

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA: CIÊNCIAS
MÉDICAS**

**ANÁLISE DE FERRAMENTAS IN SILICO PARA VARIANTES
SINÔNIMAS EM GENES DE PREDISPOSIÇÃO AO CÂNCER**

MARINA ROBERTA SCHEID

PORTO ALEGRE

2018

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA: CIÊNCIAS
MÉDICAS

**ANÁLISE DE FERRAMENTAS *IN SILICO* PARA VARIANTES
SINÔNIMAS EM GENES DE PREDISPOSIÇÃO AO CÂNCER**

MARINA ROBERTA SCHEID

Orientadora: Prof. Dra. Patricia Ashton Prolla.

Dissertação de Mestrado apresentada como requisito parcial para a obtenção do grau de Mestre em Medicina: Ciências Médicas, da Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação Medicina: Ciências Médicas.

PORTO ALEGRE

2018

DEDICATÓRIA

*Dedico esse trabalho para o Francisco, que veio me
acompanhar no finalzinho dessa jornada.*

AGRADECIMENTOS

À Professora Patricia Ashton-Prolla, agradeço pela confiança em me aceitar no seu grupo de pesquisa e incentivo para desenvolver esse projeto. Ao meu co-orientador, Gabriel Macedo, pelo suporte e auxílio em todas as etapas do projeto.

À bioinformata Martiela Freitas, que sempre se mostrou muito disponível para solucionar todas as minhas dúvidas e auxiliar na execução dos programas.

A todos os meus colegas do Laboratório de Medicina Genômica, e em especial, Tiago, Larissa, Isabel e Igor, que me apoiaram e ajudaram em diversos momentos, ao Guilherme, pela ajuda nas coletas de dado, e a Thayne, pela troca de conhecimento ao longo do projeto.

A todos os professores da UFRGS com os quais tive as disciplinas ao longo do mestrado, agradeço pela troca de conhecimento.

Aos funcionários administrativos, Everaldo, Elmo e Vera, pela ajuda com questões burocráticas.

Por fim, agradeço a minha base, minha família, meu pai e minha mãe, que sempre me apoiaram e estiveram presentes ao longo de toda a jornada. A minha irmã, Melissa, por todo o suporte e carinho. Ao meu namorado, Leonardo, que sempre acreditou em mim, me apoiou a seguir esse caminho e esteve do meu lado sempre que precisei.

RESUMO

Com a introdução do sequenciamento em massa nos testes de diagnóstico através de painéis para genes de predisposição ao câncer (GPC) houve um aumento significativo no número de variantes detectadas. Uma grande parcela destas variantes são variantes sinônimas (VS), e a maioria inferem-se serem benignas, sem passar por uma avaliação de patogenicidade. Contudo, estima-se que um menor percentual de VS é capaz de alterar a expressão gênica se estiverem inseridas em regiões conservadas, como sítios de *splicing*, ilhas CpG, ou ainda interferir na conformação secundária do RNA mensageiro (mRNA). Nesse trabalho, foi avaliada a acurácia de ferramentas *in silico* gratuitas que analisam VS. Ademais, foram avaliadas VS obtidas do banco de dados público (ExAC) e de indivíduos avaliados para genes associados a câncer hereditário. As ferramentas *in silico* selecionadas para as análises foram divididas em três grupos: ferramentas para análise de conservação (PhasCons e PhiloP), ferramentas de múltiplas anotações (PredictSNP, CADD, DANN, FATHAMM, FUNSEQ2, GWAVA e o software Silent Variant Analyzer (SilVA)), e ferramentas para análise de *splicing* (SSF, NNSplice, MaxEntScan e GeneSplicer). Para os testes de sensibilidade e especificidade foi realizada uma curva-ROC (*Receiver Operating Characteristic Curve*) para cada ferramenta utilizando o *software* SPSS. Variantes depositadas no ClinVar e sabidamente patogênicas foram utilizadas como padrão-ouro. Foram selecionadas VS de dois grupos: o primeiro composto por variantes depositadas no banco de dados ExAC, enquanto que o segundo, era composto por variantes identificadas em indivíduos com critérios clínicos para câncer hereditário e que realizaram sequenciamento completo de *TP53*, *BRCA1* e *BRCA2* no Hospital de Clínicas de Porto Alegre. Todas as VS foram avaliadas pelos preditores que obtiveram melhores valores de sensibilidade-especificidade. A análise de estrutura secundária (ES) e diferença da energia de dobramento (Δ MEF) do mRNA foi realizada através do RNAfold e a presença de ilhas CpG pelo programa MethPrimer. Como resultados, o software SilVA apresentou a maior acurácia entre todas as ferramentas de múltiplas anotações avaliadas, e apenas o preditor GWAVA não apresentou valores significativos de área abaixo da curva (AAC) para as VS avaliadas. Todas as ferramentas de conservação e *splicing* apresentaram valores significativos AAC. Do total de 4.324 VS selecionadas a partir do ExAC, aproximadamente 1% destas foram preditas como provavelmente ou possivelmente patogênicas pelo software SilVA. Análises adicionais demonstraram que alteração no *splicing* deve ser o principal

mecanismo de patogenicidade presente nesse grupo. Em relação as VS identificadas no segundo grupo de estudo, foram selecionadas VS cuja frequência populacional era inferior a 0,05. Quatorze variantes sinônimas em *TP53* (n=3), *BRCA1* (n=2) e *BRCA2* (n=9) foram selecionadas para a análise por ferramentas *in silico*. As análises estruturais de dobramento do mRNA demonstraram que duas VS identificadas em *TP53* (c.66A>G e c.108G>A) apresentaram alterações na energia termodinâmica de dobramento ou na estrutura do mRNA. Também identificamos que c.1395A>C e c.3756A>T, ambas em *BRCA2*, podem resultar em alterações no *splicing*. Este foi o primeiro estudo avaliando a acurácia de ferramentas *in silico* para VS identificadas em um grupo abrangente de genes de predisposição ao câncer. Estudos funcionais complementares são necessários para a confirmação do potencial patogênico destas alterações moleculares.

Palavras-chave: Variantes sinônimas, mutações sinônimas, mutações silenciosas, câncer hereditário, genes de predisposição ao câncer hereditário, Síndrome de Li-Fraumeni, Síndrome de câncer de mama e ovário hereditários, *TP53*, *BRCA1* e *BRCA2*.

ABSTRACT

After the introduction of the massive parallel sequencing for genetic diagnostic tests through gene panels for hereditary cancer predisposing genes there was a significant increasing in the number of detected variants. A great proportion of these variants are synonymous variants (SVs), and most of them are inferred to be benign, without undergoing pathogenicity assessment. However, it is estimated that a small but significant percentage of SVs are capable of altering gene expression whether they are inserted in conserved regions, such as splicing sites, CpG islands, or altering the secondary mRNA conformation. In this work, we evaluated the accuracy of free-access *in silico* tools that analyze SVs. We also evaluated SVs obtained from a public database (ExAC) and from genetic testing results from a cancer risk assessment clinic. *In silico* tools selected for the analysis belong to three groups: tools to evaluate nucleotide conservation (PhasCons and PhyloP), multiple annotation (PredictSNP, CADD, DANN, FATHAMM, FUNSEQ2, GWAVA and SilVA), and splicing (SSF, NNSplice, MaxEntScan and GeneSplicer). The Receiver Operating Characteristic (ROC) curve was carried out for sensibility and specificity analysis of each tool using SPSS software. Analysis of secondary structure and difference of mRNA energy folding (Δ MEF) were performed by RNAfold and the presence of CpG island by MethPrimer. The software SilVA presented the highest accuracy between all multiple annotation *in silico* selected tools, and only GWAVA did not show significant values of area under the curve (AUC). All conservation and splicing tools selected presented significant values of AUC. Of the 4,324 SVs selected from ExAC, SilVA classified approximately 1% as likely or potentially pathogenic. Analysis by other *in silico* predictors demonstrated that splicing alterations could be the main mechanism involved in the pathogenicity mechanism of this group of variants. Regarding variants identified in the second group, SVs were filtered by global minor allelic frequency lower than 0.05 and fourteen SVs were selected in *TP53* (n=3), *BRCA1* (n=2) and *BRCA2* (n=9). Structural analyses demonstrated that two SVs in the *TP53* gene (c.66A>G and c.108G>A) were predicted to change the folding energy or structure of the mRNA. We also found that c.1395A>C and c.3756A>T, both in the *BRCA2* gene, could result in splicing disruption according to *in silico* splicing tools. This was the first study to evaluate the accuracy of *in silico* tools for SVs in a comprehensive set of cancer predisposing genes. However, further analysis are still required to confirm the pathogenic potential of SVs.

Key words: Synonymous variants, synonymous mutation, silent mutation, hereditary cancer, hereditary cancer predisposing syndromes, Li-Fraumeni Syndrome, Hereditary breast and ovarian cancer, *TP53*, *BRCA1* and *BRCA2*.

LISTA DE FIGURAS

Figura 1. Fluxograma da revisão sistemática da literatura. Modelo obtido do PRISMA Statement	17
Figura 2. Tipos de mutações que ocorrem nas regiões gênicas ou intergênicas.	19

LISTA DE TABELAS

Tabela 1. Principais genes de predisposição ao câncer	21
Tabela 2. Critérios Clínicos da Síndrome de Li-Fraumeni e da variante Li-Fraumeni-Like	22
Tabela 3. Critérios clínicos para a Síndrome de predisposição hereditária ao câncer de mama e ovário, de acordo com a <i>National Comprehensive Cancer Network</i> (NCCN)..	24
Tabela 4. Alguns estudos da literatura que descreveram variantes sinônimas germinativas associadas a câncer hereditário e/ou esporádico.....	32

LISTA DE ABREVIACÕES

AAC - Área abaixo da curva

ABraOM - Arquivo Brasileiro Online de Mutações

APC – Adenomatous Polyposis coli gene

ATM - Serine-protein kinase ATM gene

AUC- *Area under the curve*

BCL2L12 – BCL2-like 12

BRCA1 - Breast cancer susceptibility gene 1

BRCA2 - Breast cancer susceptibility gene 2

CDH1 – Cadherin 1 gene

CDKN2A - Cyclin-dependent kinase Inhibitor 2A

CHEK2 - Checkpoint kinase 2

CRTF - CAR1 transcription factor

CADD – Combined annotation dependent depletion tool

DNA – Ácido desoxirribonucleico

DDIG-SN - Discriminating disease-causing synonymous variants

ExAC - Exome agregation consortium

FATHMM – Functional analysis through hidden Markov models tool

GPC - Genes de predisposição ao câncer

GWAVA – Genome wide annotation of variants tool

HBOC - Hereditary breast and ovarian cancer

MDR1 – Multi-Drug resistance gene

MES - MaxEntScan

MEN1 – Menin 1 gene

miRNA – microRNA

MLH1 – Human mutL homolog 1

mRNA - RNA mensageiro

MSH2 – MutS homolog 2 gene

MSH6 - MutS homolog 6 gene

MUTYH – MutY homolog

NCCN - National comprehensive cancer network

NGS – *Next generation sequencing*
NF1 – *Neurofibromatosis type I gene*
NF2 - *Neurofibromatosis type II gene*
NNS – *Neutral Network Splice tool*
PALB2 - *Partner and localizer of BRCA2 gene*
PMS2 – *PMS1 homolog 2 gene*
PTEN – *Phosphatase and tensin homolog gene*
RB1 – *RB transcriptional corepressor 1*
RET – *Ret proto-oncogene*
RNA – *Ácido ribonucleico*
ROC - *Receiver operating characteristic*
SiVA - *Software Silent Variant Analyzer*
SLF – *Síndrome de Li-Fraumeni*
SLL – *Síndrome de Li-Fraumeni Like*
SNP - *Single nucleotide polimorphism*
SSF - *Site Splice Finder*
SVs – *Synonymous variants*
TP53 – *Tumor protein p53 gene*
tRNA - *RNA transportador*
TSC1 – *TSC complex subunit 1*
TSC2 - *TSC complex subunit 2*
UTR – *Untranslated region*
VCF - *Variant call format*
VS - *Variantes sinônimas*

SUMÁRIO

1. INTRODUÇÃO	14
2. REVISÃO SISTEMÁTICA DA LITERATURA	16
2.1 Mutações no genoma e câncer	18
2.2 Síndromes de predisposição hereditária ao câncer	20
2.2.1 Síndrome de Li-Fraumeni	222
2.2.2 Síndrome de predisposição hereditária ao câncer de mama e ovário	24
2.3 Mutações sinônimas e mecanismos de patogenicidade	26
3. MARCO CONCEITUAL TEÓRICO	31
4. STROBE STATEMENT	32
5. JUSTIFICATIVA	35
6. OBJETIVOS	36
6.1 Objetivo geral:	36
6.2 Objetivos específicos:	36
7. REFERÊNCIAS	37
8. ARTIGO ORIGINAL	41
Abstract	41
Introduction	42
Methodology	43
Results	45
Discussion	47
Conclusion	53
10. CONSIDERAÇÕES FINAIS	60
11. PERSPECTIVAS FUTURAS	61

1. INTRODUÇÃO

O câncer é o resultado de uma interação complexa entre fatores ambientais e genéticos. Contudo, estima-se que de 5–10% das neoplasias diagnosticadas tenham um forte componente hereditário envolvido. Nestes casos, o desenvolvimento de câncer caracteriza-se pela presença de mutações germinativas em genes supressores tumorais, os quais são genes chave no controle do ciclo celular, apoptose, proliferação celular, dentre outros. Uma vez inativados, as células perdem o controle da proliferação celular e resultam no acúmulo progressivo de mutações no DNA. A inativação desses genes deve-se à presença de mutações de perda de função, as quais podem ser mutações do tipo *missense* (de sentido trocado), *nonsense* (formam um códon de parada precoce), sinônimas (ou silenciosas, que não alteram o aminoácido), *frameshift* (que alteram a fase de leitura), mutações de sítio de *splicing* e mutações em promotores ou regiões regulatórias (McLaren et al. 2010).

Com o surgimento do sequenciamento em massa, ou sequenciamento de nova geração (NGS, do inglês *Next Generation Sequencing*) nos testes de diagnóstico através de painéis multigênicos, houve um aumento significativo no número de variantes detectadas. Uma grande parcela destas variantes são variantes sinônimas (VS), em que a maioria inferem-se serem benignas sem passar por uma avaliação de patogenicidade, uma vez que não alteram o aminoácido da proteína. Contudo, Supek e colaboradores avaliaram VS somáticas de tumores, e perceberam que estas variantes, especialmente no gene *TP53*, apresentam alterações no mecanismo de *splicing* e participam na progressão tumoral (Supek et al. 2014).

Mutações germinativas no gene *TP53* resultam na Síndrome de Li-Fraumeni (SLF). Os tumores mais recorrentes em indivíduos portadores da síndrome são sarcomas de partes moles, tumores de sistema nervoso central, sarcomas ósseos, carcinoma adrenocortical e câncer de mama. Nos últimos anos, critérios clínicos menos restritivos que incluíam indivíduos com câncer de próstata, câncer de pulmão, melanoma e pâncreas foram postulados juntamente com os descritos anteriormente, configurando a Síndrome de Li-Fraumeni Like (LFL), também caracterizada por mutações germinativas em *TP53* (McBride et al, 2014). Já mutações germinativas em *BRCA1* e *BRCA2* levam a síndrome de predisposição hereditária ao câncer de mama e ovário, caracterizada pelo risco de múltiplos tumores ao longo da vida, em particular câncer de mama e ovário em idades precoces (Walsh et al. 2011).

Diversos relatos na literatura mostraram que VS podem apresentar um possível significado patogênico em condições relacionadas ao câncer hereditário, sendo que, na maioria dos casos revisados, as VS resultavam na excisão de éxons ou não-remoção de íntrons, afetando diretamente *splicing* do mRNA (Varley et al, 2001, Auclair et al, 2006, e Aretz et al, 2004). Contudo, de forma mais abrangente, outros estudos relataram que VS podem apresentar mecanismos alternativos para a inativação de genes, sendo que, entre os principais, podemos descrever: conservação da posição da mutação, alteração do códon de uso preferencial (*codon usage bias*), alterações na sequência do gene (a mutação pode afetar um sítio CpG, por exemplo), alterações que podem influenciar no mecanismo de *splicing* e alteração da conformação secundária do mRNA (Buske et al. 2013).

Dessa forma, grande parte dos estudos já publicados e que avaliaram VS em genes de predisposição ao câncer demonstraram que algumas destas variantes podem estar diretamente envolvidas na predisposição ao câncer, fato este principalmente ligado a alterações no mecanismo de *splicing* do mRNA. Contudo, há inúmeros relatos na literatura demonstrando que VS podem apresentar mecanismos alternativos para a inativação gênica.

2. REVISÃO SISTEMÁTICA DA LITERATURA

Esta revisão da literatura está focada nos aspectos relacionados aos potenciais mecanismos de patogenicidade de VS em genes associados ao câncer hereditário. Foram revisados trabalhos que descreviam aspectos gerais da carcinogênese e câncer hereditário, ferramentas para análise de VS e trabalhos que descreveram estas variantes no contexto das síndromes de predisposição a câncer hereditário, especialmente a síndrome de Li-Fraumeni ligada ao gene *TP53* e a síndrome de predisposição hereditária ao câncer de mama e ovário ligada aos genes *BRCA1* e *BRCA2*. A estratégia de busca envolveu as seguintes bases de dados: PubMed e *Web Of Science*, no período de 1960 a 2018. As palavras-chave utilizadas para a busca nas bases de dados foram feitas utilizando os termos: “Silent Mutation” [MeSH] e variações, e “Neoplastic Syndromes, Hereditary” [MeSH] e variações. O fluxograma abaixo representa de forma mais detalhada a maneira como foi conduzida a revisão sistemática para esta dissertação.

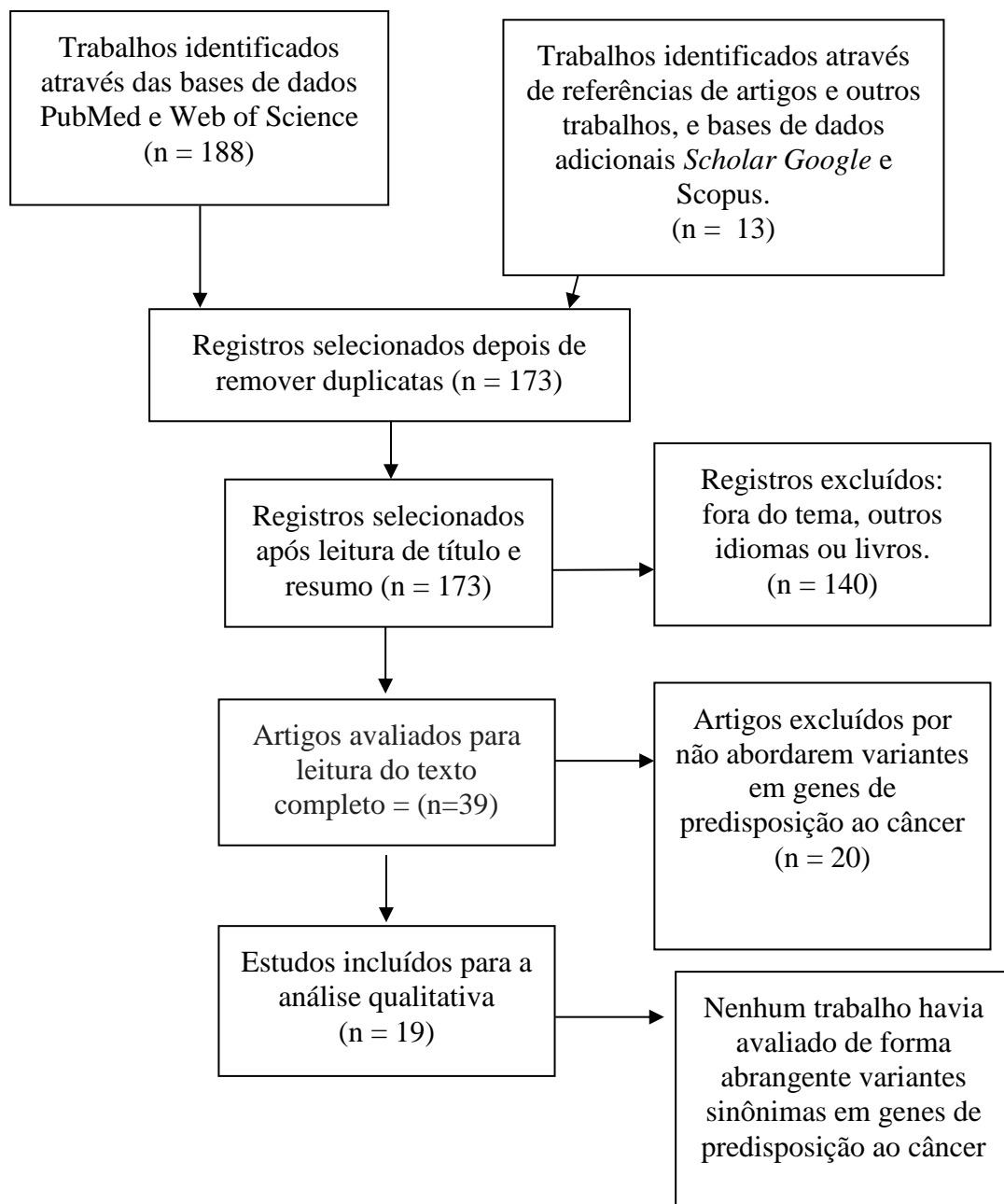


Figura 1. Fluxograma da revisão sistemática da literatura. Modelo obtido do PRISMA Statement (Moher et al, 2009).

2.1 Mutações no genoma e câncer

A carcinogênese é um processo resultante do acúmulo de mutações em uma célula, que ocorre devido a um desequilíbrio nos processos regulatórios de morte e proliferação celular. Essas mutações permitem que a célula acumule funções que não estão presentes em células normais, entre elas: sustentar sinais de proliferação, bloquear agentes supressores de crescimento, alterar a capacidade de indução de angiogênese, invasão e formação de metástase e evadir a apoptose (Hanahan and Weinberg 2011). O desenvolvimento da maioria dos tumores tem origem multifatorial, o que significa que tanto fatores genéticos quanto ambientais estão envolvidos na etiologia da doença, ocasionando um acúmulo de mutações em genes cruciais que participam dos processos de replicação, reparo do DNA, divisão e proliferação celular.

As variantes que resultam no processo carcinogênico podem ser classificadas de diferentes maneiras. Uma forma de classificá-las é quanto à sua associação com a doença. Neste contexto, as variantes podem ser descritas como patogênicas ou provavelmente patogênicas, benignas ou provavelmente benignas ou ainda de significado incerto (Richards et al. 2015). Além disso, as alterações podem ser classificadas quanto à sua relevância para a progressão do tumor. As variantes com impacto direto no processo carcinogênico são conhecidas como mutações *drivers*. Já mutações que não apresentam efeitos nesse processo são chamadas de *passengers*.

As mutações que ocorrem no nosso genoma podem ser tanto de uma troca de base única – *Single Nucleotide Polymorphism* (SNPs), ou mutações que afetam múltiplos pares de base, como as inserções, deleções e duplicações. Outra forma de classificar as mutações é de acordo com a região em que elas ocorrem no genoma. Elas podem ser classificadas como mutações em regiões codificantes ou não-codificantes. As mutações de regiões codificantes, por sua vez, podem ser subdivididas de acordo com seu o efeito na tradução. Nessa subdivisão constam as mutações do tipo *missense* (mutações de sentido trocado), onde ocorre alteração de um nucleotídeo que leva à troca de um aminoácido; as mutações do tipo *nonsense* (mutações truncadas), onde a troca de um nucleotídeo resulta em um códon de parada; mutações *frameshift*, que resultam em mudança na fase de leitura; e mutações sinônimas (ou silenciosa), onde a alteração do nucleotídeo não provoca alteração do aminoácido. Já as mutações presentes em regiões não-codificantes são subdivididas de acordo com o local onde a mutação se encontra. Essas mutações podem ser do tipo intrônicas ou mutações de regiões regulatórias

5'UTR ou 3'UTR. Mutações que ocorrem na região codificante do gene podem acarretar em defeitos de função na proteína. Contudo, embora mutações nas regiões não-codificantes não resultem em alterações diretas para a proteína, elas também devem ser consideradas, uma vez que podem resultar em alterações na expressão gênica (McLaren et al. 2010).

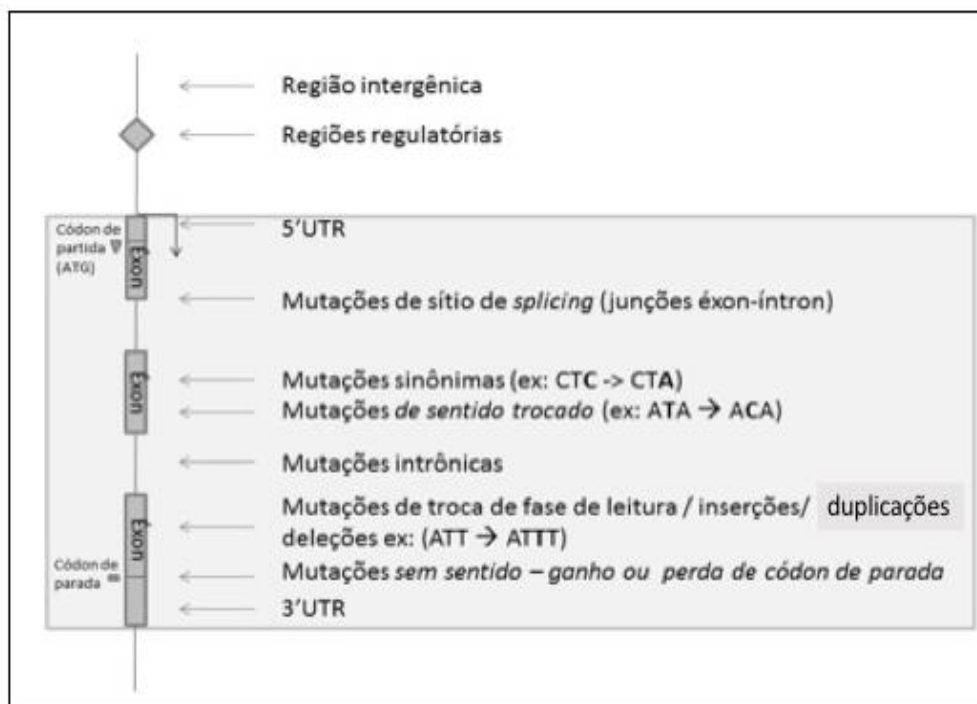


Figura 2. Tipos de mutações que ocorrem nas regiões gênicas ou intergênicas (figura adaptada de McLaren et al. 2010).

Na maior parte dos casos, as mutações que levam para a progressão tumoral são de origem somática, ou seja, estão presentes apenas nas células tumorais. Em uma menor parcela, essas mutações podem ter uma origem germinativa, estando presentes em todas as células do indivíduo. Mutações que predispõe ao mecanismo carcinogênico estão presentes em genes ligados ao controle do ciclo celular, como genes supressores de tumor, genes de reparo e oncogenes. O primeiro grupo, genes supressores de tumor, contribuem para a carcinogênese uma vez em que ocorre a perda da função destes genes, e na maioria destes, é necessário que ambas as cópias apresentem uma mutação. Já mutações em genes de reparo afetam genes que participam do reparo do DNA, e como consequência, estas resultam no acúmulo de mutações no DNA. Oncogenes resultam na carcinogênese quando superexpressos, e agem preferencialmente, induzindo a proliferação celular ou através de mecanismos anti-apoptóticos.

2.2 Síndromes de predisposição hereditária ao câncer

Entre todos os casos de tumores malignos diagnosticados, é estimado que aproximadamente 5 a 10% desses tumores tenham uma origem hereditária, onde se encontram mutações em genes de predisposição ao câncer de alta ou média penetrância, que confere uma incidência muito aumentada de câncer na família. Nesses casos, a grande maioria dos genes envolvidos tem um mecanismo de herança autossômica dominante. Além disso, de 20 a 30% dos pacientes com tumores malignos apresentam uma maior incidência na família, porém menos acentuada, e nesse caso, essa incidência aumentada pode ser resultante de uma interação complexa de genes de baixa penetrância e o meio ambiente, caracterizando assim, casos familiares (Tung et al. 2016).

A maioria das síndromes de predisposição ao câncer caracterizam-se por mutações em genes supressores tumorais. Um dos principais genes responsáveis por controlar a maior parte das vias ligadas a carcinogênese é o gene *TP53*, que é capaz de atuar tanto como gene supressor de tumor quanto como oncogene, e mutações que resultam na inativação ou mutações dominante-negativas estão frequentemente associadas a maior parte de tumores esporádicos. Mutações germinativas em *TP53* resultam na síndrome de Li-Fraumeni. Além de *TP53*, outros genes frequentemente associados a casos esporádicos e hereditários de câncer são os genes *BRCA1* e *BRCA2*, que resultam na síndrome de predisposição ao câncer de mama e ovário (Zhang et al. 2015). Atualmente, diversos genes associados a diferentes síndromes de predisposição hereditária ao câncer já foram descritos, e alguns desses genes estão listados na tabela abaixo (Tabela 1).

Tabela 1. Principais genes de predisposição ao câncer (Garber and Offit 2005)

Genes	Síndromes hereditárias	Neoplasia associadas	Penetrância
<i>APC</i>	Polipose Adenomatosa Familiar (PAF)	Adenoma periampular, Câncer colorretal, câncer gástrico, hepatoblastomas, tumores cerebrais e tumores desmóides.	Moderada
<i>ATM</i>	Risco aumentado para câncer de mama associado a ATM	Câncer de mama, linfoma	Moderada
<i>BRCA1</i>	Síndrome de Predisposição ao Câncer de Mama e Ovário	Câncer de mama, Câncer de ovário, câncer de próstata, câncer de pâncreas*	Alta

	Hereditários (HBOC)		
<i>BRCA2</i>		Câncer de mama, Câncer de ovário, câncer de próstata, câncer de pâncreas, melanoma *	Alta
<i>CDH1</i>	Câncer gástrico difuso hereditário	Câncer gástrico	Moderada
<i>CDKN2A</i>	Melanoma Familiar	Melanoma	-
<i>CHEK2</i>	Câncer de mama e colon-retal hereditário (HBOC)	Câncer de mama e colorretal *	Moderada
<i>MEN1</i>	Neoplasia endócrina múltipla	Carcinoma medular de tireóide, feocromocitoma	Moderada
<i>MLH1</i>	Síndrome de Câncer Colorretal Hereditário não-Polipomatoso (HNPCC ou Síndrome de Lynch)		Alta
<i>MSH2</i>		Câncer colorretal, endométrio, ovário, pelve renal e ureter, gástrico e de intestino delgado.	Alta
<i>MSH6</i>			Alta
<i>MUTYH</i>	Polipose Adenomatosa Familiar	Múltiplos pólipos, câncer de cólon retal, adenocarcinoma endometrial	-
<i>NF1</i>	Neurofibromatose-1	Neurofibroma, neurofibrossarcoma, glioma óptico, feocromocitoma e meningioma.	Alta
<i>NF2</i>	Neurofibromatose 2	Schwannoma vestibular bilateral e de outros nervos cranianos e periféricos, meningiomas e ependimomas.	Alta
<i>PALB2</i>	PALB2- Associated Cancer Risk	Câncer de mama	Moderada
<i>PMS2</i>	Lynch Syndrome	Câncer colorretal, endométrio, ovário, pelve renal e ureter, gástrico e de intestino delgado.	Alta
<i>PTEN</i>	Síndrome de Cowden e Bannayan-Riley-Ruvalcaba	Câncer de mama, tireoide, endométrio, renal e meningioma.	Alta
<i>RB1</i>	Retinoblastoma Hereditário	Retinoblastoma	Alta
<i>RET</i>	Neoplasia endócrina múltipla	Carcinoma medular de tireóide, feocromocitoma	Moderada
<i>TP53</i>	Síndrome de Li-Fraumeni (SLF) / Síndrome de Li-Fraumeni Like (LFL)	Sarcoma de tecidos moles, osteossarcomas, câncer de mama, tumores do sistema nervoso central, leucemia e tumores adrenocorticais.	Alta
<i>TSC1</i>	Esclerose Tuberosa		Alta
<i>TSC2</i>	Esclerose Tuberosa	Angiofibromas faciais, fibroma subungual, rabdomiomas cardíacos, angiomiolipomas e cistos renais.	Alta

Tumores hereditários podem ser diferenciados de tumores esporádicos por algumas características específicas, como: (1) diagnóstico de câncer em idade precoce; (2) histórico familiar que inclui vários casos de um mesmo tipo de tumor ou diversos tipos de tumores de um mesmo espectro; (3) ocorrência de tumores primários múltiplos em um mesmo indivíduo; (4) multifocalidade ou bilateralidade da neoplasia (Narod e Offit, 2005). Na maioria dos casos, o diagnóstico da síndrome é feito analisando o

histórico individual e familiar do paciente, e posteriormente confirmado através da identificação da variante patogênica no gene relacionado à síndrome.

2.2.1 Síndrome de Li-Fraumeni

A Síndrome de Li-Fraumeni (SLF) foi inicialmente descrita por Li e Fraumeni (Li and Fraumeni 1969) ao observar grupos de famílias com casos de sarcomas na infância e câncer de mama em idade jovem. Foram identificados tumores específicos e recorrentes em indivíduos portadores da síndrome, sendo mais comum a presença de sarcomas de partes moles, sarcomas ósseos, tumores de sistema nervoso central, carcinoma adrenocortical e câncer de mama em idade precoce (antes dos 45 anos) em indivíduos de uma mesma família. Este tumores são considerados “tumores centrais” (*core tumors*), cuja presença é critério para diagnóstico da Síndrome de Li-Fraumeni Clássica (Li and Fraumeni 1982, Malkin et al. 1990).

Com a descoberta do gene causador das SLF e com o intuito de identificar famílias com critérios para a síndrome, critérios clínicos menos restritivos foram postulados, configurando a Síndrome de Li-Fraumeni Like (LFL) que, entre outras diferenças em relação à SLF, considera também outros tumores além dos tumores centrais, tais como câncer de próstata, pulmão, melanoma e pâncreas. Os primeiros critérios de LFL foram descritos por Birch em 1994 (Birch et al. 1994). Posteriormente, foram postulados outros critérios mais abrangentes: os critérios de Elles, Chompret e Chompret Modificado. Esses critérios estão detalhados na Tabela 2 e compõem o diagnóstico inicial para a Síndrome.

Tabela 2. Critérios Clínicos da Síndrome de Li-Fraumeni e da variante Li-Fraumeni-Like (McBride et al, 2014).

Critério Clínico	Descrição
SLF – Clássico (Li, Fraumeni <i>et al.</i> , 1988)	I- Sarcoma na infância ou em idade jovem (≤ 45 anos) g II- Familiar de 1º grau com qualquer câncer em idade jovem (≤ 45 anos) g III- Familiar de 1º ou 2º graus com diagnóstico de câncer em idade jovem (≤ 45 anos) ou sarcoma em qualquer idade;
LFL- Birch (Birch, Hartley <i>et al.</i> , 1994)	I- Câncer na infância (qualquer idade) ou sarcoma, tumor SNC ou CAC (≤ 45 anos) g II- Familiar de 1º ou 2º graus com câncer típico da SLF (sarcoma, CM, tumor SNC, CAC ou leucemia) em qualquer idade g III- Familiar de 1º e 2º graus com qualquer câncer < 60 anos;
LFL – Elles (Eeles 1995)	I- Presença de 2 familiares de 1º e 2º graus com tumor típico da SLF em qualquer idade (sarcoma, CM, tumor SNC, CAC, leucemia, melanoma, câncer de próstata, câncer de pâncreas); e II- Diagnóstico de sarcoma em qualquer idade g

	<p>III- Diagnóstico de pelo menos outros 2 tumores em familiares de 1° e 2° graus (podendo estar presentes no mesmo indivíduo): CM < 50 anos, tumor de SNC, leucemia, CAC, melanoma, câncer de próstata, câncer de pâncreas < 60 anos <u>ou</u> sarcoma em qualquer idade;</p>
<p>LFL Chompret(Frebourg et al. 2001)</p>	<p>I- Diagnóstico de sarcoma, tumor SNC, CM ou CAC < 36 anos <u>e</u> II- Familiar de 1° e 2° graus com câncer (qualquer um dos descritos acima, exceto CM se o probando teve CM) <u>ou</u> familiar com múltiplos tumores primários em qualquer idade <u>ou</u> III- Múltiplos tumores primários, incluindo dois tumores que sejam do tipo: sarcoma, tumor SNC, CM ou CAC, com o primeiro tumor diagnosticado < 36 anos independente da história familiar <u>ou</u> IV- CAC em qualquer idade, independente da história familiar;</p>
<p>LFL - Chompret Modificado (Bougeard et al. 2008)</p>	<p>I- Caso índice com câncer típico da SLF (sarcoma, CM, tumor SNC, CAC, leucemia e carcinoma brônquio alveolar de pulmão < 46 anos <u>e</u> II- Familiar de 1° e 2° graus com câncer típico da SLF < 56 anos (exceto CM se o caso índice tiver esse tumor) <u>ou</u> múltiplos tumores <u>ou</u> III- Paciente índice com múltiplos tumores, sendo pelo menos dois do espectro da SLF e o primeiro com < 46 anos <u>ou</u> IV- CAC ou carcinoma de plexo coróide em qualquer idade ou CM < 36 anos sem mutação nos genes <i>BRCA1</i> ou <i>BRCA2</i>;</p>
<p>Legenda: SNC- sistema nervoso central; CM- câncer de mama; CAC- carcinoma adrenocortical; SLF- Síndrome de Li-Fraumeni; LFL- Li-Fraumeni-Like;</p>	

Pacientes que preenchem os critérios descritos acima (Tabela 2) tem indicação para realizar os testes moleculares para diagnóstico da Síndrome, que são as análises do gene *TP53* (Olivier et al. 2003). Considerando que os critérios clínicos para SLF/LFL vão desde os mais restritos, como o clássico, Birch e Chompret, até os menos restritos como os de Eeles 1, a identificação de mutações germinativas é mais frequente em famílias com critérios mais restritos (aproximadamente 70%), e menos frequente em famílias com critérios mais abrangentes e menos rigorosos. A prevalência de mutações patogênicas em *TP53* para os critérios menos abrangentes é de 25% para Chompret (original e modificado), 25% para os critérios de Birch e 14% e 8% para os critérios Eeles 1 e Eeles 2, respectivamente. A SLF é uma síndrome de alta penetrância com fenótipo variável de acordo com o gênero. Estima-se que pacientes portadores de mutações patogênicas em *TP53* apresentem 50% de chance de desenvolver tumores antes dos 40 anos de idade, comparado a ocorrência de 1% em indivíduos da população geral na mesma faixa etária, e que 90% dos portadores desenvolvam câncer até 60 anos de idade (Birch et al. 2001). A incidência de SLF/LFL na Europa e na América do Norte é estimada entre 1 a cada 2.000-5.000 nascidos vivos (Garber and Offit 2005). Contudo, SLF apresenta uma prevalência mais elevada nas regiões Sul e Sudeste do Brasil devido à presença da mutação fundadora p.Arg337His, cuja frequência estimada é de 1 a cada 300 nascidos vivos

(Giacomazzi et al. 2014).

2.2.2 Síndrome de predisposição hereditária ao câncer de mama e ovário

Estima-se que 10-15% de todos os novos casos de câncer de mama tenham algum componente hereditário. A principal síndrome de predisposição hereditária ao câncer de mama é provocada principalmente por mutações nos genes *BRCA1* e *BRCA2*, que quando mutados caracterizam a Síndrome de predisposição hereditária ao Câncer de Mama e Ovário (HBOC, *Hereditary breast and ovarian cancer syndrome*) (Claus et al. 1996). HBOC é uma síndrome de alta penetrância e herança autossômica dominante, e é responsável por 5-7% de todos os casos de câncer mama, e 8-13% dos casos de câncer epitelial de ovário (Roy, Chun and Powell 2011). A identificação de pacientes com predisposição a câncer de mama é feita através da análise do histórico familiar do paciente. A rede NCCN (National Comprehensive Cancer Network) desenvolveu um *guideline* com os critérios clínicos que indicam que o indivíduo pode ter a Síndrome HBOC e, portanto, deve ser testado para os genes *BRCA1* e *BRCA2*. Os critérios detalhados estão apresentados na tabela abaixo (Tabela 3).

Tabela 3. Critérios clínicos para a Síndrome de predisposição hereditária ao câncer de mama e ovário, de acordo com a *National Comprehensive Cancer Network* (NCCN).

NCCN *Guideline* – Critérios clínicos para testagem genética para síndrome de câncer de mama e ovário hereditário

- Indivíduo proveniente de uma família com mutação deletéria conhecida em *BRCA1* ou *BRCA2*.
 - História pessoal de câncer de mama mais um dos seguintes critérios:
 - Diagnóstico ≤ 45 anos
 - Diagnóstico ≤ 50 com:
 - Outro tumor primário
 - ≥ 1 familiar próximo* com câncer de mama em qualquer idade;
 - Histórico familiar limitado ou desconhecido
 - Diagnóstico ≤ 60 anos com:
 - Câncer de mama triplo negativo
 - Diagnóstico de câncer de mama a qualquer idade com:
 - ≥ 1 familiar próximo com câncer de mama diagnosticado < 50 anos
 - ≥ 2 familiares próximos com câncer de mama a qualquer idade
 - ≥ 1 familiar próximo com câncer epitelial de ovário
 - ≥ 2 familiares próximos com câncer de pâncreas ou câncer de próstata (*Gleason score* ≥ 7) a qualquer idade
 - Familiar próximo masculino com câncer de mama
 - Um indivíduo com etnia associada a altas frequências mutacionais (ex. Judeus Ashkenazi) não requer histórico familiar
 - Histórico pessoal de câncer epitelial de ovário
 - Histórico pessoal de câncer de mama masculino
 - Histórico pessoal de câncer de pâncreas ou próstata (*Gleason score* ≥ 7) a qualquer idade com ≥ 2 parentes próximos com câncer de mama, ovário, pâncreas ou próstata (*Gleason score* ≥ 7) a qualquer idade
-

-
- Indivíduos com câncer de pâncreas e ancestralidade Ashkenazi, apenas um familiar afetado e necessário
 - Indivíduos apenas com histórico familiar (limitações significativas interpretando os resultados dos testes devem ser discutidas para um indivíduo não-afetado):
 - Familiar de primeiro ou segundo grau com os critérios listados a cima
 - Familiar de terceiro grau com câncer de mama e/ou ovário com 2 familiares próximos com câncer de mama (um com câncer de mama <50 anos) e/ou câncer de ovário
 - Outros critérios clínicos deve ser utilizados para determinar se o paciente tem possibilidade de carregar mutação, considerando a idade do paciente não-afetado e a idade de demais mulheres não-afetadas na família.
 - Teste de indivíduos não-afetados devem ser considerados somente se não for possível a testagem de indivíduos afetados.
-

*Familiar próximo: considerado até terceiro grau (pai, mãe, filhos, irmãos, avós, tios, primos, sobrinhos, bisavó), proveniente do mesmo lado da família;

Aproximadamente 10% dos pacientes com critérios clínicos dessa síndrome apresentam mutações germinativas em *BRCA1* ou *BRCA2* (Walsh et al. 2011), de forma que o gene responsável pela predisposição hereditária ao câncer de mama permanece desconhecido na maior parte dos pacientes. Dessa forma, passou-se a buscar outros genes de supressão tumoral que pudessem estar associados com risco de câncer de mama aumentado (Angèle and Hall 2000). Com o desenvolvimento de técnicas de sequenciamento em larga escala, foi possível analisar o perfil mutacional de diversos outros genes envolvidos na predisposição ao câncer hereditário, uma vez que esta tecnologia permite o sequenciamento de centenas a milhares de pares de bases simultaneamente, a um custo acessível. Com isso, diversos estudos têm investigado outros genes além de *BRCA1* e *BRCA2* (genes de alta penetrância) em pacientes com câncer de mama hereditário, utilizando painéis. Dessa maneira, ampliou-se o conhecimento sobre a contribuição de mutações em outros genes, especialmente os com penetrância moderada como *CHEK2*, *PALB2* e *ATM*, que conferem um risco de até 4 vezes para o desenvolvimento de câncer de mama (Maxwell et al. 2015).

Atualmente, são considerados diferentes mecanismos de inativação de genes supressor de tumor que, conseqüentemente, levam ao desenvolvimento de Síndrome de predisposição ao câncer. Uma das explicações está na inativação dos produtos proteicos desses genes através de mutações do tipo *missense* ou *nonsense*, que levam a produção de uma proteína anormal devido à alteração de aminoácido ou a inserção de um códon de parada prematuro, respectivamente. Outro tipo, porém, mais raro de mutações são as inserções e deleções dos genes envolvidos ou parte destes. No entanto, muitos dos indivíduos com critérios clínicos para a predisposição de síndromes hereditária não apresentam mutações germinativas nos genes das síndromes envolvidas. Outros tipos de mutações, como mutações em promotores, 5'UTR, 3'UTR e mutações sinônimas, que

usualmente eram descartadas das análises por não apresentarem efeito na proteína, por exemplo, também podem estar envolvidas no mecanismos carcinogênicos (McLaren et al. 2010).

2.3 Mutações sinônimas e mecanismos de patogenicidade

Mutações sinônimas são aquelas que alteram a sequência de nucleotídeos sem alterar o aminoácido da proteína. Estas também podem ser denominadas variantes sinônimas (VS), quando ainda se desconhece o significado patogênico da mutação. O código genético é degenerado, isto é, o mesmo amino ácido pode ser codificado por vários códons, sendo estes denominados códons sinônimos. Contudo, apesar de codificarem o mesmo amino ácido, estes códons não são utilizados na mesma frequência em diferentes genomas, e mesmo em um único genoma o padrão de uso dos códons sinônimos pode variar muito entre os genes, ou ainda ao longo de um único gene. Isso foi primeiramente estudada em 1980 por Grantham e colaboradores, onde foi demonstrado em organismos unicelulares que a taxa de utilização de alguns códons sinônimos para cada aminoácido é variável, resultando assim, no viés de uso de códon (*codon usage bias*). Uma das possíveis explicações para haver um códon preferencial para cada aminoácido seria que a velocidade de tradução do ribossomo estaria diretamente relacionada com a concentração de tRNA. Assim, havendo um códon preferencial para cada aminoácido, resultaria na maior concentração de um aminoacil-tRNA específico, aumentando assim a eficiência da tradução. Esse fenômeno foi correlacionado com os níveis de expressão gênica em organismos como *E. coli*, *S. cerevisiae*, *C. elegans*, entre outros, e foi demonstrado que o uso de alguns códons em particular podem aumentar a expressão gênica em até 1000 vezes. Isso explica, em partes, a capacidade de VS alterarem a taxa de tradução do gene, alterando assim a sua concentração sem alterar a conformação da proteína (Novoa and Ribas de Pouplana 2012).

Com a tecnologia do sequenciamento de nova geração, aumentou-se muito o número de novas variantes identificadas, incluindo VS. E embora isso tenha permitido identificar um maior número de variantes, o grande desafio agora é conseguir associar e classificar como variantes causadoras de doença ou variantes neutras (Ng et al. 2010). Através de estudos de associação, foi possível mapear novos polimorfismos de base única no genoma humano que poderiam estar possivelmente associados com doenças

hereditárias. Em um estudo realizado por Chen e colaboradores (2010), foi avaliado mais de 2.000 publicações de estudos de associações, onde foram avaliadas mais de 20.000 variantes, sendo estas variantes sem sentido, não-sinônimas, sinônimas, variantes de 5'UTR, 3'UTR, intrônicas ou intergências. Neste estudo, foi demonstrado que a média da Odds Ratio de VS não apresentava diferença significativa em relação a variantes não-sinônimas. Assim, o estudo sugere que ambas as classes apresentam efeito similar para risco aumentado de doenças (Chen et al. 2010).

Diversos relatos na literatura mostraram que VS podem apresentar um possível significado patogênico em condições relacionadas ao câncer hereditário, sendo que, na maioria dos casos revisados, as mesmas resultavam na excisão de éxons ou não-remoção de íntrons, afetando diretamente *splicing* do mRNA (Tabela 4). Atualmente, já foram descritos na literatura trabalhos demonstrando significado clínico para VS na Síndrome de Li-Fraumeni (*TP53*) (Varley et al, 2001), Síndrome de Lynch (*MLH1*) (Auclair et al, 2006), e polipose adenomatosa familiar (*APC*) (Aretz et al, 2004), entre outros. Além disso, a variante rs180032 do gene *TP53*, que é uma variante atualmente classificada como benigna na base de dados ClinVar, foi associada com risco de metástase em pacientes com Osteossarcoma esporádico. Recentemente, Karakostis e colegas (2018) estudaram os efeitos *in vitro* da VS em *TP53* c.66A>G. Nesse estudo, foi demonstrado que a presença desta variante impede a fosforilação da serina na posição 15 (S15) de p53. Este processo de fosforilação ocorre durante a tradução do mRNA, e é realizado pela fosfatase ATM. A ausência da fosforilação de S15 previne a estabilização do complexo p53-MDM2, o que impede a ação deste complexo no estresse genotóxico.

Contudo, existem outros mecanismos, além da interferência em sítios de *splicing*, com os quais VS podem estar envolvidas em doenças mendelianas. Sauna e Kimichi-Sarfaty revisaram VS de estudos de associação envolvendo mais de 50 doenças e identificaram o mecanismo pelos quais essas variantes contribuíam para a inativação dos genes, e os mais comuns foram através da alteração do *splicing* e estrutura do mRNA. Assim, ficou mais claro que VS podem estar envolvidas em mecanismos moleculares, e que algumas variantes, de fato, podem influenciar na expressão da proteína de diferentes maneiras (Sauna and Kimichi-Sarfaty 2011). Para isso, é indispensável o desenvolvimento de ferramentas de bioinformática que auxiliam no processo de classificação de VS.

Tabela 4. Alguns estudos da literatura que descreveram variantes sinônimas germinativas associadas a câncer hereditário/ esporádico.

Autor	Condição	Gene	Mutação	Proteína	Achados
(Varley et al. 2001)	Síndrome de Li-Fraumeni	<i>TP53</i>	c.375G>A	p.Thr125Thr	Afeta o sítio doador de <i>splicing</i> do éxon 4 do gene <i>TP53</i> , e resulta na não-remoção do íntron 4.
(Montera et al. 2001)	Polipose adenomatosa familiar	<i>APC</i>	c.1869G>T	p.Arg623Arg	A variante c.1869G>T está associada com a completa remoção do éxon 14, levando a instabilidade e truncamento da proteína <i>APC</i> .
(Colapietro et al. 2003)	Neurofibromatose do tipo 1	<i>NF1</i>	c.20075G>A	p.Glu315Glu	Resulta na excisão do éxon 7, quando presente juntamente com a mutação adjacente c.20076C>A.
(Aretz et al. 2004)	Polipose adenomatosa familiar	<i>APC</i>	c.1956C>T e c.1957A>C;	p.His652His; p.Arg653Arg	Resultam na deleção do éxon 14 do gene <i>APC</i>
(Auclair et al. 2006)	Câncer de colon-retal hereditário não-poliposo	<i>MLH1</i>	C.882C>T; C.1731G>A;	p.Leu294Leu; p.Ser577Ser	Deleção dos éxons 15 e 17, respectivamente;
(Pagenstecher et al. 2006)	Câncer de colon-retal hereditário não-poliposo	<i>MLH1</i> e <i>MSH2</i>	c.1275A>G e MLH1: c.1731G>A	p.Glu425Glu e c.Ser577Ser, respectivamente	Nesse estudo foi avaliado efeito na alteração do <i>splicing</i> , e os autores não chegaram a conclusão quanto ao efeito da variante <i>MSH2</i> :c.1275A>G; e <i>MLH1</i> : c.1731G>A foi considerada possivelmente patogênica, causando a perda do éxon 15.
(Duraturo et al. 2011)	Câncer de colon-retal hereditário não-poliposo	<i>MSH3</i>	c.693G>A	p.Pro231Pro	A variante sinônima, quando presente em conjunto com variantes benignas em <i>MSH2</i> , apresentam possível efeito sinérgico.
(Mirabello et al. 2015)	Osteossarcoma esporádico	<i>TP53</i>	c.639A>G	p.Arg213Arg	Mutação sinônima associada com risco aumentado de Osteossarcoma
(Karkostis et al, 2018)	Síndrome de Li-Fraumeni	<i>TP53</i>	c.66A>G	p.Ser15Ser	Presença da variante sinônima impede a fosforilação da Serina 15 em durante a tradução de p53.

Em 2007, Kimchi-Sarfaty e colaboradores avaliaram um polimorfismo sinônimo do gene *MDRI*, que apresentava um haplótipo ligado a um códon raro, que resultava em alteração da função do produto proteico. Nesse estudo, o grupo verificou que o polimorfismo sinônimo apresentava a capacidade de alterar o dobramento co-translacional da proteína, alterando a estrutura proteica, afetando a capacidade de inserção da glicoproteína na membrana plasmática, a ligação da proteína-substrato e locais de interação do inibidor (Kimchi-Sarfaty et al. 2007). A partir de então, outros estudos também vem demonstrando que VS poderiam interferir não somente no dobramento co-translacional da proteína, mas também poderia afetar a taxa de transcrição do mRNA ao interferir na estrutura secundária do mRNA e alterar a energia de dobramento da proteína. Um exemplo é a mutação sinônima no códon 507 do gene *CFTR*, que precede a mutação patogênica $\Delta F508$ da fibrose cística. Essa mutação foi estudada por Bartoszeski e colaboradores (2010), onde foi demonstrado que a mutação sinônima acentuava o efeito da mutação patogênica, reduzindo ainda mais os níveis da expressão proteica (Bartoszewski et al. 2010).

VS podem estar localizadas no último éxon dos genes, fazendo parte da região 3'UTR, a qual é alvo para ligação de miRNA. Como exemplo, um estudo que avaliou uma amostra de sequenciamento de exoma em melanoma, e neste foi identificada uma variante somática sinônima recorrente em *BCL2L12*, F17F. Após investigação, determinou-se que essa variante era capaz de afetar sítio de ligação para o miRNA hsa-miR-671-5p, impossibilitando a ligação deste, resultando na expressão aumentada da proteína a qual participa da via de sinalização anti-apoptótica. Em um outro exemplo, uma variante sinônima patogênica foi identificada na doença de Crohn no gene *IRGM*, alterando o sítio de ligação para o miRNA-196, causando uma desregulação na função deste gene (Brest et al. 2011).

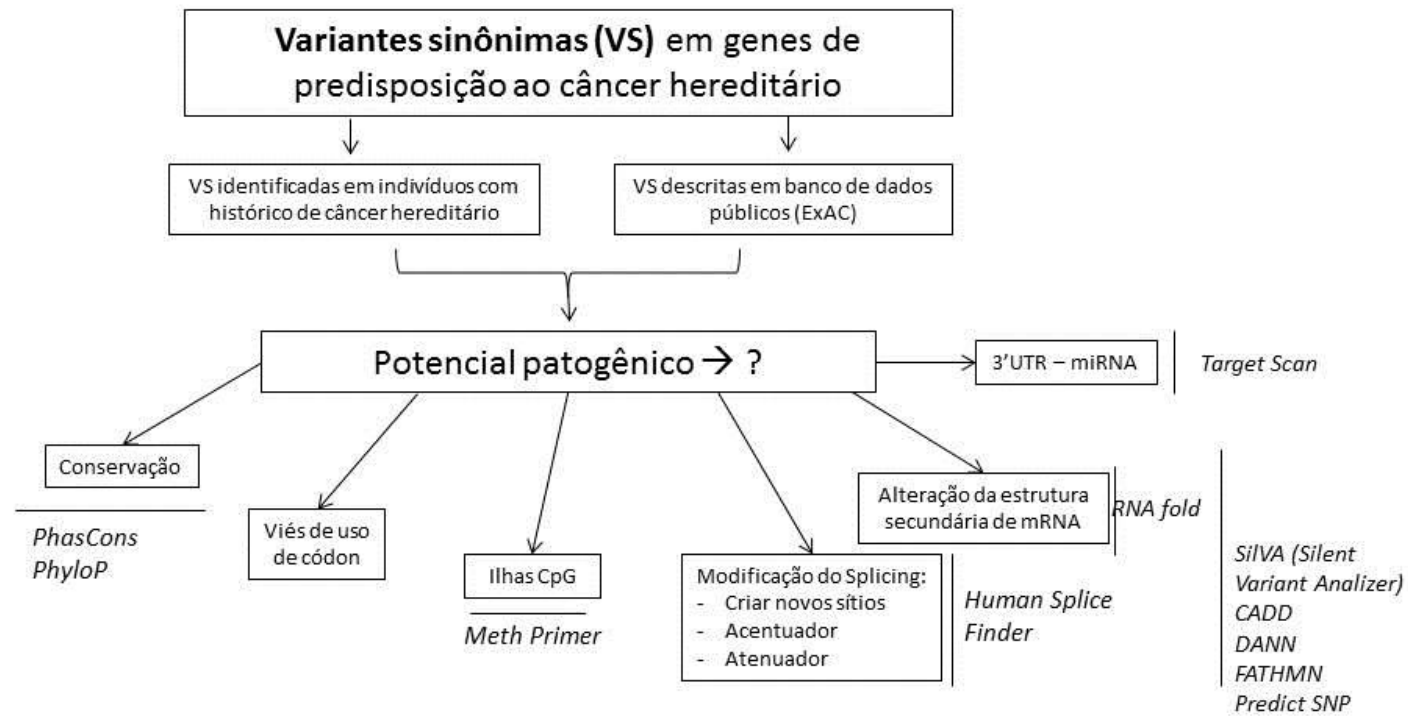
Nesse sentido, Hunt e colaboradores (2014) revisaram de forma abrangente demais ferramentas que auxiliam na identificação de VS com impacto funcional. Nesse mesmo trabalho, foram descritas ferramentas que auxiliam na predição de sítios de splicing e elementos regulatórios de *splicing*, softwares que analisam a formação de sítios de ligação de miRNA e ferramentas que avaliam a conformação secundária do mRNA (Hunt et al. 2014).

Tendo em vista que existem diversos mecanismos pelos quais VS podem alterar a expressão ou conformação da proteína, para se obter uma previsão mais acurada, é importante utilizar essas ferramentas de forma combinada, com o intuito de obter uma

melhor previsão da predição. Recentemente, o software SilVA (Silent Varinat Analyzer) foi desenvolvido e treinado para reclassificar VS em possivelmente patogênicas ou possivelmente benignas, utilizando ferramentas que avaliam diferentes funções possivelmente associada a patogenicidade de VS (Buske et al. 2013). De acordo com Buske e colaboradores (2013), as VS podem interferir na expressão gênica, e estes danos podem ser devido a cinco diferentes fatores, todos avaliados pelo preditor: (1) conservação da posição da mutação/ frequência alélica; (2) *Codon usage bias*, ou frequência com que o códon é utilizado pelo genoma; (3) alterações na sequência do gene, como por exemplo, a mutação pode afetar um sítio CpG; (4) potencial acentuador/supressor de um sítio de splicing, e criar um novo sítio ou destruir um sítio já existente; (5) alteração da conformação do mRNA, podendo afetar a tradução uma vez que é capaz de aumentar ou diminuir a energia de dobramento do mRNA ou alterar a conformação da estrutura secundária deste (Buske et al. 2013).

Em 2015, a *American College of Medical Genetics and Genomics* desenvolveu um *guideline* para a classificação de variantes genéticas (Richards et al, 2015). Nesse guia, estão descritos os principais níveis de evidência que apoiam a patogenicidade de uma variante: (1) dados populacionais, (2) dados preditivos *in silico*, (3) dados funcionais, (4) dados de segregação, (5) dados “*de novo*” e (6) dados alélicos. Nesse trabalho, são abordadas diversas ferramentas computacionais para avaliação de variantes, principalmente ferramentas utilizadas para variantes *missense*. Entre esses aspectos, essas ferramentas *in silico* que foram desenvolvidas para avaliação de variantes avaliam a conservação evolutiva e impacto na estrutura da proteína. Segundo o *guideline*, as ferramentas *in silico* devem ser utilizadas de forma conjunta, e não isoladamente, podendo constituir evidências fraca, moderada ou forte de patogenicidade. Em se tratando de VS, é ressaltado que há fortes evidências para a variante se apresentar como benigna se a variante não apresenta impacto no *splicing*. Contudo, demais abordagens *in silico* para avaliação de VS de uma forma mais abrangente e específica não são mencionadas, e, embora haja evidências que VS podem interferir na expressão gênica ao influenciar na conformação secundária do mRNA, criar sítios alternativos para mRNAs (se inseridas na 3’UTR), esses mecanismos alternativos ainda não foram apontados pelos *guidelines* de classificação de variantes (Richards et al. 2015).

3. MARCO CONCEITUAL TEÓRICO



4. STROBE STATEMENT

Checklist of items that should be included in reports of *cohort studies*

	Item No	Recommendation	Page
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	Pg. 1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	Pg. 6 - 7
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Pg. 14 - 33
Objectives	3	State specific objectives, including any prespecified hypotheses	Pg. 36
Methods			
Study design	4	Present key elements of study design early in the paper	Pg. 45 -48
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Pg. 45 -48
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Pg. 45 -48
		(b) For matched studies, give matching criteria and number of exposed and unexposed	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Pg. 45 -48
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Pg. 45 -48
Bias	9	Describe any efforts to address potential sources of bias	Pg. NA
Study size	10	Explain how the study size was arrived at	Pg. 34
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Pg. NA
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	Pg. 45 -48

		(b) Describe any methods used to examine subgroups and interactions	Pg. 45 -48
		(c) Explain how missing data were addressed	NA
		(d) If applicable, explain how loss to follow-up was addressed	NA
		(e) Describe any sensitivity analyses	NA
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Pg. 48
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	NA
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) Summarise follow-up time (eg, average and total amount)	NA
Outcome data	15*	Report numbers of outcome events or summary measures over time	Pg. 49
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Pg. 49 - 51
		(b) Report category boundaries when continuous variables were categorized	Pg. 49 - 51
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Pg. 49 - 58
Discussion			
Key results	18	Summarise key results with reference to study objectives	Pg. 51 - 52
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Pg. 51 - 52

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pg. 51 - 52
Generalisability	21	Discuss the generalisability (external validity) of the study results	Pg. 51 - 52
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	NA

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at <http://www.strobe-statement.org>.

5. JUSTIFICATIVA

A grande maioria dos estudos que avaliaram variantes sinônimas (VS) em genes de predisposição ao câncer demonstrou que uma parcela destas alterações pode estar associada ao fenótipo por causar, principalmente, modificações no mecanismo de *splicing* do mRNA. Contudo, há relatos na literatura demonstrando que VS podem apresentar mecanismos alternativos de patogenicidade. Este é o primeiro trabalho a avaliar de forma abrangente VS em um conjunto de genes de predisposição ao câncer. Inúmeros estudos em diferentes contextos de doença demonstram que VS podem apresentar um papel importante na carcinogênese, agindo como drivers germinativos ou somáticos para o desenvolvimento de tumores. Dessa forma, o estudo visa avaliar se VS germinativas em genes de predisposição ao câncer poderiam atuar como drivers para a carcinogênese.

6. OBJETIVOS

6.1 Objetivo geral:

Avaliar e caracterizar variantes sinônimas (VS) em um conjunto abrangente de genes de predisposição ao câncer através do uso de diferentes ferramentas *in silico*.

6.2 Objetivos específicos:

1. Testar a sensibilidade e especificidade de ferramentas *in silico* que podem auxiliar no processo de classificação de VS;
2. Caracterizar VS de genes de predisposição ao câncer depositadas no banco de dados público ExAC ou identificadas em indivíduos que realizaram sequenciamento completo dos genes *TP53*, *BRCA1* e *BRCA2*, a partir do uso de diferentes ferramentas de bioinformática;
3. Avaliar a conformação secundária e a diferença na energia de dobramento do RNA mensageiro contendo VS que apresentaram maior probabilidade de patogenicidade pelo uso de ferramentas *in silico*.

7. REFERÊNCIAS

- Angèle, S. & J. Hall (2000) The ATM gene and breast cancer: is it really a risk factor? *Mutat Res*, 462, 167-78.
- Aretz, S., S. Uhlhaas, Y. Sun, C. Pagenstecher, E. Mangold, R. Caspari, G. Möslein, K. Schulmann, P. Propping & W. Friedl (2004) Familial adenomatous polyposis: aberrant splicing due to missense or silent mutations in the APC gene. *Hum Mutat*, 24, 370-80.
- Auclair, J., M. P. Buisine, C. Navarro, E. Ruano, G. Montmain, F. Desseigne, J. C. Saurin, C. Lasset, V. Bonadona, S. Giraud, A. Puisieux & Q. Wang (2006) Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Human Mutation*, 27, 145-154.
- Bartoszewski, R. A., M. Jablonsky, S. Bartoszezka, L. Stevenson, Q. Dai, J. Kappes, J. F. Collawn & Z. Bebok (2010) A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem*, 285, 28741-8.
- Bendl, J., J. Stourac, O. Salanda, A. Pavelka, E. D. Wieben, J. Zendulka, J. Brezovsky & J. Damborsky (2014) PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol*, 10, e1003440.
- Birch, J. M., R. D. Alston, R. J. McNally, D. G. Evans, A. M. Kelsey, M. Harris, O. B. Eden & J. M. Varley (2001) Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene*, 20, 4621-8.
- Birch, J. M., A. L. Hartley, K. J. Tricker, J. Prosser, A. Condie, A. M. Kelsey, M. Harris, P. H. Jones, A. Binchy & D. Crowther (1994) Prevalence and diversity of constitutional mutations in the p53 gene among 21 Li-Fraumeni families. *Cancer Res*, 54, 1298-304.
- Bougeard, G., R. Sesboüé, S. Baert-Desurmont, S. Vasseur, C. Martin, J. Tinat, L. Brugières, A. Chompret, B. B. de Paillerets, D. Stoppa-Lyonnet, C. Bonaïti-Pellié, T. Frébourg & F. L. w. group (2008) Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. *J Med Genet*, 45, 535-8.
- Brest, P., P. Lapaquette, M. Souidi, K. Lebrigand, A. Cesaro, V. Vouret-Craviari, B. Mari, P. Barbry, J. F. Mosnier, X. Hébuterne, A. Harel-Bellan, B. Mograbi, A. Darfeuille-Michaud & P. Hofman (2011) A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet*, 43, 242-5.
- Buske, O. J., A. Manickaraj, S. Mital, P. N. Ray & M. Brudno (2013) Identification of deleterious synonymous variants in human genomes. *Bioinformatics*, 29, 1843-50.
- Chen, R., E. V. Davydov, M. Sirota & A. J. Butte (2010) Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS One*, 5, e13574.
- Claus, E. B., J. M. Schildkraut, W. D. Thompson & N. J. Risch (1996) The genetic attributable risk of breast and ovarian cancer. *Cancer*, 77, 2318-24.

- Colapietro, P., C. Gervasini, F. Natacci, L. Rossi, P. Riva & L. Larizza (2003) NF1 exon 7 skipping and sequence alterations in exonic splice enhancers (ESEs) in a neurofibromatosis 1 patient. *Hum Genet*, 113, 551-4.
- Duraturò, F., R. Liccardo, A. Cavallo, M. De Rosa, M. Grosso & P. Izzo (2011) Association of low-risk MSH3 and MSH2 variant alleles with Lynch syndrome: probability of synergistic effects. *Int J Cancer*, 129, 1643-50.
- Eeles, R. A. (1995) Germline mutations in the TP53 gene. *Cancer Surv*, 25, 101-24.
- Frebourg, T., A. Abel, C. Bonaiti-Pellie, L. Brugières, P. Berthet, B. Bressac-de Paillerets, A. Chevrier, A. Chompret, O. Cohen-Haguénauer, O. Delattre, J. Feingold, J. Feunteun, D. Frappaz, J. P. Fricker, P. Gesta, P. Jonveaux, C. Kalifa, C. Lasset, B. Leheup, J. M. Limacher, M. Longy, C. Nogues, D. Oppenheim, D. Sommelet, F. Soubrier, C. Stoll, D. Stoppa-Lyonnet & H. Tristant (2001) [Li-Fraumeni syndrome: update, new data and guidelines for clinical management]. *Bull Cancer*, 88, 581-7.
- Garber, J. E. & K. Offit (2005) Hereditary cancer predisposition syndromes. *J Clin Oncol*, 23, 276-92.
- Giacomazzi, J., M. S. Graudenz, C. A. Osorio, P. Koehler-Santos, E. I. Palmero, M. Zagonel-Oliveira, R. A. Michelli, C. Scapulatempo Neto, G. C. Fernandes, M. I. Achatz, G. Martel-Planche, F. A. Soares, M. Caleffi, J. R. Goldim, P. Hainaut, S. A. Camey & P. Ashton-Prolla (2014) Prevalence of the TP53 p.R337H mutation in breast cancer patients in Brazil. *PLoS One*, 9, e99893.
- Hanahan, D. & R. A. Weinberg (2011) Hallmarks of cancer: the next generation. *Cell*, 144, 646-74.
- Hunt, R. C., V. L. Simhadri, M. Iandoli, Z. E. Sauna & C. Kimchi-Sarfaty (2014) Exposing synonymous mutations. *Trends Genet*, 30, 308-21.
- Kimchi-Sarfaty, C., J. M. Oh, I. W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar & M. M. Gottesman (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, 315, 525-8.
- Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper & J. Shendure (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, 46, 310-5.
- Li, F. P. & J. F. Fraumeni (1969) Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med*, 71, 747-52.
- (1982) Prospective study of a family cancer syndrome. *JAMA*, 247, 2692-4.
- Lorenz, R., S. H. Bernhart, C. Höner Zu Siederdisen, H. Tafer, C. Flamm, P. F. Stadler & I. L. Hofacker (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, 6, 26.
- Malkin, D., F. P. Li, L. C. Strong, J. F. Fraumeni, C. E. Nelson, D. H. Kim, J. Kassel, M. A. Gryka, F. Z. Bischoff & M. A. Tainsky (1990) Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science*, 250, 1233-8.
- Maxwell, K. N., B. Wubbenhorst, K. D'Andrea, B. Garman, J. M. Long, J. Powers, K. Rathbun, J. E. Stopfer, J. Zhu, A. R. Bradbury, M. S. Simon, A. DeMichele, S. M. Domchek & K. L. Nathanson (2015) Prevalence of mutations in a panel of

- breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. *Genet Med*, 17, 630-8.
- McLaren, W., B. Pritchard, D. Rios, Y. Chen, P. Flicek & F. Cunningham (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, 26, 2069-70.
- Mirabello, L., M. Yeager, P. L. Mai, J. M. Gastier-Foster, R. Gorlick, C. Khanna, A. Patiño-Garcia, L. Sierrasesúмага, F. Lecanda, I. L. Andrulis, J. S. Wunder, N. Gokgoz, D. A. Barkauskas, X. Zhang, A. Vogt, K. Jones, J. F. Boland, S. J. Chanock & S. A. Savage (2015) Germline TP53 variants and susceptibility to osteosarcoma. *J Natl Cancer Inst*, 107.
- Montera, M., F. Piaggio, C. Marchese, V. Gismondi, A. Stella, N. Resta, L. Varesco, G. Guanti & C. Marenzi (2001) A silent mutation in exon 14 of the APC gene is associated with exon skipping in a FAP family. *J Med Genet*, 38, 863-7.
- Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure & M. J. Bamshad (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 42, 30-5.
- Novoa, E. M. & L. Ribas de Pouplana (2012) Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet*, 28, 574-81.
- Olivier, M., D. E. Goldgar, N. Sodha, H. Ohgaki, P. Kleihues, P. Hainaut & R. A. Eeles (2003) Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype. *Cancer Res*, 63, 6643-50.
- Pagenstecher, C., M. Wehner, W. Friedl, N. Rahner, S. Aretz, N. Friedrichs, M. Sengteller, W. Henn, R. Buettner, P. Propping & E. Mangold (2006) Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. *Hum Genet*, 119, 9-22.
- Pollard, K. S., M. J. Hubisz, K. R. Rosenbloom & A. Siepel (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*, 20, 110-21.
- Reese, M. G., F. H. Eeckman, D. Kulp & D. Haussler (1997) Improved splice site detection in Genie. *J Comput Biol*, 4, 311-23.
- Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm & A. L. Q. A. Committee (2015) 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*, 17, 405-24.
- Ritchie, G. R., I. Dunham, E. Zeggini & P. Flicek (2014) Functional annotation of noncoding sequence variants. *Nat Methods*, 11, 294-6.
- Roy, R., J. Chun & S. N. Powell (2011) BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer*, 12, 68-78.
- Sauna, Z. E. & C. Kimchi-Sarfaty (2011) Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*, 12, 683-91.

- Schwarz, J. M., D. N. Cooper, M. Schuelke & D. Seelow (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*, 11, 361-2.
- Shapiro, M. B. & P. Senapathy (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*, 15, 7155-74.
- Shihab, H. A., M. F. Rogers, J. Gough, M. Mort, D. N. Cooper, I. N. Day, T. R. Gaunt & C. Campbell (2015) An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, 31, 1536-43.
- Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller & D. Haussler (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, 15, 1034-50.
- Supek, F., B. Miñana, J. Valcárcel, T. Gabaldón & B. Lehner (2014) Synonymous mutations frequently act as driver mutations in human cancers. *Cell*, 156, 1324-1335.
- Tung, N., S. M. Domchek, Z. Stadler, K. L. Nathanson, F. Couch, J. E. Garber, K. Offit & M. E. Robson (2016) Counselling framework for moderate-penetrance cancer-susceptibility mutations. *Nat Rev Clin Oncol*, 13, 581-8.
- Varley, J. M., C. Attwooll, G. White, G. McGown, M. Thorncroft, A. M. Kelsey, M. Greaves, J. Boyle & J. M. Birch (2001) Characterization of germline TP53 splicing mutations and their genetic and functional analysis. *Oncogene*, 20, 2647-54.
- Walsh, T., S. Casadei, M. K. Lee, C. C. Pennil, A. S. Nord, A. M. Thornton, W. Roeb, K. J. Agnew, S. M. Stray, A. Wickramanayake, B. Norquist, K. P. Pennington, R. L. Garcia, M. C. King & E. M. Swisher (2011) Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A*, 108, 18032-7.
- Yeo, G. & C. B. Burge (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, 11, 377-94.
- Zhang, J., M. F. Walsh, G. Wu, M. N. Edmonson, T. A. Gruber, J. Easton, D. Hedges, X. Ma, X. Zhou, D. A. Yergeau, M. R. Wilkinson, B. Vadodaria, X. Chen, R. B. McGee, S. Hines-Dowell, R. Nuccio, E. Quinn, S. A. Shurtleff, M. Rusch, A. Patel, J. B. Beckfort, S. Wang, M. S. Weaver, L. Ding, E. R. Mardis, R. K. Wilson, A. Gajjar, D. W. Ellison, A. S. Pappo, C. H. Pui, K. E. Nichols & J. R. Downing (2015) Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med*, 373, 2336-2346.

8. ARTIGO ORIGINAL

Title: *In silico* analysis of synonymous variants in cancer predisposition genes

Marina Roberta Scheid (1), Martiela Vaz de Freitas (2), Gabriel Macedo (3) e Patricia Ashton-Prolla(1-3).

- (1) Programa de pós-graduação em medicina: ciências médicas, Universidade Federal do Rio Grande do Sul.
- (2) Programa de pós-graduação em genética e biologia molecular, Universidade Federal do Rio Grande do Sul.
- (3) Hospital do Clínicas de Porto Alegre.

Abstract

The biological impact and clinical significance of synonymous variants (SVs) is still seen as a big challenge in understanding their contribution in genetic diseases. It has been shown that SVs are capable of altering gene expression, as well as function and conformation of proteins. Moreover, SVs can be inserted in conserved regions, such as splicing sites, CpG islands, or can alter the secondary mRNA conformation. In this study, we evaluated the accuracy of free-access *in silico* tools that interrogate the biological impact of SVs, as well as, we analyzed SVs from cancer predisposing genes. Thirteen *in silico* tools were selected and sensibility and specificity analyses were performed. SVs identified in cancer predisposition genes were obtained from a public database and from genetic testing results from a cancer risk assessment clinic. Analyses of secondary structure and difference of mRNA energy folding (Δ MEF) were carried out using RNAfold and the presence of CpG island was estimated by MethPrimer. The software SilVA presented the highest accuracy between all multiple annotation *in silico* tools selected. SilVA classified approximately 1% of SVs obtained from ExAC as likely or potentially pathogenic. Splicing alterations and the mRNA thermodynamic folding were the main mechanisms potentially involved in SVs pathogenicity. Structure

analyses in mRNA folding demonstrated that two SVs in *TP53* (c.66A>G and c.108G>A) resulted in abnormal mRNA folding energy or structure. We also found that the SVs c.1395A>C and c.3756A>T in *BRCA2* could result in splicing disruption according. To our knowledge this is the first study evaluating the efficacy of *in silico* tools and VSs in a comprehensive set of cancer predisposing genes.

Key-words: Synonymous variants, silent mutations, hereditary cancer, *TP53*, *BRCA1* and *BRCA2*.

Introduction

Hereditary cancer is characterized by the presence of germline mutations in tumor suppressor genes or oncogenes (Tung, et al., 2016). Most of these mutations are point mutations or *single nucleotide polymorphisms*, which can be classified as *synonymous* or *non-synonymous*, based on their ability to change the encoded amino acids and affect protein function. With the emergence of massively parallel sequencing there has been a significant increase in the number of variants detected, and among them, synonymous variants (SVs). SVs, commonly known as silent mutations, are variants that change the coding sequence without altering protein sequence (McLaren, et al., 2010). For a long time, it was thought that SVs had no biological and clinical impact, since they have no apparent functional consequences. More recently, however, several studies have shown that SVs may have clinical implications, once it is shown that they can alter gene expression by different mechanisms (Sauna and Kimchi-Sarfaty, 2011).

The role of SVs in underlying different diseases, such as hereditary cancer, has been studied by several groups (Sauna and Kimchi-Sarfaty, 2011; Supek, et al., 2014). Currently, great part of the articles describing pathogenic SVs as associated to hereditary cancer relates these variant's disease-causing mechanism with splicing alterations (Aretz, et al., 2004; Auclair, et al., 2006; Varley, et al., 2001). However, alternative mechanisms of pathogenicity have been described for SVs. Recently, Karakostis et al (2018) have shown that the synonymous *TP53* variant c.66A>G changes *TP53* pre-mRNA translation and impairs p53 interaction with MDM2, a p53 negative regulator (Karakostis, et al., 2018). Moreover, SVs can interfere in codon usage, affecting gene expression, translation velocity and folding efficiency (Plotkin and

Kudla, 2011). Studies in other Mendelian diseases have demonstrated that SVs can modify cytosine-phosphate-guanine (CpG) sites, influencing in DNA methylation patterns. Finally, interference in microRNA binding sites, mRNA folding and mRNA conformation energy have been demonstrated (Hunt, et al., 2014).

The American College of Medical Genetics and Genomics guideline for variant classification highlights that SVs which are not inserted in conserved positions over evolution and which present no impact in splicing assessment prediction algorithms have strong indication of being benign (Richards, et al., 2015). According to the guideline, *in silico* tools should be used with the purpose of composing an evidence of pathogenicity. Population allelic frequency, segregation analysis, and functional studies must be considered to infer the pathogenicity state of these variants. The methods and *in silico* tools that can be used to understand the effect of SVs, such as tools for splicing, RNA folding and miRNA binding sites, were reviewed by Hunt and colleagues (Hunt, et al., 2014). Specifically for SVs, the software SilVA (Silent Variant Analyzer) was trained to reclassify and analyze different mechanisms of pathogenicity able to affect SVs. SilVA evaluates several parameters, including allelic frequency and conservation, codon usage bias, sequence features (CpG and relative mRNA position), exon splicing enhancer and suppressor motifs, splicing sites motifs and pre-mRNA folding energy (Buske, et al., 2013).

In this work, we sought to understand the mechanisms of pathogenicity of SVs in genes predisposing to hereditary cancer through *in silico* tools. We selected a range of *in silico* tools capable of evaluating SVs obtained from a public database and from a single institution derived cohort of individuals with suspected hereditary cancer.

Methodology

Selection of in silico tools for SV analysis

Freely available *in silico* tools for SVs analysis were selected and grouped into three categories: Multiple Annotation *in silico* tools, Conservation *in silico* tools, and Splice Site *in silico* tools. The tools selected were: SilVA, Mutation Taster, FATHMM, Predict SNP, CADD, GWAVA, as multiple annotation tools; PhastCons and Phylo P, as conservation tools; and Site Splice Finder, MaxEntScan, NNSplice, GeneSplicer, as

splicing tools. All selected tools are listed in Supplementary Table 1. Splice site tools were accessed by Alamut Visual Software (Alamut, Interactive Biosoftware, Roven France). Beyond these three categories, we also included *MethPrimer* for CpG island evaluation and *RNAfold* for secondary mRNA structure analysis.

Sensibility and specificity analysis

Confirmed pathogenic and benign SVs were retrieved from ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>). These variants were analyzed by *in silico* tools, and all raw scores were computed. For splicing tools, it was necessary to transform the data obtained by the tools to a single value, and we decided to use the difference module. After, we performed a Receiver-operating characteristic (ROC) curve using IBM SPSS statistic 20 version software. ClinVar classification and functional evidence for pathogenic presented in the literature were considered as gold standard in the ROC curve performance. With the analysis of sensitivity and specificity, it was possible to verify that SVs located in splicing sites (or that alter the splicing) result, in most cases, in a complete loss of a donor site. For further analysis, we selected only *in silico* tools with significant values of area under the curve (sensibility and specificity).

Selection of synonymous variants

SVs studied here were selected from two different cohorts. In the first cohort, SVs of 22 genes linked to hereditary cancer syndromes (*APC*, *ATM*, *BARD1*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *MLH1*, *MSH2*, *MSH6*, *NF1*, *NF2*, *PALB2*, *PTEN*, *RAD50*, *RAD51*, *RET*, *STK11*, *TP53*, *TSC1* and *TSC2*) were selected from ExAC (Exome Agregation Consortium) public database. In the second cohort, SVs were those identified in genetic testing reports from a single institution cancer risk assessment clinic, including reports from *TP53*, *BRCA1* and *BRCA2* analyses. Genetic testing was performed after signature of informed consent by the patients and the protocol of analysis was approved by the institutional review board (Hospital de Clínicas de Porto Alegre, Brazil, protocol number 10/0521). In this group we only selected variants with minor allelic frequencies lower than 5%.

In silico tools analysis

We evaluated all selected variants from both groups using *in silico* tools that presented significant values of area under the curve in the sensibility and specificity test. The analysis of the second group included MethPrimer and RNAfold. MethPrimer was used to verify the presence of CpG islands along the gene, and to estimate if CpG islands occur in the region where the variants are inserted. RNAfold was used to assess mRNA folding structure and the minimum free energy (MEF) of the thermodynamic ensemble molecule. First, we computed all raw scores obtained after *in silico* tools analysis. Then, for each parameter evaluated, we attributed a new score based on the significance of the tool. Scores were assigned according to cut-off points provided by the predictors themselves. Greater scores were attributed to the SilVA tool and clinical evidence of pathogenicity accessed by ClinVar. The tools with less evidence, such as conservation tools, were assigned smaller scores. The scores chosen for each parameter are described in Supplementary table 2. For both groups, the results were presented in heat maps constructed with GGPlot package (R Development Core Team, 2008).

Results

Sensitivity and specificity of in silico tools

Thirty-two pathogenic and 180 benign SVs were selected in ClinVar database. The area under the curve was calculated for the 13 tools initially selected and results are summarized in figure 1 and table 2. In the category of multiple annotation tools (Figure 1A), the predictor SilVA showed higher sensitivity and specificity (0,906 and 0,866, respectively), and an area under the curve of $0,973 \pm 0,026$ ($p < 0,001$). GWAVA was the only tool with non-significant values of sensitivity and specificity, and, therefore, was not included in subsequent analyses. Sensitivity and specificity values were similar when considering evolutionary and conservation tools (Figure 1B). For splicing tools, we excluded two pathogenic variants in which the mechanism of pathogenicity was not clearly related to splicing alterations. MaxEnt presented slightly higher values of sensitivity and specificity between all analyzed tools, however, the results from all tools were very similar (Figure 3A).

Synonymous variants from ExAC

We selected 4,324 variants from 22 hereditary cancer predisposition genes retrieved from the ExAC database. Approximately 1% (42/4,324) of the variants were

classified as likely (n=2) or potentially (n=40) pathogenic using SilVA. All these SVs were evaluated by the other tools and results are described in Figure 2. We also evaluated the global minor allelic frequency (Global MAF) and the proposed classification found in ClinVar. The two likely pathogenic SVs according to SilVA were c.96G>A in *RET* and c.375G>A in *TP53*. Two additional SVs were also classified as pathogenic in ClinVar: c.2709A>G in *NF1* and c.375G>A in *TP53*, based on functional studies, but only c.375G>A were classified as pathogenic by both ClinVar and SilVA. Eight SVs were classified as variants of uncertain significance, 22 were benign and 10 were not described in ClinVar. All variants classified as probably pathogenic presented a global MAF lower than 0.01. Predict SNP and DANN classified all 40 SVs as benign, and only 10 SVs were classified as potentially pathogenic by CADD. Seventeen SVs were classified with significant alterations in splicing by at least two splicing predictors. The highest final score was obtained by the SV c.375G>A in the *TP53* gene.

Synonymous variants from individuals tested in the context of a high risk clinic.

We analyzed sequencing results of the coding regions of *TP53* (n=279) and *BRCA1/BRCA2* (n=82) genes from individuals tested in the context of a high risk clinic, and we selected all SVs with minor allelic frequency lower than 0.05. After filtering variants by population frequency, 14 SVs were selected for evaluation using the *in silico* tools (Figure 3). Here we also included RNAfold analysis in order to infer secondary structure of mRNA, and MethPrimer to evaluate if the SVs are inserted in CpG islands. Three different variants were identified in *TP53*: c.66A>G (n=1), c.108G>A (n=3) and c.639T>G (n=12), two variants in *BRCA1*: c.366T>G (n=1) and c.1971A>G (n=1) and nine in *BRCA2*: c.1365A>G (n=4), c.1395A>C (n=1), c.1680T>C (n=1), c.2145A>G (n=1), c.2229T>C (n=4), c.3786A>T (n=1), c.5199C>T (n=1) and c.9720T>C (n=2). All individuals carrying c.1365A>G in the *BRCA1* gene also carried c.2229T>C. The *TP53* variant c.66A>G has been described as VUS and the *BRCA2* SVs c.1680T>C and c.3786A>T have not been described yet by ClinVar. All of these variants were predicted to be benign using SilVA, Predict SNP and DANN.

Analysis of secondary mRNA structure was performed using RNAfold software and using this strategy two *TP53* SVs, c.66A>G and c.108G>A, presented significant alterations in the secondary structure of mRNA (Figure 3). Variant c.66A>G showed a

structural alteration in the molecule and Δ MEF value of 0,9Kcal/mol and was identified in an individual with a familial history of cancer, but no clinical criteria for Li-Fraumeni or Li-Fraumeni like syndrome (Figure 5A). Variant c.108G>A demonstrated no structural alteration, but an increase in the thermodynamic energy and Δ MEF of 5,2Kcal/mol. The individual who carried this variant had a personal and familial history of cancer (Figure 5B) consistent with Modified Chompret criteria for Li-Fraumeni syndrome (Bougeard et al. 2008). In this case, the reduced size of the *TP53* mRNA molecule could be an alternative pathogenic mechanism associated with this particular SV.

We observed two variants in the *BRCA2* gene, c.1395A>C and c.3756A>T, with possible splice site alterations. The variant c.1395A>C was classified as benign by ClinVar and the patient carrying this variant was a male with personal history of breast cancer at age 46 years (Figure 5C). Variant c.3756A>T has not been described in ClinVar or dbSNP. Moreover, the individual carrying this variant also had a personal and familial history of cancer, with a diagnosis of breast and ovarian cancer and two first-degree relatives with breast cancer (Figure 5D). Both variants identified in *BRCA2* are not found in canonical exon-intron junction positions.

Discussion

In this study we investigate possible biological consequences of SVs in cancer predisposition genes using different *in silico* tools. Software SilVA, developed specifically for evaluation of SVs, showed the highest accuracy between all selected multiple annotation *in silico* tools. SilVA allowed us to infer that around 1% of all SVs obtained from ExAC are classified as likely or potentially pathogenic. We found that splicing alteration and the mRNA thermodynamic folding are the main mechanisms potentially involved in the pathogenicity of SVs in cancer predisposing genes.

Recently, another study assessed the sensitivity and specificity of *in silico tools* for SV analysis and compared the assessment to a new methodology, the Discriminating disease-causing synonymous variants (DDIG-SN). The tools evaluated by the group included SilVA, phyloP, FATHMM-MKL, SPIDEX, CADD and Mutation Taster. Apart from DDIG-SN, SilVA presented the highest values of area under the curve, showing the superiority of this tool (Livingstone, et al., 2017), a result that we

confirmed in the current study. In addition, the accuracy of splicing *in silico* tools was evaluated for variants of uncertain significance (VUS) in *BRCA1* and *BRCA2* that are predicted to affect splicing and all tools used showed presented significant values of AUC. The authors showed that the combined use of Splice Site Finder and MasEntScan resulted in high sensitivity (96%) and specificity (83%) in the analyses of VUS (Houdayer, et al., 2012). Analyses of splicing effects are important since several studies have demonstrated that SVs can result in a functional compromise in splicing, even if the variant location is far from the exon-intron junction (Defesche, et al., 2008; Korvatska, et al., 2013; Liu, et al., 1997). In this study, we used the ClinVar classification and clinical information presented in the literature as gold standard for synonymous variants to perform the ROC curve. However, the necessity of further information that confirm this classification is still necessary to increase the reliability of the test.

Finally, the *TP53* c.396A>G SV, the most frequent variant observed in the patient cohort, was classified as possible benign. In another study, this variant was associated with metastasis risk in patients with osteosarcoma (Mirabello et al. 2015). SV c.1395A>C in *BRCA2* was also studied previously and found to be benign using a mini gene splicing assay (Théry et al, 2011). Furthermore, we observed that the *TP53* variant c.66A>G is capable of altering the secondary mRNA structure. Recently, Karakostis and colleagues studied the effects of this variant *in vitro* and showed that it is able to prevent phosphorylation of serine 15 (S15) of p53. The absence of S15 phosphorylation prevents the stabilization of p53-MDM2 complex, and prevents the complex action during genotoxic stress (Karakostis et al, 2018). Although this study shows that SVs may interfere in the protein translation process, it is not clear if such interference results from the alteration of the secondary mRNA conformation. In the Figure 3, global MAF used for the local cohort was obtained in ExAC, and although ExAC is one of the most complete database, this information might present bias, once ExAC do not include individuals from Brazilian population. However, the Brazilian database AbraOM did not present great part of the population frequency of the SVs analyzed in this work.

Conclusion

This is the first study evaluating the efficacy of *in silico* tools in the assessment of SVs in a comprehensive set of cancer predisposition genes. *In silico* tools are a cost-effective alternative for the initial analysis of SVs, once the combined use of these methodologies is able to enhance our understanding of a potential deleterious effect. Currently, several bioinformatics pipelines resort filters for the exclusion of SVs, since their clinical significance is inferred as benign. In this study, we demonstrate that a low percentage of SVs could be associated with a pathogenic potential of hereditary cancer predisposition and thus, simply excluding them from the analysis could have as consequence a false-negative result. Among the mechanisms of pathogenicity identified in SVs, splicing alterations have shown to be very significant, but other mechanisms should also be considered.

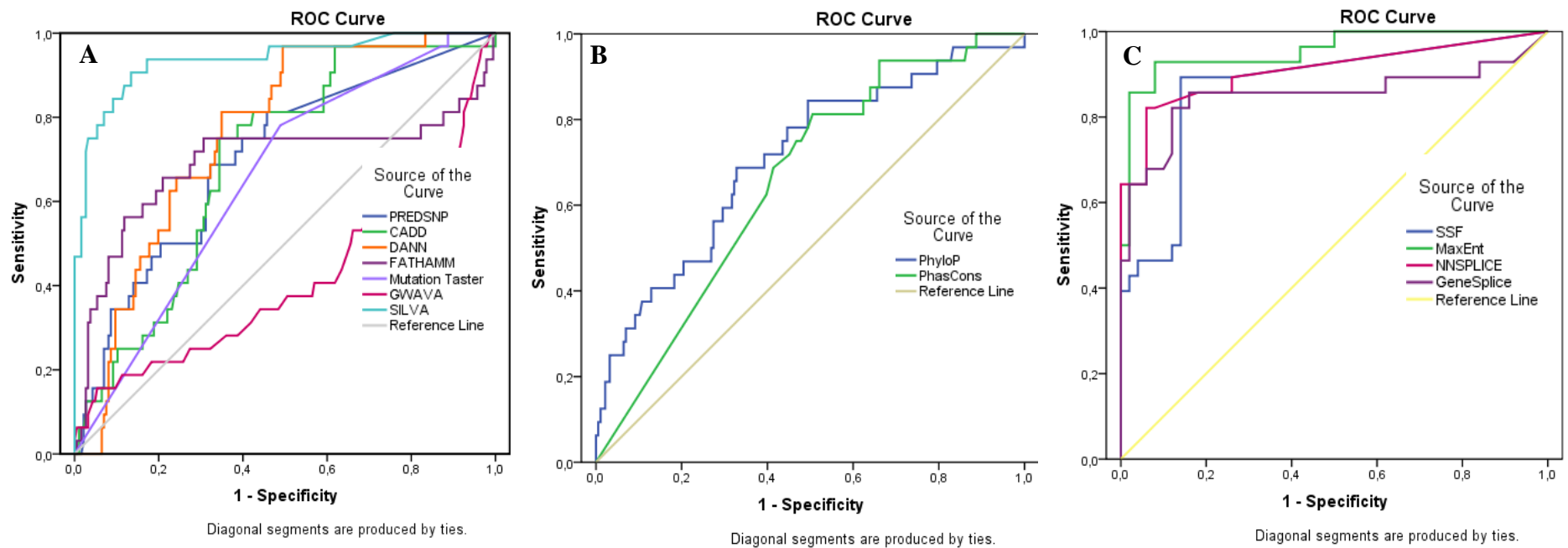


Figure 1. ROC curves for three different categories of *in silico* tools. (A) Multiple annotation *in silico* tools; (B) Conservation *in silico* tools; and (C) Splicing *in silico* tools.

Table 2. Receive operating characteristic curve values for *in silico* tools evaluated.

<i>In silico</i> tool	AUC	Std.		95% CI	
		Error	Significance	Lower	Upper
SILVA	0,937	0,026	<0,001	0,855	0,989
DANN	0,759	0,038	<0,001	0,684	0,834
PredictSNP	0,703	0,05	<0,001	0,605	0,801
CAAD	0,695	0,045	<0,001	0,607	0,783
FATHAMM	0,693	0,057	<0,001	0,582	0,804
Mutation Taster	0,654	0,048	0,006	0,560	0,747
GWAVA	0,415	0,062	0,125	0,293	0,537
PhyloP	0,708	0,501	<0,001	0,605	0,809
PhasCons	0,656	0,047	0,005	0,564	0,748
MaxEnt	0,954	0,025	<0,001	0,901	1,000
GeneSplicer	0,908	0,041	<0,001	0,827	0,988
SSF	0,871	0,045	<0,001	0,783	0,959
NNSplicer	0,851	0,057	<0,001	0,739	0,693

AUC: Area under the curve. Std Error: Standard Error; CI: Confidence interval;

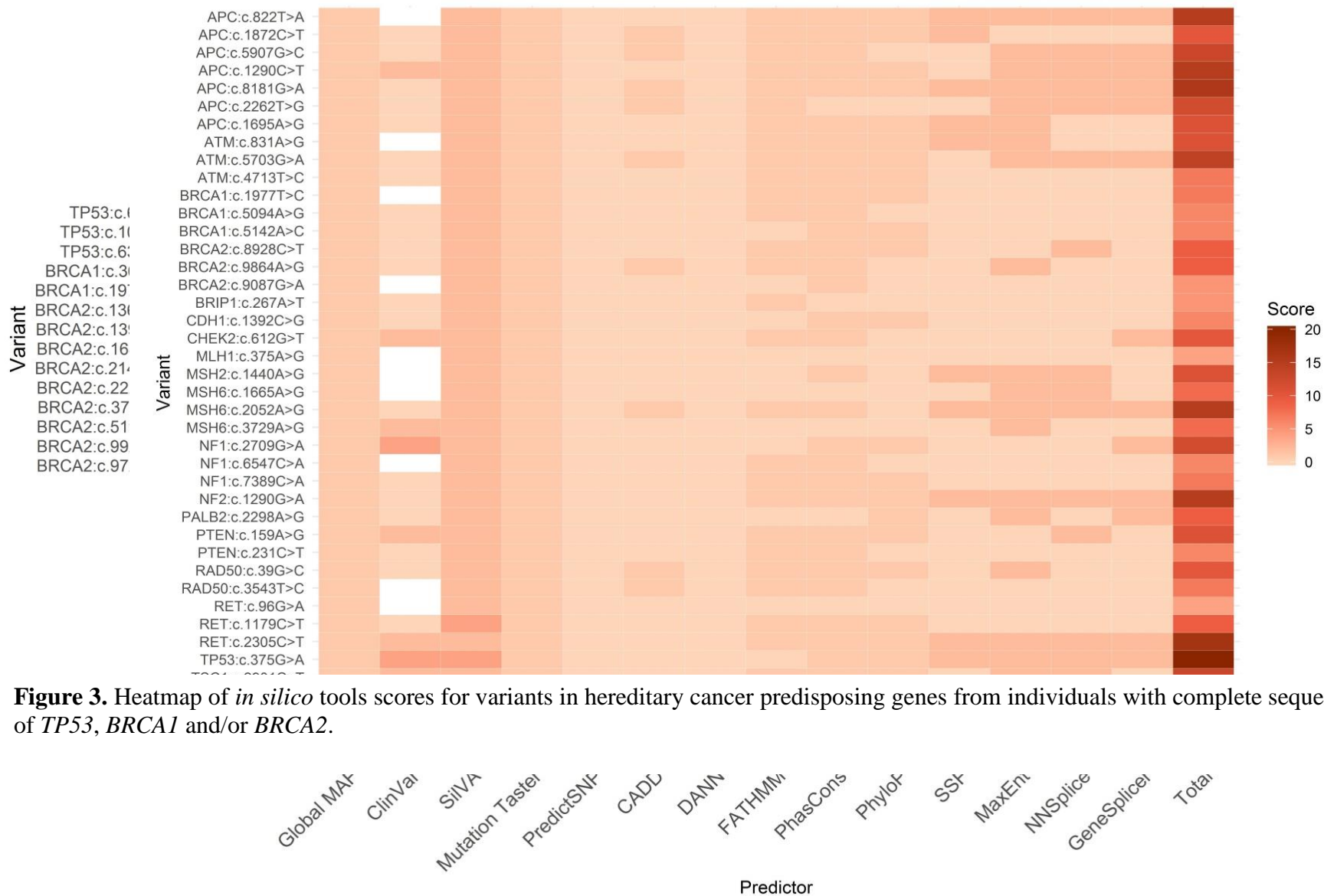


Figure 3. Heatmap of *in silico* tools scores for variants in hereditary cancer predisposing genes from individuals with complete sequencing of *TP53*, *BRCA1* and/or *BRCA2*.

Figure 2. Heatmap of *in silico* tools scores for variants in hereditary cancer predisposing genes classified as possible and probably pathogenic by SilVA selected from ExAC.

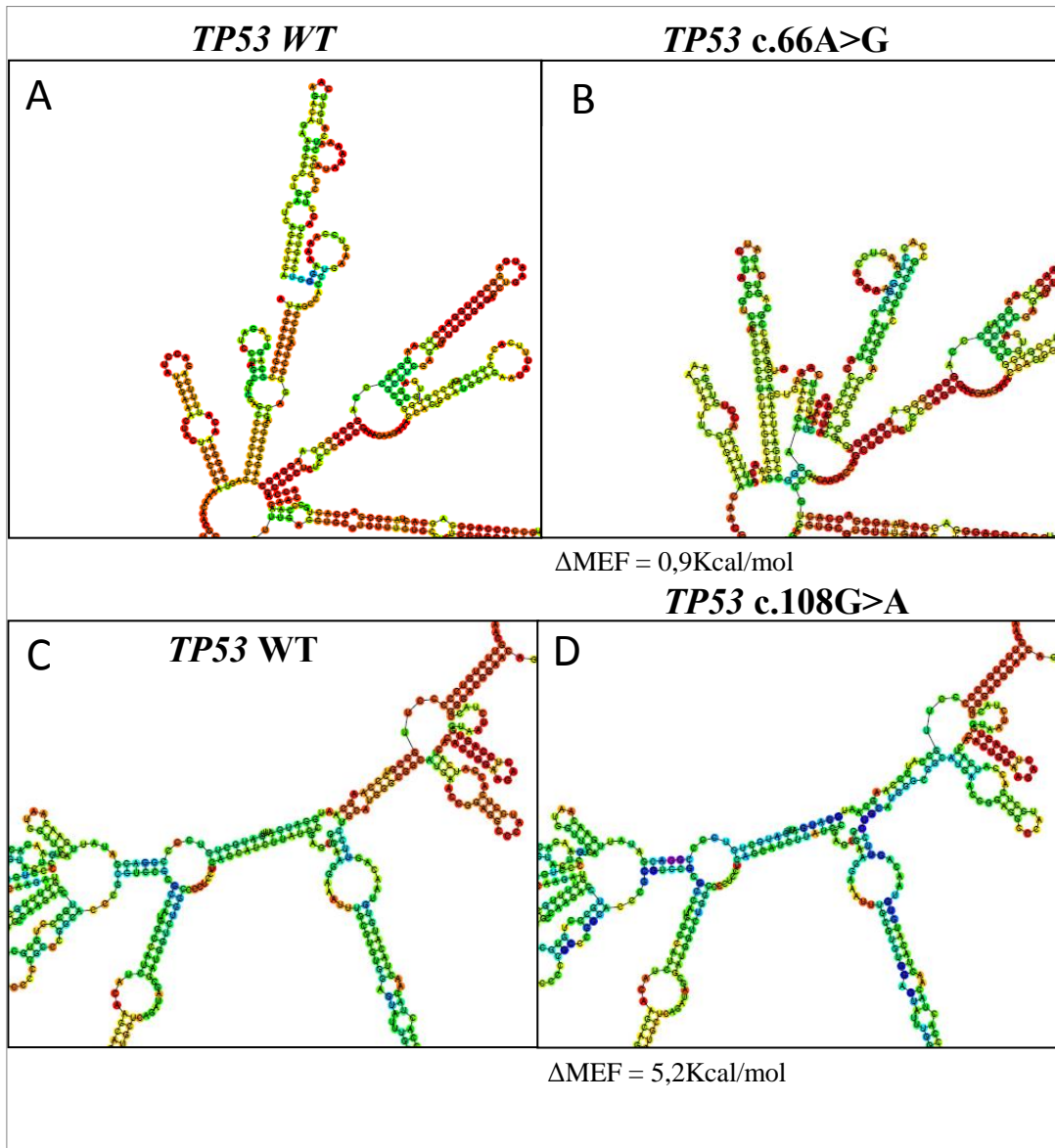


Figure 4. Secondary structure of *TP53* mRNA performed by RNAfold represented by the difference of minimum free energy (MEF) and molecule thermodynamic ensemble.

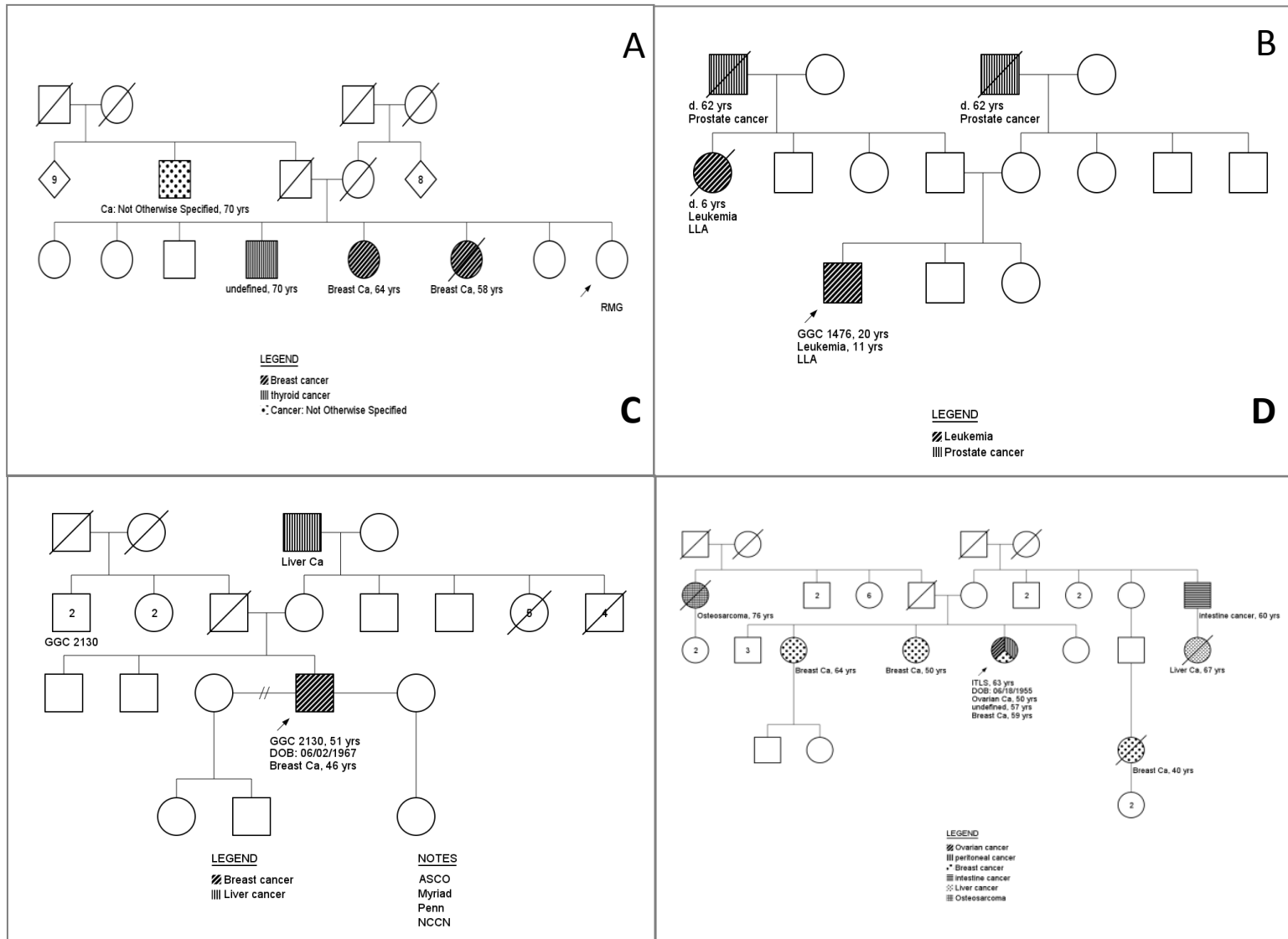


Figure 5. Family history of individuals with synonymous variants identified with significance alterations by *in silico* tools. (A) *TP53* c.66A>G; (B) *TP53* c.108G>A; (C) *BRCA2* c.1395A>C; (D) *BRCA2* c.3786A>T.

Supplementary Table 1. *In silico* tools for SVs analysis.

Tool category	Reference
Multiple annotation tools	
SilVA	(Buske et al. 2013)
Mutation Taster	(Schwarz et al. 2014)
FATHMM	(Shihab et al. 2015)
Predict SNP	(Bendl et al. 2014)
CADD	(Kircher et al. 2014)
GWAVA	(Ritchie et al. 2014)
Conservation tools	
PhastCons	(Siepel et al. 2005)
PhyloP	(Pollard et al. 2010)
Splicing tools	
Site Splice Finder	(Shapiro and Senapathy 1987)
MaxEntScan	(Yeo and Burge 2004)
NNSplice	(Reese et al. 1997)
GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml
CpG island	
MethPrimer	(Li and Dahiya, 2002)
Secondary structure of mRNA	
RNAfold	(Lorenz et al. 2011)

Supplementary table 2. Scores used for values obtained through in silico analyzes.

Predictor	Alteration	Score
Global MAF	<0,01	+1
	>0,01	0
Clinvar	Pathogenic	+4
	VUS	+2
	Benign/ LB	0
SILVA	Likely Pathogenic	+4
	Potentially Pathogenic	+2
	Benign	0
Mutation Taster	Disease causing	+1
	Polymorphism	0
PhasCons (0-1)	>0,5	+1
	<0,5	0
PhyloP (-14 - +6)	Positive values	+1
	Negative values	0
Predict SNP	Neutral	0
	Pathogenic	+1
FATHMM	Pathogenic	+1
	Neutral	0
CADD	Pathogenic	+1
	Neutral	0
DANN	Pathogenic	+1
	Neutral	0
SSF (0-100)	Strong alteration (donor/acceptor)	+2
	Weak or absense of alterations	0
MaxEnt (0-12)	Strong alteration (donor/acceptor)	+2
	Weak or absense of alterations	0
NNSplice (0-1)	Strong alteration (donor/acceptor)	+2
	Weak or absense of alterations	0
GeneSplicer (0-24)	Strong alteration donor/acceptor)	+1
	Weak or absense of alterations	0
MethPrimer (CpG Island)	Presence	+1
	Absence	0
RNA Fold	Modify stability of mRNA	+2
	Modify conformation of mRNA	+2
	No alteration	0

References

- Aretz S, Uhlhaas S, Sun Y, Pagenstecher C, Mangold E, Caspari R, Möslein G, Schulmann K, Propping P, Friedl W. 2004. Familial adenomatous polyposis: aberrant splicing due to missense or silent mutations in the APC gene. *Hum Mutat* 24(5):370-80.
- Auclair J, Buisine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S and others. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Human Mutation* 27(2):145-154.
- Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J. 2014. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol* 10(1):e1003440.
- Buske OJ, Manickaraj A, Mital S, Ray PN, Brudno M. 2013. Identification of deleterious synonymous variants in human genomes. *Bioinformatics* 29(15):1843-50.
- Defesche JC, Schuurman EJ, Klaaijzen LN, Khoo KL, Wiegman A, Stalenhoef AF. 2008. Silent exonic mutations in the low-density lipoprotein receptor gene that cause familial hypercholesterolemia by affecting mRNA splicing. *Clin Genet* 73(6):573-8.
- Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P and others. 2012. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 33(8):1228-38.
- Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. 2014. Exposing synonymous mutations. *Trends Genet* 30(7):308-21.
- Karakostis K, Vadivel Gnanasundram S, López I, Thermou A, Wang L, Nylander K, Olivares-Illana V, Fåhræus R. 2018. A single synonymous mutation determines the phosphorylation and stability of the nascent protein. *J Mol Cell Biol*.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46(3):310-5.
- Korvatska O, Strand NS, Berndt JD, Strovast T, Chen DH, Leverenz JB, Kiianitsa K, Mata IF, Karakoc E, Greenup JL and others. 2013. Altered splicing of

- ATP6AP2 causes X-linked parkinsonism with spasticity (XPDS). *Hum Mol Genet* 22(16):3259-68.
- Li LC, Dahiya R. 2002. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18(11):1427-31.
- Liu W, Qian C, Francke U. 1997. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. *Nat Genet* 16(4):328-9.
- Livingstone M, Folkman L, Yang Y, Zhang P, Mort M, Cooper DN, Liu Y, Stantic B, Zhou Y. 2017. Investigating DNA-, RNA-, and protein-based features as a means to discriminate pathogenic synonymous variants. *Hum Mutat* 38(10):1336-1347.
- Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* 6:26.
- McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26(16):2069-70.
- Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 12(1):32-42.
- Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20(1):110-21.
- Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* 4(3):311-23.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E and others. 2015. 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* 17(5):405-24.
- Ritchie GR, Dunham I, Zeggini E, Flicek P. 2014. Functional annotation of noncoding sequence variants. *Nat Methods* 11(3):294-6.
- Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 12(10):683-91.
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 11(4):361-2.

- Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 15(17):7155-74.
- Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR, Campbell C. 2015. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* 31(10):1536-43.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S and others. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8):1034-50.
- Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* 156(6):1324-1335.
- Théry JC, Krieger S, Gaildrat P, Révillion F, Buisine MP, Killian A, Duponchel C, Rousselin A, Vaur D, Peyrat JP and others. 2011. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur J Hum Genet* 19(10):1052-8.
- Tung N, Domchek SM, Stadler Z, Nathanson KL, Couch F, Garber JE, Offit K, Robson ME. 2016. Counselling framework for moderate-penetrance cancer-susceptibility mutations. *Nat Rev Clin Oncol* 13(9):581-8.
- Varley JM, Attwooll C, White G, McGown G, Thorncroft M, Kelsey AM, Greaves M, Boyle J, Birch JM. 2001. Characterization of germline TP53 splicing mutations and their genetic and functional analysis. *Oncogene* 20(21):2647-54.
- Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11(2-3):377-94.

9. CONSIDERAÇÕES FINAIS

Este trabalho avaliou a sensibilidade e especificidade de preditores *in silico* de múltiplas anotações, conservação e *splicing* para variantes sinônimas (VS). Nossos resultados demonstram que ferramentas *in silico* podem ser uma abordagem interessante para a avaliação inicial desta classe de alterações.

Para alcançar nossos objetivos tanto variantes obtidas de um banco de dados populacional (ExAC) quanto de indivíduos com critérios clínicos para síndrome de Li-Fraumeni e Síndrome do câncer de mama e ovário hereditários, e que realizaram sequenciamento completo de *TP53* ou *BRCA1/BRCA2*, foram avaliadas. Para tal, foram utilizadas ferramentas que avaliavam a conservação do nucleotídeo, alterações no *splicing* e ferramentas que avaliam mais de uma característica em conjunto (múltiplas anotações). O impacto na conformação secundária bem como diferença na energia de dobramento do RNA mensageiro (mRNA) selvagem e mutado também foram avaliados para as VS com maior probabilidade de impacto funcional. A análise abrangente de diferentes preditores nos permitiu concluir que alterações no mecanismo de *splicing* e conformação secundária do mRNA estão entre os principais mecanismos potenciais de patogenicidade de VS identificadas em genes de predisposição ao câncer.

Ao nosso conhecimento, este é o primeiro estudo que se propõe a utilizar ferramentas *in silico* de diferentes categorias para a avaliação de VS nos principais genes associados ao câncer hereditário.

10. PERSPECTIVAS FUTURAS

Abordagens *in silico* demonstraram ser uma alternativa custo-efetiva para uma análise inicial de mutações. Contudo, estudos adicionais são necessários para confirmar o mecanismo de patogenicidade dessas mutações.

Entre as perspectivas futuras desse trabalho, se encontram:

- 1) Abordagens complementares para avaliação das variantes encontradas como possivelmente patogênica. Estudos de segregação de mutações na família, avaliação da perda de heterozigotidade no tecido tumoral e ensaios funcionais poderiam corroborar para maiores evidências de patogenicidade.
- 2) Ampliação do número de VS provenientes de indivíduos com critérios clínicos para síndromes de predisposição ao câncer a ser avaliada, incluindo variantes identificadas em outros genes de predisposição ao câncer, bem como, expandir a avaliação dos mecanismos alternativos de patogenicidade (splicing, ilhas CpG, conformação secundária do mRNA) desenvolvidos nesse trabalho através de ferramentas *in silico* para variantes *missense* de significado incerto.
- 3) Realizar a análise de *codon usage bias* para genes de predisposição ao câncer em humanos, visto que VS apresentam um papel fundamental para avaliação evolução e conservação de genes.