no rMATS com esse paciente contra outros seis pacientes (RNAP-118, RNAP-119, RNAP-121, RNAP-122, RNAP-123 e RNAP-124), em que o RNA extraído dessas amostras também era de mama normal e não possuíam variantes VUS nem P/PP no gene *MLH1*. Foi encontrado o evento de salto do E10 com éxons flanqueadores E9 e E11, em que a paciente inclui menos o E10 (84%) comparado com os controles que incluem de ~94% a 98%. Outro evento também significativo foi identificado, em que ocorre o salto dos éxons 9, 10 e 11 e éxons flanqueadores 8 e 12, sendo menos incluso na paciente com 50% das leituras e nos controles tirando uma amostra discrepante, tem inclusão de 77% a 100%. No entanto, como o número de pacientes controles é pequeno (6), essa análise não é conclusiva sobre o potencial efeito em *splicing* dessa variante, sendo mantida a classificação de VUS com aplicação dos critérios PM2_sup, BP4, BP7_Strong (RNA_NA).

A paciente RNAP-124, com variante *BRCA1*:c.4096+1G>A localizada no íntron 10 apresentou um salto do éxon 10 *in-frame* levando a perda de 1.142 aminoácidos do domínio BRCT. Os depósitos dessa variante no ClinVar são conflitantes para classificações de variante de significado incerto e provavelmente patogênica por laboratórios de referência, que descrevem que ocorre uma perda parcial do E10, a qual ocorre de forma natural nos controles, porém é retida uma função residual mínima na proteína. A ferramenta rMATS detectou um evento SE do éxon 10 completo, com a diferença de PSI entre caso e controles de 10%, e outro evento com salto do éxon 10 parcial retendo os 118 pb iniciais do E10 com a diferença de PSI de 11%. Essa diferença é pequena, porém atinge o valor que estabelecemos para considerar o evento como significativo. A ferramenta LeafCutterMD apresenta um resultado semelhante, em que os 122 pb iniciais do E10 se ligam diretamente ao E11, essa junção acontece em apenas 2 CTRLs externos.

Nesse paciente foram encontrados cinco SNPs marcadores, quatro variantes no E10 (c.1067A>G, c.2612C>T, c.3113A>G e c.3548A>G), ilustradas na Figura 59, e uma no E15 (c.4837A>G). A variante *BRCA1*: c.1067A>G estava a 394 pb do início do E10, tinha 90% de expressão do alelo referência e 10% do alelo alternativo. As próximas variantes c.2612C>T e c.3113A>G, que estão a 502 pb de distância uma da outra estavam com expressões semelhantes, alelo referência 34% e alelo alternativo 66%, indicando estarem em *cis* entre si e com um certo desbalanço. O último SNP marcador desse éxon tinha proporções diferentes de expressão: alelo referência 17% e alelo alternativo 83%. Se tivesse ocorrido um salto de *splicing* completo (deleção completa [100% das reads] do restante do E10 na isoforma que abriga o alelo alterado),

todos esses SNPs marcadores só teriam a expressão ou do alelo referência ou do alelo alternativo em *cis* com o alelo alterado. Porém, nesse caso vemos a expressão de ambos os alelos de forma desbalanceada, indicando a ocorrência de *leaky splicing* (termo utilizado quando a variante leva a uma produção parcial de transcritos alterados e de transcritos *full length*).

Figura 59: Visão geral dos 4 SNPs marcadores encontrados no E10 da paciente RNAP-124



Fonte: Autoria própria.

Legenda: Os SNPs marcadores são: *BRCA1*:c.1067A>G (em vermelho); *BRCA1*:c.2612C>T (em azul); *BRCA1*:c.3113A>G (em verde); e *BRCA1*:c.3548A>G (em roxo). O trecho de salto parcial do E10 encontrado no rMATS, o qual retém apenas os 118 pb iniciais está representado em laranja.

Não foi possível realizar a análise direcionada ao transcrito neste paciente, pois o éxon 10 é muito extenso, o qual possui 3.426 nucleotídeos. Tentamos desenhar *primers* no final do E9 e início do E11 para conseguir amplificar a isoforma com o salto parcial do E10 que geraria um *amplicon* de aproximadamente 230 pb, porém não houve amplificação na paciente nem nos controles, e no sequenciamento só apareceu cobertura próximo das regiões em que estavam os *primers*, com nenhuma leitura mapeada no E10. Devido as evidências insuficientes, essa variante continua como VUS, com a aplicação dos critérios PVS1_mod (RNA), PM2_sup, PP3.

6. DISCUSSÃO

Em nosso estudo, implementamos e validamos um conjunto de ferramentas computacionais, genômicas e experimentais para aprimorar a priorização e interpretação clínica de VUS preditas como spliceogênicas em genes de predisposição hereditária ao câncer. A integração de múltiplos métodos, como preditores de *splicing*, sequenciamento de RNA por captura e por transcrito direcionado permitiu não apenas a identificação de potenciais eventos aberrantes, mas também a criação de um pipeline automatizado robusto para a seleção e avaliação de variantes, tanto codificantes quanto não codificantes.

Este projeto descreveu a análise de RNA como uma ferramenta complementar ao teste genético de DNA para investigar o *splicing* aberrante através do sequenciamento de captura de mRNA em 50 pacientes, a partir de amostras de leucócitos ou tecido normal. Avaliamos variantes de nucleotídeo único exônicas e intrônicas até 20 pares de base do limite éxon-intron com a adição de algumas deleções/duplicações de interesse, assim como variantes presentes nas regiões 5' e 3' UTR nos 27 genes de predisposição hereditária ao câncer com intuito de auxiliar na reclassificação das VUS.

Não existe ainda um protocolo padrão único de ponta a ponta para as análises de RNA, especificamente para a análise de *splicing*, que inclui desde a coleta do material biológico, armazenamento, extração, construção de biblioteca de transcriptoma total ou de captura, pipelines robustos para processamento dos FASTQs e mapeamento de RNA, parâmetros adequados às ferramentas para detecção de *splicing* diferencial e/ou aberrante.

O que temos atualmente nesse contexto são recomendações para triagem de variantes de *splice* encontradas a partir do sequenciamento de DNA e interpretação das variantes intrônicas presentes nos sítios canônicos e não canônicos. Essas recomendações foram adaptadas aos critérios da ACMG e desenvolvida pelo grupo de especialistas de *splice* da ClinGen em 2023 (*ClinGen SVI Splicing Subgroup*) (Walker *et al.*, 2023). Neste trabalho, foi recomendado o uso do SpliceAI como a ferramenta padrão para triagem de variantes de *splice* no cenário clínico com o *cut-off* de $\geq 0,20$, o qual indica risco moderado de causar erro de *splicing*, além de apresentar uma árvore de decisão para ponderar o peso de maior força, o 'PVS1', nas variantes nos dinucleotídeos conservados intrônicos $\pm 1,2$ e aplicação de outros códigos em outras variantes intrônicas para ressaltar a evidência de impacto no RNA, quando existir dados de estudos robustos (Walker *et al.*, 2023).

Em relação a triagem de variantes com potencial de serem spliceogênicas (preditas de

serem spliceogênicas), cerca de 14,30% dos 2.076 pacientes totais desse estudo tinham variantes preditas de causar erro de *splice*, valor próximo ao encontrado em outro estudo que foi de 17,5% (Kamps-Hughes *et al.*, 2023). Utilizamos os preditores dbSNP, MaxEntScan e SpliceAI por serem os mais utilizados para o diagnóstico clínico e em outros estudos para o mesmo tipo de análise, além de estarem disponíveis de forma gratuita (Casadei *et al.*, 2019; Wai *et al.*, 2020; Horton *et al.*, 2022; Kamps-Hughes *et al.*, 2023). Entraram para o estudo apenas aquelas variantes preditas de causar erro de *splicing* em pelo menos um dos três preditores, que pode ter levado a exclusão de variantes que tem a capacidade de alterar o *splice*, mas que não foram preditas por nenhum dos preditores. Uma vantagem desse estudo foi termos capturado variantes sinônimas para análise do impacto de *splicing*, as quais geralmente são negligenciadas no cenário de diagnóstico clínico por não alterarem o aminoácido, mas que podem ter a capacidade de alterar o *splice*, sendo que representam 8,9% das variantes preditas spliceogênicas.

É importante destacar que esses três preditores são desenhados para prever variantes de nucleotídeo único, sendo que o SpliceAI consegue prever também algumas *indels*, por isso a predição de variantes de grandes deleções/inserções (CNVs) não geram pontuações. Dessa forma, ao serem identificadas CNVs no sequenciamento convencional de DNA, todas devem ser avaliadas de forma minuciosa e individual para validar se elas causam ou não alteração de *splice* relevantes. Nesse estudo demonstramos através do RNA-cap o efeito em *splicing* e na regulação da transcrição de duas CNVs: a variante patogênica em *PALB2* que leva à duplicação *em tandem* do E11 seguida do salto do E12; e a variante em *MSH2* (classificada inicialmente como VUS) reclassificada para patogênica por deletar a região promotora do gene e impedir sua transcrição, através da confirmação da expressão monoalélica. Nesse sentido, estudos já destacaram que a análise de RNA auxilia a determinar se grandes duplicações ocorreram *em tandem* em relação a sequência referência (Horton *et al.*, 2022).

Em relação a pequenas deleções e inserções (*indels*) o único preditor que conseguiu gerar escores para essas variantes foi o SpliceAI, no qual três variantes P/PP e duas VUS foram preditas como spliceogênicas nos 27 genes. Vale destacar que conseguimos validar efeito de *splice* anormal de duas dessas três variantes *indels* patogênicas, *CHEK2*:c.846+4_846+7del e *RAD51C*:c.905-2_905-1delAG. A predição de *indels* por essa ferramenta é possível, pois ela possui um banco de dados de pequenas deleções/inserções de 1 a 4 bases em genes específicos. Logo, *indels* maiores talvez não sejam bem previstas por esse preditor.

O SpliceAI, por não se basear nas regiões canônicas, mas sim no impacto de *splice* cinco

mil nucleotídeos antes e depois da variante em questão (independentemente da sua posição) foi a ferramenta com melhor desempenho para a triagem das variantes de *splice*, uma vez que detectou mais variantes potenciais de serem spliceogênicas P/PP e VUS que entraram para o estudo. Outra vantagem desse preditor é gerar pontuações indicativas de erro de *splice* em variantes mais internas do éxon (variantes exônicas distantes da borda – *deep exonic*), como por exemplo, VUS exônicas a partir da posição 5 até 191 pb dentro do éxon e VUS intrônicas com ± 10 pb de distância do éxon. Por outro lado, as outras ferramentas geram escores de predição para variantes exônicas localizadas até no máximo os três últimos nucleotídeos próximos da borda e para variantes intrônicas até as posições D+8 e A-20.

É importante ressaltar que, antes da recomendação da ClinGen (Walker *et al.*, 2023) e antes da criação do SpliceAI em 2019 era mais comum os estudos utilizarem a ferramenta paga Alamut que agrega 7 preditores de *splicing* (SpliceSiteFinder-like, MaxEntScan, GeneSplicer, NNSPLICE, ESEFinder, RESCUE-ESSE e EX-SKIP), na qual não estava incluso o SpliceAI (Casadei *et al.*, 2019; Wangenstein *et al.*, 2019; Gelli *et al.*, 2019; Morak *et al.*, 2019; Rofes *et al.*, 2020; Ryu *et al.*, 2020; Agiannitopoulos *et al.*, 2021; Castillo-Guardiola *et al.*, 2022; Bozsik *et al.*, 2022; Li *et al.*, 2022; Li *et al.*, 2022). Diversos estudos compararam a performance de múltiplas ferramentas de bioinformática de predição de *splice* e encontraram o SpliceAI como apresentando o melhor sensibilidade e especificidade (Findlay *et al.*, 2018; Ha *et al.*, 2021, Riepe *et al.*, 2021; Jang *et al.*, 2022). Ainda, essa ferramenta demonstrou ser razoavelmente eficaz para a prever ganhos de sítios AG, embora não tão assertivo em prever a perdas do sítio acceptor com o ganho de variantes AG (Zhang *et al.*, 2023). Por todas essas evidências discutidas e vantagens identificadas nesse estudo, ela é atualmente a ferramenta mais indicada para fazer a triagem das variantes spliceogênicas no contexto do cenário clínico.

Em nosso estudo, as variantes P/PP preditas de serem spliceogênicas estavam localizadas em sua maioria nas regiões canônicas de *splice* $\pm 1,2$ pb do íntron e nos dois últimos nucleotídeos do éxon. Este mesmo padrão foi visto no estudo de Kamps-Hughes e colaboradores (2023), no qual dentre as variantes potenciais de causar alteração no *splice* classificadas como P/PP, a maior parte estavam localizadas nas regiões canônicas em relação as VUS. No estudo de Horton e colaboradores (2022), das variantes que tinham classificação P/PP, 95,6% estavam localizadas nesses dinucleotídeos conservados e 62,7% no último nucleotídeo do éxon. Além disso, encontramos uma frequência maior de variantes P/PP e VUS na região doadora (D+3, D+4 e D+5 pb) após o sítio canônico doador (D+1,2 pb) do que nas regiões aceptoras (A-3, A-4 e A-5 pb) após o sítio canônico acceptor (A-1,-2 pb). No estudo de

Kamps-Hughes e colaboradores (2023), os autores também relataram que houve um enriquecimento nas posições D+3, D+4 e D+5 do sítio de *splicing* do doador canônico, embora no estudo de Horton e colaboradores (2022) mais variantes foram encontradas na região acceptora (A-3 ao A-5) do que na região doadora (D+3 ao D+5). Porém nesse estudo, a maior parte das variantes da região acceptora tinham classificação PB, enquanto as da região doadora tinham mais classificações VUS ou PP/P, ou seja, variantes reportáveis. O estudo de Wai e colaboradores (2020), encontrou que 77% das SNVs que apresentaram *splicing* anormais estavam localizadas na região doadora consenso (3 últimos nt do éxon até o 8º nt do íntron) e 13/19 SNVs na região acceptora (os últimos 8 nt do íntron até 3 nt do éxon) alteraram o *splice*, por isso pode-se afirmar que variantes na região doadora podem ter uma probabilidade maior de ocasionar erros de *splice* se comparados com a região acceptora.

Sabe-se que as posições canônicas de *splice* $\pm 1,2$ pb são altamente conservadas com uma média no phyloP de 6,34 (Blakes *et al.*, 2022), e que cerca de 89% das variantes encontradas nessa região são classificadas como P/PP e apenas 10% como VUS (Walker *et al.*, 2023). As variantes próximas do sítio canônico também possuem certo grau de conservação, sendo suas posições e sua média de conservação no phyloP de D+5 (3,44), D+4 (2,39), D+3 (1,97), D+6 (1,29) e a menos conservada A-4 (0,076) (Blakes *et al.*, 2022), isso pode explicar o porquê variantes nessas regiões são mais preditas de causar dano em *splice* do que variantes em outras posições intrônicas, e por isso foram mais enriquecidas em nossa coorte e nos outros 3 estudos (Wai *et al.*, 2020; Horton *et al.*, 2022; Kamps-Hughes *et al.*, 2023).

Posições de *splicing* não canônicas podem abrigar variantes deletérias. No estudo de Blakes e colaboradores (2022) foi relatado que a posição D+5 e D-1 (último nt do éxon) provavelmente podem abrigar variantes prejudiciais, e que estimativas de trabalhos anteriores indicam que pode haver um aumento de 90% de variantes patogênicas na posição D+5 em variantes *de novo* próximas da região de *splice* em genes dominantes de doenças raras (Lord *et al.*, 2019; Blakes *et al.*, 2022). Ainda no estudo de Horton e colaboradores (2022) os autores relataram que houve maior atualização de variantes localizadas na região D+5 ou D+3 após a análise de RNA. Em nosso estudo, tivemos quatro variantes na região D+5 que foram sequenciadas pela técnica de RNA-cap: 2 variantes P/PP (RNAP-101 *MLH1*:c.588+5G>C e RNAP-113 *MLH1*:c.116+5G>C) e 2 VUS (RNAP-103 *BRCA1*:c.547+5G>A e RNAP-114 *SMARCA4*:c.859+5G>A) sendo observado diferentes eventos de *splice*. A primeira variante em *MLH1* gerou diferentes saltos E6, E7 ou E6 + E7, já a segunda variante em *MLH1* foi visto nesse estudo o uso reduzido do sítio alternativo 5' de *splice*, embora outros estudos relatarem

alteração de *splice*. Já para ambas as variantes VUS, não foram encontradas evidências de impactos no *splice* pelo RNA-cap, e pela análise do transcrito direcionado. Isso demonstra que os efeitos esperados podem ser diversos para essa posição.

Das variantes intrônicas encontradas na região aceptora, no estudo Kamps-Hughes e colaboradores (2023) observou-se que as posições A-11, A-10 e A-9 do sítio acceptor canônico na maioria das vezes estavam propensas a se tornar novos sítios aceptores de *splice*. Já Horton e colaboradores (2022) observaram que, depois das variantes encontradas nos sítios D+5 e D+3, as variantes com distância de 10 nucleotídeos do éxon foram as que mais sofreram reclassificações após a análise de RNA. Isso foi visto no nosso estudo com a variante *BRCA2*:c.7618-10T>G (VUS → P) que se tornou um sítio acceptor mais forte que o sítio canônico ocorrendo a retenção de 9 pb do íntron com adição de um códon de parada prematuro e alteração do *splice*. Já para a variante P *BRCA1*:c.213-11T>G, houve a retenção de aproximadamente 59 pb do íntron, devido à diminuição da força do sítio acceptor canônico e a utilização de um sítio acceptor críptico de *splice* na posição -60 e -61.

As variantes localizadas fora da região reportável em painéis multigenes de DNA (± 20 pb do éxon) em geral, carecem de evidências científicas para interpretação e possível reclassificação. Alguns estudos já reportaram que variantes que causam alteração no *splicing* podem estar localizadas numa região de íntron profundo, com uma distância do éxon > 200 pb (Casadei *et al.*, 2019; Canson *et al.*, 2020; Horton *et al.*, 2022; Kamps-Hughes *et al.*, 2023; Horton *et al.*, 2024; Jaramillo *et al.*, 2024), mas essas áreas são dificilmente cobertas por painéis gênicos, possuindo baixa cobertura e qualidade de sequenciamento, sendo inviáveis para a detecção no sequenciamento de DNA sozinho. Dois estudos que fizeram o teste pareado DNA e RNA em grandes coortes tiveram pouca representatividade de variantes intrônicas profundas (*deep intronic*) na reclassificação para aumento da taxa de diagnóstico positivo. No primeiro estudo, apenas 8 de 20.317 pacientes totais (0,04%) tiveram variantes fora da zona reportável reclassificadas de VUS para P/PP (Kamps-Hughes *et al.*, 2023), as quais estavam a mais de 200 pb de distância do éxon (*FH*:c.1391-269A>G, *APC*:c.531+1482A>G, *ATM*:c.2639-384A>G, *PALB2*:c.49-1123C>G e *NF1*:c.1642-449A>G). Em um segundo estudo, 28 de 43.524 pacientes totais (0,06%) tiveram uma nova variante intrônica profunda P/PP detectada exclusivamente pela adição do sequenciamento de RNA (Horton *et al.*, 2024). Ainda, o estudo de Kamps-Hughes e colaboradores (2023), as variantes intrônicas profundas representaram 5,6% de um total de 605 variantes spliceogênicas, das quais apenas 2/34 variantes presentes em dois pacientes foram classificadas como benignas e 27/34 variantes foram classificadas como

VUS (79,4%), ou seja, uma vez encontradas variantes preditas de serem spliceogênicas nessas regiões a maioria será reclassificada para VUS. Em suma, mesmo com a adição do sequenciamento de RNA, variantes além de 20 pb são ainda um desafio para interpretação clínica. Por último, cerca de 93% dessas 605 variantes estavam dentro do intervalo reportável de variantes. Neste presente trabalho não selecionamos variantes além de ± 20 pb de distância do éxon, visto que a cobertura dos painéis de DNA para além desse limite é irregular e conforme apresentado nos estudos acima, existe uma probabilidade pequena dessas variantes sofrerem reclassificação de VUS para P/PP.

Em relação às variantes das regiões não traduzidas, realizamos apenas a filtragem das que estavam localizadas na região 3'UTR sem utilizar nenhuma estratégia de priorização, pois atualmente faltam ferramentas de predição e abordagens experimentais validadas clinicamente para reclassificar essas variantes. Já as variantes na região 5' UTR foram filtradas e preditas pelo UTR Annotator, porém a predição de impacto na maior parte das vezes estava anotada em transcritos não-MANE, dificultando ainda mais a interpretação dessas variantes. Ainda, poucas variantes foram preditas de ter efeitos em 5' UTR nos 27 genes de nosso estudo, e essas variantes serão mais exploradas em um projeto futuro do grupo.

A análise de RNA para as variantes de *splice* começou de forma desafiadora desde a etapa de validação técnica, na qual buscamos validar os efeitos de *splice* já conhecidos das variantes P/PP spliceogênicas. Observamos que das três ferramentas de análise de *splice*, o rMATS foi a que mais identificou eventos esperados (11/14), seguido do LeafCutterMD e FRASER2 com 7/14 eventos cada. No entanto, com essas ferramentas não foi possível identificar o efeito de *splice* em três variantes P/PP: *BRCA2*:c.156_157insAlu (RNAP-03), *MUTYH*:c.933+3A>C (RNAP-116) e *BRCA1*:c.213-11T>G (RNAP-118). Em outras variantes foi identificada uma alteração de *splice*, mas não do evento esperado já descrito por outros estudos, como para: *MLH1*:c.588+5G>C (RNAP-101) e *MLH1*:c.116+5G>C (RNAP-113). Percebemos então que essas ferramentas tem como limitação não encontrar eventos de *splicing* que ocorrem em pequenas ou grandes inserções de íntron (*BRCA1*:c.213-11T>G), variantes que estão no éxon e/ou íntron 1 (*MLH1*:c.116+5G>C), e quando ocorrem múltiplos eventos de *splice* ao mesmo tempo em uma mesma região (*MLH1*:c.588+5G>C e *MUTYH*:c.933+3A>C), ou seja, quando um gene possui muitas isoformas devido a grande quantidade de eventos alternativos de *splice*, tendo como exemplo o gene *MUTYH*, no qual seus íntrons são muito curtos e possui múltiplas isoformas. O trabalho de Jaramillo e colaboradores (2024) ao avaliar quatro ferramentas análise de *splice* (rMATS, MAJIQ, FRASER2 e LeafCutterMD) identificou

que o rMATS também foi a ferramenta que detectou mais eventos de *splicing* alternativos (11/14 eventos) sendo o mais sensível, identificando exclusivamente 7 eventos, seguido da ferramenta MAJIQ e FRASER2 (4/14 eventos), por último LeafCutterMD (2/14 eventos). Em todos esses 14 eventos alguma anormalidade já tinha sido encontrada a partir da inspeção visual dos BAMs gerados do RNA-seq transcriptoma total. Nesse estudo, foram identificadas três variantes que foram perdidas por essas quatro ferramentas, dessa forma descreveram as principais limitações encontradas: eventos de retenção de íntron tem maior probabilidade de serem perdidos comparados com outros eventos, e que as ferramentas conseguem detectar eventos anormais de *splice* com no mínimo 5X de cobertura de leituras na nova junção. Essas limitações de detecção também já foram vistas e descritas em estudos e revisão anteriores, relacionando-as com os dados gerados do RNA-seq de leituras curtas (Brandão *et al.*, 2019; Wang *et al.*, 2023), os quais descreveram que o comprimento curto das leituras torna mais difícil o mapeamento das isoformas, não sendo suficiente para fasear eventos de *splicing* individuais, e detectar transcritos originados de diferentes eventos. Nesse sentido, utilizar o sequenciamento de leituras longas poderia solucionar a complexidade da estrutura exônica e intrônica de transcritos completos/inteiros, pelo fato de não precisar reconstruir sequências curtas na etapa de bioinformática (Walker *et al.*, 2023; Wang *et al.*, 2023).

A complementação das análises do RNA-cap com a avaliação do transcrito direcionado e desbalanço alélico foram essenciais para compreender melhor os eventos de *splicing* complexos tanto em variantes P/PP como VUS, quando as ferramentas não encontraram nenhum evento de *splice* anormal a partir dos BAMs gerados no RNA-cap, ou para complementar eventos de *splicing* anormais detectados no RNA-cap. Alguns exemplos são os pacientes RNAP-118 e RNAP-05, em que só foi possível observar a retenção de íntron através da inspeção visual das bandas na eletroforese capilar e no IGV. A análise do desbalanço alélico foi muito importante para avaliar de forma indireta a perda da expressão completa do alelo alterado que tinha uma grande deleção que envolveu a região 5'UTR e a deleção parcial do éxon 1 no gene *MSH2*, nos pacientes RNAP-09 e RNAP-09.02. Outro caso semelhante foi o do paciente RNAP-04 com uma variante no final do E1 *RAD51D*:c.80C>A, o que a torna não passível de análise pelas ferramentas, porém como é considerada um SNP marcador, permitiu observar que houve perda do alelo alterado, indicando que nesse alelo o E1 é saltado e não expresso. Essa complementação por RT-PCR também foi crucial no trabalho de Jaramillo e colaboradores (2024) para as três variantes que foram perdidas pelas ferramentas de análise de *splice*, conseguindo validar os eventos anormais, lembrando que alguma anormalidade já havia

sido detectada previamente no IGV, a partir da análise de seus BAMs oriundos do RNA-seq. Todas essas evidências demonstram que apenas as ferramentas de análise de *splicing* diferencial e de *splicing* discrepante não são suficientes para caracterizar um evento de *splicing* anormal, principalmente aqueles eventos parciais de retenção/inserção de íntron/éxon, uma vez que essas ferramentas conseguem identificar melhor eventos de retenção/inserção de éxon/íntron completos. Por outro lado, essas ferramentas são essenciais para fornecer os valores quantitativos que demonstram se o evento de *splice* é ou não significativo, os quais não são possíveis de se extrair apenas a partir da inspeção visual, como as métricas de P valor, FDR e PSI.

Tendo como base as evidências achadas no nosso estudo e de Jaramillo e colaboradores (2024), é possível afirmar que a inspeção visual dos BAMs gerados no RNA-seq é muito importante, pois nessa inspeção podem identificar anormalidades de *splice* que não são detectadas pelas ferramentas. Sendo assim, a inspeção visual pode ser considerada como a primeira linha de análise, em complemento às ferramentas de *splice* para apoiar o resultado com suas métricas de significância, e as variantes que não forem detectadas pelas ferramentas precisam da validação do evento através do RT-PCR (análise do transcrito direcionado) seguido de sequenciamento NGS.

Utilizamos amostras de leucócitos de indivíduos saudáveis como controles negativos e amostras de tecido congelado para a maioria das amostras dos pacientes que eram retrospectivos, tanto para o RNA-cap como para as análises do transcrito direcionado. Ao utilizar amostras de leucócitos nos controles assumiu-se que o padrão de *splicing* é parecido com as amostras de tecido sólido dos pacientes, porém sabe-se que para alguns genes esse padrão pode ser diferente. Levando em conta o contexto clínico dos testes genéticos, coletar amostras de sangue de indivíduos é considerada uma técnica menos invasiva, rápida e custo efetiva. Existem estudos que já compararam perfis de *splice* entre amostras de tecido normal da mama e de sangue e constataram que 11 genes relacionados ao HBOC (*BARD1*, *BRCA1*, *BRCA2*, *BRIP1*, *CHEK2*, *PALB2*, *RAD51B*, *RAD51C*, *RAD51D*, *XRCC2* e *XRCC3*) apresentaram perfis de *splice* parecidos entre esses dois tecidos (Davy *et al.*, 2017; Wangenstein *et al.*, 2019). No nosso estudo, a partir da análise do transcrito direcionado, no paciente RNAP-115, foi possível constatar que o perfil de *splicing* do gene *MLH1* é diferente no tecido da mama em relação ao sangue. No paciente o RNA foi extraído da mama normal, e observou-se que seu padrão de bandas era idêntico às amostras de outros pacientes com amostra de mama normais (que apresentaram poucas bandas), porém diferente das amostras controles

de leucócitos (que apresentaram mais bandas). Essa evidência demonstra que a variante não causou uma perda da diversidade de isoformas, mas que o tecido de mama possui menos isoformas que os leucócitos, logo possui menos eventos de *splicing* alternativos. O ideal seria ter a amostra de leucócito do paciente para fazer uma avaliação mais adequada contra os controles, e por esse motivo a variante permaneceu como VUS.

A integridade do RNA (RIN) é um fator determinante no NGS, pois quanto mais fragmentada a molécula de mRNA mais vieses são acrescentados para as análises de expressão gênica e de *splicing*. Nesse estudo não limitamos as análises das amostras: RNAP-03, RNAP-07, RNAP-110, RNAP-112, RNAP-113, RNAP-114 e RNAP-116 que tinham RIN baixo entre 3,1 e 4,6 para realização do RNA-cap nem para o transcrito direcionado. Em todas essas amostras foi possível a realização do RNA-cap, as quais apresentaram boas métricas de qualidade de sequenciamento, exceto para a amostra RNAP-07, em que a expressão gênica do gene *TP53* foi baixa no paciente e controles, assim como o valor de cobertura, com menos de 10X no éxon que abrangia a variante. Já a amostra RNAP-06 com variante no gene *ATM* teve RIN médio de 5,6 e cobertura também abaixo do estabelecido (82,4%), porém como esse gene é altamente expresso nos tecidos dos controles e do paciente, a cobertura dos éxons foi homogênea, a amostra foi viável para as análises posteriores. Para a análise de transcrito direcionado não foi possível realizar RT-PCR nas amostras RNAP-110 e RNAP-07, por falha na PCR, devido seu material que estava altamente fragmentado. Com isso pode-se estabelecer que amostras com RIN acima de 5 são adequadas para as análises de RNA-cap e transcrito direcionado, embora amostras com RIN de no mínimo 3,0 podem ser viáveis apenas para RNA-cap em genes em que a expressão gênica é adequada no tecido de origem, mas essas amostras não são viáveis para análise do transcrito direcionado.

É importante destacar que uma variante ser spliceogênica (que altera o *splicing*) não é sinônimo de ser uma variante patogênica, mesmo quando esse evento de *splice* é anormal (ocorrendo apenas no paciente). Ao detectar a alteração de *splicing* é preciso identificar se, por exemplo, eventos de salto de éxon são *in-frame* ou *out-frame* e, quando *in-frame*, se acontecem em um domínio crítico para a função da proteína. Isso ocorreu com a variante *MUTYH*:c.934-2A>G nos pacientes RNAP-105 e RNAP-106, em que foi vista a deleção *in-frame* de 9 nucleotídeos no início do E11, num domínio não crítico para proteína. Ainda, é preciso avaliar se a região excluída representa mais ou menos que 10% do comprimento total da proteína, e nos casos com interrupção da janela de leitura, se o transcrito sofre degradação via NMD. Existem ainda variantes candidatas a “resgate”, por exemplo, se ocorre uma variante com perda

de função em um éxon que é naturalmente saltado, o código PVS1 não deve ser aplicado, ou se num evento *in-frame* a expressão gênica mantida é >10% com perda parcial de um domínio funcionalmente relevante para proteína, ou em um domínio/região não conhecido removendo menos que 10% do comprimento da proteína (Walker *et al.*, 2023). Em resumo, é preciso avaliar de forma individual, minuciosa e criteriosa todos esses aspectos para determinar se houve de fato uma alteração de *splicing* prejudicial para a proteína, para poder caracterizar a variante spliceogênica como patogênica/provavelmente patogênica ou não.

Utilizamos as novas recomendações (Walker *et al.*, 2023) para interpretação de variantes de *splice* nesse estudo e 3 variantes foram reclassificadas para P/PP, 5 para B/PB e 6 continuaram VUS. A partir do escore no SpliceAI $\geq 0,20$ foi possível adicionar o código PP3, e de forma contrária retirar o código PP3 das que estavam com escore entre $>0,1$ e $<0,2$, e ainda adicionar o código BP7 àquelas com escore $\leq 0,1$, que não estavam na posição $\pm 1,2$ de *splice*, o que contribuiu na reclassificação da variante *SMARCA4:c.859+5G>A* para PB. Ainda, seguindo as recomendações de Walker e colaboradores (2023), conseguimos dar peso PVS1 cheio para variantes fora dos sítios canônicos, as quais foram comprovadas de causar o erro completo de *splice* e levar a transcritos NMD com códon de parada prematuro através da nossa análise de RNA, o que não era possível anteriormente. Nesse aspecto, a força da adição desse critério foi decisiva para as mudanças de classificação das três variantes VUS para P/PP (*BRCA2:c.7618-10T>G*, *ATM:c.7273G>A* e *MSH2:del 5'UTR + del parcial do E1*). Além disso, conseguimos remover ou reduzir a força de PVS1 de variantes exônicas e intrônicas dentro ou fora do sítio canônico, as quais foram vistas de causar alteração de *splicing*, porém de forma não danosa à proteína no ensaio de RNA (*MUTYH:c.934-2A>G* e *NBN:c.2071-4A>G*), ou quando ocorreu uma alteração de *splice* de forma parcial (*leaky splicing*) (*CHEK2:c.320-5T>A*, *MLH1:c.790C>G* e *BRCA1:c.4096+1G>A*).

Existe uma propensão das VUS preditas como spliceogênicas serem mais reclassificadas para PB/B do que para PP/P. No estudo de Horton e colaboradores (2024), dos casos VUS, 45,5% foram reclassificadas para PB/B, e no nosso estudo cerca de 36% das variantes sofreram esse rebaixamento. A aplicação do código BP7_strong só foi possível através dessa nova recomendação, para variantes que não se observou nenhuma alteração de *splice* pelo ensaio de RNA, sendo essencial para a mudança de classificação das variantes de estudo *BRCA1:c.547+5G>A*, *MUTYH:c.1434+2C>T* e *SMARCA4:c.859+5G>A*.

No nosso estudo, variantes no gene *ATM*, *BRCA2* e *MSH2* foram elevadas de VUS para P/PP, enquanto variantes no gene *MUTYH* (em dois pacientes) e em *CHEK2* (em três pacientes)

foram rebaixadas para PB. Resultados semelhantes foram encontrados no estudo de Horton e colaboradores (2024) que relata que o aumento da taxa positiva de diagnóstico (variantes reclassificadas para PP/P) foi superior a 5% nos genes *ATM*, *CDH1*, *MLH1* e *NF1*, enquanto a taxa de decaimento de VUS aconteceu em quase todos os genes do estudo, porém de forma mais relevante nos genes *CHEK2*, *MUTYH*, *NF1* e *TP53*.

Em relação às variantes que permaneceram como VUS, temos duas variantes (*TP53*:c.114A>G e *NBN*:c.2071-4A>G) nos pacientes RNAP-07 e RNAP-110, respectivamente, que suas amostras não tinham RIN adequado, por isso não conseguimos gerar evidências de RNA para impactar em suas classificações. Ainda, temos a variante no paciente RNAP-04 (*RAD51D*:c.80C>A) que está no E1, a qual não é possível realizar também a análise direcionada ao transcrito. No paciente RNAP-03 (*RAD51C*:c.571+4A>G) não vimos grandes diferenças na inspeção visual dos BAMs e no padrão de bandas da eletroforese capilar, embora alguns laboratórios já tenham descrito o evento de salto do E3. No paciente RNAP-115 (*MLH1*:c.790C>G) não foi visto alteração de *splice*, visto que o alelo alternativo estava sendo expresso normalmente, comprovando que não houve perda do E9, porém o efeito da troca de aminoácidos na proteína ainda é inconclusivo. Por último, tem as variantes que possuem eventos de *splice* incompletos (*leaky splice*) nas amostras RNAP-110 e RNAP-124 (*NBN*:c.2071-4A>G e *BRCA1*:c.4096+1G>A). Esse termo significa que eventos anormais de *splicing* ocorrem em uma proporção pequena, de forma parcial, reduzindo as taxas de *splicing* normal, mas não o excluindo, dessa maneira não se sabe se essa pequena proporção de *splicing* anormal é suficiente para desencadear o fenótipo associado ao gene.

No caso da variante *NBN*:c.2071-4A>G, foi identificado um salto *in-frame* do E13 pelo rMATS, em que o nível de inclusão no paciente desse éxon é um pouco menor (85%) que a média dos controles (98%). Por apresentar um evento incompleto, não é possível aplicar PVS1_mod (RNA), mantendo sua classificação como VUS. No caso da variante *BRCA1*:c.4096+1G>A foi visto o salto parcial e total do E10 com nível de inclusão parecido com os controles, porém levemente reduzido com uma diferença de PSI de 11% e 10%, respectivamente, sendo que na literatura é relatado que mesmo com a perda do E10, a proteína mantém uma função mínima suficiente para seu desempenho. E o salto parcial do E10 é ocasionado pela utilização um sítio de *splice* dentro do éxon, o qual também é utilizado por isoformas alternativas. De acordo com o CSpec de *BRCA1* publicado pela ClinGen/ENIGMA deve-se adicionar o critério PVS1_mod (RNA) para salto do E10, porém mesmo com a adição desse critério a variante continua sendo VUS. É importante destacar que segundo a última

revisão do ENIGMA em 2018 essa variante permanece VUS, e na recomendação do CSpec do gene existe uma nota para alterações de salto deste éxon, que menciona que “a possibilidade de variantes de risco intermediário não deve ser descartada”.

Eventos incompletos (ou parciais) ainda são considerados grandes desafios atuais para as análises de *splicing* utilizando o RNA, pois a proporção desse evento anormal pode estar relacionada com a expressão e severidade do fenótipo do paciente e/ou redução de penetrância de um fenótipo clínico esperado. Estudos demonstraram que variantes no gene *BRCA1* que conseguem manter a expressão gênica entre 20-30% do transcrito funcional não devem ser classificadas como patogênicas em indivíduos com HBOC (Walker *et al.*, 2023). Por isso, devem ser realizados mais estudos que avaliam eventos de *splicing* utilizando a métrica do PSI, ou através da expressão alélica específica (desbalanço alélico), correlacionando os fenótipos dos pacientes para poder estabelecer pontos de corte mínimos de expressão de transcritos funcionais para cada gene.

Estudos que avaliaram grandes coortes de indivíduos suspeitos em serem portadores de síndromes de predisposição hereditária (Horton *et al.*, 2022; Kamps-Hughes *et al.*, 2023, Horton *et al.*, 2024) concluíram que o teste genético de RNA foi significativamente mais relevante para ao aumento das taxas de classificações definitivas e redução de VUS em indivíduos autodeclarados negros/afro-americanos, asiáticos ou hispânicos, comparados com indivíduos autodeclarados brancos não hispânicos. Dessa forma, essa metodologia consegue mitigar as discrepâncias no diagnóstico e tratamento de saúde de populações subrepresentadas nos bancos de dados populacionais de variantes genéticas (Ambry Genetics, 2025). Nosso estudo contribuiu na reclassificação de três variantes spliceogênicas VUS para patogênicas, *BRCA2*:c.7618-10T>G, *ATM*:c.7273G>A, *MSH2*:c.del 5'UTR + del parcial do E1, as quais não estão presentes em nenhum dos bancos de frequências populacionais, como gnomAD. A primeira variante possui apenas três depósitos no ClinVar (Mendelics, Ambry e Invitae) e já foi identificada em outro estudo do nosso grupo em um indivíduo masculino com câncer de mama; a segunda possui com apenas um depósito do laboratório Mendelics e a terceira não possui evidências clínicas no ClinVar. Isso indica que essas variantes são raras e encontradas na população brasileira, e uma vez classificadas como patogênicas pode-se estender essa classificação para outros pacientes tanto aparentados como não aparentados portadores da mesma variante, funcionando como um efeito cascata (Horton *et al.*, 2024; Ambry Genetics, 2025). Além disso, essas reclassificações podem solucionar casos com resultados inconclusivos de predisposição ao câncer familiar e promover oportunidades profiláticas e terapêuticas

(Horton *et al.*, 2024).

A aplicação do sequenciamento de RNA como complemento ao teste genético de DNA nas condições atuais, seria mais indicado/adequado para pacientes que apresentem resultado negativo ou inconclusivo (VUS) no teste genético de DNA, os quais também possuem alto risco de câncer hereditário com um forte histórico familiar e/ou pessoal. Porém, deve-se destacar que é uma metodologia considerada de alto custo e de alta complexidade, que depende do trabalho conjunto de uma equipe multidisciplinar composta por especialistas capacitados e atualizados para identificação/interpretação dos eventos de *splice* em relação a uma coorte robusta de controles para realizar a interpretação de variantes spliceogênicas. Ainda, em relação ao acesso ao teste, o teste de RNA não é coberto por planos de saúde brasileiros, e provavelmente o paciente que teria que arcar com os custos adicionais para complementação do seu diagnóstico. Atualmente apenas o laboratório Fleury, com parceria com a Ambry Genetics, oferece um painel comercial de câncer expandido com análise de RNA a partir de amostras de sangue, chamado “RNAinsight”.

Dentre as limitações deste estudo está o fato de que, para fins de padronização de parâmetros para filtragem de eventos anormais de *splicing*, avaliamos na validação clínica apenas pacientes que possuíam VUS preditas de serem spliceogênicas e direcionamos a investigação das variantes de *splicing* somente no gene da VUS de cada caso. Com esses parâmetros definidos e validados, poderemos utilizar as ferramentas de maneira exploratória com a aplicação de filtros ajustados (mais ou menos restritivos) para descobertas secundárias de *splicing* anormal em outros genes de predisposição hereditária (sem variante candidata *a priori*), que precisarão ter seu efeito confirmado com análise direcionada ao transcrito e/ou desbalanço alélico. No entanto, o processo de descobrir um novo diagnóstico a partir de um paciente sem nenhuma variante candidata tem um rendimento inferior daqueles pacientes com uma VUS candidata já identificada (Jaramillo *et al.*, 2024). Outra limitação do estudo é que não incluímos em nossas análises variantes em regiões pouco conservadas, como sítios de ponto de ramificação (*branchpoints*) (que geralmente se localizam nas posições -20 a -40 nucleotídeos de distância do sítio aceptor AG). Também não avaliamos o impacto nas regiões de intensificadores/silenciadores intrônicos/exônicos, os quais são regiões muito curtas, mal definidas, que variam muito na distância dos sítios de *splice*, podendo sobrepor éxons ou íntrons, e por isso são difíceis de serem identificados e, caso uma variante seja identificada nessas regiões, o desafio será entender se afetará a afinidade de ligação dos fatores regulatórios de *splicing* neles (Barbosa *et al.*, 2023). Por fim, nesse estudo não realizamos o tratamento com

puromicina previamente à extração de RNA (que tem como intuito impedir a degradação de transcritos com códon de parada prematuro via *nonsense mediated decay* [NMD]). Optamos por essa abordagem, pois estudos recentes que fizeram captura de RNA com amostras tratadas (P+) e não tratadas com puromicina (P-) conseguiram capturar os mesmos transcritos em ambas as condições, porém amostras P- tiveram uma menor intensidade e frequência da variante alterada. É possível detectar essas variantes mesmo em baixa frequência, pelo fato de alcançarem alta cobertura para as regiões alvo através do método de captura, logo demonstraram que o tratamento de puromicina não é essencial se for utilizada essa abordagem de sequenciamento de RNA (Schwenk et al 2023).

De forma resumida, podemos considerar como desafios que permanecem mesmo com a adição do teste de RNA: **(i)** ferramentas de predição de *splice* devem ser continuamente atualizadas e treinadas (Walker *et al.*, 2023), pois tem dificuldade em identificar e prever impacto em sítios não conservados e redundantes (como *branchpoints*, ESE/ESS e ISE/ISS), assim como prever variantes que ativam eventos de pseudoexonização em regiões intrônicas profundas ou variantes que impactam sítios canônicos atípicos (não GT-AG) nas posições $\pm 1,2$ pb (Canson *et al.*, 2020; Walker *et al.*, 2023), uma vez que o desempenho dessas ferramentas diminui em variantes > 10 pb de distância das junções de *splice* (Barbosa *et al.*, 2023); **(ii)** difícil interpretação de variantes em íntron profundo e das que causam eventos de *splicing* incompletos, devido a falta ainda de estudos de análises de RNA que usem a métrica de PSI ou expressão alélica correlacionando com o fenótipo do paciente (Kamps-Hughes *et al.*, 2023; Walker *et al.*, 2023); **(iii)** limitações inerentes ao sequenciamento de leituras curtas para variantes que causam múltiplos eventos de *splicing* ou de genes que possuem muitas isoformas (múltiplos eventos alternativos) (Wang *et al.*, 2023); **(iv)** limitação de detecção de eventos de retenção parcial de íntrons pelas ferramentas de análise de *splicing* diferencial como rMATS e *splicing* outlier como o FRASER2 (Wang *et al.*, 2023; Jaramillo *et al.*, 2024); **(v)** amostras de leucócitos podem não representar de forma fiel o perfil de *splice*, ou expressão do transcrito no tecido de origem da doença (Wai *et al.*, 2020; Kamps-Hughes *et al.*, 2023); **(vi)** esforços para treinamentos contínuos de cientistas na interpretação de variantes, assim como o compartilhamento de dados e a colaboração entre pesquisadores, laboratórios e médicos no objetivo comum que é oferecer melhores resultados para os pacientes (Horton *et al.*, 2024).

7. CONCLUSÃO

Conseguimos concluir no presente estudo que:

- Desenvolvemos um *pipeline* com preditores *in silico* e parâmetros de filtragem para priorizamos variantes raras, codificantes ou não codificantes e preditas como spliceogênicas ou com efeito em UTRs. Através da revisão de testes genéticos prévios identificamos 182/3.236 (5,6%) variantes intrônicas e exônicas preditas como spliceogênicas, assim como 32/384 (8,3%) variantes com possível impacto na região 5'UTR, e realizamos um levantamento de 898 variantes que estavam na região 3'UTR.
- Validamos uma metodologia para sequenciamento alvo de RNA de 27 genes de predisposição ao câncer, com a qual obtivemos boas métricas de cobertura para as regiões alvo. A comparação entre ferramentas de análise de *splice* evidenciou que o rMATS foi a ferramenta mais sensível para a detecção de eventos no grupo de pacientes com variantes P/PP spliceogênicas, identificando 78% dos eventos desse grupo (11/14).
- Identificamos limitações das ferramentas de análise de *splice* para detectar eventos de retenção parcial de íntron, múltiplos eventos de *splicing* alternativos concomitantes e eventos envolvendo o primeiro éxon e/ou íntron. Por isso, complementar o RNA-cap com o sequenciamento de RNA direcionado e análise de desbalanço alélico foi essencial nas análises de algumas variantes. Essas análises também contribuíram com a identificação de novas evidências ou confirmação de evidências que tinham sido encontradas via ferramentas de *splice* e/ou inspeção visual dos BAMs do RNA-cap, apoiando as conclusões das análises.
- Na validação clínica para avaliação de VUS, a combinação de abordagens de análise de RNA permitiu a reclassificação de 57% (8/14) das variantes, sendo três variantes reclassificadas para P/PP e 5 para B/PB. Das 6 VUS na qual a classificação foi mantida como VUS, duas apresentaram alteração de *splice* parcial e potencialmente prejudicial (*NBN*:c.2071-4A>G e *BRCA1*:c.4096+1G>A), porém são necessários outros estudos funcionais ou populacionais para identificar qual a real consequência fenotípica da redução da isoforma normal e o impacto para o desenvolvimento da doença.

- A combinação da utilização do SpliceAI para triagem de variantes preditas de serem spliceogênicas, rMATS para análise do perfil de *splice* entre o caso vs. controles, e inspeção visual dos BAMs do RNA-cap foram complementares entre si para identificar as variantes, compreender os eventos anormais de *splicing* de forma quantitativa e qualitativa, e contribuir para a classificação de variantes.

8. REFERÊNCIAS

- Agiannitopoulos K, Pepe G, Papadopoulou E, Tsaousis GN, Kampouri S, Maravelaki S, Fassas A, Christodoulou C, Iosifidou R, Karageorgopoulou S, Markopoulos C, Natsiopoulou I, Papazisis K, Vasilaki-Antonatou M, Venizelos V, Ozmen V, Tansan S, Kaban K, Eniu DT, Chiorean A, Nasioulas G. Clinical Utility of Functional RNA Analysis for the Reclassification of *Splicing* Gene Variants in Hereditary Cancer. *Cancer Genomics Proteomics*. 2021 May-Jun;18(3):285-294. doi: 10.21873/cgp.20259. PMID: 33893081; PMCID: PMC8126331.
- Ambry Genetics, 2025. The Growing Impact of Concurrent DNA and RNA Sequencing. Disponível em: https://www.ambrygen.com/file/material/view/2226/WhitePaper_GrowingImpactOfConcurrentDNAandRNASequencing_1224_v4_Final%20%281%29.pdf. Acesso em: Abr. 2025.
- Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015 Jan 15;31(2):166-9. doi: 10.1093/bioinformatics/btu638. Epub 2014 Sep 25. PMID: 25260700; PMCID: PMC4287950.
- Barbosa C, Peixeiro I, Romão L. Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet*. 2013;9(8):e1003529. doi: 10.1371/journal.pgen.1003529. Epub 2013 Aug 8. PMID: 23950723; PMCID: PMC3738444.
- Barbosa P, Savisaar R, Carmo-Fonseca M, Fonseca A. Computational prediction of human deep intronic variation. *Gigascience*. 2022 Dec 28;12:giad085. doi: 10.1093/gigascience/giad085. Epub 2023 Oct 25. PMID: 37878682; PMCID: PMC10599398.
- Blakes AJM, Wai HA, Davies I, Moledina HE, Ruiz A, Thomas T, Bunyan D, Thomas NS, Burren CP, Greenhalgh L, Lees M, Pichini A, Smithson SF, Taylor Tavares AL, O'Donovan P, Douglas AGL; Genomics England Research Consortium, *Splicing* and Disease Working Group; Whiffin N, Baralle D, Lord J. A systematic analysis of *splicing* variants identifies new diagnoses in the 100,000 Genomes Project. *Genome Med*. 2022 Jul 26;14(1):79. doi: 10.1186/s13073-022-01087-x. PMID: 35883178; PMCID: PMC9327385.
- Bozsik A, Papp J, Grolmusz VK, Patócs A, Oláh E, Butz H. Reclassification of Five BRCA1/2 Variants with Unknown Significance Using Complex Functional Study. *Cancer Res Treat*. 2022 Oct;54(4):970-984. doi: 10.4143/crt.2021.1078. Epub 2022 Feb 8. PMID: 35167739; PMCID: PMC9582465.
- Brandão RD, Mensaert K, López-Perolio I, Tserpelis D, Xenakis M, Lattimore V, Walker LC, Kvist A, Vega A, Gutiérrez-Enríquez S, Díez O; KConFaB Investigators; de la Hoya M, Spurdle AB, De Meyer T, Blok MJ. Targeted RNA-seq successfully identifies normal and pathogenic *splicing* events in breast/ovarian cancer susceptibility and Lynch syndrome genes. *Int J Cancer*. 2019 Jul 15;145(2):401-414. doi: 10.1002/ijc.32114. Epub 2019 Feb 7. PMID: 30623411; PMCID: PMC6635756.
- Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl Acad. Sci. USA* 106, 7507–7512 (2009).
- Canson D, Glubb D, Spurdle AB. Variant effect on *splicing* regulatory elements, branchpoint usage, and pseudoexonization: Strategies to enhance bioinformatic prediction using hereditary cancer genes as exemplars. *Hum Mutat*. 2020 Oct;41(10):1705-1721. doi:

10.1002/humu.24074. Epub 2020 Jul 17. Erratum in: Hum Mutat. 2022 Dec;43(12):2328. doi: 10.1002/humu.24500. PMID: 32623769.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic *splicing* enhancers. Nucleic Acids Res. 2003 Jul 1;31(13):3568-71. doi: 10.1093/nar/gkg616. PMID: 12824367; PMCID: PMC169022.

Casadei S, Gulsuner S, Shirts BH, Mandell JB, Kortbawi HM, Norquist BS, Swisher EM, Lee MK, Goldberg Y, O'Connor R, Tan Z, Pritchard CC, King MC, Walsh T. Characterization of splice-altering mutations in inherited predisposition to cancer. Proc Natl Acad Sci U S A. 2019 Dec 26;116(52):26798-26807. doi: 10.1073/pnas.1915608116. Epub 2019 Dec 16. PMID: 31843900; PMCID: PMC6936554.

Castillo-Guardiola V, Rosado-Jiménez L, Sarabia-Meseguer MD, Marín-Vera M, Macías-Cerrolaza JA, García-Hernández R, Zafra-Poves M, Sánchez-Henarejos P, Moreno-Locubiche MÁ, Cuevas-Tortosa E, Arnaldos-Carrillo M, Ayala de la Peña F, Alonso-Romero JL, Noguera-Velasco JA, Ruiz-Espejo F. Next step in molecular genetics of hereditary breast/ovarian cancer: Multigene panel testing in clinical actionably genes and prioritization algorithms in the study of variants of uncertain significance. Eur J Med Genet. 2022 Apr;65(4):104468. doi: 10.1016/j.ejmg.2022.104468. Epub 2022 Mar 1. PMID: 35245693.

Clift K, Macklin S, Halverson C, McCormick JB, Abu Dabrh AM, Hines S. Patients' views on variants of uncertain significance across indications. J Community Genet. 2020 Apr;11(2):139-145. doi: 10.1007/s12687-019-00434-7.

Davy G, Rousselin A, Goardon N, Castéra L, Harter V, Legros A, Muller E, Fouillet R, Brault B, Smirnova AS, Lemoine F, de la Grange P, Guillaud-Bataille M, Caux-Moncoutier V, Houdayer C, Bonnet F, Blanc-Fournier C, Gaildrat P, Frebourg T, Martins A, Vaur D, Krieger S. Detecting *splicing* patterns in genes involved in hereditary breast and ovarian cancer. Eur J Hum Genet. 2017 Oct;25(10):1147-1154. doi: 10.1038/ejhg.2017.116. Epub 2017 Jul 26. PMID: 28905878; PMCID: PMC5602017.

de Angelis de Carvalho N, Niitsuma BN, Kozak VN, Costa FD, de Macedo MP, Kupper BEC, Silva MLG, Formiga MN, Volc SM, Aguiar Junior S, Palmero EI, Casali-da-Rocha JC, Carraro DM, Torrezan GT. Clinical and Molecular Assessment of Patients with Lynch Syndrome and Sarcomas Underpinning the Association with MSH2 Germline Pathogenic Variants. Cancers (Basel). 2020 Jul 9;12(7):1848. doi: 10.3390/cancers12071848.

de Melo Costa VR, Pfeuffer J, Louloui A, Ørom UAV, Piro RM. SPLICE-q: a Python tool for genome-wide quantification of *splicing* efficiency. BMC Bioinformatics. 2021 Jul 15;22(1):368. doi: 10.1186/s12859-021-04282-6. PMID: 34266387; PMCID: PMC8281633.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PMID: 23104886; PMCID: PMC3530905.

Ellingford, J. M., Ahn, J. W., Bagnall, R. D., Baralle, D., Barton, S., Campbell, C., Downes, K., Ellard, S., Duff-Farrier, C., FitzPatrick, D. R., Greally, J. M., Ingles, J., Krishnan, N., Lord, J., Martin, H. C., Newman, W. G., O'Donnell-Luria, A., Ramsden, S. C., Rehm, H. L., Richardson, E., ... Whiffin, N. (2022). Recommendations for clinical interpretation of variants found in non-coding regions of the genome. Genome medicine, 14(1), 73. <https://doi.org/10.1186/s13073-022-01073-3>.

- Farh KK. Predicting *Splicing* from Primary Sequence with Deep Learning. *Cell*. 2019 Jan 24;176(3):535-548.e24. doi: 10.1016/j.cell.2018.12.015. Epub 2019 Jan 17. PMID: 30661751.
- Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. *J Exp Clin Cancer Res*. 2020 Mar 4;39(1):46. doi: 10.1186/s13046-020-01554-6.
- Ferraro NM, Strober BJ, Einson J, Abell NS, Aguet F, Barbeira AN, Brandt M, Bucan M, Castel SE, Davis JR, Greenwald E, Hess GT, Hilliard AT, Kember RL, Kotis B, Park Y, Peloso G, Ramdas S, Scott AJ, Smail C, Tsang EK, Zekavat SM, Ziosi M, Aradhana; TOPMed Lipids Working Group; Ardlie KG, Assimes TL, Bassik MC, Brown CD, Correa A, Hall I, Im HK, Li X, Natarajan P; GTEx Consortium; Lappalainen T, Mohammadi P, Montgomery SB, Battle A. Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science*. 2020 Sep 11;369(6509):eaaz5900. doi: 10.1126/science.aaz5900. Epub 2020 Sep 10. PMID: 32913073; PMCID: PMC7646251.
- Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J. Accurate classification of BRCA1 variants with saturation genome editing. *Nature*. 2018 Oct;562(7726):217-222. doi: 10.1038/s41586-018-0461-z. Epub 2018 Sep 12. PMID: 30209399; PMCID: PMC6181777.
- Garcia FAO, de Andrade ES, Palmero EI. Insights on variant analysis in silico tools for pathogenicity prediction. *Front Genet*. 2022 Nov 29;13:1010327. doi: 10.3389/fgene.2022.1010327. PMID: 36568376; PMCID: PMC9774026.
- Garutti M, Foffano L, Mazzeo R, Michelotti A, Da Ros L, Viel A, Miolo G, Zambelli A, Puglisi F. Hereditary Cancer Syndromes: A Comprehensive Review with a Visual Tool. *Genes (Basel)*. 2023 Apr 30;14(5):1025. doi: 10.3390/genes14051025. PMID: 37239385; PMCID: PMC10218093.
- Gelli E, Colombo M, Pinto AM, De Vecchi G, Foglia C, Amitrano S, Morbidoni V, Imperatore V, Manoukian S, Baldassarri M, Lo Rizzo C, Catania L, Frullanti E, Tagliafico E, Cortesi L, Spaggiari F, Mencarelli MA, Trevisson E, Radice P, Renieri A, Ariani F. Usefulness and Limitations of Comprehensive Characterization of mRNA *Splicing* Profiles in the Definition of the Clinical Relevance of *BRCA1/2* Variants of Uncertain Significance. *Cancers (Basel)*. 2019 Mar 1;11(3):295. doi: 10.3390/cancers11030295. PMID: 30832263; PMCID: PMC6468917.
- GenomeAsia100K Consortium. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019 Dec;576(7785):106-111. doi: 10.1038/s41586-019-1793-z. Epub 2019 Dec 4. PMID: 31802016; PMCID: PMC7054211.
- Gimeno-Valiente F, López-Rodas G, Castillo J, Franco L. The Many Roads from Alternative *Splicing* to Cancer: Molecular Mechanisms Involving Driver Genes. *Cancers (Basel)*. 2024 Jun 1;16(11):2123. doi: 10.3390/cancers16112123. PMID: 38893242; PMCID: PMC11171328.
- Ha C, Kim JW, Jang JH. Performance Evaluation of SpliceAI for the Prediction of *Splicing* of *NF1* Variants. *Genes (Basel)*. 2021 Aug 25;12(9):1308. doi: 10.3390/genes12091308. PMID: 34573290; PMCID: PMC8472818.
- Halperin RF, Hegde A, Lang JD, Raupach EA; C4RCD Research Group; Legendre C, Liang WS, LoRusso PM, Sekulic A, Sosman JA, Trent JM, Rangasamy S, Pirrotte P, Schork NJ.

- Improved methods for RNAseq-based alternative *splicing* analysis. *Sci Rep*. 2021 May 24;11(1):10740. doi: 10.1038/s41598-021-89938-2. PMID: 34031440; PMCID: PMC8144374.
- Hernandez F, Conner BR, Richardson ME, LaDuca H, Chao E, Pesaran T, Karam R. Classification of the canonical *splice* alteration MUTYH c.934-2A > G is likely benign based on RNA and clinical data. *Cold Spring Harb Mol Case Stud*. 2022 Jan 10;8(1):a006152. doi: 10.1101/mcs.a006152. PMID: 34716202; PMCID: PMC8744492.
- Horton C, Cass A, Conner BR, Hoang L, Zimmermann H, Abualkheir N, Burks D, Qian D, Molparia B, Vuong H, LaDuca H, Grzybowski J, Durda K, Pilarski R, Profato J, Clayback K, Mahoney M, Schroeder C, Torres-Martinez W, Elliott A, Chao EC, Karam R. Mutational and *splicing* landscape in a cohort of 43,000 patients tested for hereditary cancer. *NPJ Genom Med*. 2022 Aug 25;7(1):49. doi: 10.1038/s41525-022-00323-y. PMID: 36008414; PMCID: PMC9411123.
- Horton C, Hoang L, Zimmermann H, *et al*. Diagnostic Outcomes of Concurrent DNA and RNA Sequencing in Individuals Undergoing Hereditary Cancer Testing. *JAMA Oncol*. 2024;10(2):212-219. doi:10.1001/jamaoncol.2023.5586.
- Iyevleva, A.G., Imyanitov, E.N. Cytotoxic and targeted therapy for hereditary cancers. *Hered Cancer Clin Pract* 14, 17 (2016). <https://doi.org/10.1186/s13053-016-0057-2>.
<https://doi.org/10.1186/s13053-016-0057-2>.
- Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ,
- Jamuar, S.S., Tan, E. Clinical application of next-generation sequencing for Mendelian diseases. *Hum Genomics* 9, 10 (2015). <https://doi.org/10.1186/s40246-015-0031-5>.
- Jang W, Park J, Chae H, Kim M. Comparison of *In Silico* Tools for Splice-Altering Variant Prediction Using Established Spliceogenic Variants: An End-User's Point of View. *Int J Genomics*. 2022 Oct 13;2022:5265686. doi: 10.1155/2022/5265686. PMID: 36275637; PMCID: PMC9584665.
- Jaramillo Oquendo C, Wai HA, Rich WI, Bunyan DJ, Thomas NS, Hunt D, Lord J, Douglas AGL, Baralle D. Identification of diagnostic candidates in Mendelian disorders using an RNA sequencing-centric approach. *Genome Med*. 2024 Sep 9;16(1):110. doi: 10.1186/s13073-024-01381-w. PMID: 39252027; PMCID: PMC11382415.
- Jian X, Boerwinkle E, Liu X. In silico prediction of *splice*-altering single nucleotide variants in the human genome. *Nucleic Acids Res*. 2014 Dec 16;42(22):13534-44. doi: 10.1093/nar/gku1206. PMID: 25416802; PMCID: PMC4267638.
- Kamps-Hughes N, Carlton VEH, Fresard L, *et al*. A Systematic Method for Detecting Abnormal mRNA *Splicing* and Assessing Its Clinical Impact in Individuals Undergoing Genetic Testing for Hereditary Cancer Syndromes. *J Mol Diagn*. 2023;25(3):156-167. doi:10.1016/j.jmoldx.2022.12.002.
- Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*. 2010 Dec;7(12):1009-15. doi: 10.1038/nmeth.1528. Epub 2010 Nov 7. PMID: 21057496; PMCID: PMC3037023.
- Landrith T, Li B, Cass AA, Conner BR, LaDuca H, McKenna DB, Maxwell KN, Domchek S, Morman NA, Heinlen C, Wham D, Koptiuch C, Vagher J, Rivera R, Bunnell A, Patel G, Geurts JL, Depas MM, Gaonkar S, Pirzadeh-Miller S, Krukenberg R, Seidel M, Pilarski R, Farmer M, Pyrtel K, Milliron K, Lee J, Hoodfar E, Nathan D, Ganzak AC, Wu S, Vuong H,

Xu D, Arulmoli A, Parra M, Hoang L, Molparia B, Fennessy M, Fox S, Charpentier S, Burdette J, Pesaran T, Profato J, Smith B, Haynes G, Dalton E, Crandall JR, Baxter R, Lu HM, Tippin-Davis B, Elliott A, Chao E, Karam R. *Splicing* profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. NPJ Precis Oncol. 2020 Feb 24;4:4. doi: 10.1038/s41698-020-0109-y.

Li HD, Funk CC, McFarland K, Dammer EB, Allen M, Carrasquillo MM, Levites Y, Chakrabarty P, Burgess JD, Wang X, Dickson D, Seyfried NT, Duong DM, Lah JJ, Younkin SG, Levey AI, Omenn GS, Ertekin-Taner N, Golde TE, Price ND. Integrative functional genomic analysis of intron retention in human and mouse brain with Alzheimer's disease. Alzheimers Dement. 2021 Jun;17(6):984-1004. doi: 10.1002/alz.12254. Epub 2021 Jan 21. PMID: 33480174; PMCID: PMC8248162.

Li J, Wang P, Zhang C, Han S, Xiao H, Liu Z, Wang X, Liu W, Wei B, Ma J, Li H, Guo Y. Characterization of Synonymous *BRCA1*:c.132C>T as a Pathogenic Variant. Front Oncol. 2022 Jan 11;11:812656. doi: 10.3389/fonc.2021.812656. PMID: 35087763; PMCID: PMC8789006.

Li J, Wang Z, Han Y, Jin C, Cheng D, Zhou YA, Zhen J. Genetic and functional analyses detect an EXT1 *splicing* pathogenic variant in a Chinese hereditary multiple exostosis (HME) family. Mol Genet Genomic Med. 2022 Mar;10(3):e1878. doi: 10.1002/mgg3.1878. Epub 2022 Feb 1. PMID: 35106951; PMCID: PMC8922959.

Liu J, Lin CX, Zhang X, Li Z, Huang W, Liu J, Guan Y, Li HD. Computational approaches for detecting disease-associated alternative *splicing* events. Brief Bioinform. 2023 May 19;24(3):bbad106. doi: 10.1093/bib/bbad106. PMID: 36987778.

Liu L, Dilworth D, Gao L, Monzon J, Summers A, Lassam N, Hogg D. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. Nat Genet. 1999 Jan;21(1):128-32. doi: 10.1038/5082. PMID: 9916806.

Lord J, Gallone G, Short PJ, McRae JF, Ironfield H, Wynn EH, Gerety SS, He L, Kerr B, Johnson DS, McCann E, Kinning E, Flinter F, Temple IK, Clayton-Smith J, McEntagart M, Lynch SA, Joss S, Douzgou S, Dabir T, Clowes V, McConnell VPM, Lam W, Wright CF, FitzPatrick DR, Firth HV, Barrett JC, Hurles ME; Deciphering Developmental Disorders study. Pathogenicity and selective constraint on variation near splice sites. Genome Res. 2019 Feb;29(2):159-170. doi: 10.1101/gr.238444.118. Epub 2018 Dec 26. PMID: 30587507; PMCID: PMC6360807.

Matthes T, Aguilar-Martinez P, Pizzi-Bosman L, Darbellay R, Rubbia-Brandt L, Giostra E, Michel M, Ganz T, Beris P. Severe hemochromatosis in a Portuguese family associated with a new mutation in the 5'-UTR of the HAMP gene. Blood. 2004 Oct 1;104(7):2181-3. doi: 10.1182/blood-2004-01-0332. Epub 2004 Jun 15. PMID: 15198949.

Mertes C, Scheller IF, Yépez VA, Çelik MH, Liang Y, Kremer LS, Gusic M, Prokisch H, Gagneur J. Detection of aberrant *splicing* events in RNA-seq data using FRASER. Nat Commun. 2021 Jan 22;12(1):529. doi: 10.1038/s41467-020-20573-7. Erratum in: Nat Commun. 2022 Jun 16;13(1):3474. doi: 10.1038/s41467-022-31242-2. PMID: 33483494; PMCID: PMC7822922.

Morak M, Schaefer K, Steinke-Lange V, Koehler U, Keinath S, Massdorf T, Mauracher B, Rahner N, Bailey J, Kling C, Haeusser T, Laner A, Holinski-Feder E. Full-length transcript amplification and sequencing as universal method to test mRNA integrity and biallelic

expression in mismatch repair genes. *Eur J Hum Genet.* 2019 Dec;27(12):1808-1820. doi: 10.1038/s41431-019-0472-8. Epub 2019 Jul 22. PMID: 31332305; PMCID: PMC6870986.

Morak M, Pineda M, Martins A, Gaildrat P, Tubeuf H, Drouet A, Gómez C, Dámaso E, Schaefer K, Steinke-Lange V, Koehler U, Laner A, Hauchard J, Chauris K, Holinski-Feder E, Capellá G. *Splicing* analyses for variants in MMR genes: best practice recommendations from the European Mismatch Repair Working Group. *Eur J Hum Genet.* 2022 Sep;30(9):1051-1059. doi: 10.1038/s41431-022-01106-w. Epub 2022 Jun 9. PMID: 35676339; PMCID: PMC9437034.

Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, Cox E, Davidson C, Ermolaeva O, Farrell CM, Fatima R, Gil L, Goldfarb T, Gonzalez JM, Haddad D, Hardy M, Hunt T, Jackson J, Joardar VS, Kay M, Kodali VK, McGarvey KM, McMahon A, Mudge JM, Murphy DN, Murphy MR, Rajput B, Rangwala SH, Riddick LD, Thibaud-Nissen F, Threadgold G, Vatsan AR, Wallin C, Webb D, Flicek P, Birney E, Pruitt KD, Frankish A, Cunningham F, Murphy TD. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature.* 2022 Apr;604(7905):310-315. doi: 10.1038/s41586-022-04558-8. Epub 2022 Apr 6. PMID: 35388217; PMCID: PMC9007741.

Murff HJ, Spigel DR, Syngal S. Does This Patient Have a Family History of Cancer An Evidence-Based Analysis of the Accuracy of Family Cancer History. *JAMA.* 2004;292(12):1480–1489. doi:10.1001/jama.292.12.1480.

Nagy, R., Sweet, K. & Eng, C. Highly penetrant hereditary cancer syndromes. *Oncogene* 23, 6445–6470 (2004). <https://doi.org/10.1038/sj.onc.1207714>.

Norton SS, Vaquero-Garcia J, Lahens NF, Grant GR, Barash Y. Outlier detection for improved differential *splicing* quantification from RNA-Seq experiments with replicates. *Bioinformatics.* 2018 May 1;34(9):1488-1497. doi: 10.1093/bioinformatics/btx790. PMID: 29236961; PMCID: PMC6454425.

Occhi G, Regazzo D, Trivellin G, Boaretto F, Ciato D, Bobisse S, Ferasin S, Cetani F, Pardi E, Korbonits M, Pellegata NS, Sidarovich V, Quattrone A, Opocher G, Mantero F, Scaroni C. A novel mutation in the upstream open reading frame of the CDKN1B gene causes a MEN4 phenotype. *PLoS Genet.* 2013 Mar;9(3):e1003350. doi: 10.1371/journal.pgen.1003350. Epub 2013 Mar 21. PMID: 23555276; PMCID: PMC3605397.

Piñero TA, Soukariéh O, Rolain M, Alvarez K, López-Köstner F, Torrezan GT, Carraro DM, De Oliveira Nascimento IL, Bomfim TF, Machado-Lopes TMB, Freitas JC, Toralles MB, Sandes KA, Rossi BM, Junior SA, Meira J, Dominguez-Valentin M, Møller P, Vaccaro CA, Martins A, Pavicic WH. MLH1 intronic variants mapping to + 5 position of *splice* donor sites lead to deleterious effects on RNA *splicing*. *Fam Cancer.* 2020 Oct;19(4):323-336. doi: 10.1007/s10689-020-00182-5.

Postel MD, Culver JO, Ricker C, Craig DW. Transcriptome analysis provides critical answers to the "variants of uncertain significance" conundrum. *Hum Mutat.* 2022;43(11):1590-1608. doi:10.1002/humu.24394.

Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* 505, 302–308 (2014). <https://doi.org/10.1038/nature12981>.

RAMALHO, Rodrigo Fernandes. **Análise dos sinais de seleção natural em reguladores de *splicing* exônicos do genoma humano.** Orientador: Dr.Diogo Meyer. 2013. Tese (Doutorado em Ciências na área de Genética e Biologia Evolutiva – Universidade de São Paulo (USP), São Paulo, 2012. doi:10.11606/T.41.2012.tde-15102012-105825. Acesso em: 24 jan. 2025.

- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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. doi: 10.1038/gim.2015.30.
- Riepe TV, Khan M, Roosing S, Cremers FPM, 't Hoen PAC. Benchmarking deep learning splice prediction tools using functional splice assays. *Hum Mutat*. 2021 Jul;42(7):799-810. doi: 10.1002/humu.24212. Epub 2021 May 20. PMID: 33942434; PMCID: PMC8360004.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011 Jan;29(1):24-6. doi: 10.1038/nbt.1754. PMID: 21221095; PMCID: PMC3346182.
- Rofes P, Menéndez M, González S, Tornero E, Gómez C, Vargas-Parra G, Montes E, Salinas M, Solanes A, Brunet J, Teulé A, Capellá G, Feliubadaló L, Del Valle J, Pineda M, Lázaro C. Improving Genetic Testing in Hereditary Cancer by RNA Analysis: Tools to Prioritize *Splicing* Studies and Challenges in Applying American College of Medical Genetics and Genomics Guidelines. *J Mol Diagn*. 2020 Dec;22(12):1453-1468. doi: 10.1016/j.jmoldx.2020.09.007. Epub 2020 Oct 1. PMID: 33011440.
- Ryu JS, Lee HY, Cho EH, Yoon KA, Kim MK, Joo J, Lee ES, Kang HS, Lee S, Lee DO, Lim MC, Kong SY. Exon *splicing* analysis of intronic variants in multigene cancer panel testing for hereditary breast/ovarian cancer. *Cancer Sci*. 2020 Oct;111(10):3912-3925. doi: 10.1111/cas.14600. Epub 2020 Sep 2. PMID: 32761968; PMCID: PMC7540976.
- Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative *Splicing* Signatures in RNA-seq Data: Percent *Spliced* in (PSI). *Curr Protoc Hum Genet*. 2015 Oct 6;87:11.16.1-11.16.14. doi: 10.1002/0471142905.hg1116s87. PMID: 26439713.
- Schulz J, Mah N, Neuenschwander M, Kischka T, Ratei R, Schlag PM, Castaños-Vélez E, Fichtner I, Tunn PU, Denkert C, Klaas O, Berdel WE, von Kries JP, Makalowski W, Andrade-Navarro MA, Leutz A, Wethmar K. Loss-of-function uORF mutations in human malignancies. *Sci Rep*. 2018 Feb 5;8(1):2395. doi: 10.1038/s41598-018-19201-8. Erratum in: *Sci Rep*. 2021 Jan 28;11(1):2917. PMID: 29402903; PMCID: PMC5799362.
- Schuster SL, Hsieh AC. The Untranslated Regions of mRNAs in Cancer. *Trends Cancer*. 2019 Apr;5(4):245-262. doi: 10.1016/j.trecan.2019.02.011. Epub 2019 Mar 22. PMID: 30961831; PMCID: PMC6465068.
- Schwenk V, Leal Silva RM, Scharf F, Knaust K, Wendlandt M, Häusser T, Pickl JMA, Steinke-Lange V, Laner A, Morak M, Holinski-Feder E, Wolf DA. Transcript capture and ultradeep long-read RNA sequencing (CAPLRseq) to diagnose HNPCC/Lynch syndrome. *J Med Genet*. 2023 Aug;60(8):747-759. doi: 10.1136/jmg-2022-108931. Epub 2023 Jan 2. PMID: 36593122; PMCID: PMC10423559.
- Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. rMATS: robust and flexible detection of differential alternative *splicing* from replicate RNA-Seq data. *Proc Natl Acad Sci U S A*. 2014 Dec 23;111(51):E5593-601. doi: 10.1073/pnas.1419161111. Epub 2014 Dec 5. PMID: 25480548; PMCID: PMC4280593.
- Silva FCC, Torrezan GT, Ferreira JRO, Oliveira LP, Begnami MDFS, Aguiar S Junior, Carraro DM. Germline Mutations in MLH1 Leading to Isolated Loss of PMS2 Expression in

Lynch Syndrome: Implications for Diagnostics in the Clinic. *Am J Surg Pathol*. 2017 Jun;41(6):861-864. doi: 10.1097/PAS.0000000000000827.

Steri M, Idda ML, Whalen MB, Orrù V. Genetic variants in mRNA untranslated regions. *Wiley Interdiscip Rev RNA*. 2018 Jul;9(4):e1474. doi: 10.1002/wrna.1474. Epub 2018 Mar 26. PMID: 29582564; PMCID: PMC6002891.

Tadaka S, Katsuoka F, Ueki M, Kojima K, Makino S, Saito S, Otsuki A, Gocho C, Sakurai-Yageta M, Danjoh I, Motoike IN, Yamaguchi-Kabata Y, Shiota M, Koshiha S, Nagasaki M, Minegishi N, Hozawa A, Kuriyama S, Shimizu A, Yasuda J, Fuse N; Tohoku Medical Megabank Project Study Group; Tamiya G, Yamamoto M, Kinoshita K. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. *Hum Genome Var*. 2019 Jun 18;6:28. doi: 10.1038/s41439-019-0059-5. PMID: 31240104; PMCID: PMC6581902.

Tadaka S, Kawashima J, Hishinuma E, Saito S, Okamura Y, Otsuki A, Kojima K, Komaki S, Aoki Y, Kanno T, Saigusa D, Inoue J, Shiota M, Takayama J, Katsuoka F, Shimizu A, Tamiya G, Shimizu R, Hiratsuka M, Motoike IN, Koshiha S, Sasaki M, Yamamoto M, Kinoshita K. jMorp: Japanese Multi-Omics Reference Panel update report 2023. *Nucleic Acids Res*. 2024 Jan 5;52(D1):D622-D632. doi: 10.1093/nar/gkad978. PMID: 37930845; PMCID: PMC10767895.

Tavtigian SV, Harrison SM, Boucher KM, Biesecker LG. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat*. 2020 Oct;41(10):1734-1737. doi: 10.1002/humu.24088. Epub 2020 Aug 30. PMID: 32720330; PMCID: PMC8011844.

Truty R, Ouyang K, Rojahn S, *et al*. Spectrum of *splicing* variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *Am J Hum Genet*. 2021;108(4):696-708. doi:10.1016/j.ajhg.2021.03.006

von Bohlen AE, Böhm J, Pop R, Johnson DS, Tolmie J, Stücker R, Morris-Rosendahl D, Scherer G. A mutation creating an upstream initiation codon in the SOX9 5' UTR causes acampomelic campomelic dysplasia. *Mol Genet Genomic Med*. 2017 Mar 21;5(3):261-268. doi: 10.1002/mgg3.282. PMID: 28546996; PMCID: PMC5441400.

Wai HA, Lord J, Lyon M, Gunning A, Kelly H, Cibir P, Seaby EG, Spiers-Fitzgerald K, Lye J, Ellard S, Thomas NS, Bunyan DJ, Douglas AGL, Baralle D; *Splicing* and disease working group. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med*. 2020 Jun;22(6):1005-1014. doi: 10.1038/s41436-020-0766-9. Epub 2020 Mar 3. Erratum in: *Genet Med*. 2020 Jun;22(6):1129. doi: 10.1038/s41436-020-0789-2. PMID: 32123317; PMCID: PMC7272326.

Walker LC, Hoya M, Wiggins GAR, Lindy A, Vincent LM, Parsons MT, Canson DM, Bis-Brewer D, Cass A, Tchourbanov A, Zimmermann H, Byrne AB, Pesaran T, Karam R, Harrison SM, Spurdle AB; ClinGen Sequence Variant Interpretation Working Group. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on *splicing*: Recommendations from the ClinGen SVI *Splicing* Subgroup. *Am J Hum Genet*. 2023 Jul 6;110(7):1046-1067. doi: 10.1016/j.ajhg.2023.06.002. Epub 2023 Jun 22. PMID: 37352859; PMCID: PMC10357475.

Wang J, Wang X, Bhat A, Chen Y, Xu K, Mo YY, Yi SS, Zhou Y. Comprehensive Network Analysis Reveals Alternative *Splicing*-Related lncRNAs in Hepatocellular Carcinoma. *Front*

Genet. 2020 Jul 15;11:659. doi: 10.3389/fgene.2020.00659. PMID: 32760422; PMCID: PMC7373802.

Wang R, Helbig I, Edmondson AC, Lin L, Xing Y. *Splicing* defects in rare diseases: transcriptomics and machine learning strategies towards genetic diagnosis. Brief Bioinform. 2023 Sep 20;24(5):bbad284. doi: 10.1093/bib/bbad284. PMID: 37580177; PMCID: PMC10516351.

Wangensteen T, Felde CN, Ahmed D, Mæhle L, Ariansen SL. Diagnostic mRNA *splicing* assay for variants in *BRCA1* and *BRCA2* identified two novel pathogenic *splicing* aberrations. Hered Cancer Clin Pract. 2019 May 22;17:14. doi: 10.1186/s13053-019-0113-9. PMID: 31143303; PMCID: PMC6532242.

Ward, Laura Sterian. (2002). Entendendo o Processo Molecular da Tumorigênese. Arquivos Brasileiros de Endocrinologia & Metabologia, 46(4), 351-360. <https://doi.org/10.1590/S0004-27302002000400006>.

Wen Y, Liu Y, Xu Y, Zhao Y, Hua R, Wang K, Sun M, Li Y, Yang S, Zhang XJ, Kruse R, Cichon S, Betz RC, Nöthen MM, van Steensel MA, van Geel M, Steijlen PM, Hohl D, Huber M, Dunnill GS, Kennedy C, Messenger A, Munro CS, Terrinoni A, Hovnanian A, Bodemer C, de Prost Y, Paller AS, Irvine AD, Sinclair R, Green J, Shang D, Liu Q, Luo Y, Jiang L, Chen HD, Lo WH, McLean WH, He CD, Zhang X. Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. Nat Genet. 2009 Feb;41(2):228-33. doi: 10.1038/ng.276. Epub 2009 Jan 4. Erratum in: Nat Genet. 2009 Jun;41(6):762. PMID: 19122663.

Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, Roberts AM, Quaife NM, Schafer S, Rackham O, Alföldi J, O'Donnell-Luria AH, Francioli LC; Genome Aggregation Database Production Team; Genome Aggregation Database Consortium, Cook SA, Barton PJR, MacArthur DG, Ware JS. Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. Nat Commun. 2020 May 27;11(1):2523. doi: 10.1038/s41467-019-10717-9. Erratum in: Nat Commun. 2021 Feb 2;12(1):839. PMID: 32461616; PMCID: PMC7253449.

Wieder N, D'Souza EN, Martin-Geary AC, Lassen FH, Talbot-Martin J, Fernandes M, Chothani SP, Rackham OJL, Schafer S, Aspden JL, MacArthur DG, Davies RW, Whiffin N. Differences in 5'untranslated regions highlight the importance of translational regulation of dosage sensitive genes. Genome Biol. 2024 Apr 29;25(1):111. doi: 10.1186/s13059-024-03248-0. PMID: 38685090; PMCID: PMC11057154.

Yalamanchili HK, Li Z, Wang P, Wong MP, Yao J, Wang J. SpliceNet: recovering *splicing* isoform-specific differential gene networks from RNA-Seq data of normal and diseased samples. Nucleic Acids Res. 2014 Sep;42(15):e121. doi: 10.1093/nar/gku577. Epub 2014 Jul 17. PMID: 25034693; PMCID: PMC4150760.

Yang C, Arnold AG, Catchings A, Rai V, Stadler ZK, Zhang L. The RAD51D c.82G>A (p.Val28Met) variant disrupts normal *splicing* and is associated with hereditary ovarian cancer. Breast Cancer Res Treat. 2021 Feb;185(3):869-877. doi: 10.1007/s10549-020-06066-7. Epub 2021 Jan 16. PMID: 33452952; PMCID: PMC8414856.

Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA *splicing* signals. J Comput Biol. 2004;11(2-3):377-94. doi: 10.1089/1066527041410418. PMID: 15285897.

Zemankova P, Cerna M, Horackova K, *et al.* A deep intronic recurrent *CHEK2* variant c.1009-118_1009-87delinsC affects pre-mRNA *splicing* and contributes to hereditary breast cancer predisposition. *Breast*. 2024;75:103721. doi:10.1016/j.breast.2024.103721.

Zhang P, Chaldebas M, Ogishi M, Al Qureshah F, Ponsin K, Feng Y, Rinchai D, Milisavljevic B, Han JE, Moncada-Vélez M, Keles S, Schröder B, Stenson PD, Cooper DN, Cobat A, Boisson B, Zhang Q, Boisson-Dupuis S, Abel L, Casanova JL. Genome-wide detection of human intronic AG-gain variants located between *splicing* branchpoints and canonical splice acceptor sites. *Proc Natl Acad Sci U S A*. 2023 Nov 14;120(46):e2314225120. doi: 10.1073/pnas.2314225120. Epub 2023 Nov 6. PMID: 37931111; PMCID: PMC10655562.

Zhang P, Philippot Q, Ren W, Lei WT, Li J, Stenson PD, Palacín PS, Colobran R, Boisson B, Zhang SY, Puel A, Pan-Hammarström Q, Zhang Q, Cooper DN, Abel L, Casanova JL. Genome-wide detection of human variants that disrupt intronic branchpoints. *Proc Natl Acad Sci U S A*. 2022 Nov;119(44):e2211194119. doi: 10.1073/pnas.2211194119. Epub 2022 Oct 28. PMID: 36306325; PMCID: PMC9636908.

Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5'untranslated region variants with the UTRannotator. *Bioinformatics*. 2021 May 23;37(8):1171-1173. doi: 10.1093/bioinformatics/btaa783. PMID: 32926138; PMCID: PMC8150139.

Zhang Z, Pan Z, Ying Y, Xie Z, Adhikari S, Phillips J, Carstens RP, Black DL, Wu Y, Xing Y. Deep-learning augmented RNA-seq analysis of transcript *splicing*. *Nat Methods*. 2019 Apr;16(4):307-310. doi: 10.1038/s41592-019-0351-9. Epub 2019 Mar 25. PMID: 30923373; PMCID: PMC7605494.

APÊNDICE A - Painéis de genes utilizados para teste genético

Em verde estão os genes presentes em cada painel e em azul estão os genes comuns nos 6 painéis.

Genes	26G (DNA)	27G* (RNA)	35G (DNA)	39G (DNA)	94G (DNA)	112G (DNA)	GENES COMUNS
<i>ACD</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>AIP</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>AKT1</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>ALK</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>APC</i>	X	X	X	X	X	X	X
<i>ATM</i>	X	X	X	X	X	X	X
<i>AXIN2</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>BAP1</i>	(-)	(-)	X	X	X	X	(-)
<i>BARD1</i>	X	X	X	(-)	(-)	X	(-)
<i>BLM</i>	(-)	(-)	(-)	X	X	X	(-)
<i>BMPR1A</i>	(-)	X	X	X	X	X	(-)
<i>BRCA1</i>	X	X	X	X	X	X	X
<i>BRCA2</i>	X	X	X	X	X	X	X
<i>BRIP1</i>	X	X	X	X	X	X	X
<i>BUB1B</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>CASR</i>	(-)	(-)	(-)	(-)		X	(-)
<i>CDC73</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>CDH1</i>	X	X	X	X	X	X	X
<i>CDK4</i>	(-)	X	(-)	X	X	X	(-)
<i>CDKN1B</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>CDKN1C</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>CDKN2A</i>	(-)	X	X	X	X	X	(-)
<i>CEBPA</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>CEP57</i>	(-)	(-)	(-)	(-)	X	(-)	(-)
<i>CHEK2</i>	X	X	X	X	X	X	X
<i>CREBBP</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>CTNNA1</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>CYLD</i>	(-)	(-)	(-)	(-)	X	(-)	(-)
<i>DDB2</i>	(-)	(-)	(-)	(-)	X	(-)	(-)
<i>DICER1</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>DIS3L2</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>EGFR</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>EGLN1</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>EPCAM</i>	X	X	X	X	X	X	X
<i>ERBB2</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>ERCC2</i>	(-)	(-)	(-)	(-)	X	X	(-)

ERCC3	(-)	(-)	(-)	(-)	X	(-)	(-)
ERCC4	(-)	(-)	(-)	(-)	X	(-)	(-)
ERCC5	(-)	(-)	(-)	(-)	X	(-)	(-)
EXT1	(-)	(-)	(-)	(-)	X	X	(-)
EXT2	(-)	(-)	(-)	(-)	X	X	(-)
EZH2	(-)	(-)	(-)	(-)	X	(-)	(-)
FAM175A	X	(-)	(-)	(-)	(-)	(-)	(-)
FANCA	(-)	(-)	(-)	X	X	(-)	(-)
FANCB	(-)	(-)	(-)	X	X	(-)	(-)
FANCC	(-)	(-)	(-)	X	X	X	(-)
FANCD2	(-)	(-)	(-)	X	X	(-)	(-)
FANCE	(-)	(-)	(-)	X	X	(-)	(-)
FANCF	(-)	(-)	(-)	X	X	(-)	(-)
FANCG	(-)	(-)	(-)	X	X	X	(-)
FANCI	(-)	(-)	(-)	X	X	(-)	(-)
FANCL	(-)	(-)	(-)	X	X	(-)	(-)
FANCM	(-)	(-)	(-)	(-)	X	X	(-)
FH	(-)	(-)	(-)	(-)	X	X	(-)
FLCN	(-)	(-)	(-)	(-)	X	X	(-)
GALNT12	(-)	(-)	(-)	(-)	(-)	X	(-)
GATA2	(-)	(-)	(-)	(-)	X	X	(-)
GPC3	(-)	(-)	(-)	(-)	X	(-)	(-)
GREM1	(-)	(-)	X	(-)	(-)	X	(-)
HNF1A	(-)	(-)	(-)	(-)	X	X	(-)
HOXB13	(-)	(-)	(-)	(-)	(-)	X	(-)
HRAS	(-)	(-)	(-)	(-)	X	X	(-)
KIF1B	(-)	(-)	(-)	(-)	(-)	X	(-)
KIT	(-)	(-)	(-)	(-)	X	X	(-)
LZTR1	(-)	(-)	(-)	(-)	(-)	X	(-)
MAX	(-)	(-)	(-)	(-)	X	X	(-)
MC1R	(-)	(-)	(-)	(-)	(-)	X	(-)
MEN1	(-)	(-)	(-)	X	X	X	(-)
MET	(-)	(-)	(-)	(-)	X	X	(-)
MITF	(-)	(-)	(-)	(-)	(-)	X	(-)
MLH1	X	X	X	X	X	X	X
MLH3	(-)	(-)	(-)	(-)	(-)	X	(-)
MRE11	X	(-)	X	(-)	(-)	X	(-)
MSH2	X	X	X	X	X	X	X
MSH3	(-)	(-)	X	(-)	(-)	X	(-)
MSH6	X	X	X	X	X	X	X
MUTYH	X	X	X	X	X	X	X
NBN	X	X	X	X	X	X	X
NF1	(-)	X	(-)	(-)	X	X	(-)
NF2	(-)	(-)	(-)	(-)	X	X	(-)

<i>NSD1</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>NTHL1</i>	(-)	(-)	X	(-)	(-)	X	(-)
<i>PALB2</i>	X	X	X	X	X	X	X
<i>PHOX2B</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>PIK3CA</i>	X	(-)	X	(-)	(-)	(-)	(-)
<i>PMS1</i>	(-)	(-)	(-)	X	X	(-)	(-)
<i>PMS2</i>	X	X	X	X	X	X	X
<i>POLD1</i>	(-)	(-)	X	(-)	(-)	X	(-)
<i>POLE</i>	(-)	(-)	X	(-)	(-)	X	(-)
<i>POLH</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>POT1</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>PRF1</i>	(-)	(-)	(-)	(-)	X	(-)	(-)
<i>PRKAR1A</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>PRSS1</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>PTCH1</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>PTEN</i>	X	X	X	X	X	X	X
<i>RAD50</i>	X	(-)	X	(-)	(-)	X	(-)
<i>RAD51C</i>	X	X	X	X	X	X	X
<i>RAD51D</i>	X	X	X	X	X	X	X
<i>RB1</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>RECQL4</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>RET</i>	(-)	X	(-)	X	X	X	(-)
<i>RHBDF2</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>RUNX1</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>SBDS</i>	(-)	(-)	(-)	(-)	X	(-)	(-)
<i>SCG5</i>	(-)	(-)	X	(-)	(-)	(-)	(-)
<i>SDHA</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>SDHAF2</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>SDHB</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>SDHC</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>SDHD</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>SLX4</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>SMAD4</i>	(-)	(-)	X	X	X	X	(-)
<i>SMARCA4</i>	(-)	X	(-)	(-)	(-)	X	(-)
<i>SMARCB1</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>SMARCE1</i>	(-)	(-)	(-)	(-)		X	(-)
<i>STK11</i>	X	X	X	X	X	X	X
<i>SUFU</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>TERC</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>TERF2IP</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>TERT</i>	(-)	(-)	(-)	(-)	(-)	X	(-)
<i>TMEM127</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>TP53</i>	X	X	X	X	X	X	X
<i>TSC1</i>	(-)	(-)	(-)	(-)	X	X	(-)

<i>TSC2</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>VHL</i>	(-)	(-)	(-)	X	X	X	(-)
<i>WRN</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>WT1</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>XPA</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>XPC</i>	(-)	(-)	(-)	(-)	X	X	(-)
<i>XRCC2</i>	X	(-)	X	(-)	(-)	X	(-)
<i>YAP1</i>	(-)	(-)	(-)	(-)	(-)	X	(-)

X: presente no painel; (-): não presente no painel.

*Painel de captura de 27 genes utilizado para o sequenciamento alvo de RNA.

APÊNDICE B - Tabela com a relação Gene-Doença

Gene	Classificação	Doença relacionada	Tipo de herança
APC	Definitiva	Polipose adenomatosa familiar tipo 1	AD
	Definitiva	Tumor desmoide	AD
	Definitiva	Adenocarcinoma gástrico e Polipose proximal do estômago	AD
	Definitiva	Polipose adenomatosa familiar clássica ou atenuada	AD
	Moderada	Sarcoma	AD
ATM	Definitiva	Ataxia Telangiectasia	AR
	Definitiva	Carcinoma de mama hereditário	AD
	Moderada	Sarcoma	AD
	Moderada	Câncer colorretal não-poliposo hereditário	AD
BARD1	Definitiva	Carcinoma de mama hereditário	AD
BRCA1	Definitiva	Suscetibilidade para Câncer de mama e ovário familiar	AD
	Definitiva	Anemia de Fanconi, grupo de complementação S	AR
	Moderada	Suscetibilidade para câncer de pâncreas	AD
BRCA2	Definitiva	Suscetibilidade para Câncer de mama e ovário familiar	AD
	Definitiva	Complementação de anemia de Fanconi grupo D1	AR
	Definitiva	Carcinoma de próstata familiar	AD
	Forte	Suscetibilidade para câncer de pâncreas	AD
	Moderada	Sarcoma	AD
BRIP1	Definitiva	Anemia de Falconi	AR
	Definitiva	Complementação de anemia de Fanconi grupo J	AR
	Definitiva	Câncer de ovário familiar	AD
	Forte	Carcinoma de mama hereditário	AD
	Moderada	Adenoma colorretal	AD
CDH1	Definitiva	Adenocarcinoma gástrico difuso hereditário	AD
	Definitiva	Carcinoma de mama hereditário	AD

	Definitiva	Síndromes de câncer gástrico difuso relacionado com <i>CDH1</i> e câncer de mama lobular	AD
	Definitiva	Síndrome blefaroqueilodontia 1	AD
	Moderada	Fenda do palato mole	AD
<i>CHEK2</i>	Definitiva	Síndrome Li-Fraumeni 2	AR
	Definitiva	Carcinoma de mama hereditário	AD
	Definitiva	Síndrome Li-Fraumeni	AD
	Moderada	Leucemia mielóide aguda	AD
	Moderada	Câncer colorretal	AD
<i>EPCAM</i>	Definitiva	Diarreia congênita 5 com enteropatia em tufos	AR
	Definitiva	Síndrome de Lynch	AD
<i>MLH1</i>	Definitiva	Síndrome de câncer do reparo de mal pareamento 1	AR
	Definitiva	Síndrome de Lynch 2	AD
	Definitiva	Síndrome de Muir-Torre	AD
	Forte	Câncer de ovário	AD
	Moderada	Rabdomiossarcoma	AR
	Moderada	Neoplasia maligna do pâncreas	AD
<i>MSH2</i>	Definitiva	Síndrome de Lynch	AD
	Definitiva	Síndrome de câncer do reparo de mal pareamento	AR
	Definitiva	Síndrome de Muir-Torre	AD
	Forte	Câncer de ovário	AD
	Moderada	Rabdomiossarcoma	AR
	Moderada	Câncer de próstata	AD
	Moderada	Neoplasia maligna do pâncreas	AD
<i>MSH6</i>	Definitiva	Câncer de endométrio	AD
	Definitiva	Síndrome de Lynch	AD
	Definitiva	Síndrome de câncer do reparo de mal pareamento	AR
	Forte	Câncer de ovário	AD
	Moderada	Rabdomiossarcoma	AR
	Moderada	Neoplasia maligna do pâncreas	AD

	Moderada	Síndrome de Muir-Torre	AD
MUTYH	Definitiva	Polipose adenomatosa familiar 2	AR
NBN	Definitiva	Síndrome de ruptura de Nijmegen	AR
	Moderada	Rabdomiossarcoma	AR
PALB2	Definitiva	Complementação de anemia de Fanconi grupo N	AR
	Definitiva	Carcinoma de mama hereditário	AD
	Forte	Suscetibilidade a câncer de pâncreas	AD
	Moderada	Câncer de ovário familiar	AD
PMS2	Definitiva	Síndrome de Lynch	AD
	Definitiva	Síndrome de câncer do reparo de mal pareamento	AR
	Moderada	Rabdomiossarcoma	AR
	Moderada	Câncer de Ovário	AD
	Moderada	Neoplasia maligna do pâncreas	AD
	Moderada	Síndrome de Muir-Torre	AR
PTEN	Definitiva	Síndrome de Cowden	AD
	Definitiva	Síndrome de Proteus	AD
	Definitiva	Síndrome do tumor hamartoma PTEN	AD
	Definitiva	Síndrome de macrocefalia-autismo	AD
	Moderada	Carcinoma de células renais	AD
	Moderada	Leiomiiossarcoma	AR
RAD51C	Definitiva	Suscetibilidade ao câncer de mama e ovário familiar	AD
	Forte	Complementação de anemia de Fanconi grupo O	AR
RAD51D	Definitiva	Suscetibilidade ao câncer de mama e ovário familiar	AD
STK11	Definitiva	Carcinoma pancreático familiar	AD
	Definitiva	Síndrome de Peutz-Jeghers	AD
TP53	Definitiva	Síndrome de Li-Fraumeni	AD
	Definitiva	Câncer de mama	AD
	Forte	Sarcoma	AD
	Forte	Carcinoma adrenocortical hereditário	AD

Fonte: <https://search.thegencc.org/>

AD: Autossômica dominante; **AR:** Autossômica recessiva.