



Universidade Estadual Paulista “Júlio de Mesquita Filho”

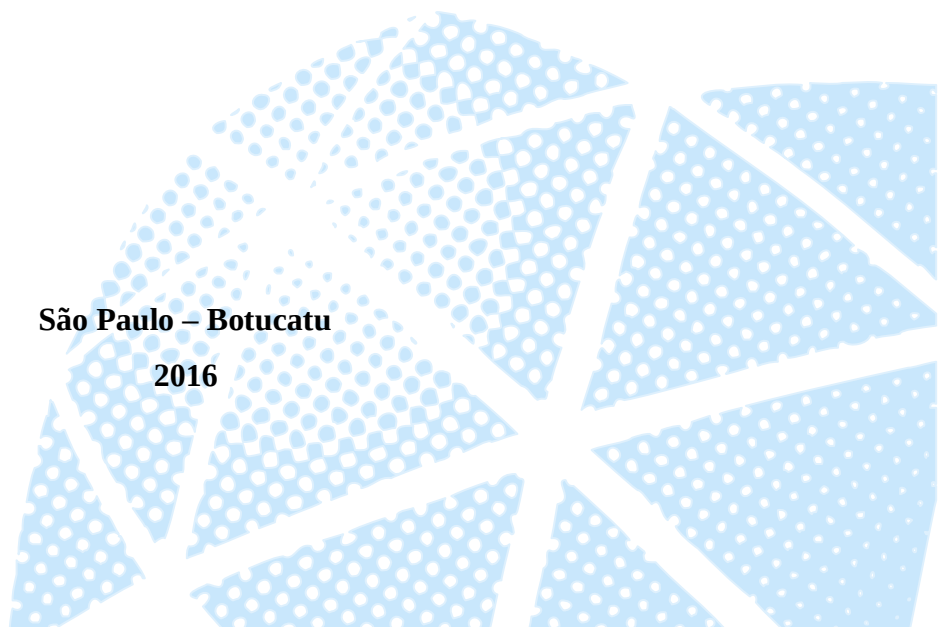
Instituto de Biosciências de Botucatu

**IDENTIFICAÇÃO DE GENES DE PREDISPOSIÇÃO À SÍNDROME DO
CARCINOMA DE MAMA E TIREOIDE POR SEQUENCIAMENTO
TOTAL DO EXOMA**

MAÍSA PINHEIRO

São Paulo – Botucatu

2016



**IDENTIFICAÇÃO DE GENES DE PREDISPOSIÇÃO À SÍNDROME DO
CARCINOMA DE MAMA E TIREOIDE POR SEQUENCIAMENTO
TOTAL DO EXOMA**

MAÍSA PINHEIRO

Orientadora: Profª Dra Silvia Regina Rogatto

Tese apresentada ao Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, para obtenção do título de Doutor em Ciências Biológicas – Área de concentração: Genética.

São Paulo – Botucatu

2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Pinheiro, Maísa.

Identificação de genes de predisposição à síndrome do carcinoma de mama e tireoide por sequenciamento total do exoma / Maísa Pinheiro. - Botucatu, 2016

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Silvia Regina Rogatto

Capes: 20205007

1. Mamas - Câncer. 2. Tireoide - Câncer. 3. Diátese.
4. Exoma.

Palavras-chave: Cancer de mama; Cancer de tireoide;
Sequenciamento do exoma total; Síndromes de predisposição
ao câncer.

“Only one who devotes himself to a cause with his whole strength and soul can be a true master. For this reason mastery demands all of a person.”

Albert Einstein

AGRADECIMENTOS

Primeiramente à Dra. Silvia Rogatto, minha orientadora querida, muitas vezes mãe me dando bronca nas horas certas e me acolhendo quando havia lágrimas em meus olhos. Agradeço imensamente os ensinamentos e a confiança em mim depositada para desenvolver este trabalho lindo. Espero ter conseguido retribuir da melhor maneira possível. Vai fazer muita falta em minha vida!

Ao meu namorado Bruno por estar ao meu lado sempre, independente de como, quando e onde, com amor e atenção. Fortalecendo minhas bases sempre quando eu achava que eu não conseguiria.

Agradeço minha família, mãe, pai, irmãos que são a razão do meu viver e de todo meu esforço. Às minhas tias que participaram da minha educação em todas as etapas aconselhando e aconchegando, sempre.

Ao Fábio que, além de “chefinho”, é um amigo de coração, que nunca me disse não e sempre fez de tudo pra me ajudar no que fosse.

Agradeço à Sandra por sempre me guiar da forma mais simples e verdadeira, ajustando e refinando minhas habilidades para hoje eu ser o que sou.

Aos colaboradores deste trabalho Dra. Maria Isabe Achatz, Dr. Paulo Kowalski e Dr. Luiz Coutinho que contribuíram imensamente e apostaram na realização deste trabalho.

Ao Dr. Igor Jurisica que me recebeu tão bem em seu laboratório em Toronto – CA, e por todo o acolhimento de Chiara, Tomas e Mike.

À Dra. Sonia Andrade e Leandro que me ensinaram tudo o que sei hoje de bioinformática e que, diga-se de passagem, não foi nada fácil!

Aos amigos e companheiros de laboratório Às minhas amigas Carol (de Minas), Hellen, Luísa, Carol (de São Paulo), Mariana Fragoso, Sara, Flávia, Isabela, Mariana, Priscila, Marco, Mateuzinho, André Guolo, André Rolando, Graziela, Tati e Julia. Ao Kelvin que se tornou um irmão nessa caminhada de pós-graduação.

Às equipe da Oncogenética que sem as quais eu não poderia ter realizado este trabalho: Fernanda, Amanda, Andreia, Gislene, Juliana, Renata e principalmente Karina.

Agradeço aos companheiros de Botucatu Tainara, Natália, Márcio, Jovita, Naiara, Cadu e Talita. Às pesquisadoras e amigas Dra. Patrícia Reis e Dra. Reneé Amorin.

Aos órgãos de financiamento Conselho Nacional de Pesquisa do Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – processo 2014/03983-8) pelas bolsas cedidas.

À Deus e meu anjo da guarda.

RESUMO

A identificação de fatores de predisposição e a caracterização de fatores genéticos de risco para o desenvolvimento do câncer são essenciais para o diagnóstico, prognóstico e aconselhamento genético dos pacientes afetados e seus familiares. Há uma alta incidência de pacientes que possuem carcinomas de mama (CM) e/ou carcinoma de tireoide (CT) e com história familiar destes tumores, sugerindo o envolvimento de gene(s) associado(s) com a predisposição hereditária ao câncer. O objetivo deste trabalho foi investigar a presença de genes de predisposição ao desenvolvimento de CM e CT por meio do sequenciamento total do exoma, em pacientes que sejam afetados por pelo menos um destes tumores e possuam história familiar de câncer. Foram incluídas seis pacientes que apresentavam CM, 12 com CT e seis com ambos os tumores, todos com história destes tumores em suas famílias. O DNA de sangue periférico foi sequenciado pela plataforma HiSeq 1000 (Illumina). Após análise de bioinformática foram identificadas alterações em genes de alta e baixa penetrância comuns e raras. As alterações raras destes genes foram consideradas como patogênicas quando presentes em pelo menos três bancos de dados e em até dois indivíduos. A paciente 14 teve diagnóstico de síndrome de Cowden confirmado pela presença da mutação p.R233X no gene *PTEN*. Também foram investigadas outras alterações em novos genes o que resultou em 911 SNVs (*single nucleotide variations*), sendo 287 não-sinônimas e 50 alterações raras consideradas patogênicas em pelo menos três bancos de dados. Entre as 50 alterações patogênicas, 13 mostraram interação proteica com produtos codificados por genes de alta ou baixa penetrância previamente descritos em síndromes hereditárias clássicas. Também foram identificadas alterações em genes recentemente descritos como associados à CT e ou CM hereditários (*HABP2*, *PARP4* e *SRRM2*). Foi verificado que as vias de produção de hormônios tireoidianos e de sinalização RAP1 podem ter suas funções comprometidas por alterações deletérias que foram detectadas em 10 e 14 pacientes deste estudo, respectivamente. Especificamente a via de síntese de hormônio tireoidiano revelou estar potencialmente desregulada nas Famílias I e II por meio das alterações em *PLCB3* e *PAX8*, respectivamente. As alterações consideradas mais promissoras incluem a rs35169799 (*PLCB3*), as classificadas como de manutenção do genoma (rs61997220: *ZC3H12D*, rs61902276: *DDIAS*, rs61754785: *MUS81* e rs62619782: *RNASEH2A*) e as potencialmente desreguladas em vias importantes (rs3188996: *PAX8*). As variantes rs61754785 (*MUS81*) e rs7080536 (*HABP2*) foram confirmadas e validadas nos pacientes índice e em seus familiares; a última dando suporte adicional a dados da literatura como não envolvida em casos com CT não papilífero hereditário e CM. Este estudo revelou genes candidatos com potencial para contribuir para o diagnóstico molecular e aconselhamento genético destes pacientes e suas famílias.

ABSTRACT

The identification of predisposing genes and characterization of genetic risk factors for cancer development are essential for the diagnosis, prognosis and genetic counseling of affected patients and their families. A significant number of patients with breast (BC) and papillary thyroid carcinomas (TC) and family history of these tumors suggest the involvement a hereditary cancer predisposition gene. The aim of this study was to evaluate new predisposing genes related to BC and TC by whole exome sequencing, in patients who are affected by at least one of these tumors and have cancer family history. Six patients with CM, 12 with CT and six with both tumors were included in the study, all of them with family history of these tumors. DNA from peripheral blood sample was sequenced by HiSeq 1000 platform (Illumina). After the bioinformatics analysis, variations in high and low-penetrance genes were identified; most of them were considered common according to 1000 Genomes and 6500 Exomes. Furthermore, rare variations were described in these genes, which were considered pathogenic in at least three databases and observed in up to two patients. Patient 14 presented the p.R233X mutation in PTEN confirming the diagnosis of Cowden Syndrome. Other genes were investigated in order to characterize new variations related to the phenotype studied. This analysis resulted in 911 SNVs (single nucleotide variations), 287 non-synonymous alterations and 50 rare pathogenic variations (detected in at least three databases). From these, 13 alterations were mapped in genes that interact with high and low-penetrance genes previously described in literature. Alterations were also identified in genes recently described as associated with hereditary PTC and BC (HABP2, PARP4 and SRRM2). Enrichment analysis revealed compromised functions in the Thyroid hormone Synthesis and RAP1 Signaling pathways, involving deleterious alterations in 10 and 14 patients, respectively. Specifically, the thyroid hormone synthesis pathway may be potentially deregulated in individuals from Families I and II through PLCB3 and PAX8 alterations, respectively. Promising SNP candidates included the rs35169799 (PLCB3), one of the most frequent alterations, those classified as genome maintenance (rs61997220: ZC3H12D, rs61902276: DDIAS, rs61754785: MUS81, rs62619782: RNASEH2A), and mapped in potentially deregulated pathways (rs3188996: PAX8). Confirmation and validation analysis were performed for two variants rs61754785 (MUS81) and rs7080536 (HABP2) in 7 and 9 index patients, and also in their family members; the last one gives an additional support for literature data as not involved in hereditary PTC and BC. Furthermore, the data generated in this study could be incorporated into the general knowledge of cancer genetics, allowing the development of prevention strategies, screening and genetic counseling in these families.

LISTA DE FIGURAS

Figura 1	Esquema representativo da etapa de clusterização realizada no equipamento cBot	41
Figura 2	Esquema representativo da etapa de sequenciamento.	43
Figura 3	Fluxograma de análises de bioinformática, ilustrando as etapas básicas das análises.	44
Figura 4	Classificação dos 46 genes com alterações deletérias e que possuem interação proteica.	66
Figura 5	Rede de interação entre os genes identificados com alterações deletérias e suas conexões com genes de alta e baixa penetrância descritos em síndromes hereditárias de predisposição ao câncer.	67
Figura 6	Via de sinalização Rap1.	71
Figura 7	Via de síntese de hormônios da tireoide.	72
Figura 8	Heredogramas das Famílias I e II cujas alterações estão descritas na Tabela 12.	74
Figura 9	Família do paciente 7 (IV.11), portador da alteração homozigota. GG: Homozigoto para o alelo normal; GA: Heterozigoto; AA: Homozigoto para a alteração.	76
Figura 10	Confirmação e validação da alteração <i>HABP2</i> c.1601G>A (rs7080536). Eletroferogramas demonstrando a presença da mutação em homozigose e heterozigose nos membros da família do paciente 7 no gene <i>HABP2</i> (NM_004132:exon13).	77
Figura 11	Família da paciente 9, portadora heterozigota da alteração no gene <i>HABP2</i> e no <i>MUS81</i> .	78
Figura 12	Confirmação e validação da alteração <i>HABP2</i> c.1601G>A e <i>MUS81</i> c.1292G>A.	78
Figura 13	Eletroferogramas resultantes da análise de sequenciamento de Sanger para a confirmação da alteração rs61754785 no gene <i>MUS81</i> presente nos pacientes avaliados 6, 11, 20 e 22.	79

LISTA DE TABELAS

Tabela 1	Dados clínicos e história familiar dos pacientes que foram analisados por sequenciamento do exoma.	33
Tabela 2	Genes de alta e baixa penetrância investigados em detalhes entre os 24 pacientes avaliados por NGS.	50
Tabela 3	Sequência de iniciadores que flanqueiam o SNP rs7080536 mapeado no gene <i>HABP2</i> (Gara et al, 2015) e rs61754785 mapeado no gene <i>MUS81</i> .	52
Tabela 4	Resumo dos resultados de limpeza, mapeamento dos <i>reads</i> e número de variantes detectadas após o sequenciamento do exoma dos pacientes com carcinomas de mama e tireoide.	55
Tabela 5	Relação dos SNPs mapeados em genes de alta penetrância entre os 24 casos avaliados.	57
Tabela 6	Alterações mapeadas em genes de baixa penetrância previamente descritas como associadas com CM e CT.	58
Tabela 7	Alterações mapeadas em genes clássicos que são preditas como patogênicas pela análise <i>in silico</i> , porém com associação fenotípica desconhecida de acordo com os bancos de dados CLINVAR, LOVD e BIC.	60
Tabela 8	Detalhamento das 50 alterações classificadas como patogênicas em pelo menos três dos nove bancos de dados investigados.	62
Tabela 9	Alterações encontradas nos genes recentemente descritos como associados aos CT e CM hereditários.	64
Tabela 10	Descrição dos 13 genes identificados com alterações deletérias, sua distribuição entre as amostras avaliadas (três membros da família I, dois membros da família II e os demais indivíduos não aparentados) e as interações com os genes associados com as síndromes de predisposição hereditária ao câncer.	68
Tabela 11	Duas das 99 vias enriquecidas por PathDIP e IPA com potencial significado em carcinomas de mama e tireoide.	69
Tabela 12	Genes e variantes genéticas detectadas nas famílias I e II e sua respectiva função.	72

LISTA DE SIGLAS E ABREVIATURAS

AR	Receptor de Andr�geno
ASCO	<i>American Society Of Clinical Oncology</i>
CCR	C�ncer Colorretal
Cld-4	Claudina 4
CM	C�ncer de Mama
CNVs	<i>Copy Number Variation</i>
CPT	Carcinoma Papil�fero da Tire�ide
CT	C�ncer de Tireoide
DNA	�cido Desoxirribonucl�ico
EE1	<i>Enrichment Elution Buffer 1</i>
EEX	<i>Expanded Exome Oligos</i>
EGA	<i>European Genome–Phenome Archive</i>
EHB	<i>Enrichment Hybridization Buffer</i>
ER	Receptor de Estr�geno
ERK	<i>Extracellular Signal-Regulated Kinases</i>
ET2	<i>Elute Target Buffer 2</i>
EtOH	Etanol
EWS	<i>Enrichment Wash Solution</i>
F	Feminino
FAP	Polipose Adenomatosa Familiar
FEN1	<i>Flap Structure-Specific Endonuclease</i>
FI	Fam�lia 1
FII	Fam�lia 2
GEO	<i>Gene Expression Omnibus</i>
GTP	Trifosfato de Guanosina
GWAs	<i>Genome Wide Association Studies</i>
HBOC	<i>Hereditary Breast and Ovary Carcinomas</i>
HER2	Receptor do Fator de Crescimento Epid�rmico
HT	Hormonioterapia
IPA	<i>Ingenuity Pathway Analysis</i>
K5/14/17	Queratinas 5/14/17
HPV	<i>Human Papiloma Virus</i>
HBV	<i>Hepatitis B Virus</i>
HCV	<i>Hepatitis C Virus</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
Ki67	Marcador de Prolifera��o Celular
LOH	<i>Loss Of Heterozygosity</i>
M	Masculino
MAF	<i>Minor Allele Frequency</i>
MEN	Neoplasia End�crina M�ltipla
MEN1	<i>Multiple Endocrine Neoplasia</i>
MEN2A	Neoplasias End�crinas M�ltiplas do Tipo 2A
MMR	<i>Mismatch repair genes</i>
MT	Carcinoma de Mama e Tireoide
NaOH	Hidr�xido de S�dio
NCCN	<i>National Comprehensive Cancer Network</i>

ND	Nada Consta
NEM	<i>Nextera Enrichment Amplification Mix</i>
NIH	<i>National Institute of Health</i>
PCR	<i>Polymerase Chain Reaction</i>
PLC	Enzima Fosfolipase
PPC	<i>PCR Primer Cocktail</i>
PR	Receptor de Progesterona
qPCR	<i>Quantitative Polymerase Chain Reaction</i>
QT	Quimioterapia
RIT	Radioiodoterapia
RNA	Ácido Ribonucléico
RSB	<i>Resuspension Buffer</i>
RT	Radioterapia
SAV	<i>Software Analysis Viewer</i>
SEER	<i>Surveillance, Epidemiology, and End Results</i>
SET	Sequenciamento do Exoma Total
SMB	<i>Streptavidin Magnetic Beads</i>
SNC	Sistema Nervoso Central
SNP	<i>Single Nucleotide Polymorphism</i>
SNV	Sistema Nervoso Central
SPB	<i>Sample Purification Beads</i>
SPC	Síndromes de Predisposição ao Câncer
TCGA	<i>The Cancer Genome Atlas</i>
TU	Tumor
UTRs	<i>Untranslated Regions</i>
VCF	<i>Variant Call Format</i>
VDR	Receptor de Vitamina D

ÍNDICE

1.	INTRODUÇÃO	11
1.1	Síndromes de predisposição ao câncer (SPC)	11
1.2	Sequenciamento de exoma total (SET)	13
1.3	Câncer de mama (CM)	16
1.4	Câncer de tireoide (CT)	11
1.5	Associação entre o câncer de mama e tireoide	24
2.	JUSTIFICATIVA	29
3.	OBJETIVOS	30
3.1	Objetivo geral	30
3.2	Objetivos específicos	30
4.	MATERIAL E MÉTODOS	31
4.1	Pacientes.....	31
4.2	Extração do DNA genômico	36
4.3	Construção das bibliotecas de exoma	36
4.4	Pré-sequenciamento e sequenciamento do exoma	40
4.5	Análise de bioinformática	43
4.6	Estratificação e seleção das variantes	50
4.7	Confirmação das alterações	52
5.	RESULTADOS	54
5.1	Análise dos dados brutos	54
5.2	Genes de alta e baixa penetrância associados ao CT e CM hereditários	57
5.3	Alterações compartilhadas entre os casos	62
5.4	Investigação de genes recentemente descritos como associados com CT hereditários ou com CM/CT	64
5.5	Análise de interação proteica	65
5.6	Análise de enriquecimento	70
5.7	Análise das famílias I e II	73
5.8	Confirmação e Validação	75
6.	DISCUSSÃO	80
7.	CONCLUSÃO	107
8.	REFERÊNCIAS	109
9.	SUPLEMENTOS	131

1. INTRODUÇÃO

1.1 Síndromes de Predisposição ao Câncer (SPC)

O câncer, por se tratar de uma doença com alto índice de mortalidade, é considerado um dos maiores problemas de saúde pública do mundo. A ocorrência mundial de novos casos de câncer estimada em 2012 foi de 14,1 milhões, sendo detectados 8,2 milhões de óbitos devido à doença. Segundo os dados do Globocan (2012), estima-se cerca de 20 milhões de novos casos até 2020 em todo o mundo (Torre *et al.*, 2015).

A maioria dos cânceres que acomete a população mundial é considerada esporádica, tendo como fatores de risco a exposição ambiental e fatores genéticos. Porém, ainda há uma parcela da população que possui um alto risco de desenvolver tumores no decorrer da vida, associado com fatores genéticos de predisposição herdáveis que, em conjunto, caracterizam as Síndromes de Predisposição ao Câncer (Nagy *et al.*, 2004). Os tumores envolvidos nas SPC podem diferir significativamente em relação aos mecanismos moleculares quando comparados com os mesmos tipos tumorais esporádicos, indicando que as SPC devem receber uma atenção especial até mesmo quanto à estratégia de tratamento (Imyanitov and Byrski, 2013).

O aparecimento de tumores em idade jovem, sincrônicos ou metacrônicos, bilaterais ou multifocais, afetando vários membros da mesma família são características que, comumente, fazem parte do espectro das SPC. Contudo, estas características variam de acordo com os critérios para cada síndrome. A orientação do *National Comprehensive Cancer Network* (NCCN) e da *American Society of Clinical Oncology* (ASCO) inclui os critérios clínicos adotados para as SPC. Quando há um gene de predisposição associado com uma determinada SPC, é necessária a busca de características que auxiliem na identificação dos pacientes acometidos incluindo os dados da histologia do tumor, expressão de proteínas específicas, análise de marcadores moleculares, penetrância do gene em questão bem como abordagens de prevenção e sobrevida (Garber and Offit, 2005).

Agrupamentos familiares ou síndromes de predisposição podem evidenciar tumores hereditários. O primeiro caso é consequência de interações envolvendo genes de baixa ou moderada penetrância, e/ou interação gênica e ambiental cujo risco é

classificado como baixo ou moderado. As síndromes de predisposição hereditária ao câncer são resultado de mutações germinativas em genes de alta penetrância com risco aumentado de desenvolvimento da doença (Nagy *et al.*, 2004; Hampel *et al.*, 2004). Os genes associados às síndromes de predisposição hereditária ao câncer pertencem frequentemente ao grupo de genes codificadores e classificados como supressores tumorais que, quando alterados, não desempenham adequadamente seu papel regulador do crescimento celular ou do mecanismo de reparo a danos no DNA (Davies, 2007).

Apesar do perfil de hereditariedade esperado, é comum para as SPC que uma parcela de indivíduos com diagnóstico clínico de determinada síndrome não possua mutações em genes clássicos conhecidos como associados com este fenótipo. Um exemplo conhecido é o descrito em casos de pacientes com Câncer Colorretal (CCR) familiar que não apresentam alterações nos genes de reparo a erros de pareamento (MMR – *mismatch repair genes*). Com o avanço das tecnologias de sequenciamento, foram descritas novas alterações de predisposição ao carcinoma colorretal (Gylfe *et al.*, 2013; Chubb *et al.*, 2015). Gylfe *et al.* (2013) analisaram o exoma de 96 indivíduos e identificaram 11 genes incluindo 14 alterações, sendo que cinco indivíduos apresentavam apenas uma alteração, oito indivíduos apresentavam duas alterações e apenas um caso apresentava três alterações simultaneamente. Os pacientes não apresentavam mutações nos genes comumente descritos no CCR hereditário, incluindo *APC* e *MUTYH*. Os autores sugeriram que as variantes detectadas poderiam estar associadas com o aparecimento da doença. As alterações foram validadas por sequenciamento convencional e pela avaliação de segregação entre indivíduos da mesma família (Gylfe *et al.*, 2013). Estudos semelhantes foram descritos para casos com câncer de mama e ovário hereditários sem mutações nos genes *BRCA1* e *BRCA2* (Gracia-Aznarez *et al.*, 2013; Cybulski *et al.*, 2015) e câncer gástrico difuso (Donner *et al.*, 2015).

Os agrupamentos familiares de cânceres se tornaram um aspecto importante na oncologia clínica devido a implementação bem sucedida de métodos de triagem e testes genéticos para muitas síndromes. A identificação de genes de predisposição e a caracterização de fatores de risco genético são fundamentais para o diagnóstico, prognóstico e aconselhamento genético de pacientes afetados e seus familiares.

1.2 Sequenciamento de Exoma Total (SET)

Os estudos genômicos do câncer vêm evoluindo rapidamente, principalmente com o aparecimento das técnicas de sequenciamento de alto desempenho. Estudos que focam em síndromes de predisposição ao câncer são beneficiados pela técnica de sequenciamento do exoma total (SET) (Backes *et al.*, 2015; Cascón *et al.*, 2015; Esteban-Jurado *et al.*, 2015; Gara *et al.*, 2015). Esta técnica consiste na captura de sequências codificadoras de proteínas – éxons – por meio de hibridação (Bamshad *et al.*, 2011). As regiões codificadoras de proteínas correspondem de 1 a 2% do genoma e, na maioria das doenças de caráter monogênico já descritas até o momento, as alterações de predisposição à doenças comumente estão localizadas nestas regiões (Majewski *et al.*, 2011; Stranneheim and Wedell *et al.*, 2016).

O processo de análise de dados provenientes de SET em estudos exploratórios passa por algumas etapas clássicas. Geralmente a primeira etapa é a aplicação do *discrete filtering*, que consiste na comparação entre as sequências de exoma geradas e um grupo controle, podendo este ser constituído por bancos de dados públicos e/ou sequências de indivíduos normais, não afetados pela doença. Este filtro tem como objetivo eliminar variantes que possivelmente não são causais, uma vez que a casuística estudada não está representada no grupo controle, reduzindo ao mínimo o número de genes candidatos que possam efetivamente estar associados ao fenótipo de estudo (Bamshad *et al.*, 2011).

Geralmente as alterações de base única identificadas são classificadas conforme um gradiente de maior para menor relevância a seguir: mutações sem sentido (*nonsense*), que resultam em códon de parada; mutações que alteram quadro de leitura; mutações localizadas em sítios de *splicing*; mutações de sentido trocado, que alteram o aminoácido; e por último as mutações silenciosas, que não alteram o aminoácido. Com SET é possível também identificar pequenas inserções e deleções, bem como variação do número de cópias genômicas, denominadas CNVs (do inglês, *Copy Number Variation*). A classificação das alterações também pode ser feita de acordo com a função biológica, quando sua ação molecular é considerada, a qual permite a identificação de redes de interação gênica e proteica em que tais variantes estão inseridas (Ku *et al.*, 2012).

Estas análises resultam comumente numa vasta quantidade de alterações candidatas. No caso de doenças hereditárias, a seleção da variação patogênica de interesse também deve ser norteadas pelas informações obtidas a partir de heredogramas e histórico familiar, levando em conta número de gerações afetadas e grau de parentesco entre os indivíduos afetados de uma mesma família, ou até mesmo fenótipos compartilhados entre família diferentes. Segundo Gilissen *et al.* (2012) devem ser consideradas algumas estratégias para estas análises: (1) Análise de ligação: pode ser realizada quando acredita-se que a doença é monogênica e vários membros da mesma família são avaliados. À medida que um indivíduo com grau de parentesco mais distante é incorporado à análise, o número de mutações compartilhadas diminui, auxiliando na redução dos genes candidatos para posterior análise. (2) Análise de homozigose: no caso de suspeita de doença recessiva com história de consanguinidade e quando apenas um indivíduo acometido é avaliado, as regiões que possuem alterações homozigotas devem ser priorizadas; (3) Análise de double-hit: a qual consiste na busca de alterações em homozigose contidas no DNA de sangue periférico de apenas um indivíduo com suspeita de doença recessiva sem o histórico de consanguinidade. Esta estratégia é um pouco mais difícil de ser aplicada pela falta de dados; (4) Análise de sobreposição: é realizada quando são avaliados indivíduos não aparentados que compartilham o mesmo fenótipo. Neste caso suspeita-se que a doença seja autossômica dominante, com base na investigação dos heredogramas das famílias; (5) Estratégia de novo: avalia-se o probando e seus pais com o intuito de identificar uma nova mutação na geração F₁ (Gilissen *et al.*, 2012).

Um dos primeiros estudos realizados envolvendo identificação de genes de predisposição em síndromes hereditárias foi realizado por Ng *et al.*, em 2010. Utilizando a estratégia de sobreposição, os autores avaliaram 10 indivíduos com a síndrome de Kabuki por sequenciamento de exoma e posteriormente 43 indivíduos passaram por sequenciamento direto na etapa de validação. Esta síndrome é caracterizada por múltiplas malformações faciais, anomalias cardíacas, anormalidades ósseas, sistema imunológico deficiente e retardo mental moderado. As sequências obtidas passaram por uma série de análises dentre a estratificação das alterações, quanto à sua natureza e função, e comparação entre pacientes que compartilhavam fenótipos semelhantes. Ao final, foi demonstrado que alterações no gene *MLL2* estavam presentes na maioria dos pacientes. Aproximadamente 60% dos casos apresentavam mutações

sem sentido e *indels* no gene *MLL2*, sugerindo fortemente que alterações neste gene são a principal causa da síndrome de Kabuki. Os autores ressaltam a necessidade do desenvolvimento de estratégias de análises para a melhor classificação dos dados, tanto em nível fenotípico quanto genotípico, de modo que as tecnologias de sequenciamento de genoma e exoma total possam ser mais bem exploradas (Ng *et al.*, 2010).

Foram investigadas 34 famílias com melanoma cutâneo envolvendo 87 indivíduos e uma nova alteração no gene *POLE* foi identificada em três indivíduos da mesma família por sequenciamento de genoma e exoma. A alteração c.1041G>T, responsável por uma mutação *missense* e troca de aminoácido p.Trp347Cys ocorre no domínio exonuclease da subunidade maior da proteína (Aoude *et al.*, 2015). Este gene codifica uma subunidade catalítica da polimerase ϵ que atua no reparo do DNA e duplicação dos cromossomos. Outra alteração c.1270C>G que também causa uma troca de aminoácido p.Leu424Val, mapeada também no gene *POLE* foi relatada em pacientes com adenomas e carcinomas colorretais, por meio de sequenciamento de exoma (Palles *et al.*, 2013). Interessantemente, Palles *et al.* (2013) encontraram uma série de bases incorporadas erroneamente nos tumores nos pacientes avaliados, o que comprova o efeito do mau funcionamento do gene em sua atividade de conferência e excisão de bases, o que foi comprovado por ensaio funcional em *C.elegans*. Estes achados evidenciam que diferentes alterações, mesmo que localizadas no mesmo gene, podem conferir risco ao desenvolvimento de diferentes tipos tumorais.

Analizando o exoma de sete famílias com histórico de câncer de mama, sem a presença de outros tipos de tumores e sem mutações nos genes *BRCA1* e *BRCA2*, Gracia-Aznarez *et al.* (2012) identificaram 67 SNPs e 14 *indels* como potenciais alvos a predisposição para o CM. Estas variantes foram divididas em um grupo que possuía principalmente alterações que resultavam em proteína truncada ou que compartilhavam alterações em um maior número de membros da família, e outro grupo com variantes consideradas raras. No primeiro grupo foi relatada uma alteração já conhecida no gene *CHEK2* (c.1100delC), não tendo sido observado outros genes de alta penetrância. Para o segundo grupo foram identificados sete variantes em genes associados à angiogênese, progressão do ciclo celular, adesão celular e outras funções associadas ao desenvolvimento tumoral (*MAPKAP1*, *TNFSF8*, *PTPRF*, *UBA3*, *TIMP3*, *CNTROB*, *S1PR3*), porém os autores não puderam concluir qual é o papel destas alterações na predisposição ao CM familiar (Gracia-Aznarez *et al.*, 2013).

O SET tem sido utilizado em estudos exploratórios com sucesso, na busca de novos alelos de predisposição. Na prática clínica tem sido aplicado na realização de testes para auxiliar no diagnóstico de pacientes com suspeita de doença Mendeliana rara (Yang *et al.*, 2013). A utilização de sequenciamento de exoma pode contribuir para o aconselhamento genético dos pacientes e suas famílias, contudo deve-se ter cautela quando são identificadas variantes de significado incerto ou achados acidentais, que podem dar indícios de outras doenças. Uma equipe multidisciplinar envolvida nas análises, interpretação dos resultados e no aconselhamento genético tem potencial para trazer benefícios para o paciente e suas famílias em tempo real (Stranneheim and Wedell, 2016).

Em resumo, é evidente que a metodologia de sequenciamento de exomas permite identificar mutações e variações no genoma, as quais podem estar associadas com as síndromes de predisposição hereditária ao risco de desenvolvimento de tumores, além de contribuir com a prática clínica na forma de painéis de diagnóstico e úteis no aconselhamento genético.

1.3 Câncer de Mama (CM)

O CM é o segundo tumor mais incidente no mundo, correspondendo a 11,9% dos casos, e o primeiro mais incidente em mulheres, contudo é o quinto responsável pelas causas de morte pela doença, devido aos avanços em seu tratamento. Nos Estados Unidos, é o mais comum entre as mulheres, excluindo os tumores de pele não-melanoma, ocorrendo aproximadamente em um a cada três casos de câncer no país (DeSantis *et al.*, 2016). Segundo a última estimativa do GLOBOCAN (2012), o CM foi estimado em 25% dos casos de câncer que acometem a o sexo feminino (Ferlay *et al.*, 2014). Foram estimados para o ano de 2012 mais de 52 mil novos casos de CM no Brasil, sendo o mais incidente na população feminina, exceto na região norte onde ocupa o segundo lugar (INCA, 2014).

Vários fatores de risco têm sido associados com o CM incluindo ganho de peso após os 18 anos de idade, obesidade e sobrepeso em qualquer idade, o uso combinado de estrógeno e progesterona no tratamento da menopausa, sedentarismo e inatividade física, consumo de álcool, tabagismo iniciado antes do primeiro filho, padrão de sono irregular, diabetes tipo 2, presença de tecido mamário denso e maior quantidade de ciclos

menstruais (associada à menarca precoce e/ou menopausa tardia), uso de contraceptivos orais (como a exposição cumulativa a estrógenos e/ou progesterona no epitélio mamário), nuliparidade ou primeiro filho após os 30 anos de idade, dieta alimentar pouco saudável (baseada em gorduras e proteínas de origem animal), baixa ingestão de frutas, vegetais e cereais, além da presença de mutações em genes de predisposição hereditária (*American Cancer Society*, 2015).

Os carcinomas de mama mais comuns que acometem a população feminina são os ductais, correspondendo de 70 a 80% dos casos. Estes são classificados como *in situ*, os quais não ultrapassam as primeiras camadas celulares do ducto, podendo ser classificados quanto ao aspecto histológico como sólido, cribiforme, papilar e micropapilar podendo também ser de grau tumoral alto, intermediário ou baixo (Burstein and Polyak, 2004; Ponti *et al.*, 2014; Yerushalmi *et al.*, 2009). A recidiva dos carcinomas *in situ* na forma de carcinoma ductal invasivo é de 10% a 24% após 10 anos, e esta manifestação de recorrência pode ser aumentada quando o tumor é HER2+, atingindo até 50% dos casos (Ponti *et al.*, 2014). Os carcinomas invasivos são mais incidentes quando comparados aos carcinomas *in situ* (DeSantis *et al.*, 2016). Contudo essa progressão tumoral linear tem sido discutida e infere-se que seria possível o desenvolvimento de um tumor invasivo já nos primeiros estágios da doença (Virnig *et al.*, 2009). O CM lobular acomete a porção lobular da glândula mamária e, geralmente, possui baixa taxa de proliferação. Este tipo tumoral é pouco frequente, sendo detectado de 0,5% a 3,8% das biópsias benignas de mama. O subtipo lobular pleomórfico, considerado mais agressivo e com alta taxa de proliferação, é relatado em 8% a 14% dos carcinomas invasivos de mama (Al-Baimani *et al.*, 2015; Morrow *et al.*, 2015).

Muitos estudos de integração de dados provenientes de diferentes plataformas de análises e envolvendo um grande número de amostras, têm proporcionado a base para a classificação molecular dos tumores da mama. Estes estudos tem contribuído para o delineamento de uma nova abordagem para o diagnóstico clínico e tratamentos mais adequados das pacientes acometidas por câncer de mama. Os primeiros estudos evidenciaram subtipos moleculares os quais foram detectados por meio de análises de dados globais de expressão gênica e de alteração do número de cópias genômicas. Entretanto, na prática clínica, os testes de expressão proteica por imuno-histoquímica mostraram-se apropriados para a classificação dos subtipos tumorais (Goldhirsch *et al.*, 2013). Os principais marcadores utilizados na prática clínica são os receptores de

estrógeno (ER), progesterona (PR), receptor de fator de crescimento epidérmico (*HER2*) e o marcador de proliferação Ki-67. Os principais subtipos descritos como resultantes desses estudos moleculares são: Luminal A, Luminal B-*HER2* negativo, Luminal B-*HER2* positivo, *HER2* positivo não-luminal e *basal-like* ou também conhecido como triplo negativo (Dawson *et al.*, 2013; Sonnenblick *et al.*, 2014). Contudo, o CM é uma doença com alta heterogeneidade e com uma grande variedade em suas características clínicas e moleculares. Dawson *et al.* (2013) descreveram a existência de 10 subtipos tumorais de CM associados com evolução clínica distintas, reforçando que estes achados proporcionam conhecimento para entender os mecanismos celulares envolvidos na carcinogênese e na identificação de potenciais genes *drivers*. Além disso, a heterogeneidade intra-tumoral é uma questão que também tem sido levantada, pois a combinação de diferentes agentes pode contribuir para o tratamento dos diferentes alvos que são passíveis da ação de medicamentos utilizados hoje na prática clínica (Zardavas *et al.*, 2015).

Mais recentemente, Santagata *et al.* (2014) propuseram outra classificação dos tumores de mama considerando as características histológicas de 11 tipos celulares e quatro tipos de combinações de receptores hormonais. Esta classificação foi realizada com base em células luminais normais e a metodologia foi baseada em marcadores de imuno-histoquímica como ER, AR (receptor de andrógeno), VDR (receptor de vitamina D), K5/14/17 (queratinas 5/14/17), Ki67, Cld-4 (claudina 4), K7/8/18 (queratinas 7/8/18) e CD10/SMA/p63 (marcadores de célula tronco, mioepitelial e supressor tumoral). Na análise da camada luminal do lóbulo de 3157 tumores de mama utilizando estes marcadores foi possível identificar 11 padrões de diferenciação, que vão de L1 a L11. Além disso, foi possível combinar a esta classificação o número de receptores hormonais que estão presentes nos tecidos, podendo ser H0 (negativos para os receptores), H2 (pelo menos dois receptores positivos) e H3 (com a presença de 3 receptores). Também foi possível classificar as camadas mioepiteliais, podendo estas ser divididas em My1 (K5/14/17-) e My2 (K5/14/17+). Esta classificação também permitiu a identificação das pacientes com melhor prognóstico, como por exemplo, as pacientes classificadas como HR3, as quais são positivas para todos os receptores hormonais ER+AR+VDR+ (Santagata *et al.*, 2014).

Aproximadamente 10% dos casos de CM são familiares com uma proporção significativa de tumores hereditários cujos pacientes apresentam mutações germinativas

em genes altamente penetrantes (Paradiso and Formenti, 2011). Em casos familiares, quanto maior o número de indivíduos afetados, maior é o risco de desenvolver a doença, podendo ser de 2 a 4 vezes maior quando o grau de parentesco é menor (Majdak-Paredes and Fatah, 2009).

A maioria dos carcinomas de mama hereditários são ductais invasivos (80% dos casos) e, comumente, estão associados com mutação nos genes *BRCA1* e *BRCA2*. Aproximadamente 70% e 16% dos tumores *BRCA1* e *BRCA2* positivos, respectivamente, são negativos para os receptores hormonais ER e PR e possuem baixa expressão HER2, portanto triplo-negativos (Mavaddat *et al.*, 2012). Os tumores associados ao *BRCA1* são considerados mais agressivos por elevado número de mitoses, áreas de necrose, margens comprometidas e infiltração linfocítica, além de possuírem mutações adquiridas no gene *TP53*. Os carcinomas mamários associados ao *BRCA2* também possuem alta taxa mitótica e margens comprometidas (Larsen *et al.*, 2014). Já para os CM hereditários que não possuem alteração nos genes clássicos *BRCA1* e *BRCA2*, a heterogeneidade é maior, porém parecem ser menos agressivos quando comparados aos tumores esporádicos (Larsen *et al.*, 2014). Alterações localizadas em diferentes locais dos genes *BRCA1* e *BRCA2* podem contribuir com diferentes riscos para o desenvolvimento de tumores de mama e de ovário (Rebbeck *et al.*, 2015).

A síndrome HBOC (do inglês, *Hereditary Breast and Ovary Carcinomas*) é mais frequentemente descrita. Durante a avaliação do paciente podem surgir evidências para o diagnóstico da síndrome HBOC, como: dois ou mais casos de câncer de mama e/ou ovário em idade precoce (pré-menopausa), múltiplos casos na família, câncer de mama bilateral e padrão de transmissão vertical. Os genes *BRCA1* e *BRCA2* foram os primeiros genes de predisposição hereditária descritos na síndrome HBOC (Miki *et al.*, 1994; Tavtigian *et al.*, 1996). Desde então, testes genéticos para a triagem de mutações germinativas têm sido utilizados na identificação destes genes, em famílias clinicamente caracterizadas como portadoras da síndrome. Além de *BRCA1* e *BRCA2*, outros genes considerados de alta penetrância já foram relatados como associados a predisposição ao CM hereditário, como *PTEN* (síndrome de Cowden), *TP53* (síndrome de Li-Fraumeni), *CDH1* (câncer gástrico difuso hereditário) e *STK11* (síndrome de Peutz-Jeghers). Também foram relatados genes com penetrância moderada associados ao CM hereditário, tais como *CHEK2*, *ATM*, *BRIP1* e *PALB2* (Narod, 2004; Zhang *et al.*, 2011).

Em uma meta análise, foram descritos 10 alelos consideradas de alto risco para câncer de mama hereditário em seis genes, incluindo *ATM*, *CASP8*, *CHEK2*, *CTLA4*, *NBN* e *TP53* (Zhang *et al.*, 2011). Outras alterações envolvendo os genes *ATM*, *CYP19A1*, *TERT* e *XRCC3* foram consideradas como de risco moderado (Zhang *et al.*, 2011). Recentemente, Shiovitz e Korde (2014) relataram um maior número de alelos de risco para câncer de mama em estudos de GWA (*Genome Wide Association*) envolvendo genes de baixa penetrância, mesmo sendo frequentes na população em geral. Apesar da grande quantidade de dados, o avanço das tecnologias de sequenciamento tem hoje proporcionado a elaboração de painéis comerciais contendo genes de penetrância alta, moderada e baixa, relacionados ao diagnóstico e úteis para o aconselhamento genético das pacientes acometidas por CM familiar. A maioria dos testes comerciais atuais para avaliar mutação em genes de predisposição hereditária ao câncer de mama inclui também genes associados com o câncer de ovário e colorretal hereditários (Easton *et al.*, 2015; Shiovitz and Korde, 2015). Um dos benefícios da aplicação destes painéis é o auxílio na detecção de pacientes portadores de mutações extremamente raras, bem como polimorfismos comuns, além de contribuírem para o diagnóstico clínico de HBOC não tão evidente, como no caso de pacientes com poucos indivíduos acometidos na família. A realização destes testes auxilia ainda na confirmação do diagnóstico de uma síndrome específica e de casos suspeitos, mesmo quando associados com outro espectro de tumores (Pinheiro *et al.*, 2015; Shiovitz and Korde, 2015). Os painéis de análise de mutações em genes específicos podem auxiliar na identificação de vários *loci* que estão associados ao desenvolvimento da doença em questão, ao invés de apenas uma alteração específica. Contudo a sua interpretação é difícil, sendo necessária a busca dos achados em bancos de dados específicos, investigação fenotípica e até mesmo estudos funcionais para comprovar de fato a causa das mutações reportadas nos testes (Biesecker and Green, 2014).

Mutações germinativas em *TP53* estão associadas à síndrome Li-Fraumeni, uma síndrome complexa, autossômica dominante caracterizada por agrupamento familiar de sarcomas, cânceres de mama, tumores cerebrais e carcinomas adrenocorticais, diagnosticados antes dos 45 anos de idade (Li and Fraumeni, 1969; Malkin *et al.*, 1990). Algumas famílias apresentam incidência aumentada de outros tipos de cânceres como de pulmão, cólon, estômago, próstata e melanomas. Apesar de tumores da tireoide não serem indicadores diagnósticos, estes já foram descritos em famílias com a síndrome

(Li and Frameni, 1969; Achatz *et al.*, 2007; Masciari *et al.*, 2008). Uma variante do gene *TP53* responsável pela troca de aminoácido R337H foi detectada em uma coorte de pacientes brasileiros, nas regiões sul e sudeste do país estando associada com os tumores de mama, sistema nervoso central, sarcoma de partes moles, pulmão, estômago e tumores adrenocorticais, sendo este último mais evidente em pacientes com esta mutação específica quando comparado àqueles com outras mutações no gene *TP53* (Achatz *et al.*, 2007).

A identificação de mutações que ocorrem em genes de baixa penetrância tem permitido avanços em todas as áreas de conhecimento da genética básica e aplicada. Isto porque estas mutações e/ou polimorfismos podem desempenhar funções biológicas diretamente associadas ao fenótipo de estudo, e/ou por demonstrarem forte associação com determinadas doenças como o câncer, auxiliando na predição do risco e prevenção de pacientes suscetíveis. A identificação de um acúmulo de alterações pode potencializar a predição de risco quando combinamos os dados de diversos *loci*, mesmo que o efeito das alterações seja considerado moderado.

1.4 Câncer de Tireoide (CT)

O CT acometeu 298 mil pessoas em todo o mundo no ano de 2012 de acordo com a estimativa mundial Globocan em 2012, sendo mais comum em e correspondendo a 3,5% dos cânceres que acometem o sexo feminino (Ferlay *et al.*, 2014). Apesar de sua taxa de mortalidade ser considerada baixa (< 1:100.000 e 2: 100.000 para homens e mulheres, respectivamente), tem se notado um aumento da incidência da doença em função da utilização de melhores estratégias de rastreamento e detecção e, conseqüentemente, do diagnóstico das lesões (Ferlay *et al.*, 2014). No Brasil, para o ano de 2014, o CT foi estimado em 1.150 casos novos para o sexo masculino e 8.050 para o sexo feminino, sendo o quarto mais frequente em mulheres da região sul do país (INCA, 2014).

O CT é o mais comum dentre os tumores que atingem o sistema endócrino. Muitos fatores de risco estão relacionados à doença como aumento de exposição à radiação nuclear, mesmo em baixas doses; a ocorrência de câncer da tireoide na família; maior uso de diagnósticos por radiação, como exemplo o raio X e, maior exposição à tecnologia de comunicação do tipo *wireless* (Yu *et al.*, 2010; Nachalon *et al.*, 2015). Os

acidentes nucleares, envolvendo vazamento de material radioativo, são muitas vezes acompanhados de casos de carcinoma papilar de tireoide, como os ocorridos em Hiroshima e Nagasaki, em 1945, e em Chernobyl, em 1986. Este último demonstrou um aumento na incidência do carcinoma papilífero de tireoide em crianças expostas a isótopos radioativos de iodo (Kazakov *et al.*, 1992; Yip and Carty, 2015). Dietas deficientes em iodo estimulam a formação do bócio, como um mecanismo de compensação. Paralelamente, a falta do elemento está também relacionada ao aumento da incidência de carcinomas foliculares de tireoide (Kondo *et al.*, 2006; Zimmermann and Boelaert, 2015). O desenvolvimento de bócio coloide e nódulos benignos, bem como adenomas, é outro forte fator de risco para CT (Vecchia *et al.*, 2014).

A maioria dos carcinomas tem origem nas células foliculares, que revestem os colóides tireoidianos contendo os hormônios T3 e T4 enquanto outros têm origem em células C ou parafoliculares, adjacentes às células foliculares, que secretam calcitonina (LiVolsi, 2011). Os carcinomas de tireoide mais comuns são os diferenciados, provenientes de células foliculares e correspondem a mais de 90% dos casos. Dentre os carcinomas diferenciados, o subtipo mais comum é o carcinoma papilífero, com prevalência de cerca de 50% a 80%, enquanto que o subtipo folicular tem prevalência de cerca de 10% a 40% (INCA, 2014). O carcinoma medular originário de células parafoliculares ocorre em 2% a 3% dos casos e possui patogenicidade melhor estabelecida. O carcinoma indiferenciado anaplásico, apesar de originado de células foliculares, é considerado o CT mais agressivo, ocorrendo em 2 a 3% dos casos, com expectativa de sobrevida média de quatro meses e quase 100% de óbito (Nikiforov and Nikiforova, 2011; Xing, 2013).

As síndromes genéticas associadas ao CT são caracterizadas por múltiplos tumores heterogêneos, multicentricidade e bilateralidade, cujos pacientes compartilham o fato de apresentarem a doença em idade precoce. Os carcinomas medulares e não-medulares podem ser hereditários. Estima-se que 25% dos carcinomas medulares estão associados com a síndrome da Neoplasia Endócrina Múltipla (MEN) do tipo II, onde 85% dos casos são devidos a mutações em ponto no proto-oncogene *RET*, responsável por codificar um receptor de tirosina quinase (Nose, 2011).

Algumas SPC contém em seu espectro os carcinomas não-medulares da tireoide incluindo a Polipose Adenomatosa Familiar (FAP) ou Síndrome de Gardner, Complexo de Carney, Síndrome de Werner, Síndrome DICER1 e Síndrome de Cowden (Gara *et*

al., 2015). A FAP está associada ao desenvolvimento de vários pólipos adenomatosos na parede do trato intestinal e porção colorretal. É comum a manifestação subsequente de carcinomas colorretais, de sistema nervoso central, fígado, ducto biliar e tireoide. A síndrome de Gardner é uma variante de FAP quanto à manifestação clínica de tumores extracolônicos como osteomas e tumores de pele. Nesta síndrome são mais comumente diagnosticados os tumores de tecidos moles e desmóides (García and Knoers, 2009). A síndrome de Gardner é autossômica dominante e decorrente da mutação germinativa no gene *APC*, contudo manifestações atenuadas do fenótipo clínico estão relacionadas com alterações no gene *MUTYH* (Half *et al.*, 2009).

A síndrome de Complexo de Carney possui 100% de penetrância e está associada com pigmentação e lesão de pele e mucosa, mixomas (tumores cardíacos benignos) e tumores endócrinos como glândula pituitária, doença adrenal nodular pigmentada, tumores de células de Sertoli e de Leydig além de tumores da tireoide, principalmente foliculares (Nosé, 2010; Correa *et al.*, 2015). Mutações no gene *PRKAR1A* ocorrem em mais de 70% dos casos (Correa *et al.*, 2015).

A síndrome de Werner tem herança autossômica recessiva com mutação no gene *WRN* e está associada com o envelhecimento precoce, bem como com o aparecimento de CT (Goto *et al.*, 1996; Nose, 2011). A síndrome de DICER1 acomete indivíduos que possuem mutações no gene *DICER1*, uma ribonuclease responsável por clivar pequenas moléculas de RNA que darão origem aos miRNAs, importantes para regulação gênica pós-transcricional. Nesta síndrome são relatados o blastoma pleuropulmonar, nefroma cístico, tumores de células de Leydig e Sertoli, rabdomiossarcoma embrionário, tumores de Wilms e bócio multinodular, que muitas vezes correspondem a cistos e hiperplasia tireoideana (De Kock *et al.*, 2014).

A síndrome de Cowden está relacionada com mutações no gene supressor tumoral *PTEN*. Esta síndrome é caracterizada pelo aparecimento do bócio e adenoma folicular, com alto risco de manifestação de um tumor maligno da glândula além da presença de tumores mamários (Yeager *et al.*, 2007).

Algumas mutações de baixa penetrância foram recentemente descritas como associadas à CT familiar, como exemplo os polimorfismos de base única (SNPs) rs1801516 (G>A, Asp1853Asn), rs17879961 (T>C, Ile157Thr) e rs16941 (A>G, Glu1038Gly), localizados nos genes *ATM*, *CHEK2* e *BRCA1*, respectivamente (Wojcicka *et al.*, 2014). É interessante ressaltar que estes alelos já foram descritos como

associados à leucemia linfocítica e CM, e pacientes diagnosticados com CM possuem um risco aumentado para o desenvolvimento de CT papilífero (Wojcicka *et al.*, 2014). Outros cinco SNPs foram associados com o risco aumentando de desenvolvimento de CT em pacientes que possuem histórico familiar deste tumor: rs966423 (C), rs944289 (T), rs2439302 (G), rs116909374 (T) e rs965513 (A) (Liyanarachchi *et al.*, 2013).

Recentemente os genes *HABP2* (Gara *et al.*, 2015) e *SRRM2* (Tomsic *et al.*, 2015) foram reportados como associados aos carcinomas de tireoide não medulares hereditários. Recentemente, Tomsic *et al.* (2016) relataram que a mutação neste gene não se segregava com o carcinoma papilífero da tireoide em famílias com este fenótipo. Desta forma, o papel do *HABP2* em carcinomas de tireoide hereditários precisa ser melhor avaliado para comprovar esta associação.

1.5 Associação entre o câncer de mama e tireoide

Há muito se busca entender os mecanismos que estão envolvidos na associação de doenças da mama e tireoide (Hedley *et al.*, 1981). A principal SPC que envolve estes dois tumores é a síndrome de Cowden. Lloyd e Dennis (1963), pela primeira vez, relataram a história da família Cowden e os achados clínicos de múltiplos adenomas da tireoide e fibrose da mama seguido de manifestação maligna (Lloyd and Dennis, 1963). Em 1997, foi descrita pela primeira vez a presença de mutações no gene *PTEN* (Liaw *et al.*, 1997), revelando que este era um gene supressor tumoral importante. Outras características clínicas foram sendo incorporadas ao espectro da síndrome e hoje os critérios clínicos utilizados para diagnóstico dos pacientes são preconizados nas recomendações para a síndrome presentes no NCCN. Os critérios são divididos em principais e secundários, e dentre os principais estão: câncer de mama, câncer de endométrio, câncer folicular da tireoide, múltiplos pólipos hamatomatosos, macrocefalia, pigmentação macular da glândula, lesões mucocutâneas. Dentre os critérios secundários encaixam-se: espectro autista, câncer de cólon, acantose glicogênica do esôfago, lipomas, carcinoma papilífero da tireoide ou sua variante folicular, lesões tireoidianas, carcinoma renal, lipomatose testicular e anomalias vasculares (NCCN, 2015). São recomendados testes genéticos para o diagnóstico da síndrome de Cowden por sua importante aplicação no aconselhamento genético das famílias acometidas.

Outras associações entre lesões de mama e tireoide têm sido relatadas. Em 1997, Galanti *et al.* relataram um aumento da incidência de CM em membros das famílias, na maioria das vezes mães, de pacientes acometidos com CT papilífero e folicular, quando comparados à taxa de incidência destes tumores na população local (Galanti *et al.*, 1997). O desenvolvimento de um segundo tumor primário em pacientes diagnosticados com CT ocorre em 16% dos indivíduos, sendo o tumor mamário o segundo tumor primário mais frequente nesta parcela de indivíduos (Garner *et al.*, 2007). Nos casos de CM, o segundo tumor primário detectado com frequência é o CT (Garner *et al.*, 2007).

Apesar da alta frequência de pacientes que apresentam diagnóstico clínico para lesões de ambos os tecidos glandulares, ainda está por ser elucidado quais os principais mecanismos moleculares que de fato estão envolvidos nesta associação. É descrito em literatura que para pacientes que já desenvolveram CT, tratados com cirurgia e radioiodoterapia, é comum o aumento de incidência de um segundo tumor primário na mama. Consorti *et al.* (2011) acompanharam pacientes previamente tratadas para CT com o intuito de investigar a necessidade de uma política de prevenção para o CM. Contudo a relação patofisiológica entre CM e CT ainda não está clara, havendo hipóteses de desregulação do sistema imune, causa hormonal e predisposição genética (Consorti *et al.*, 2011).

Em uma coorte de 1064 pacientes (no período 2000 a 2006), 60 foram identificados com CM e CT sendo 43% e 39% diagnosticados primeiramente com CT e CM, respectivamente (Garner *et al.*, 2007). Estes dados revelaram que pacientes que desenvolvem CT como primeiro tumor primário tem um risco aumentado de desenvolver CM de 1,7%. Neste mesmo estudo, em uma segunda coorte de análise, 820 pacientes tiveram seus registros avaliados entre os anos de 1995 a 2006. Vinte e três indivíduos desenvolveram CT e CM. Interessantemente, em 9% destes casos havia histórico familiar de CT e 26% de CM, porém apenas um destes pacientes foi diagnosticado com ambos os tumores, CT medular e CM, o único indivíduo do sexo masculino. Este estudo apontou a importância da investigação sobre a história familiar de CT e/ou CM em membros das famílias e sua relevância para o aconselhamento genético apropriado. Nestes casos, testes de triagem são úteis para a prevenção do aparecimento de um segundo tumor primário ou a detecção de outros membros acometidos pela mesma mutação dos probandos (Garner *et al.*, 2007).

Uma meta-análise confirmou o risco aumentado para o desenvolvimento de um segundo tumor primário em pacientes com CM e ou CT. Joseph *et al.* (2015) relataram que o risco de desenvolver um segundo tumor primário em geral é de 21% para pacientes que já tiveram CT, e de 42% para pacientes que já tiveram CM. Levando em conta apenas os tumores tireoidianos e mamários, o risco de desenvolver CT seguido de CM é maior, cerca de 17%, quando comparado aos casos de CM após CT, sendo apenas de 2 a 3%, o que pode ser consequência dos tratamentos aplicados nos casos de CM incluindo de radioterapia. No entanto, os carcinomas de tireoide e mama foram considerados mais incidentes na forma de segundo tumor primário em relação a outros tipos tumorais, o que leva a concluir que outros mecanismos moleculares e genéticos possam estar implicados na associação entre estes dois tumores (Joseph *et al.*, 2015).

Identificar a ocorrência de um segundo tumor primário mais frequente pode ser importante para o acompanhamento, vigilância e prevenção. Para a investigação do segundo tumor primário mais frequente seguido de CT, foi realizada uma análise retrospectiva utilizando os dados do *Surveillance, Epidemiology, and End Results* (SEER-9 database). Este é um importante registro populacional de câncer dos Estados Unidos, envolvendo mulheres acima de 18 anos de idade, no período de 1973 a 2011 (Kuo *et al.*, 2015). Foi observado que na maioria dos casos as pacientes diagnosticadas com CT desenvolveram CM dentro de cinco anos, além de terem sido consideradas mais jovens quando comparadas às pacientes com diagnóstico de CM da população em geral. Outro achado interessante foi que na maior parte dos casos os tumores de mama eram do tipo lobular ductal misto, positivas para os receptores de estrógeno e progesterona. Os autores concluíram que é importante o acompanhamento de pacientes que desenvolvem CT em idade jovem, evidenciando a necessidade do estabelecimento de *guidelines* específicos para incentivar a realização de exames de rastreamento para CM não só para o paciente em si, mas para a família como um todo, pois esta associação entre CM e CT demonstra um fator genético ainda não explicado (Kuo *et al.*, 2015).

Recentemente, An *et al.* (2015) relataram que os receptores de estrógeno e progesterona podem estar relacionados com o risco aumentado de desenvolvimento de CT e CM. Avaliando um grupo de pacientes da *Seoul National University*, os autores identificaram 55 pacientes com CT seguido de CM (Grupo-1), 81 pacientes com CM seguido de CT (Grupo 2) e 45 pacientes com ambos os tumores sincrônicos (Grupo-3).

O estudo identificou a expressão de PR⁺ e ER⁺ nos tumores de mama dos Grupos-1 e 3 sugerindo que estes receptores podem ter resposta hormonal associada ao desenvolvimento de CT embora o real mecanismo ainda não seja esclarecido (An *et al.*, 2015).

Utilizando sequenciamento do exoma, Ikeda *et al.* (2015) investigaram 14 pacientes com CM e CT diagnosticados antes dos 50 anos, com histórico familiar destes tumores, sem mutação no gene *PTEN*. Os autores priorizaram as alterações raras, patogênicas e com função de reparo a danos no DNA. Baseado nesta premissa, foram identificadas como mais frequentes as alterações p.G496V e p.T1170I no gene *PARP4* (Ikeda *et al.*, 2015). A função deste gene ainda não é totalmente esclarecida, porém por meio de estudo funcionais em linhagens celulares de mama, os autores demonstraram o potencial papel de supressor tumoral uma vez que este gene pertence à família PARP que atua em processos importantes como reparo do DNA, regulação da transcrição e interferência por RNA (Gibson and Kraus, 2012). Os autores concluíram mais estudos são necessários para comprovar o real impacto de alterações do gene *PARP4* e sua associação com uma nova síndrome de predisposição hereditária envolvendo CT e CM.

Bardhan *et al.* (2012) também investigaram a associação entre CT e CM em dois pacientes do sexo masculino que haviam desenvolvido CM com 63 e 68 anos, respectivamente, sem histórico destes tumores na família e com amplificação do gene *HER2*. Concomitantemente ao tratamento do carcinoma de mama, foi identificado um carcinoma papilífero de tireóide nos dois pacientes. Relatos em literatura evidenciam que é comum o desenvolvimento de um segundo tumor primário em homens, quando comparado ao sexo feminino (Auvinen *et al.*, 2002; Hemminki *et al.*, 2005), entretanto, os autores relataram que esse foi o primeiro estudo descrevendo a presença do tumor papilífero de tireóide como segundo tumor em homens. Os autores ainda ressaltaram uma possível causa hereditária envolvendo as síndromes de Cowden e de Peutz-Jeghers, contudo os genes específicos não foram testados para as mutações. Segundo os autores, a investigação de CT em mulheres *HER2*-positivas pode ter relevância clínica, mas é necessário a realização de estudos mais detalhados em um grande número de casos para elucidar esta associação (Bardhan *et al.*, 2012).

Em resumo, os dados descritos em literatura evidenciam que a metodologia de sequenciamento de exomas permite identificar mutações e variações no genoma, as quais podem estar associadas com as síndromes de predisposição hereditária ao risco de

desenvolvimento de tumores. A detecção e a definição de novas alterações genéticas em agrupamentos familiares de CT e CM podem ampliar o espectro de síndromes de predisposição associadas com o câncer e permitir a identificação de pacientes com risco de desenvolver estes tumores.

2. JUSTIFICATIVA

Na prática de atendimento clínico do *A.C. Camargo Cancer Center*, São Paulo, foi verificado uma alta incidência de pacientes que possuem carcinomas de mama e ou tireoide e história destes tumores em membros destas famílias. O diagnóstico atribuído a esses pacientes normalmente se enquadra no perfil da Síndrome do Câncer de Mama Hereditário (*BRCA1*, *BRCA2*, *CHEK2*), Li-Fraumeni (*TP53*), Cowden (*PTEN*), sendo que esta última é a única que associa CM e CT em seu espectro de tumores, dentre outras. Porém, em alguns casos em que foram utilizadas análises de sequenciamento de genes candidatos, estes pacientes eram negativos para as mutações testadas. Em outros casos, os pacientes não se encaixavam nos critérios clínicos das síndromes clássicas. Em decorrência disso, é relevante investigar outras alterações genéticas relacionadas ao desenvolvimento de CT e CM em famílias específicas.

Recentes estudos demonstram uma possível associação entre o câncer de mama e de tireoide mas ainda não foi bem estabelecido o envolvimento de genes relacionados a uma provável síndrome de predisposição a estes tumores conhecida, exceto a Síndrome de Cowden. Além disso, não parece ser coincidência que ambas as doenças acometam principalmente mulheres jovens. Essas observações sugerem a presença de uma doença autossômica, responsável pela manifestação dos cânceres de mama e tireoide, podendo caracterizar uma nova síndrome de predisposição genética. Os dados gerados neste estudo poderão ser incorporados ao conhecimento geral da genética do câncer, permitindo o desenvolvimento de estratégias preventivas, de triagem e aconselhamento genético para muitas famílias.

3. OBJETIVOS

3.1. Objetivo Geral

Identificar alterações genéticas que possam estar associadas ao desenvolvimento dos carcinomas de mama e tireoide em uma possível nova síndrome hereditária, a partir do sequenciamento total do exoma de pacientes diagnosticados com pelo menos um destes tumores e com história familiar destes cânceres.

3.2. Objetivos Específicos

- Identificar as alterações genéticas candidatas, a partir do sequenciamento de exoma total em pacientes diagnosticados com CT e ou CM com história familiar destes tumores, para a seleção das variantes raras, patogênicas e as mais frequentemente encontradas nessa casuística;
- Identificar as redes de interação proteica envolvendo os genes alterados com o intuito de elucidar se há integração entre as diferentes alterações identificadas neste grupo de pacientes;
- Avaliar as variantes obtidas que se mostrarem de relevância em outros membros de algumas famílias selecionadas, afetados ou não por câncer, para investigar a segregação das alterações, quando possível;
- Selecionar os programas mais adequados e implementar um *pipeline* de análise de bioinformática específico para a análise do exoma de pacientes com histórico familiar de câncer.

4. MATERIAL E MÉTODOS

4.1. Pacientes

Este projeto foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Estadual Paulista Júlio de Mesquita Filho, Botucatu, SP (FMB-PC-197/2012) e do A.C. Camargo Cancer Center (Processo CEP 1175/08 ext). Os critérios de inclusão e exclusão adotados neste estudo foram elaborados com a colaboração da Dra. Maria Isabel Achatz, responsável pelo Serviço de Oncogenética do AC Camargo Cancer Center:

Critérios de Inclusão: Probando que tenha desenvolvido câncer de tireoide (não medular) e/ou câncer de mama, sincrônicos ou metacrônicos **E** que possuam histórico familiar dos tumores **E** a presença de pelo menos dois familiares de primeiro ou segundo grau que tenham sido afetados com um dos dois tumores e pelo menos um dos casos em idade inferior a 45 anos **OU** pelo menos um familiar que tenha desenvolvido um dos tumores, sendo que este paciente e o probando devem ter desenvolvido os tumores antes dos 40 anos **E** haver outros tumores em outras gerações que não aquela acometida por CM e CT (não medular).

Critérios de Exclusão: Pacientes do sexo feminino com carcinoma de mama tratado por radioterapia seguido de câncer de tireoide não medular **E** pacientes que tenham desenvolvido câncer medular de tireoide **E** pacientes provenientes de famílias classificadas por critérios clínicos de outras síndromes, com mutações germinativas patogênicas nos genes relacionados (quando disponíveis), como Síndrome de Li-Fraumeni (*TP53*), Síndrome de Lynch (*MLH1*, *MSH2*, *MSH6*, *PMS2*), Síndrome do Câncer de Mama e Ovário Hereditário (*BRCA1/BRCA2* ou *CHEK2*), dentre outras. As informações sobre testes genéticos ou mutações de predisposição a outras síndromes hereditárias que não as descritas acima, não estavam disponíveis nos prontuários médicos para todos os pacientes selecionados, impedindo a exclusão dos mesmos.

No levantamento inicial, 125 famílias apresentavam os dois tipos de câncer. Após a aplicação dos critérios de inclusão e exclusão na tentativa de diminuir o viés de heterogeneidade, foram selecionados 24 pacientes de 21 famílias não relacionadas, sendo 12 diagnosticados com CT, seis com CM e seis com os dois tumores (Tabela 1). Alguns pacientes tiveram os seguintes testes genéticos realizados: paciente 3 negativa para *TP53* R337H; paciente 8 negativa para mutações no gene *TP53*; paciente 9 negativa para

mutações no gene *PTEN*; paciente 12 negativo para mutações no gene *TP53*. Foi possível incluir amostras de sangue periférico de membros de duas famílias (FI e FII), sendo os membros da família I diagnosticados com CM, CM e CT, respectivamente, e membros da família II diagnosticados com CT e CT/CM, respectivamente. Também foram incluídas amostras de tumores de quatro pacientes sendo todos classificados como carcinoma papilífero de tireoide. Os heredogramas dos pacientes avaliados estão em material suplementar.

Tabela 1. Dados clínicos e história familiar dos pacientes que foram analisados por sequenciamento do exoma.

Caso	Sexo	Amostra Tumor	Câncer Tireoide (idade diagnóstico)	Tratamento	Câncer Mama (idade diagnóstico)	Tratamento	Receptores	História Familiar
1	F		CPT (31a)	cirurgia (9/2008) RIT (10/2008)	-	-	-	irmã (caso 2) CPT(24a); mãe CM(49a); tio materno cólon (36a); tia materna CM(37a); tia avó materna tu desconhecido (64a); tio avô materno tu desconhecido (10a); tia avó paterna cólon (79a)
2	F	DNA/ RNA	CPT (24a)	cirurgia (8/2009) RIT (03/2010)	-	-	-	irmã (caso 1) CPT(31a); mãe CM(49a)
3	F		-	-	CM ductal invasivo (49a)/ recidiva (53a)	cirurgia (03/1998), RT (6 ciclos após cirurgia) / cirurgia (07/2002) QT (2002) HT (01/2003 + 5 anos)	ER+ PR+ ER2-	filha (caso 1) CPT(24a); filha (caso 2) CPT(32a); irmã CM(36a)
4	F		CPT (20a)	cirurgia (1987)	CM ductal <i>in situ</i> (43a)	cirurgia (2002) HT (2013)	ND	irmão (caso 5) CPT(40a); pai próstata (64a); mãe melanoma (71a); tio materno próstata (ND)
5	M		CPT (41a)	cirurgia 7/2010) RIT (8/2010)	-	-	-	irmã(caso 4) CPT(20a) e CM(43a)
6	F	DNA/ RNA	CPT (53a)	cirurgia (10/2007) RIT 12/2007)	-	-	-	mãe CM(40a); tia materna CM(79a); filha nasofaringe (30a) cabeça e pescoço (32a); avó materna útero (ND); prima materna pâncreas (58a); tio paterno esôfago (ND); tio paterno próstata (50a), primo paterno pâncreas (ND)
7	M	DNA	CPT (41a)	cirurgia (09/2008) RIT (10/2008)	-	-	-	sobrinho CPT(27a); prima materna CPT(55a); tia materna CM(87a); mãe pele (ND), tio materno próstata (70a), tio materno tu desconhecido (ND)
8	F		CPT (26a)	cirurgia (1994)	-	-	-	irmão CPT(18a); tia CM(40a); prima CM(30a)
9	F		-	-	CM (ND) (45)	RT (1996) QT (1996) HT (1996 + 5 anos)	ND	irmã CPT(68a); irmã CM (70a); irmã CM(61a); tia materna CM(ND) (todos os irmãos com alterações na tireoide (?); irmã tireoidectomia preventiva)sobrinha CM e ovário (ND); tia paterna CM(ND);

10	F	CPT (38a)	cirurgia (2007) RIT (2007)	-	-	-	tio paterno pulmão (ND) mãe CM(68a); tia materna CM(ND); avó materna CM(75a); avô materno cólon (ND); tio paterno cólon (57a); tio paterno pulmão (ND); prima paterna CM(ND), prima paterna de 2º grau leucemia
11	F	CPT (40a)	cirurgia (2006)	-	-	-	avó paterna CM (40a); irmã CPT(40a); pai sarcoma (73a) e cólon (67a); prima paterna 2º grau cabeça e pescoço (20a)
12	M	CPT (37a) Linfoma (43a)	cirurgia (2005) RIT (2005) / RT por linfoma (11/2011)	-	-	-	tia materna CM(69a) e mielomamúltiplo (69a); prima materna CM(51a); tia materna pele (65a); tio materno pele (74a); avó materna estômago (73a); pai bexiga (ND); tia paterna tu desconhecido; primo paterno pele (ND)
13	F	CPT (36a)	cirurgia (2011)	-	-	-	irmã CM(34a); irmã CM(33a); irmã CM(53a); mãe cólon (65a); avô materno cólon (41a); pai estômago (47a); tio paterno próstata (69a); tia paterna CM(36a) e cabeça e pescoço (57a); tia paterna tu desconhecido (51a); primo paterno cólon (49a); primo paterno SNC (52a); avô paterno desconhecido (ND)
14	F	CPT (52a)/ outro primário(53a)	cirurgia (10/2012) RIT (11/2012) / cirurgia (01/2013)	CMductal in situ (43a) / CM bilateral metacrônico (52a)	cirurgia (2002) RT (2002) / RT (10 e 11 de 2012, 02/2013) HT (2002 a 2013)	ER+PR+HER2-	mãe CPT(65a) e sarcoma (45a); prima materna CM(48a); tio materno estômago (60a); avó materna rim (84a); avô paterno cólon (60a)
15	F	CPT (67a)	cirurgia (05/2011)	CM ductal in situ + CM ductal invasivo (10%) (67a)	cirurgia (11/2011) RT (01/2012) HT (ND)	ER+PR+HER2-	irmã CM(65a); irmã CPT(62a); sobrinha CPT (37a); mãe CM(70a); prima materna CM(67a); duas tias materna CM(65ae 70a); avó materna estômago (63a); prima materna estômago (50a)
16	F	CPT (49a)	cirurgia (05/2005) RIT (06/2005)	-	-	-	irmã CM(52a); tia paterna CM(35a); duas tias maternas útero (ND)
17	F	-	-	CM ductal <i>in situ</i> (41a)	cirurgia (2008) HT (2008 a 2011)	ND	tia materna CPT(52a); tia materna CM(38a); tio materno pâncreas (55a); avô

18	F		CPT (52a)	cirurgia (2011)	CMductal invasivo (52a)	cirurgia (2011)	ER+PR+HER2-	materno bexiga (61a); tio paterno bexiga (79a); tia avó útero (ND) sobrinha CM(25a); irmã CPT(57a); irmã CPT(63a)
19	F		-	-	CMductal <i>in situ</i> (48a) / CM bilateral (59a) / recidiva mama direita (67a)	cirurgia (1995) QT (1995) / cirurgia (2005) QT (2005) / cirurgia (2013) HT (até 11/2010)	ND	mãe CM(67a), sobrinho CPT (25a); irmã pulmão (52a), irmão próstata (67a), pai prostata (73a), tia materna tu desconhecido (22a), tio avô tu desconhecido (73a), 4 primas maternas tu desconhecido
20	F		-	-	CMductal invasivo (49a)	cirurgia (11/2007) RT (01 e 03 de 2008) HT (2007 a 2012)	ER+PR+HER2-	sobrinha CPT (39a); sobrinha CM(48a); sobrinha útero (32a); irmão próstata (60a)
21	F	DNA/ RNA	CPT (50a)	cirurgia (12/2006) RIT (01/2007)	CM ductal <i>in situ</i> (50a)	cirurgia (07/2006)	-	prima materna CPT(53a); prima materna CM (42a); tia materna CM(78a); primo materno tu desconhecido (ND); tio materno bexiga (45a); avó materna CM(ND); prima paterna tu desconhecido (ND); tio paterno estômago (ND), tio paterno próstata (ND); avô paterno pulmão (ND)
22	F		-	-	CMductal invasivo (39a)	-	ER+PR+HER2+	irmão tireoide (36a); tia mama (65a)
23	F		CPT (50a)	cirurgia (05/2012)	CMductal invasivo (27a)/ recidiva (30a)	cirurgia (1992) QT (1992) / cirurgia (1995) HT (até 1997)	ER+PR-HER2-	meia irmã CM(50a); sobrinha paterna tireoide (31a); tio paterno estômago (63a); tio paterno pulmão (59a); avô paterno próstata (88a); primo paterno cabeça e pescoço (ND)
24	F		CPT (40a)	cirurgia (2002)	-	-	-	irmã CPT(34a); prima materna CPT(32a); prima paterna CPT(37a); tia materna CM(58a); tia materna útero (55a); pai estômago (52a); primo paterno cólon (60a); primo paterno próstata (58a); prima paterna 2º linfoma (31a)

M: masculino, **F:** feminino, **CPT:** carcinoma papilífero da tireóide, **CM:** carcinoma de mama, **ND:** nada consta, **RIT:** radioiodoterapia, **RT:** radioterapia, **QT:** quimioterapia, **HT:** hormonioterapia, **SNC:** sistema nervoso central, **ER:** receptor de estrógeno, **PR:** receptor de progesterona, **HER2:** receptor do fator de crescimento epidérmico 2, **Tu:** tumor.

4.2 Extração do DNA genômico

A extração do DNA do sangue periférico e dos tecidos foi realizada, pelo Banco de Macromoléculas, A.C. Camargo Cancer Center, SP, utilizando o Qiacube DNA Blood kit (Qiagen) e Gentra Puregene Tissue (Qiagen), por meio do protocolo automatizado utilizando o equipamento QiaCube Robot (Qiagen). O DNA extraído foi submetido à eletroforese em gel de agarose 1,5% para avaliar a sua qualidade sendo também quantificado pelo espectrofotômetro NanoDrop ND 1000 (Thermo Scientific). O método fluorimétrico utilizando Quibit 2.0 (Life Technologies) também foi realizado para uma quantificação mais precisa.

4.3 Construção das Bibliotecas de Exoma

As bibliotecas de exoma foram construídas utilizando o kit *Nextera Exome Enrichment* (Illumina) seguindo as recomendações do fabricante. O sequenciamento foi realizado a partir do equipamento *HiSeq1000* (Illumina). Este kit captura 62 Mb do genoma que corresponde ao exoma total. O protocolo inclui os seguintes passos (1) fragmentação do DNA genômico usando transposases, ligação e tagmentação de adaptadores, (2) limpeza do produto tagmentado, (3) primeira amplificação por PCR convencional, (4) limpeza do produto amplificado pela PCR, (5) duas hibridações e capturas das sondas hibridadas, (6) lavagem das bibliotecas, (7) segunda amplificação por PCR convencional, (8) limpeza do produto de PCR, (9) validação das bibliotecas usando PCR quantitativa (qPCR). Estas etapas estão descritas abaixo, em detalhes.

(1) *Fragmentação do DNA, ligação e tagmentação de adaptadores.* Esse processo utiliza uma enzima capaz de fragmentar o DNA genômico e ao mesmo tempo adicionar adaptadores nas extremidades, permitindo a amplificação pela PCR nos processos subsequentes. Inicialmente as amostras foram diluídas para uma concentração inicial de 50ng de DNA genômico (5ng/ μ L). A fragmentação do DNA e a ligação dos adaptadores ocorreu a partir da adição de 25 μ L de *Tagment DNA Buffer* e 15 μ L de *Tagment DNA Enzyme*. As amostras foram homogeneizadas a 1800 rpm por um minuto e centrifugadas a 280xg por mais um minuto. Em seguida, foram incubadas a 58°C por 10 minutos e posteriormente adicionados 15 μ L de *Stop Tagment Buffer*, que inibe a

ligação dos adaptadores. Após homogeneização e centrifugação das amostras a 1800rpm e 280xg, respectivamente, por um minuto, as amostras foram incubadas por quatro minutos a temperatura ambiente.

(2) *Limpeza do produto tagmentado.* O DNA tagmentado passou por um processo de purificação para que, em seguida, pudesse ser amplificado. Foram adicionadas e homogeneizadas 65µL de *Sample Purification Beads* (SPB). Em seguida, o produto foi incubado por oito minutos a temperatura ambiente e por dois minutos (ou até o sobrenadante ficar claro) em raque magnética. Foram descartados 130 µL do sobrenadante e, sem perturbar as *beads*, foram adicionados 200 µL de etanol 80% (EtOH 80%). Em seguida, houve incubação por 30 segundos e o etanol foi removido. Foi realizada mais uma lavagem com EtOH 80% nas mesmas condições. Após remoção do sobrenadante, as amostras foram incubadas por 10 minutos a temperatura ambiente na raque magnética para secagem das *beads*. O DNA tagmentado purificado foi ressuspenso em 22,5 µL de *Resuspension Buffer* (RSB). Em seguida, foi incubado por dois minutos a temperatura ambiente e dois minutos (ou até o sobrenadante ficar claro) na raque magnética. Foram transferidos 20 µL do sobrenadante para tubos de PCR.

(3) *Primeira amplificação por PCR convencional.* Nesta etapa o DNA tagmentado e purificado foi amplificado pela PCR sendo adicionados os adaptadores “*Index 1*” e “*Index 2*” necessários para identificação das amostras na etapa de sequenciamento. Foram adicionados 5µL de *Index 1* em todas amostra, 5µL de *Index 2* específico para cada amostra e 20µL de *Nextera Library Amplification Mix*. O produto foi homogeneizado a 1200 rpm por um minuto e centrifugado a 280xg por um minuto. Em seguida, a amplificação pela PCR ocorreu usando o programa de 72°C por três minutos, 98°C por 30 segundos, 10 ciclos (98°C por 10 segundo, 60°C por 30 segundos e 72°C por 30 segundos), 72°C por cinco minutos e 10°C ∞.

(4) *Limpeza do produto amplificado pela PCR.* A biblioteca de DNA foi purificada para remoção de produtos indesejáveis. As amostras foram removidas do termociclador e centrifugadas por um minuto a 280xg. Foram adicionados 90µL de SPB aos 50µL de amostra e foram homogeneizados por um minuto a 1800 rpm. As amostras foram incubadas a temperatura ambiente por 10 minutos e em seguida por dois minutos (ou até

o sobrenadante ficar claro) em raque magnética. Após a remoção do sobrenadante, foram adicionados 200 μ L de EtOH 80% e as amostras incubadas por 30 segundos. Em seguida, foi realizada mais uma lavagem com EtOH 80% nas mesmas condições. As amostras foram incubadas novamente a temperatura ambiente em raque magnética por 10 minutos para secagem das *beads*. O DNA amplificado e purificado foi ressuspendido em 27 μ L de RSB. As amostras foram incubadas a temperatura ambiente por dois minutos e em raque magnética por dois minutos (ou até o sobrenadante ficar claro). Foram transferidos 25 μ L do sobrenadante para um novo tubo para que a biblioteca seja quantificada novamente pelo método fluorimétrico Qubit, como citado no início deste protocolo.

(5) *Primeira hibridação*. Nesta etapa a biblioteca de DNA foi hibridada com sondas biotiniladas responsáveis por capturar as regiões alvos de interesse (exons). As amostras na concentração de 500ng foram combinadas em “*pool*”, no qual cada um deles pode conter até oito amostras e deve ter um volume final de 40 μ L. Foram adicionados 50 μ L de *Enrichment Hybridization Buffer* (EHB) e 10 μ L de *Expanded Exome Oligos* (EEX) em cada *pool* de amostras. Após homogeneização (1200rpm por um minuto) e centrifugação (280xg por um minuto), o *pool* de amostras foi incubado em termociclador a 95°C por 10 minutos, seguido de 18 ciclos de um minuto de incubação começando a 94°C e diminuindo 2°C por ciclo até chegar aos 58°C, e permaneceu nesta temperatura entre 16 e 18 horas. O tempo de hibridação recomendado garante que as regiões alvo se liguem às sondas de captura completamente.

(6) *Primeira captura*. As sondas biotiniladas hibridadas para as regiões alvo de interesse foram capturadas a partir de *beads* marcadas com estreptavidina. O produto passa por processos de lavagens para que ligações não específicas sejam removidas e a biblioteca enriquecida possa ser eluída e preparada para uma segunda etapa de hibridação. Após 16 a 18h, o *pool* de amostras foi retirado do termociclador, centrifugado por um minuto a 280xg sendo adicionados 250 μ L de *Streptavidin Magnetic Beads* (SMB). Após homogeneização por cinco minutos a 1200rpm, procede-se incubação a temperatura ambiente por 25 minutos e centrifugação a 280xg por um minuto. O *pool* de amostras foi incubado em raque magnética por dois minutos (ou até o sobrenadante ficar claro) e o sobrenadante descartado. Foram então adicionados 200 μ L de *Enrichment Wash*

Solution (EWS) aos *pools* e estes homogeneizados a 1800 rpm por quatro minutos. As amostras foram incubadas a 50°C por 30 minutos e em raque magnética por mais dois minutos (ou até o sobrenadante ficar claro). Em seguida, o sobrenadante foi descartado. Esse passo de lavagem com EWS seguido das duas incubações e descarte do sobrenadante foi repetido. Vinte e três µL de uma solução de pré-eluição (28,5µL de *Enrichment Elution Buffer 1-EE1* e 1,5µL de 2N NaOH) foram adicionados e homogeneizados ao *pool* de amostras por dois minutos a 1800 rpm. Este foi incubado a temperatura ambiente por dois minutos e em raque magnética por mais dois minutos (ou até o sobrenadante ficar claro). Vinte e um µL do sobrenadante foram transferidos para um novo tubo de *eppendorf* com 4µL de *Elute Target Buffer 2* (ET2), homogeneizados a 1200 rpm por um minuto e centrifugados por a 280xg por um minuto.

(7) *Segunda hibridação e captura.* Os passos da segunda hibridação e captura são os mesmos descritos acima. A repetição deste processo assegura alta especificidade das regiões capturadas.

(8) *Limpeza das bibliotecas enriquecidas.* Foram adicionados ao *pool* de amostras 45µL de SPB e estes foram homogeneizados a 1800 rpm por um minuto e então centrifugados a 280xg por um minuto. Em seguida, foram incubados por 10 minutos a temperatura ambiente e na raque magnética por dois minutos (ou até o sobrenadante ficar claro). O sobrenadante foi descartado e 200 µL de EtOH 80% foram adicionados e incubados por 30 segundos. Novamente o sobrenadante foi removido e esta etapa de lavagem foi repetida. Todo o excesso de etanol foi removido e as amostras foram incubadas em raque magnética por 10 minutos até secarem completamente. Em seguida, as amostras foram transferidas para raque normal e resuspendidas em 27,5µL de RSB e homogeneização a 1800 rpm por um minuto. As amostras foram incubadas em temperatura ambiente por dois minutos e em raque magnética por dois minutos (ou até o sobrenadante ficar claro), sendo então transferidos 25µL do sobrenadante para um novo tubo.

(9) *Segunda amplificação pela PCR convencional.* A biblioteca de DNA purificada foi amplificada utilizando PCR convencional. Foram adicionados ao *pool* de amostras 20µL de *Nextera Enrichment Amplification Mix* (NEM) e 5µL de *PCR Primer Cocktail*

(PPC). Após homogeneização a 1200rpm por um minuto e centrifugação a 280xg por um minuto, as amostras foram incubadas em termociclador a 98°C por 30 segundos, 12 ciclos de (98°C por 10 segundos, 60°C por 30 segundos e 72°C por 30 segundo), 72°C por cinco minutos e 10°C ∞.

(10) *Limpeza do produto amplificado pela PCR.* A biblioteca de DNA foi purificada para remoção de produtos indesejáveis. As amostras foram removidas do termociclador e centrifugadas por um minuto a 280xg. Foram adicionados 90 µL de SPB aos 50µL de amostra sendo a solução homogeneizada por um minuto a 1800 rpm. As amostras foram incubadas a temperatura ambiente por 10 minutos e em seguida por dois minutos (ou até o sobrenadante ficar claro) na raque magnética. Após a remoção do sobrenadante, foram adicionados 200 µL de EtOH 80% e as amostras incubadas por 30 segundos. Em seguida, foi realizada nova lavagem com EtOH 80% nas mesmas condições. As amostras foram incubadas a temperatura ambiente na raque magnética por 10 minutos para secagem das *beads*. O DNA amplificado e purificado foi ressuspendido a partir da homogeneização com 32 µL de RSB. As amostras foram incubadas a temperatura ambiente por dois minutos e em raque magnética por dois minutos (ou até o sobrenadante ficar claro). Foram transferidos 30 µL do sobrenadante para ser submetido à etapa de sequenciamento do exoma.

4.4 Pré-Sequenciamento e Sequenciamento do Exoma

O sequenciamento de alto desempenho utilizado o equipamento HiSeq (Illumina) foi realizado no Centro de Genômica Funcional, ESALQ – USP, Piracicaba, São Paulo com a colaboração do Prof. Dr. Luiz Lehmann Coutinho.

O sequenciador HiSeq 1000 é capaz de gerar fragmentos de 200 pb e cerca de 300 Gb de dados, por *flow-cell* (lâmina de sequenciamento), em uma única corrida (http://www.illumina.com/systems/hiseq_2000_1000.ilmn). Cada *flow-cell* é dividida em oito *lanes* (canaletas), cuja capacidade de geração de dados é de aproximadamente 37,5Gb. Com estas informações e com os dados do tamanho do exoma sequenciado (62Mb = 0,062Gb baseado no protocolo utilizado neste trabalho) é possível calcular e estimar a cobertura das sequências, como exemplificado a seguir:

$$\frac{37,5 \text{ Gb (Total gerado por lane)}}{6 \text{ (Número de amostra por lane)}} = 6,25 \text{ Gb (Total por amostra)}$$

$$\frac{6,25 \text{ Gb (Total por amostra)}}{0,062 \text{ Gb (Tamanho do exoma)}} = 100x \text{ (Cobertura média por amostra)}$$

Este cálculo é extremamente importante para a obtenção de dados fidedignos, que representem informações biologicamente corretas, reprodutíveis e que possam ser confirmados por experimentos de validação adequados.

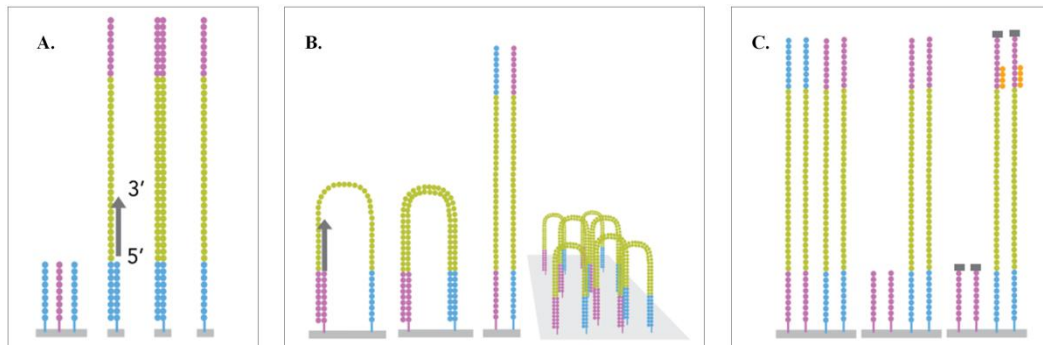
Foi realizada também a validação das bibliotecas pela PCR quantitativa (qPCR), utilizando o KAPA Illumina GA with revised primers-SYBR ABI Prism kit (Kapa Biosystems).

Após a validação das bibliotecas pela qPCR, os *pools* de amostras são incluídos na *flowcell*, seguida pela amplificação clonal em fase sólida no equipamento cBot (Illumina Inc, San Diego, CA, USA), o qual é, totalmente automatizado (Figura 1). Cada *pool* contém 6 amostras, que por sua vez são colocadas em *lanes* separadamente, de acordo com o cálculo da cobertura previamente realizado. Plotadas na *flowcell* estão as sondas correspondentes às sequências contidas nos adaptadores utilizados na construção das bibliotecas. Estas sondas funcionam como iniciadores para que a amplificação em ponte ocorra na etapa de clusterização. Deste modo, as bibliotecas são agrupadas em *clusters*, ou seja, são produzidas múltiplas cópias de fragmentos específicos por meio da desnaturação da dupla fita de DNA e incorporação de nucleotídeos na nova cadeia de DNA crescente, como acontece em uma PCR convencional. Após o término da clusterização, os fragmentos são desnaturados de modo que DNAs de fita única fiquem ligados à *flowcell* para posterior sequenciamento.

Figura 1. Esquema representativo da etapa de clusterização realizada no equipamento cBot (cBot samplesheet; https://www.illumina.com/documents/products/datasheets/datasheet_cbot.pdf).

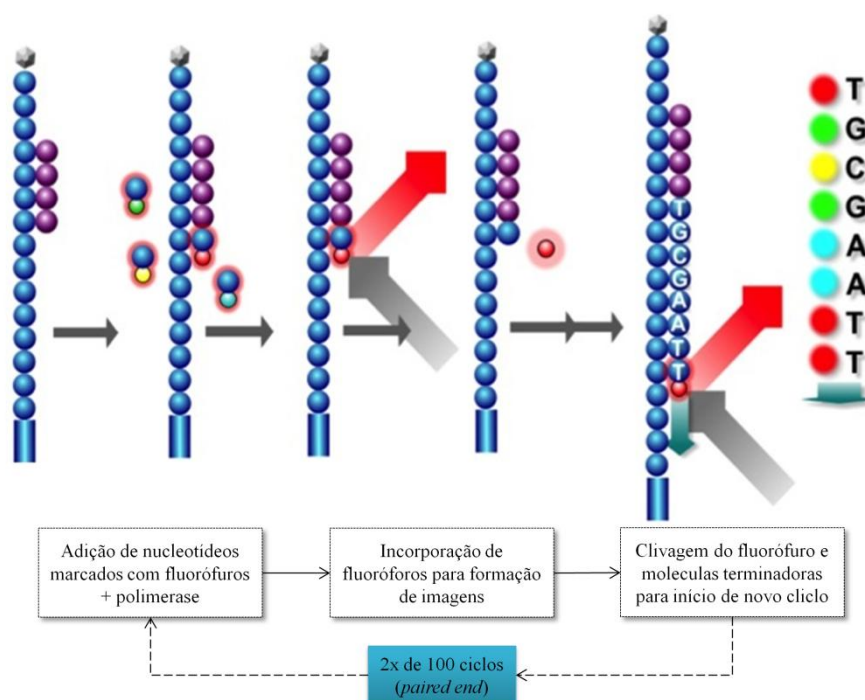
A. Ligação dos fragmentos da biblioteca à *flowcell* por meio de anelamento com iniciadores plotados na lâmina, e amplificação do fragmento. **B.** Novo anelamento com iniciadores plotados na lâmina seguido de amplificação em ponte e formação do *cluster*. **C.** Eliminação da fita complementar e bloqueio dos iniciadores da lâmina, possibilitando o arranjo dos fragmentos do *cluster* em apenas um sentido único para

posterior sequenciamento dos fragmentos idênticos contidos no *cluster*.



O sequenciamento ocorre logo após a clusterização, onde a flowcell é colocada no sequenciador. Na etapa de sequenciamento também há amplificação de moléculas de DNA contidas nos *clusters*, porém os nucleotídeos incorporados possuem bases marcadas com fluoróforos, sendo cada base correspondente a uma cor diferente (Figura 2). A cada ciclo uma base com fluoróforo é incorporada, a imagem do cluster é registrada e é retirado o fluoróforo juntamente com o agente terminador da cadeia crescente de DNA para que um novo ciclo se inicie. À medida que cada base é incorporada à cadeia crescente de DNA, o *cluster* emite uma fluorescência correspondente à base e assim, sucessivamente até o completo sequenciamento dos fragmentos (200 pares de base em corrida *pair-end*). Os sinais fluorescentes são lidos e interpretados por *software* específico, possibilitando a construção da sequência. Todos os procedimentos seguiram as recomendações do fabricante.

Figura 2. Esquema representativo da etapa de sequenciamento (fonte: Introduction to SBS technology, <http://www.illumina.com>).



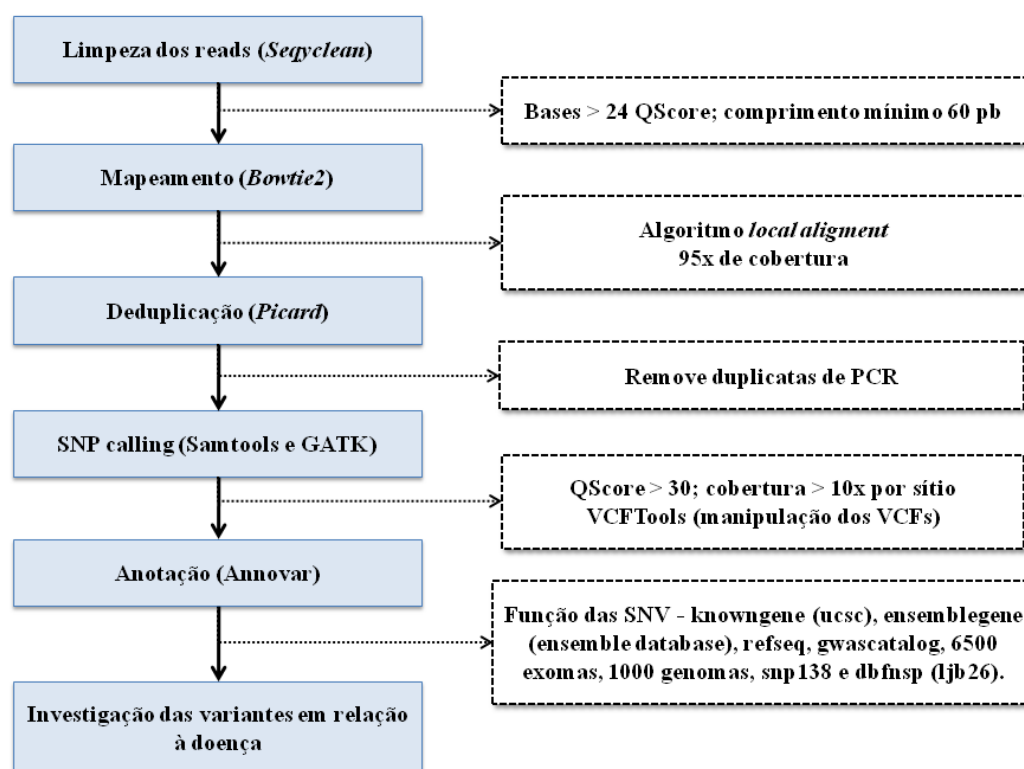
Após o sequenciamento, as bases foram nomeadas (*base-calling*) e os dados extraídos utilizando os programas SAV (Software Analysis Viewer; https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.html) e CASAVA (https://support.illumina.com/sequencing/sequencing_software/casava.html). Em seguida, as amostras foram identificadas e separadas dos *pools* de amostras sequenciadas em cada *lane* (denominado de demultiplexação). Após essa etapa, os dados foram compactados e 164 Gb de dados foram transferidos e armazenados via *download* para o servidor do Laboratório NeoGene, Faculdade de Medicina, UNESP, Botucatu, para a análise de bioinformática.

4.5 Análise de Bioinformática

Há uma enorme variedade de *pipelines* descritos atualmente em literatura e desenvolvidos para a análise de dados provenientes de sequenciamento de alto desempenho. Por este motivo, foi testada uma série de programas e parâmetros diferentes, com o intuito de melhor atender as necessidades deste estudo. O pipeline utilizado foi construído em colaboração com a Dra. Sonia Andrade, ESALQ-US. Aqui

está descrito o procedimento final adotado, sendo demonstradas as linhas de comandos utilizadas, incluindo a limpeza dos *reads*, mapeamento, retirada dos *reads* duplicados, chamada de variantes até a etapa de anotação dos dados. Após isso houve estratificação das alterações, a qual está detalhada na próxima seção.

Figura 3. Fluxograma de análises de bioinformática, ilustrando as etapas básicas das análises.



O primeiro procedimento adotado foi a limpeza dos dados brutos, de acordo com a qualidade das bases sequenciadas, comprimento dos *reads* e possíveis moléculas contaminantes utilizando o programa Seqclean (versão 1.9.10) (<https://bitbucket.org/izhbannikov/seqclean>). Foram descartados os *reads* com QScore abaixo de 26 e/ou 24 e de comprimento inferior a 60 pb. O banco de dados UniVec (<http://www.ncbi.nlm.nih.gov/tools/vecscreen/univac/>) foi utilizado para excluir as sequências provenientes de moléculas de DNA contaminantes, como adaptadores e vetores. Abaixo estão apresentados os comandos utilizados para a limpeza dos *reads*:

```
seqclean
  / -v UniVec.fa
  / -qual 26 24
```

```

/-minimum_read_length 60
/-1 $name-R1.fastq.gz
/-2 $name-R2.fastq.gz
/-o $name-cleaned

```

Posteriormente os *reads* foram mapeados utilizando o algoritmo de alinhamento local do programa Bowtie2 (versão 2.1.0) (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Com esta estratégia há maior número de alinhamento dos *reads*, pois as sequências não são descartadas quando não há o pareamento exato em ambas as extremidades. Os comandos utilizados para o alinhamento das sequências ao genoma de referência foram:

```

bowtie2-align \
  --very-sensitive-local \
  -p 1 \
  -x path_to/bowtie2_index/genome \
  -1 $name-cleaned_PE1.fastq \
  -2 $name-cleaned_PE2.fastq \
  -U $name-cleaned_SE.fastq \
  -S $name-mapped.sam

```

O arquivo gerado é convertido em “.sam” (*Sequence Alignment/Map*) em sua forma binária “.bam”, necessário para continuidade do pipeline:

```

samtools view \
  -bS $name-mapped.sam > $name-mapped.bam

```

Após o alinhamento, as sequências foram ordenadas para a retirada dos *reads* duplicados. Os *reads* duplicados são compostos por uma série de fragmentos que se iniciam e terminam exatamente na mesma base, podendo ser consequência da amplificação pela PCR realizada durante o preparo das bibliotecas, fragmentação do DNA no mesmo local, artefatos devido a regiões repetitivas e *background* na detecção de *clusters* na fase do sequenciamento. Esta limpeza é necessária para que a cobertura das regiões genômicas não seja superestimada e não haja impacto na acurácia de detecção das variantes. Foi utilizado o programa PICARD (versão 1.101) (<http://broadinstitute.github.io/picard/>). As linhas de comando para ordenar os *reads* de forma a prosseguir com a etapa de deduplicação foram:

```

java -Xmx4g -jar path_to/picard/SortSam.jar \
  I=$name-mapped.bam \
  O=sorted.bam \
  SORT_ORDER=coordinate \
  VALIDATION_STRINGENCY=SILENT \

```

```
TMP_DIR=tmp \
QUIET=true \
COMPRESSION_LEVEL=0
```

Após gerar o grupo de *reads*, nomeia-se o conjunto de sequências de acordo com sua origem. Esta etapa é necessária para entrada no programa GATK. As linhas de comando utilizadas foram:

```
java -Xmx4g -jar path_to/picard/AddOrReplaceReadGroups.jar \
  I=sorted.bam \
  O=grouped.bam \
  SORT_ORDER=coordinate \
  RGID=1 \
  RGLB=name-of-bam \
  RGPL=ILLUMINA \
  RGPU=$name-of-sample \
  RGSM=$name-of-sample \
  CREATE_INDEX=true \
  VALIDATION_STRINGENCY=SILENT \
  TMP_DIR=tmp \
  QUIET=true
```

Para a retirada dos *reads* duplicados foram utilizados os comandos:

```
java -Xmx4g -jar path_to/picard/MarkDuplicates.jar \
  I=grouped.bam \
  O=dedup_gatk.bam \
  M=metrics \
  VALIDATION_STRINGENCY=SILENT \
  TMP_DIR=tmp \
  CREATE_INDEX=true \
  REMOVE_DUPLICATES=true
```

A cobertura de mapeamento alcançou uma média de 95x por amostra. A alta cobertura permite avaliar as alterações genômicas com precisão, diminuindo o número de falso-positivos. O resultado gerado com o comando seguinte, com o programa Samtools (versão 1.2.1), que é a profundidade das leitura das regiões genômicas, deve ser dividido pelo tamanho do exoma, no caso 62 Mb. Os comando para essa etapa foram:

```
samtools depth $name_sorted.bam | \
  awk '{sum+=$3;cnt++}END{print sum/cnt" "sum}'
```

As variantes de base única e pequenas inserções e deleções (*Indels*) foram identificadas e selecionadas de acordo com os seguintes parâmetros: qualidade de base acima de 30 (Qscore> 30) e cobertura mínima de 10x por sítio de alteração. Foram

utilizadas duas ferramentas para a geração das variantes, os programas GATK (versão 3.2-2-gec30cee) (McKenna *et al.*, 2010) e Samtool (versão 1.2.1) (Li *et al.*, 2009). A seguir, utilizou-se os comandos de geração das variantes com Samtools e filtrações das variantes com o conjunto de ferramentas VCFtools (versão 0.1.12) (<https://vcftools.github.io/index.html>).

Foi então realizada a comparação entre as sequências geradas, levando em conta a qualidade de mapeamento, e o genoma de referência para identificação das alterações.

Os comandos utilizados para esta etapa foram:

```
samtools mpileup \
  -q30 -Q20 -AB -C50 \
  -uf path_to_reference/genome.fa \
  $name_dedup.bam | \
  bcftoolsview -bcvg - > $name_var.raw.bcf
```

Geração do arquivo VCF (*Variant Call Format*):

```
bcftoolsview $name_var.raw.bcf | \
  vcfutils.pl varFilter \
  -d 10 \
  -D 99999 \
  > $name_var.flt.vcf
```

Filtração das variantes geradas pelo Samtools:

```
cat total.vcf | \
  vcf-annotate \
  -H \
  --filter Qual=30/MinDP=10 \
  > $name_SNP_filtered_qual.vcf
```

A geração das alterações com o programa GATK, foi realizado segundo o manual de boas práticas (https://www.broadinstitute.org/gatk/guide/bp_step.php?p=2).

Os comandos abaixo permitem a definição dos intervalos das regiões alvo para posterior realinhamento local:

```
java -Xmx6g -jar GenomeAnalysisTK.jar \
  -T RealignerTargetCreator \
  -nt 4 \
  -R path_to_reference/hg19/genome.fa \
  -I dedup_gatk.bam \
  -o sample.intervals
```

As alterações foram então realinhadas com os intervalos previamente gerados,

para um alinhamento local de maior qualidade, utilizando os comandos:

```
java -Xmx6g -jar GenomeAnalysisTK.jar \
  -T IndelRealigner \
  -R path_to_reference/hg19/genome.fa \
  -I dedup_gatk.bam \
  -o indelrealigner.bam \
  -targetIntervals sample.intervals
```

As alterações obtidas na etapa anterior foram re-calibradas utilizando o banco de dados dbSNP, conforme descrito abaixo:

```
java -Xmx6g -jar GenomeAnalysisTK.jar \
  -T BaseRecalibrator \
  -R path_to_reference/hg19/genome.fa \
  -knownSites dbsnp138.vcf \
  -knownSites indel_dbsnp138.vcf \
  -I indelrealigner.bam \
  -o recal.grp
```

As alterações identificadas após o realinhamento e recalibragem, para posterior filtragem foram utilizadas para gerar um arquivo, como descrito abaixo:

```
java -Xmx6g -jar GenomeAnalysisTK.jar \
  -T PrintReads \
  -R path_to_reference/hg19/genome.fa \
  -I indelrealigner.bam \
  -o sample_recalibrated.bam \
  -BQSR recal.grp
```

Após esta etapa, foram identificados os SNPs e *Indels*, gerando o arquivo VCFs, conforme descrito abaixo:

```
java -Xmx6g -jar GenomeAnalysisTK.jar \
  -T HaplotypeCaller \
  -R path_to_reference/hg19/genome.fa \
  -I sample_recalibrated.bam \
  -o $name.vcf \
  -stand_call_conf 30 \
  -stand_emit_conf 10 \
  -minPruning 3
```

A seguir a filtragem das variantes para SNPs e *Indels*, de acordo com o programa GATK (<http://gatkforums.broadinstitute.org/gatk/discussion/2806/howto-apply-hard-filters-to-a-call-set>):

Filtro para SNPs:

```
java -jar GenomeAnalysisTK.jar \
```

```
-T VariantFiltration \
-V SNP.recode.vcf \
-R path_to_reference/hg19/genome.fa \
--filterExpression "QD < 2.0 || DP < 10 || FS > 60.0 || MQ < 40.0 || HaplotypeScore>
13.0 || MappingQualityRankSum < -12.5 || ReadPosRankSum< -8.0" \
--filterName "Qual" \
-o filteredSNP.vcf
```

Filtro para *indels*:

```
java -jar GenomeAnalysisTK.jar \
-T VariantFiltration \
-R path_to_reference/hg19/genome.fa \
-V Indels.recode.vcf \
--filterExpression "QD < 2.0 || DP < 10 || FS > 200.0 || ReadPosRankSum < -20.0" -
filterName "QualIndel" \
-o filteredINDEL.vcf
```

Após a geração dos arquivos VCFs foi realizada a união entre as variantes provenientes do programa Samtools e GATK. Além disso, houve também a comparação entre as amostras para selecionar as variantes presentes em pelo menos quatro indivíduos. Esta manipulação de arquivos VCF foi realizada com o programa VCFtools (versão 0.1.12) (Danecek *et al.*, 2011) (<https://vcftools.github.io/index.html>).

As variantes resultantes das ferramentas Samtools e GATK foram anotadas utilizando o programa Annovar (versão 14 Dec 2015) (Wang *et al.*, 2010). As variantes foram nomeadas de acordo com os bancos de dados known Gene (UCSC), ensemble Gene (Ensemble Database), RefSeq, GWAS Catalog, clinvar, cosmic70, 1000 Genomas, 6500 Exomas, 1000 Genomas, snp138 e dbfnsp (ljb26). Este último, dbfnsp, contém as informações de patogenicidade dos seguintes bancos de dados: Sift, Polyphen2_HDIV, Polyphen2_HVAR, LRT, Mutation Taster, FATHMM, Radial SVM, LR, GERP++, PhyloP, SiPhy. Todos os bancos de dados utilizados foram baixados separadamente. Estes bancos de dados estão disponíveis para *download* no site do programa Annovar (<http://www.openbioinformatics.org/annovar>). Abaixo estão descritos os comandos para a geração de arquivo de entrada no annovar, transformando o arquivo .vcf em .avinput:

```
convert2annovar.pl \
-format vcf4 \
$name.vcf \
-out vcf.avinput \
-includeinfo
```

Anotação das variantes com o comando que resultara em um arquivo .txt que possibilita a análise de em programas como Excel.

```
table_annovar.pl \
vcf.avinput \
humandb -buildver hg19 \
-out anotação.txt \
-protocol
refGene,knownGene,ensGene,gwasCatalog,clinvar,cosmic,esp6500siv2_all,1000g2014
oct_all,snp138,ljb26_all -operation g,g,r,f,f,f,f,f,f \
-otherinfo
```

É importante ressaltar que todos os programas e bancos de dados aqui utilizados estão em constante melhoria e desenvolvimento. Procurou-se utilizar as últimas versões e bancos de dados lançados até a finalização deste estudo (Dezembro de 2015).

4.6 Estratificação e Seleção das Variantes

Para a seleção das variantes candidatas, foram realizadas duas abordagens diferentes. Na primeira, foram aplicados filtros para a seleção das variantes considerando 64 genes de alta e baixa penetrância associados aos CT e CM hereditários previamente descritos em literatura (Tabela 2). A segunda abordagem consistiu em aplicar filtros para a identificação de variantes mais frequentes e compartilhadas entre todas as amostras (sangue).

Tabela 2. Genes de alta e baixa penetrância investigados em detalhes entre os 24 pacientes avaliados por NGS.

Alta penetrância			Baixa penetrância	
Genes associados ao CM	Genes associados ao CT	Outras Síndromes	Genes associados ao CM (Shuen and Foulkes, 2011; Lalloo and Evans, 2012)	Genes associados ao CT# (OMIM:188550)
<i>BRCA1</i> , <i>BRCA2</i> , <i>PALB2</i> , <i>TP53</i>	<i>APC</i> , <i>PTEN</i> , <i>RET</i> *, <i>NTRK1</i> *	<i>MLH1</i> , <i>MSH1</i> , <i>MSH2</i> , <i>PMS1</i> , <i>PMS2</i> (Síndrome de Lynch); <i>RB</i> (Retinoblastoma); <i>CDKN2A</i> (Melanoma Familiar); <i>MEN1</i> (Neoplasia Endócrina Múltipla)	<i>ATM</i> , <i>CHEK2</i> , <i>BRP1</i> , <i>NF1</i> , <i>FGFR2</i> , <i>TOX3</i> , <i>TNRC9</i> , <i>MRPS30</i> , <i>MAP3K1</i> , <i>CASP8</i> , <i>FAM84B</i> , <i>LSP1</i> , <i>NEK10</i> , <i>COX11</i> , <i>TNP1</i> , <i>IGFBP5</i> , <i>IGFBP2</i> , <i>TNS1</i> , <i>NOTCH2</i> , <i>RAD51L1</i> , <i>ESR1</i> , <i>CDKN2a</i> , <i>RAD54L</i> , <i>BARD1</i> , <i>PIK3CA</i> , <i>HMMR</i> , <i>NQO2</i> , <i>RB1CC1</i> , <i>SLC22A1L</i> , <i>TSG101</i> , <i>KRAS</i> , <i>XRCC3</i> , <i>AKT1</i> , <i>RAD51A</i> , <i>PHB</i> , <i>PPM1D</i> , <i>BRIP1</i> , <i>AKAP9</i> , <i>ATXN7</i>	<i>C8orf4</i> , <i>TRIM33</i> , <i>TRIM24</i> , <i>PCM1</i> , <i>NCOA4</i> , <i>CCDC6</i> , <i>GOLGA5</i> , <i>PRKAR1A</i> , <i>DICER1</i>

*: Genes de alta penetrância associados ao carcinoma medular de tireoide, #: genes comumente descritos em literatura associados ao carcinoma papilífero de tireoide, tanto em síndromes hereditárias quanto em tumores esporádicos (OMIM:188550).

Esta segunda etapa de estratificação foi realizada em colaboração com o Dr. Igor Jurisica, *University Health Network*, Toronto, Canadá. Para a segunda abordagem de análise acima mencionada, foram consideradas apenas as alterações heterozigotas nas amostras do sangue periférico, uma vez que o fenótipo em questão aparenta ser de caráter dominante, revelado por meio de análises de genealogias. Segundo Gilissen *et al.* (2012), a análise de heredogramas pode ser um importante aliado na estratificação das variantes.

A realização da etapa de filtragem dos dados se fez necessária para a priorização do menor número possível de genes candidatos baseado no fenótipo da doença, pois são milhares de alterações identificadas, o que dificulta a seleção das variantes candidatas. Contudo, as alterações em homozigose também foram avaliadas em um segundo momento levando em consideração as informações sobre a função das variantes. Como estratégia inicial, foi utilizada a *Análise de Sobreposição*, onde são avaliados vários indivíduos não aparentados que compartilham o mesmo fenótipo. Para que o maior número de alterações compartilhadas possível fossem identificadas, foram consideradas as alterações presentes em pelo menos 4 pacientes. Foram considerados 23 dos 24 pacientes incluídos no projeto, pois um caso (paciente 14) foi diagnosticado com a Síndrome de Cowden a partir de uma alteração identificada no presente estudo.

Foram excluídas as variantes comuns, com MAF (*minor allele frequency*) acima de 0,05 descritas em bancos de dados como 1000 Genomas (<http://www.1000genomes.org/>) (Abecasis *et al.*, 2012) e 6500 Exomas (<http://evs.gs.washington.edu/EVS/>). A construção das bibliotecas de exoma (Nextera Exome Enrichment, Illumina) é baseada na captura de algumas regiões UTRs (*untranslated regions*), intrônicas e intergênicas e, por este motivo, foi necessária uma filtragem para manter apenas as alterações localizadas em regiões exônicas, consideradas não-sinônimas. Foi também avaliada a patogenicidade das alterações e estas foram consideradas deletérias quando descritas como patogênicas em pelo menos três dos nove bancos de dados utilizados. A partir dos genes alvo identificados, foram pesquisadas outras alterações nos mesmos genes.

A seguir, os genes com alterações consideradas deletérias foram submetidos a uma análise de interação de proteínas utilizando o banco de dados *Interologous Interaction Database* (I2D - <http://ophid.utoronto.ca/ophidv2.204/>) (Savas *et al.*, 2009) e o programa *Navigator* (versão 2.3) (<http://ophid.utoronto.ca/navigator/index.html>) (Brown *et al.*, 2009). Em adição, foi realizado um enriquecimento de vias envolvendo todas as alterações não sinônimas, com os programas *Ingenuity Pathway Analysis* (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) e a ferramenta PathDIP (em construção). Os programas PathDIP e Navigator, bem como o banco de dados I2D foram desenvolvidos pelo colaborador de nosso grupo Dr. Igor Jurisica, *University Health Network*, Toronto, Canadá.

4.7 Confirmação das Alterações

Após a etapa de estratificação e seleção das variantes, algumas destas foram selecionadas para a confirmação e validação incluindo rs61754785 e rs7080536 mapeadas nos genes *MUS81* e *HABP2*, respectivamente. Estas alterações foram investigadas pela metodologia de sequenciamento convencional (Método de Sanger) nos pacientes índice e em outros membros das famílias, conforme descrito a seguir.

Para o procedimento de sequenciamento de Sanger, foram desenhados e sintetizados iniciadores flaqueando as regiões contendo as alterações detectadas por sequenciamento do exoma (Tabela 3).

Tabela 3. Sequência de iniciadores que flanqueiam o SNP rs7080536 mapeado no gene *HABP2* (Gara *et al.*, 2015) e rs61754785 mapeado no gene *MUS81*.

HABP2	Referência	NM_001177660:exon13:c.G1523A:p.G508E NM_004132:exon13:c.G1601A:p.G534E
	SNP	rs7080536
	Tamanho do fragmento	336 pares de base
	Sequência forward	CCCTGACACCCCCTGGAGAG
	Sequência reverse	GCTCTGGAGGTGTCCATTGT
MUS81	Referência	NM_025128:exon13:c.G1292A:p.R431H
	SNP	rs61754785
	Tamanho do fragmento	282 pares de base
	Sequência forward	TGGCTGGAGTTGTTCTTCG
	Sequência reverse	AAACACTTCTCGCACCGACT

A amplificação pela PCR para as regiões genômicas de interesse foi realizada utilizando-se o kit Platinum Taq DNA Polymerase High Fidelity (Life) nas seguintes condições:

Amplificação do produto a ser sequenciado e limpeza: a reação de amplificação das regiões de interesse foi realizada com a adição de iniciadores *forward* e *reverse* à concentração final de 0,2 μ M, dNTP mix contendo dATP, dCTP, dGTP e dTTP à concentração final de 0,2 mM, MgSO₄ à concentração final de 1,5mM, enzima Platinum Taq DNA Polymerase High Fidelity à 0,5 U, tampão 1x, 25ng DNA genômico e água destilada ultra pura estéril para completar um volume final da reação de 10 μ L. As condições de ciclagem de amplificação foram: 94°C por dois minutos, 35 ciclos a 94°C por 45 segundos, 63°C por 30 segundos e 72°C por 45 segundos, 72°C por sete minutos e 10°C ∞ . Ao volume de 10 μ L do produto amplificado foi adicionado 5 μ L da Enzima ExoSap – IT (GE), com o intuito de remover os dNTPs e os iniciadores utilizados na reação de amplificação. A reação foi incubada em temperatura de 37°C por 30 minutos, seguido pela inativação enzimática a 80°C por 15 minutos.

Reação de sequenciamento direto: A reação de sequenciamento pelo método de Sanger foi realizada com a adição de 1,5 μ L 5x Sequencing Buffer BigDyeTerminator v1.1, v3.1, 5 μ M do iniciador, 1 μ L de BigDye® Terminator v3.1 (AppliedBiosystems, Foster City, CA, EUA), utilizando-se os pares de iniciadores específicos para cada fragmento amplificado separadamente. As condições da reação de sequenciamento foram: desnaturação a 95°C por 2 min, seguidos de 40 ciclos a 95°C por 18 seg, 55°C por 18 seg e 60°C por 4 min. As amostras foram ressuspendidas em 13 μ L de formamida Hi-Di (AppliedBiosystems, Foster City, CA, EUA), desnaturadas a 95°C em termociclador por 5 min e imediatamente transferidas para recipiente com gelo por 3 min. As reações de sequenciamento foram realizadas no sequenciador ABI Prism 3130 xl (AppliedBiosystems, Carlsbad, CA, EUA). Os eletroferogramas foram visualizados utilizando CLC Main Workbench 5.0.2 (CLC bio).

5. RESULTADOS

5.1 Análise dos dados brutos

Os dados gerados pelo sequenciamento resultaram em 164 gigabytes. Após a transferência via *download* para o servidor do laboratório NeoGene, na Faculdade de Medicina de Botucatu – UNESP, as análises de limpeza foram iniciadas com o programa seqclean, seguidas do mapeamento, deduplicação e identificação das variantes. Na Tabela 4 está disposto um breve resumo das análises geradas para cada etapa.

Considerando que o sequenciamento realizado é do tipo *paired-end*, há sempre dois *reads* de no máximo 100 pb para cada fragmento, de modo que para dar continuidade às análises, estes devem ser pareados com o intuito de formar fragmentos maiores de cerca de 200 pb. Isto é fundamental para o adequado alinhamento das sequências na fase do mapeamento. A porcentagem de *reads* pareados foi superior a 60% para todas as amostras, conforme descrito na Tabela 4. Também é possível observar o número de reads que foi mapeado para cada amostra, ressaltando que houve similaridade quase que total entre os dados de exoma e o genoma de referência. Há uma pequena parcela (menos de 1%) das sequências que resultaram em *singletons*, que representam aqueles reads que não foram associados a nenhuma região do genoma humano. Contudo, foram detectados *reads* que não puderam ser pareados denominados *single-end*, isto ocorre possivelmente por artefatos do sequenciamento. Em conjunto, os *reads paired-end* e *single-end* compõem o total de sequências mantidas após a etapa de limpeza dos dados. A cobertura média alcançou 95x.

Tabela 4: Resumo dos resultados de limpeza, mapeamento dos *reads* e número de variantes detectadas após o sequenciamento do exoma dos pacientes com carcinomas de mama e tireoide.

Amostra	Nº de <i>reads</i> antes da filtragem	Nº de <i>reads</i> após filtragem	% <i>reads</i> mantidos após a limpeza	Nº <i>reads</i> mapeados	% de <i>reads</i> alinhados	Cobertura	Nº de alterações pós filtragem
1	61808526	49172136	79,56	49146954	99,95	73	113183
2	113231188	86175913	76,11	86124313	99,94	125	149237
2T	88438680	66712462	75,43	66652261	99,91	96	137967
3	58937096	45540868	77,27	45507131	99,93	95	99947
4	85848314	70834261	82,51	70798169	99,95	104	137751
5	65183174	53349989	81,85	53320133	99,94	78	119248
6	113826888	91805771	80,65	91745049	99,93	134	166461
6T	84029714	65794036	78,30	65744786	99,93	95	147212
7	115187504	92709903	80,49	92635920	99,92	135	155736
7T	83624504	65349704	78,15	65298144	99,92	94	139024
8	82697560	67631040	81,78	67108908	99,94	98	128247
9	78483750	60381854	76,94	60343142	99,94	87	118862
10	95193892	71548219	75,16	71500651	99,93	103	131563
11	65021448	50857864	78,22	50719942	99,73	74	118275
12	96226998	74080057	76,98	74033413	99,94	107	132503
13	74833528	56921557	76,06	56884340	99,93	82	125113

14	67325310	52291543	77,67	51985508	99,41	75	109587
15	70411748	54770904	77,79	54495443	99,50	79	111517
16	64875638	50129780	77,27	50089143	99,92	72	110966
17	81739842	64910827	79,41	64868874	99,94	117	131950
18	86957398	65667457	75,52	65615220	99,92	94	145304
19	99002690	80599873	81,41	80559417	99,95	117	145191
20	91008000	74699499	82,08	74664175	99,95	109	139827
21	72693782	58936277	81,07	58901967	99,94	87	129096
21T	70242076	54864670	78,11	54825295	99,93	80	122193
22	69319264	56541469	81,57	56512031	99,95	83	117227
23	80907694	66070242	81,66	66034774	99,95	97	132453
24	85074351	66136801	77,74	66081473	99,92	95	121332

T: amostra do DNA tumoral.

5.2 Genes de alta e baixa penetrância associados ao CT e CM hereditários

Foram investigados 63 genes de alta e baixa penetrância, previamente descritos em literatura (Tabela 2) associados aos carcinomas de mama e tireoide familiares, nas 24 amostras de sangue periférico dos pacientes incluídos nesta etapa do estudo. Os genes de alta penetrância e associados aos CT e CM, bem como os genes descritos em síndromes de predisposição ao câncer foram os primeiros a serem avaliados. Embora algumas pacientes houvessem sido testadas para os principais genes relacionados ao câncer de mama, como *BRCA1* e *BRCA2*, estes genes foram avaliados em detalhes pela análise de SET. Na Tabela 5 estão descritas as 12 alterações mais frequentes presentes em pelo menos três indivíduos e as suas possíveis associações já descritas em literatura.

Todas as variantes descritas na Tabela 5 foram consideradas comuns de acordo com o banco de dados 1000 Genomas, uma vez que as frequências alélicas foram acima de 0,05 para todos estes SNPs. A maioria das variantes foi considerada benigna ou não patogênica, contudo dois dos SNPs mapeados no gene *BRCA1* foram caracterizados como possivelmente deletérios, rs16941 e rs799917, de acordo com os bancos de dados SIFT e Polyphen.

Tabela 5. Relação dos SNPs mapeados em genes de alta penetrância entre os 24 casos avaliados.

SNPs	Gene	Relevância Clínica	1000 Genomas	Patogenicidade (SIFT/PolyP)	Base	Frequência	Pacientes portadores
rs1799939	<i>RET</i>	CT (Santos <i>et al.</i> , 2014)	0,15	Benigna	G>A	38%	1,3,8,11,15,17, 19, 22,24
rs1799977	<i>MLH1</i>	Síndrome de Lynch e Câncer Colorretal (Nejda <i>et al.</i> , 2009)	0,17	Benigna	A>G	58%	1,2,5,7,8,10,11, 13,18,19, 20,21,23,24
rs1805321	<i>PMS2</i>	Não	0,39	Benigna	G>A	50%	5,6,8,12,13, 16,17,18,19, 22,23,24
rs2228006	<i>PMS2</i>	Não	0,88	Benigna	T>C	63%	1,3,4,6,9,10,11, 12,13,16,17,21, 22,23,24
rs2959656	<i>MEN1</i>	Não	0,84	Benigna	T>C	96%	Todos, exceto caso 18
rs1042522	<i>TP53</i>	Linfoma Não-Hodgkin (Fan <i>et al.</i> , 2014) OMIM:191170#0005)	0,62	Benigna	G>C	96%	Todos, exceto o caso 6
rs144848	<i>BRCA2</i>	Sem associação com Câncer de Mama e Ovário (Ramus <i>et al.</i> ,	0,24	Benigna	A>C	29%	2,4,5,6,12,17, 20

2008); (Hasan <i>et al.</i> , 2013)							
rs169547	BRCA2	Não	0,97	Benigna	T>C	58%	2,6,7,9,10,11,12,14,16,17,18,19,20,23
rs1799966	BRCA1	Não	0,37	Benigna	T>C	58%	4,5,6,11,12,13,14,15,17,18,20,21,22,23
rs16942	BRCA1	Associado à CM, porém com um risco diminuído (Cox <i>et al.</i> , 2011)	0,32	Benigna	T>C	42%	4,5,6,11,12,15,17,18,21,23
rs16941	BRCA1	CPT (Wojcicka <i>et al.</i> , 2014).	0,30	Possivelmente deletéria	T>C	25%	5,6,12,20,21,22
rs799917	BRCA1	Associado ao CM, porém com conclusões controversas (Qin <i>et al.</i> , 2014)	0,48	Deletéria/Benigna	G>A	50%	4,5,6,11,12,13,17,18,19,20,21,24

Em nosso estudo, também foram investigados 47 genes com baixa penetrância (Tabela 2), relatados como alterados em pacientes com história familiar de câncer, alguns também associados ao CT esporádico. Os genes de baixa penetrância, apesar de importantes, parecem ter um papel sutil na carcinogênese. Porém quando diversas alterações ocorrem de forma acumulativa, o efeito destas mutações pode ser potencializado conferindo um alto risco de desenvolvimento de doenças como o câncer. Na Tabela 6 estão descritas as alterações encontradas em 15 genes de baixa penetrância e a *C8orf4*.

Tabela 6. Alterações mapeadas em genes de baixa penetrância previamente descritas como associadas com CM e CT.

SNPs	Gene	Frequência	Relevância Clínica	1000 Genomas	Patogenicidade (SIFT/PolyP)	Troca Base	Pacientes portadores
rs918949	TNS1	67%	Não	0,62	Benigna	C>T	1,2,3,4,5,6,8,10,12,13,14,16,18,19,20,24
rs3796028	TNS1	63%	Não	0,39	Benigna	G>A	1,3,4,7,8,9,12,13,15,17,20,21,22,23,24
rs2571445	TNS1	88%	Função pulmonar, possivelmente migração celular (Repapi <i>et al.</i> , 2010)	0,66	Benigna	A>G	2,3,4,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20,21,23,24
rs3796033	TNS1	63%	Não	0,38	Deletéria/Benigna	G>A	4,7,8,9,11,13,14,15,17,18,20,21,22,23,24
rs3747479	MRPS30/L OC100506 674	50%	Não	C: 0,41 G: 0,59	Benigna	G>C - G>T(1)	1,3,5,7,10,12,15,16,20,21,23

rs832582	MAP3K1	88%	Doença de Crohn (Hancock <i>et al.</i> , 2008)	0,70	Benigna	G>A	1,3,4,5,6,7,9, 10,11,12,13, 14,15,16,18, 19,20,21,22, 23,24
rs1143684	NQO2	100%	Melhor prognóstico em CM (Jamieson <i>et al.</i> , 2011)	0,98	Benigna	C>T	Todos
rs621679	LSP1	75%	Não	0,42	Benigna	G>A	1,3,4,6,7,8,9, 10,11,13,14, 15,16,17,18, 20,21,24
rs659243	ATM	71%	Não	1	Benigna	A>G	1,2,4,6,7,9,10, 11,12,13,14, 15,17,18,22, 23,24
rs861539	XRCC3	58%	Alelo de risco para melanoma e progressão de CM (Smolarz <i>et al.</i> , 2015)	0,25	Benigna	G>A	1,4,5,6,8,9, 10,11,14,17, 18,22,23,24
rs10510592	NEK10	42%	CM (Milne <i>et al.</i> , 2014)	0,21	Benigna	A>G	2,6,7,12,13, 17,18,19,20,22
rs299295	HMMR	38%	Não	0,25	Benigna	C>T	2,3,6,9,13, 14,18,19,23
rs2229571	BARD1	50%	Neuroblastoma (Capasso <i>et al.</i> , 2009)	0,49	Benigna	C>G	4,6,7,9,11, 12,14,15,16, 17,21,22
rs1048108	BARD1	46%	Neuroblastoma - regula negativamente a apoptose (Lee <i>et al.</i> , 2014)	0,35	Benigna/ Possivelmente de letéria	G>A	3,4,6,7,9,11, 14,15,20,21,22
rs2070094	BARD1	63%	Não	0,28	Benigna	C>T	1,4,6,7,9,11, 12,13,14,1,5, 16,17,29,21,22
rs4986764	BRIP1	67%	Câncer Cervical (Ma <i>et al.</i> , 2013)	0,67	Benigna	A>G	3,5,6,7,9,10, 12,13,14,15, 16,17,18,19, 23,24
rs17337252	RB1CC1	50%	Não	0,55	Benigna	A>G	7,9,10,11,13, 14,15,17,18, 20,21,24
rs370429	PCM1	29%	Esquizofrenia (Datta <i>et al.</i> , 2010)	0,11	Benigna	C>T	1,2,4,10,11, 12,23
rs208753	PCM1	92%	Não	0,72	Benigna	A>G	Todos, exceto os casos 1e 3
rs10761581	NCOA4	71%	Próstata (Wang <i>et al.</i> , 2011)	0,41	Benigna	T>G	1,3,4,5,6,7,8, 10,14,15,16, 17,18,19,20, 21,22
rs1053266	CCDC6	83%	Não	0,46	Benigna/ Desconhecida	G>T	1,2,3,4,5,6,7,9, 10,11,12,13, 15,17,19,20, 21,22,23,24

rs1040835	<i>GOLGA5</i>	88%	Não	0,82	Benigna	T>C	1,2,3,4,5,6,7,9, 10,11,12,13, 14,15,16,18, 19,20,22,23,14
rs6474226	<i>C8orf4</i>	79%	Não	0,99	Benigna	G>A	2,4,5,6,7,9,10, 11,12,13,14, 15,16,17,18, 19,21,22,24

Em relação aos genes descritos na Tabela 2, também foram buscadas as alterações consideradas altamente deletérias de acordo com os bancos de dados de predição de patogenicidade. Nesta busca, foi detectado que a paciente 14 era portadora de uma mutação no gene *PTEN* rs121909219 (p.R233X) (<http://www.ncbi.nlm.nih.gov/clinvar/variation/7813/>), compatível com o diagnóstico da síndrome de Cowden. Por falta de informações da história completa de câncer da família, a paciente não preenchia os critérios clínicos para Cowden. Uma análise mais detalhada revelou que a prima materna da paciente apresentou câncer de mama bilateral aos 43 e 48 anos, respectivamente. A mãe da paciente tinha história de carcinoma papilífero de tireoide e tumores de partes moles aos 65 e 45 anos e a avó materna apresentou câncer renal. Em conjunto, estes dados evidenciaram que a família preenchia os critérios da Síndrome de Cowden sendo portadora da mutação no gene *PTEN*. Este achado possibilitou que as alterações raras encontradas nesta amostra fossem utilizadas como uma referência para a busca de novos genes de predisposição. As alterações raras encontradas nesta paciente foram: *BTD* (rs13078881), *HSPA4* (rs61745470), *ZFPL1* (rs35251366), *PACS2* (rs147461490), *PNPLA7* (rs61747535), *TDRKH* (rs17853082), *RBM23* (rs72681969), *MRPL28* (rs11557302), *CNGB1* (rs112002818), *RLTPR* (rs78320020), *EMP3* (rs4893), *DCHS2* (rs61741046), *DST* (rs41271862) e que, portanto, não deveriam estar associadas como o espectro de alterações esperadas para o fenótipo em análise.

Foram encontradas outras variantes preditas como patogênicas por análise *in silico* e também consideradas raras de acordo com os bancos de dados 1000 Genomas, 6500 Exomas e Exac. A maioria delas foi detectada em apenas um indivíduo, com exceção de rs4987188, rs17158558, rs28363218. Estas alterações não são utilizadas na prática clínica para diagnóstico por ainda não terem sua função e efeito esclarecido nas síndromes de predisposição ao câncer (Tabela 7).

Tabela 7. Alterações mapeadas em genes clássicos que são preditas como patogênicas pela análise *in silico*, porém com associação fenotípica desconhecida de acordo com os bancos de dados CLINVAR, LOVD e BIC.

SNP	Gene	MAF	Caso	Bancos de Dados de Predição*	Relevância Clínica em Publicações
rs28897745	BRCA2	≤ 0.0005	17	8	Pacientes com pólipos adenomatosos e CCR em idade jovem (Lefevre <i>et al.</i> , 2012)
rs28909982	CHEK2	≤ 0.0002	17	8	Desconhecida
rs4987188	MSH2	≤ 0.0114	9, 17	5	Pacientes com CM triplo negativas (Smolarz <i>et al.</i> , 2014)
rs750746034	MSH2	0.00002490	13	8	Desconhecida
rs150514959	DICER1	0.0002	21	3	Desconhecida
rs35001569	MLH1	≤ 0.0037	18	8	Afeta a função da proteína – análise <i>in silico</i> (Joy <i>et al.</i> , 2013)
rs63750449	MLH1	≤ 0.0038	18	8	Desconhecida
rs17158558	RET	≤ 0.022	11, 24	4	Desconhecida
rs143638171	APC	≤ 0.0024	20	4	Detectados em homens com CM (Deb <i>et al.</i> , 2014).
rs45551636	PALB2	≤ 0.0179	20	5	CM familiar (Leyton <i>et al.</i> , 2015)
rs200069356	PRKAR1A	0.00039	6	3	Desconhecida
rs28363218	RAD54L	≤ 0.003	6, 7	7	Desconhecida
rs730881383	ATM	0.00002471	24	7	Variante somática em câncer de ovário (Kanchi <i>et al.</i> , 2014)

*Número de bancos de dados de predição de patogenicidade dbfnsp (Sift, Polyphen2_HDIV; Polyphen2_HVAR; LRT; MutationTaster; FATHMM; RadialSVM; LR; GERP++; PhyloP; SiPhy); MAF: Frequência do alelo menor (*minor allele frequency*) de acordo com os bancos de dados 1000G, 6500E e ExAC.

Um dado interessante foi a detecção da variante rs750746034 no gene *MSH2* (p.P27L), a qual foi detectada em heterozigose no paciente 13 sendo considerada patogênica em 8 bancos de dados. A paciente 13 foi diagnosticada com CT aos 36 anos e possui histórico de carcinoma colorretal nos dois ramos da família, bem como outros tumores característicos da Síndrome de Lynch (segundo o *National Comprehensive Cancer Network*, Inc 2015) incluindo o pai com tumor gástrico aos 57 anos e um primo paterno com tumor de sistema nervoso central aos 52 anos. Esta paciente não possui outras alterações patogênicas e também não possui nenhuma das novas alterações selecionadas como candidatas. Desta forma, sugerimos que esta alteração pode ser uma nova variante associada à Síndrome de Lynch. Deverão ser realizados estudos futuros para a análise de segregação desta alteração em membros dessa família.

5.3 Alterações compartilhadas entre os casos

Foram investigadas outras variantes, além daquelas descritas em genes previamente associados às síndromes de predisposição ao câncer, com o intuito de identificar novos genes de predisposição, os quais podem contribuir com o desenvolvimento dos tumores de mama e tireoide. É importante ressaltar que estas variantes identificadas (Tabelas 5, 6 e 7) não são utilizadas na prática clínica.

Alterações presentes em pelo menos quatro amostras (17%) foram inicialmente selecionadas com o intuito de identificar aquelas compartilhadas entre os indivíduos avaliados. A paciente 14 não foi considerada por possuir a Síndrome de Cowden. Foi realizada a comparação dos resultados com aqueles depositados em bancos de dados públicos sendo excluídas as alterações com $MAF > 0,05$, com o objetivo de identificar apenas aquelas consideradas raras. Esta análise resultou na detecção de 911 SNVs. A seguir, as variantes foram classificadas de acordo com a região em que estão mapeadas e a função predita, procurando excluir todas as alterações intrônicas, intergênicas e sinônimas. Foram mantidas as alterações em regiões exônicas, em sítios de *splicing* e as classificadas como não-sinônimas ou *indels*, que alteram a matriz de leitura e são responsáveis por perda ou ganho de códon de parada (*stoploss* e *stopgain*). Esta análise resultou em 287 alterações consideradas como não-sinônimas ou *indels*. Destas 287 alterações, 50 foram classificadas como deletérias, segundo pelo menos três dos nove bancos de dados de patogenicidade utilizados nesta análise (Tabela 8).

Posteriormente, foram realizadas análises de interação proteica utilizando os 50 genes considerados deletérios, e uma análise de enriquecimento de vias utilizando os 287 com alterações não-sinônimas. Estas análises servem como uma estratégia de investigação para avaliar a interação dos genes e as vias em que atuam, auxiliando na seleção das variantes candidatas.

Tabela 8. Detalhamento das 50 alterações classificadas como patogênicas em pelo menos três dos nove bancos de dados investigados. A anotação ou as informações obtidas destas alterações foi realizada com o programa Annovar, utilizando os bancos de dados RefSeq e SNP138 (<http://annovar.openbioinformatics.org/en/latest/user-guide/download/#-for-filter-based-annotation>).

		FI		FII		Indivíduos não aparentados																				
		CT	CT	CM	MT	CT	CT	CT	CT	CM	CT	CT	CT	MT	CT	CM	MT	CM	CM	MT	CM	MT	CT			
Gene	snp138	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	21	22	23	24	Nº de indivíduos	
PLCB3	rs35169799	x	x	x							x						x					x			6	
ERAP2	rs17408150					x	x				x		x	x					x						6	
F5	rs6027				x		x				x		x					x							5	
KCNJ10	rs1130183												x	x		x	x						x		5	
DSC3	rs35630063						x						x				x					x		x	5	
OR51L1	rs61729748	x		x			x	x		x															5	
MRPL22	rs3749671	x	x	x	x									x											5	
MUS81	rs61754785						x			x		x								x		x			5	
SLC7A6OS	rs11548855						x	x									x	x					x		5	
TCN2	rs1131603	x							x	x		x											x		5	
CHST6	rs117435647	x		x	x		x											x							5	
RNASEH2A	rs62619782						x	x											x	x				x	5	
MC2R	rs28926178	x	x	x						x															4	
ISX	rs8140287													x		x	x		x						4	
SLC12A8	rs2993631							x		x				x					x						4	
ABCB8	rs17545756											x				x				x				x	4	
SLC39A12	rs72778328		x	x							x													x	4	
FARP1	rs61730892	x	x	x							x														4	
ADCK4	rs11538384				x														x			x		x	4	
DMWD	rs146577305	x	x	x										x											4	
ZC3H12D	rs61997220								x		x	x							x						4	
SAMD9L	rs151304501		x	x						x						x									4	
SVEP1	rs111245230							x					x	x				x							4	
NPHS2	rs61747728				x	x	x	x																	4	
SIGIRR	rs117739035	x		x						x								x							4	
OR51I2	rs75620804				x	x		x		x															4	
VWF	rs7962217				x		x							x						x					4	
RLTPR	rs117556162						x										x	x					x		4	
PLEKHA4	rs62127976		x				x					x								x					4	
PAX8	rs3188996				x	x																	x	x	4	
COL5A2	rs35852101				x	x		x														x			4	
RNF123	rs35620248							x	x		x											x			4	
TRERF1	rs11751765													x				x				x		x	4	
S1PR3	rs34075341		x														x	x					x		4	
IPPK	rs2277170	x	x																		x		x		4	
AVPR1B	rs35369693				x	x			x				x												4	
HSPG2	rs2229475					x		x						x						x					4	
NOC2L	rs35471880	x		x	x			x																	4	
PRKG1	rs34997494									x				x	x		x								4	
OR4D6	rs17500380	x	x	x					x																4	
DDIAS	rs61902276												x	x		x							x		4	
ERCC5	rs1047769	x		x							x			x											4	
L3HYPDH	rs34741399				x	x				x							x								4	
MAPT	rs63750072				x	x		x										x							4	
ZNF804A	rs112183442		x		x					x										x					4	
SULT6B1	rs45552433		x								x							x		x					4	
DNAJC13	rs149121829	x	x	x								x													4	
TTC21A	rs35934336	x	x											x						x					4	
PCSK1	rs6232		x										x						x					x	4	
OR2H1	rs17184086	x	x	x								x													4	

CT – Carcinoma de tireoide; CM – Carcinoma de mama; MT – Carcinoma de mama e tireoide diagnosticado no mesmo paciente; x – alterações distribuídas entre os indivíduos;

rosa: membros da família I (3 membros analisados); **azul:** membros da família II (2 membros analisados) e **verde:** apenas o propósito foi analisado.

5.4 Investigação de genes recentemente descritos como associados com CT hereditários ou com CM/CT

Recentemente dois genes foram descritos como de predisposição ao carcinoma de tireoide não medular hereditário, *SRRM2* (Tomsic *et al.*, 2015) e *HABP2* (Gara *et al.*, 2015) e um possivelmente associado com a predisposição aos carcinomas de mama e tireoide, *PARP4* (Ikeda *et al.*, 2015). Os três estudos utilizaram a técnica de SET e identificaram alterações de base única, raras e preditas *in silico* como patogênica, abrangendo p.G534E (*HABP2*), p.G496V e p.T1170I (*PARP4*) e p.S346F (*SRRM2*). Em função destes achados, estes genes foram investigados com mais detalhes em nossa casuística. Na tabela 9 estão descritas as alterações encontradas, todas preditas como patogênicas em pelo menos um banco de dados.

Seis pacientes (casos 24, 14, 7, 17, 16 e 2) apresentaram seis variantes raras no gene *SRRM2*, não tendo sido detectada a mesma variante em qualquer um deles. Destas seis variantes, c.G5801A (caso 7) não foi descrita em bancos de dados públicos.

A alteração p.G534E do gene *HABP2*, a mesma descrita em literatura, foi detectada nos casos 7 e 9, tendo sido investigadas em maiores detalhes (resultados descritos na seção 5.8).

No gene *PARP4* estavam mapeadas quatro alterações incluindo duas comuns, uma rara e uma ainda não descrita em bancos de dados públicos. A alteração comum rs2275660, com frequência alélica maior que 0,65 foi detectada em todos os pacientes avaliados, com exceção do caso 19. Já a variante nova c.A2078G, presente na paciente 20, é predita como altamente patogênica em 9 bancos de dados pela análise *in silico*. As demais variantes detectadas no gene *PARP4* foram preditas como patogênicas em apenas um banco de dados.

Tabela 9. Alterações encontradas nos genes recentemente descritos como associados aos CT e CM hereditários.

Gene	Troca de base	Troca de AA	ExAC	1000 Genomas	SNP138	Bancos de Dados de Predição*	Casos
<i>SRRM2</i>	c.G5990A	p.R1997H	0.0032	0.00199681	rs138447860	5	24
<i>SRRM2</i>	c.G6098C	p.R2033P	0.0072	0.00139776	rs117133016	5	14
<i>SRRM2</i>	c.G5801A	p.R1934H	0.000041	-	rs780832298	4	7
<i>SRRM2</i>	c.G3107C	p.C1036S	0.0028	0.00159744	rs72768765	2	17
<i>SRRM2</i>	c.C3527G	p.A1176G	0.0005	0.00179712	rs147634161	2	16
<i>SRRM2</i>	c.A986G	p.K329R	0.0022	0.00039936	rs144062317	1	2
<i>HABP2</i>	c.G1523A; c.G1601A	p.G508E; p.G534E	0.0222	0.0081869	rs7080536	9	7,9
<i>PARP4</i>	c.A2078G	p.Y693C	-	-	-	9	20
<i>PARP4</i>	c.G2695A	p.A899T	0.742	0.654153	rs2275660	5	Todas exceto 19
<i>PARP4</i>	c.G4966C	p.A1656P	0.3766	0.419928	rs7571	1	6,20
<i>PARP4</i>	c.A241G	p.I81V	0.0301	0.0123802	rs35200240	1	9

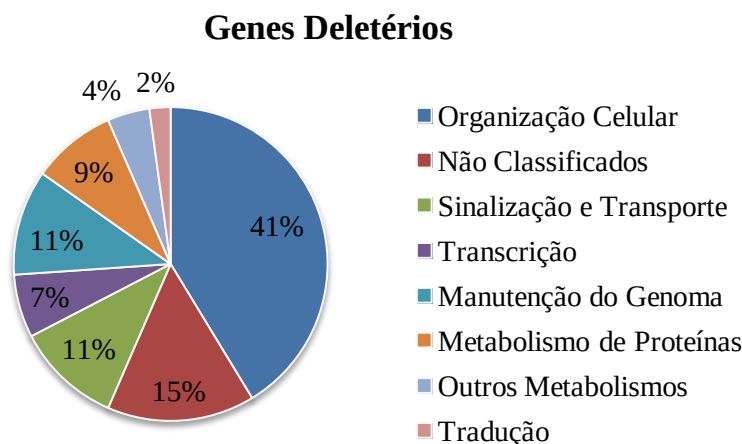
5.5 Análise de interação proteica

Para selecionar as alterações com maior impacto no fenótipo, foi realizada a análise de vias e interação proteica. Esta análise permite visualizar como as alterações identificadas se conectam, por meio das interações gênicas e proteicas em que estão inseridas. Deste modo, é possível relacionar as amostras por meio das alterações compartilhadas ou que se conectam, uma vez que a maioria das variantes não está presente em 100% dos casos.

Para esta análise foram utilizadas as 50 alterações classificados como patogênicas em pelo menos três dos nove bancos de dados investigados. O banco de dados I2D (<http://ophid.utoronto.ca/ophidv2.204>), juntamente com o programa Navigator (<http://ophid.utoronto.ca/navigator/index.html>), foram utilizados para identificar as interações proteicas. Quatro genes, *OR51I2*, *OR51L1*, *SLC39A12*, *SULT6B1*, não mostraram interação entre as proteínas por eles codificadas, resultando em 46 genes com pelo menos uma interação proteica. Estes 46 genes restantes permitiram agrupar as 22 das 23 amostras avaliadas nesta análise, com exceção da amostra 21. Os genes foram classificados em diferentes funções e processos celulares (Figura 4). Esta análise revelou que 41% dos genes atuam na organização celular e 11% na manutenção do genoma.

Figura 4. Classificação dos 46 genes com alterações deletérias e que possuem interação proteica. **A).** Porcentagem de genes em relação à sua ontologia. **B).** Descrição dos 46 genes que possuem interações de acordo com o banco de dados I2D.

A.

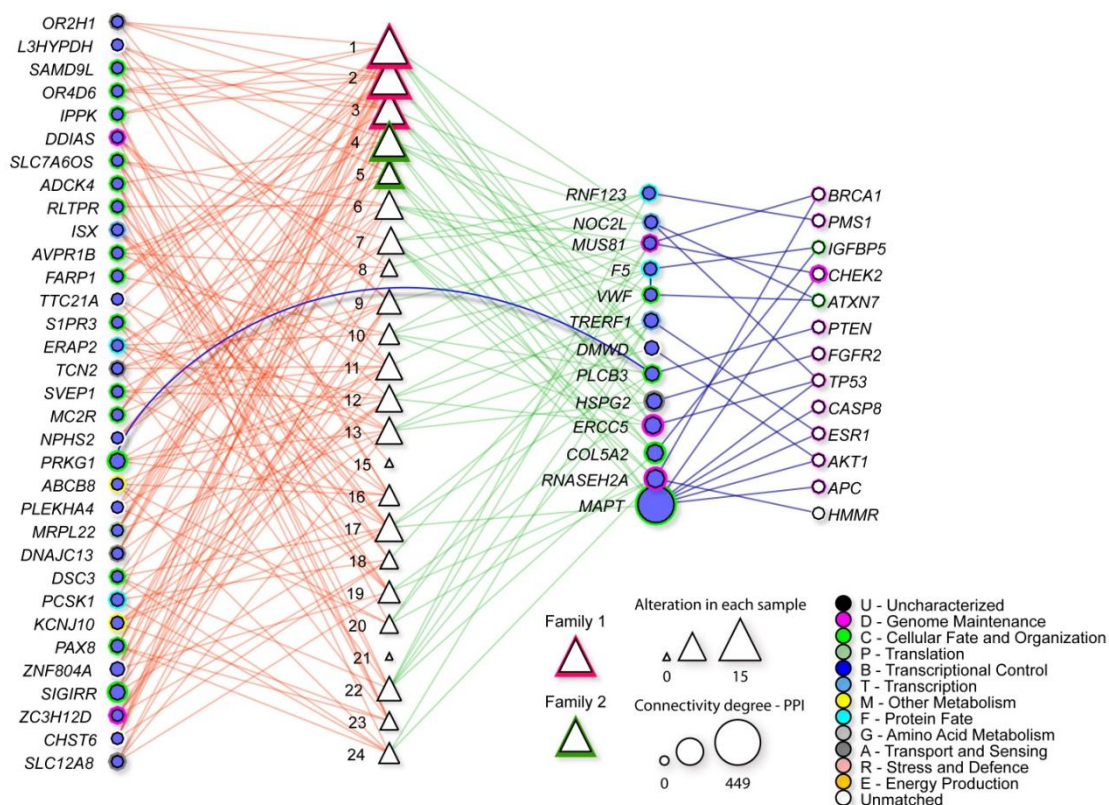


B.

Ontologia Gênica	Genes
Organização Celular	<i>IPPK, SIGIRR, PAX8, DSC3, COL5A2, OR4D6, PRKG1, MC2R, SVEP1, MAPT, FARP1, PLCB3, SLC7A6OS, ADCK4, S1PR3, SAMD9L, RLTPR, VWF, AVPR1B</i>
Não Classificados	<i>ZNF804A, DMWD, L3HYPDH, NPHS2, CHST6, PLEKHA4, TTC21A</i>
Sinalização e Transporte	<i>DNAJC13, HSPG2, SLC12A8, OR2H1, TCN2</i>
Transcrição	<i>NOC2L, ISX, TRERF1</i>
Manutenção do Genoma	<i>MUS81, ERCC5, DDIAS, ZC3H12D, RNASEH2A</i>
Metabolismo de Proteínas	<i>PCSK1, RNF123, F5, ERAP2</i>
Outros metabolismos	<i>ABCB8, KCNJ10</i>
Tradução	<i>MRPL22</i>

Com a análise de redes de interação foi possível observar que entre os 46 genes identificados, quatro possuíam interação proteica entre si. Esta interação pode ser observada na figura 5, onde o gene *PLCB3* interage com o gene *PRKG1*, da mesma forma que o gene *F5* interage com o gene *VWF*. Outra informação interessante foi que 13 das 46 variantes consideradas patogênicas possuem uma interação com genes previamente associados à tumores hereditários, sendo eles *BRCA1*, *PMS1*, *PTEN*, *TP53*, *APC*, *IGBP5*, *FGFR2*, *CHECK*, *ATXN7*, *CASP8*, *ERS1*, *AKT1* e *HMMR* (previamente descritos na Tabela 2).

Figura 5. Rede de interação entre os genes identificados com alterações deletérias e suas conexões com genes de alta e baixa penetrância descritos em síndromes hereditárias de predisposição ao câncer.



Triângulos: grupo de pacientes avaliados; **Círculos azuis:** genes que possuem alterações preditas como patogênicas em pelo menos três bancos de dados; **Círculos brancos:** genes associados às síndromes de predisposição ao câncer; **Destaques dos círculos:** classes a que pertencem os genes de acordo com GO (Gene Ontology); **Arestas verdes e vermelhas:** indicam as conexões entre as amostras e os genes que interagem ou não, respectivamente, com os genes previamente descritos em literatura associados à tumores hereditários; **Arestas azuis:** representam as conexões entre os genes encontrados alterados em nossa casuística e genes previamente descritos em literatura como associados à síndromes de predisposição ao câncer.

Os 13 genes alterados descritos como interagindo com genes de alta e baixa penetrância estavam presentes em 19 (82%) das 23 amostras avaliadas (Tabela 10). Os casos 15, 16, 21 e 23 não possuíam interações com os genes de alta e baixa penetrância.

Quatro casos (15, 16, 21 e 23) não apresentaram alterações mapeadas em qualquer um dos genes que interagem com os genes de alta a baixa penetrância, descritos na Tabela

10. A paciente 15 foi diagnosticada com dois tumores sincrônicos aos 67 anos, um CT papilífero e um CM ductal *in situ*. De maneira similar, a paciente 21 foi diagnosticada com tumores sincrônicos aos 50 anos, um CT papilífero e CM ductal *in situ*. A paciente 16 desenvolveu CT papilífero aos 49 anos de idade, e a paciente 23 desenvolveu o mesmo tipo tumoral aos 50 anos de idade.

Entre as 50 alterações patogênicas descritas na Tabela 8, a amostra 15 possuía apenas a variante rs61902276 mapeada no gene *DDIAS*. Já as pacientes 16 e 23 possuíam mais de uma alteração dentre todas as selecionadas. O caso 16 apresentou as variantes preditas como patogênicas (mínimo em três bancos de dados) nos genes *KCNJ10* (rs1130183), *ISX* (rs8140287), *L3HYPDH* (rs34741399), *PRKG1* (rs34997494), *S1PR3* (rs34075341), *SAMD9L* (rs151304501) e *ABCB8* (rs17545756) assim como a variante p.A1176G do gene *SRRM2*, descrita acima.

A paciente 23 apresentou as alterações patogênicas (mínimo em três bancos de dados) nos genes *KCNJ10* (rs1130183), *SLC7A6OS* (rs11548855), *TCN2* (rs1131603), *DDIAS* (rs61902276), *PAX8* (rs3188996) e *RLTPR* (rs117556162).

A paciente 21 foi detectada a alteração rara e predita como patogênica em 5 bancos de dados, rs4987188 do gene *MSH2*. Esta variante rara foi detectada na análise prévia realizada em genes de alta e baixa penetrância.

Tabela 10. Descrição dos 13 genes identificados com alterações deletérias, sua distribuição entre as amostras avaliadas (três membros da família I, dois membros da família II e os demais indivíduos não aparentados) e as interações com os genes associados com as síndromes de predisposição hereditária ao câncer.

Genes Alterados	Amostras	F.I			F.II		Indivíduos não aparentados																		Nº de amostras	Interação com Genes de Alta e Baixa Penetrância	Genes Interligados
		1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	21	22	23	24			
		CT	CT	CM	MT	CT	CT	CT	CT	CT	CM	CT	CT	CT	CT	MT	CT	CM	MT	CM	CM	MT	CM	MT	CT		
Genes Alterados	<i>PLCB3</i>	x	x	x							x						x					x			6	<i>PTEN</i>	-
	<i>F5</i>				x		x				x		x					x							5	<i>IGFBP5</i>	<i>VWF</i>
	<i>MUS81</i>						x			x		x								x		x			5	<i>BRCA1, PTEN</i>	-
	<i>RNASEH2A</i>						x	x											x	x				x	5	<i>BRCA1, HMMR</i>	-
	<i>HSPG2</i>					x		x						x					x						4	<i>FGFR2</i>	-
	<i>NOC2L</i>	x		x	x			x																	4	<i>TP53, ATXN7</i>	-
	<i>VWF</i>				x		x							x						x					4	<i>ATXN7</i>	<i>F5</i>
	<i>ERCC5</i>	x		x							x		x												4	<i>TP53</i>	-
	<i>MAPT</i>				x	x		x									x								4	<i>TP53, AKT1, CHEK2, CASP8, ESR1, APC</i>	-
	<i>DMWD</i>	x	x	x										x											4	<i>AKT1</i>	-
	<i>COL5A2</i>				x	x		x														x			4	<i>IGFBP5</i>	-
	<i>RNF123</i>							x	x		x											x			4	<i>PMS1</i>	-
	<i>TRERF1</i>											x					x					x		x	4	<i>ERS1</i>	-

CT – Carcinoma de tireoide; **CM** – Carcinoma de mama; **MT** – Carcinoma de mama e tireoide diagnosticado no mesmo paciente; **x** – alterações distribuídas entre os indivíduos; **rosa**: membros da família I (3 membros analisados); **azul**: membros da família II (2 membros analisados) e **verde**: apenas o propósito foi analisado.

5.6 Análise de enriquecimento

Com o intuito de evidenciar quais processos e vias metabólicas podem estar envolvidos com a possível predisposição hereditária dos carcinomas de mama e tireoide foi realizada uma análise de enriquecimento de vias utilizando os 287 genes que mostraram pelo menos uma alteração não-sinônima. Todas as alterações não sinônimas foram utilizadas para evidenciar o maior número de vias possível, as quais também podem desempenhar algum papel no desenvolvimento da doença em estudo, uma vez que a não-patogenicidade destas alterações é apenas predita em análises *in silico*. Utilizando a ferramenta PathDIP, desenvolvida pelo colaborador deste estudo, Igor Jurisica, 99 vias foram enriquecidas (p-valor < 0,05 disponível em anexo). A via de sinalização de síntese de hormônio tireoidiano também foi resultado significativo de uma análise realizada com o programa IPA (Qiagen), já a via de sinalização RAP1 foi resultado da análise gerada pelo programa IPA, porém com um p-valor de 0,064. A lista de vias de ambas as ferramentas PathDIP e IPA encontra-se no material suplementar. A Tabela 11 exemplifica as duas principais vias alteradas: via de sinalização Rap1 e a de síntese do hormônio tireoidiano, as quais foram selecionadas para estudo mais detalhado.

Tabela 11: Duas das 99 vias enriquecidas por PathDIP e IPA com potencial significado em carcinomas de mama e tireoide.

Vias metabólicas	P-valor	SNP Patogênicos	SNPs Não-sinônimos	% de genes com variantes patogênicas detectadas na mesma via
Sinalização de Rap1	0,0218	ITGAL, PLCB3, PDGFRB, SIPA1L3	ADCY2, PDGFD, TEK	57
Síntese do hormônio tireoidiano	0,0168	PLCB3, PAX8	DUOX2, NCOA1	50

A seguir, estão esquematizadas as duas vias selecionadas na Tabela 11. Os genes destacados em vermelho são os que possuem alterações patogênicas, e em azul, estão destacados os genes com alterações não-sinônimas.

Figura 6. Via de sinalização Rap1. Em vermelho e/ou azul estão indicados os complexos proteicos que possuem em sua conformação os genes com alterações patogênicas ou não-sinônimas, respectivamente.

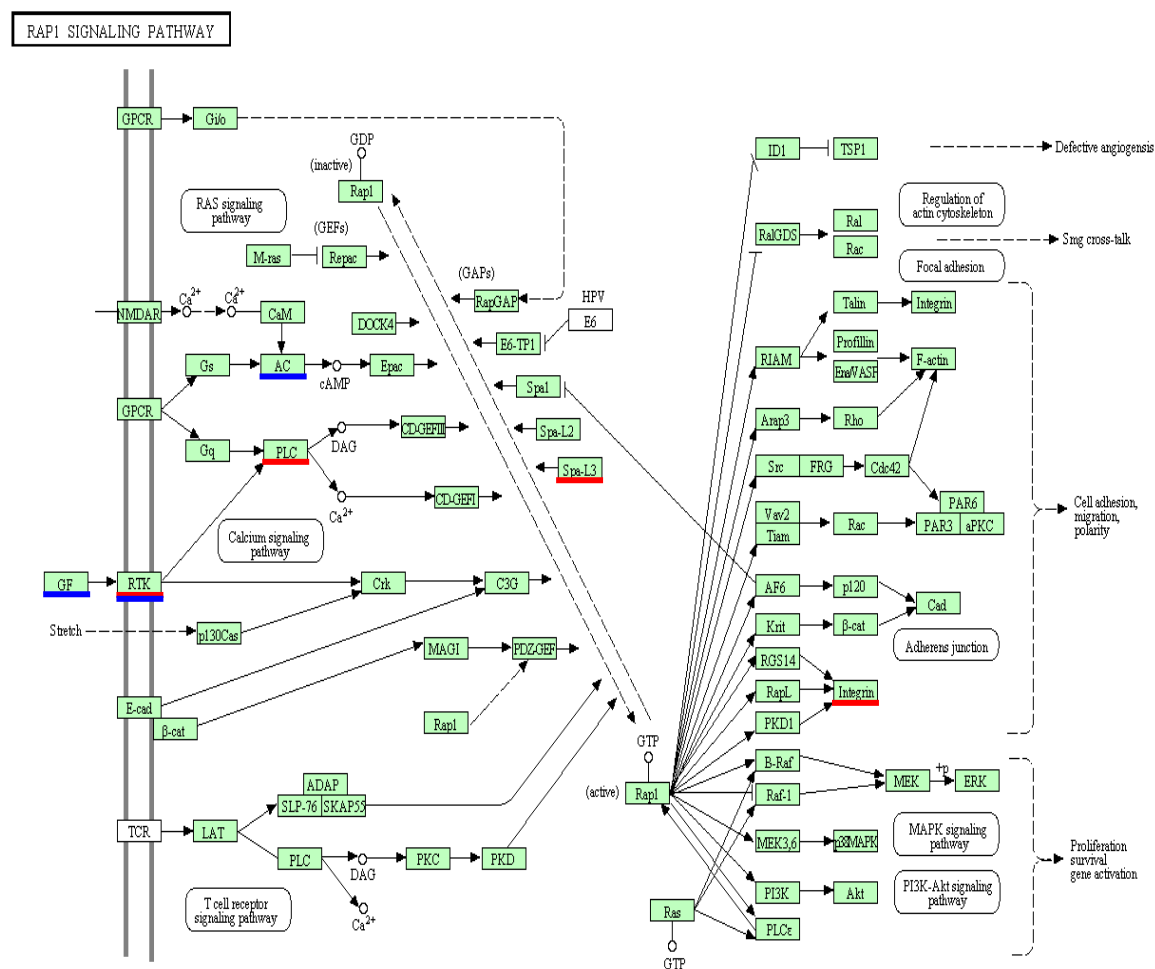
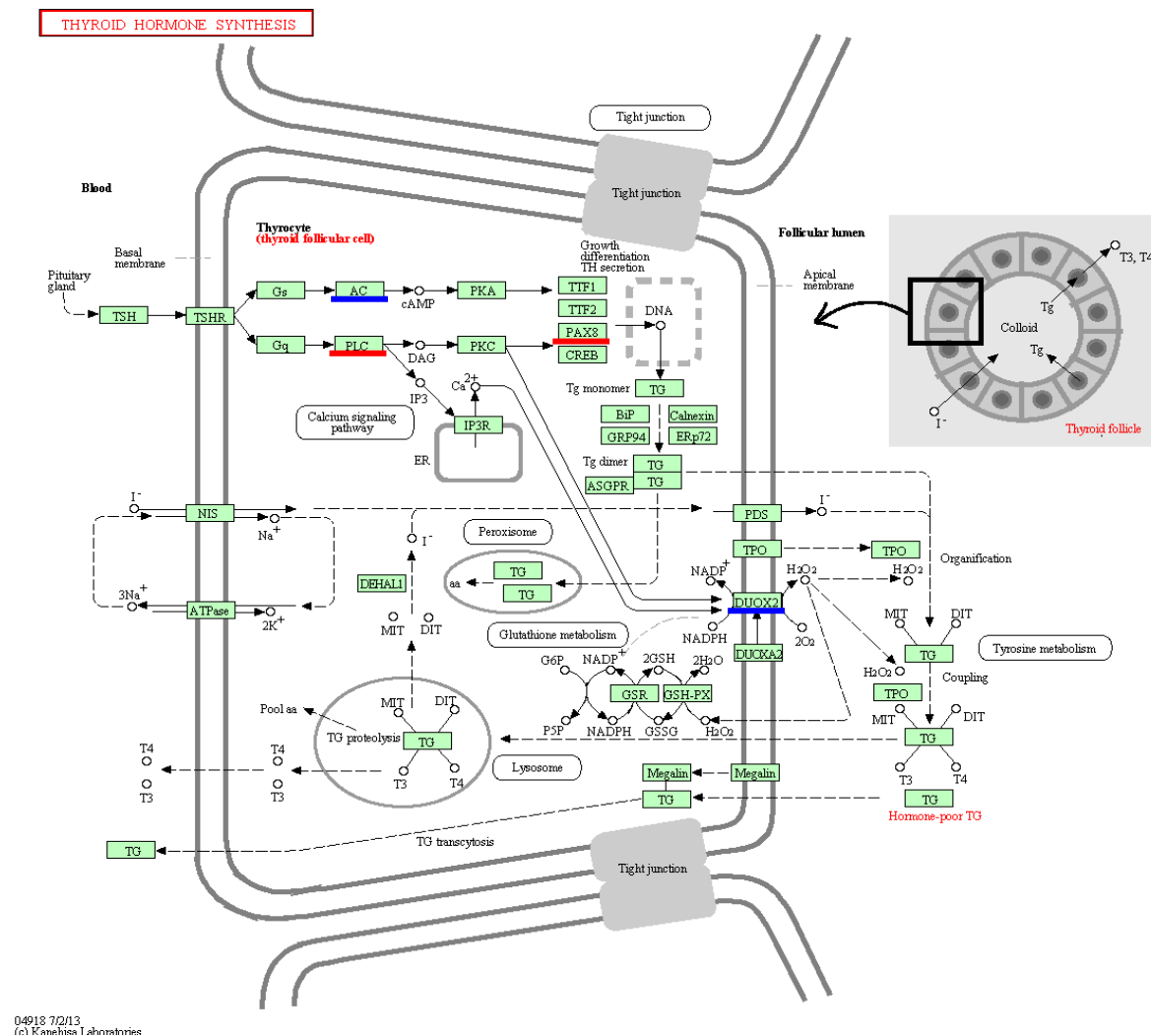


Figura 7. Via de síntese de hormônios da tireoide. Em vermelho e/ou azul estão indicados os complexos proteicos que possuem em sua conformação os genes com alterações patogênicas ou não-sinônimas, respectivamente.



É interessante ressaltar que, para a via de síntese de hormônio tireoidiano, as alterações em *PLCB3* e *PAX8* são mutuamente exclusivas, sendo que as únicas características compartilhadas entre os 10 pacientes portadores dessas alterações referem-se a vias potencialmente desreguladas e a presença de carcinomas de mama e tireoide.

5.7 Análise das famílias I e II

Foi possível realizar a análise de segregação das variantes genéticas em pacientes de duas famílias (I e II) onde mais de um parente foi também analisado por SET. As alterações foram consideradas relevantes quando foram categorizadas como patogênicos em pelo menos três bancos de dados (apresentados Tabela 8). Foi identificado um subgrupo de alterações mapeadas em 8 e 7 genes nas Famílias I e II, respectivamente (Tabela 12). Os heredogramas das famílias também estão representados na Figura 8. As alterações detectadas não foram compartilhadas entre as duas famílias.

Tabela 12: Genes e variantes genéticas detectadas nas famílias I e II e sua respectiva função.

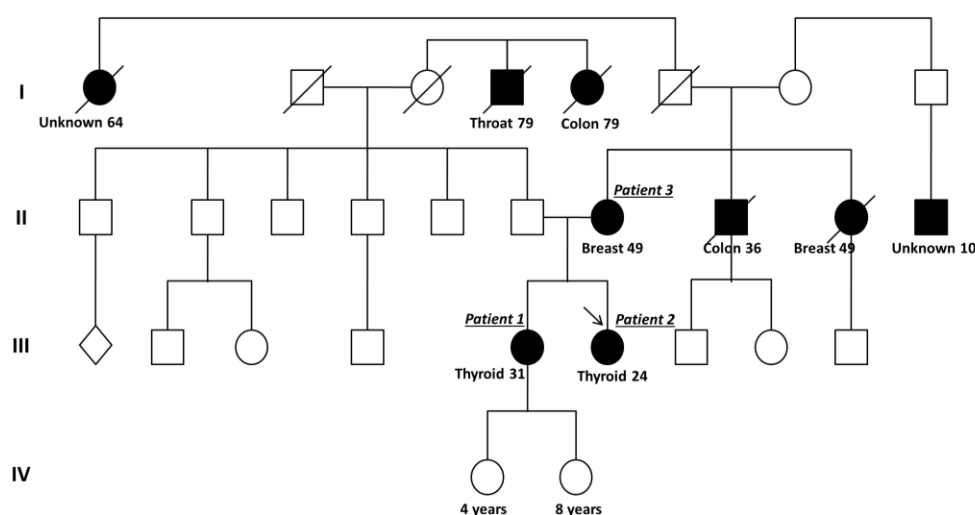
Gene	Nome	Função (NCBI)	Associação com doença
<i>PLCB3*</i> <i>rs35169799</i>	Phospholipase C-beta-3	Subunidade defosfolipase C beta (PLCβ); ativação de enzimas que abrem canais de cálcio.	Já associado a MEN1, mas excluído do espectro da síndrome (de Wit <i>et al.</i> , 1997)
<i>OR4D6*</i> <i>rs17500380</i>	Olfactory receptor, family 4, subfamily D, member 6	Interage com moléculas de odores, percepção de cheiros.	-
<i>DNAJC13*</i> <i>rs149121829</i>	DnaJ homolog subfamily C member 13	Trafego por endossomo e reciclagem de proteínas.	Doença de Parkinson (Gustavsson <i>et al.</i> , 2015).
Família I	<i>OR2H1*</i> <i>rs17184086</i>	Olfactory receptor, family 2, subfamily H, member 1	Interage com moléculas de odores, percepção de cheiros.
	<i>MRPL22*</i> <i>rs3749671</i>	Mitochondrial ribosomal protein L22	Síntese de proteína mitocondrial
	<i>MC2R*</i> <i>rs28926178</i>	Melanocortin 2 receptor	Deficiência de glicocorticoide (Aza-Carmona <i>et al.</i> , 2011).
	<i>FARP1*</i> <i>rs61730892</i>	FERM, ARH/RhoGEF and pleckstrin domain protein 1	Fator de troca de guanina, associado a proteínas que ligam citoesqueleto à membrana.
	<i>DMWD*</i> <i>rs146577305</i>	dystrophia myotonica, WD repeat containing	-
	<i>NPHS2#</i> <i>rs61747728</i>	Nephrosis 2, idiopathic, steroid-resistant (Podocin)	Regula permeabilidade glomerular.
	<i>AVPR1B</i> <i>rs35369693#</i>	Arginine vasopressin receptor 1B	Receptor de vasopressina arginina.
Família II	<i>L3HYPDH#</i> <i>rs34741399</i>	L-3-hydroxyproline dehydratase	Metabolismo de prolina (Visser <i>et al.</i> , 2012)
	<i>MAPT#</i>	Microtubule-	Polimerização dos microtubulos
			Biomarcador em CM

<i>rs63750072</i>	associated protein TAU	do citoesqueleto.	(Mihály <i>et al.</i> , 2013)
<i>PAX8</i> [#] <i>rs3188996</i>	Paired box protein Pax-8	Fator de transcrição Paired Box .	CT (Santarpia <i>et al.</i> , 2010)
<i>COL5A2</i> [#] <i>rs35852101</i>	Collagen, type V, alpha 2	Colágeno que compõem o colágeno fibrilar, localizado na matriz extracelular.	Câncer de ovário (Zhang <i>et al.</i> , 2014) Síndrome Ehlers– Danlos (Ritelli <i>et al.</i> , 2015).
<i>OR5112</i> [#] <i>rs75620804</i>	Olfactory receptor, family 51, subfamily I, member 2	Interage com moléculas de odores, percepção de cheiros.	-

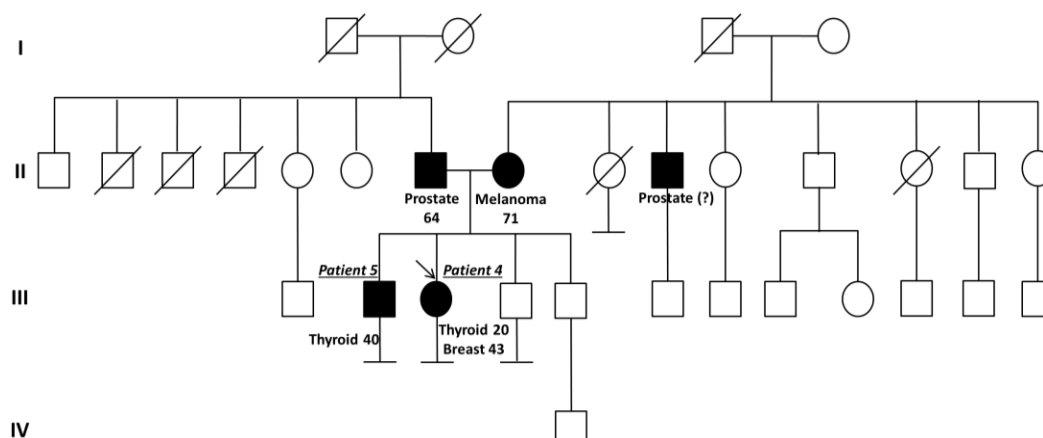
*: gene pertencente à família I, #: gene pertencente à família II.

Figura 8. Heredogramas das Famílias I e II cujas alterações estão descritas na Tabela 12

A. Família I. Foram analisados por SET os pacientes II-7, III-5 e III-6. Esta família compartilha alterações nos seguintes genes: *PLCB3*, *MRPL22*, *MC2R*, *FARP1*, *DMWD*, *OR4D6*, *DNAJC13* e *OR2H1*.



B. Família II. Foram analisados por SET as pacientes III-2 e III-3. Esta família compartilha alterações nos genes *PAX8*, *L3HYPDH*, *NPHS2*, *OR5112*, *COL5A2*, *AVPR1B* e *MAPT*.



As variantes que são compartilhadas na Família I foram encontradas nos três pacientes que tiveram seu exoma avaliado incluindo 1(III-5), 2(III-6) e 3 (II-7) e compreendem as alterações nos genes *PLCB3*, *MRPL22*, *MC2R*, *FARP1*, *DMWD*, *OR4D6*, *DNAJC13* e *OR2H1*. Na Família II, alterações nos genes *PAX8*, *L3HYPDH*, *NPHS2*, *OR51I2*, *COL5A2*, *AVPR1B* e *MAPT* foram detectadas em ambos os indivíduos avaliados incluindo paciente 4 (III-3) e 5(III-2).

5.8 Confirmação e Validação

Foram selecionadas as alterações *HABP2* c.1601G>A e *MUS81* c.1292G>A para serem investigadas em detalhes. Pelo fato de recentemente a alteração *HABP2* c.1601G>A (rs7080536), mapeada no gene *HABP2* ter sido descrita por Gara *et al.* (2015) como associada a tumores hereditários da tireoide, foi investigado em detalhes a presença desta alteração na casuística deste estudo. Esta alteração c.1601G>A é responsável pela troca de aminoácido no domínio catalítico da proteína, onde uma glicina é substituída pelo ácido glutâmico (G534E). A investigação detalhada do SNP rs7080536 mostrou a sua presença em dois indivíduos não aparentados (casos 7 e 9), cujos heredogramas estão esquematizados nas Figuras 9 e 11.

O paciente 7 apresentou homozigose para *HABP2* c.1601G>A tanto na amostra de DNA do sangue periférico quanto no tumoral e, interessantemente, o paciente relata que seus pais provém de casamento consanguíneo. O paciente do sexo masculino

desenvolveu CPT aos 41 anos tem história familiar de sobrinho com CPT aos 27 anos, prima materna com CPT aos 55 anos, tia materna com CM aos 87 anos cujas três filhas desenvolveram CT, CM e CM aos 55, 61 e 60 anos respectivamente. Sua mãe tem história de carcinoma de pele sem idade definida, tio materno com carcinoma de próstata aos 70 anos e tio materno com tumor de origem desconhecida. A irmã do probando teve bócio da tireoide aos 51 anos e um pólipó uterino benigno aos 54 anos. O paciente índice (IV.11) foi confirmado para a presença da variante *HABP2* c.1601G>A por sequenciamento de Sanger e seus familiares III-9, IV-9, IV-10, IV-13, IV-14, V-1 e V-2 também foram testados, como ilustrado na Figura 9. Como resultado para a variante *HABP2* c.1601G>A os indivíduos IV-9 e IV-11 foram homozigotos, os indivíduos III-9, IV-10, IV-13, V-1, V-2 foram heterozigotos, e o indivíduo IV-14 foi negativo para a mutação (Figuras 9 e 10).

Figura 9. Família do paciente 7 (IV.11), portador da alteração homozigota. **GG:** Homozigoto para o alelo normal; **GA:** Heterozigoto; **AA:** Homozigoto para a alteração.

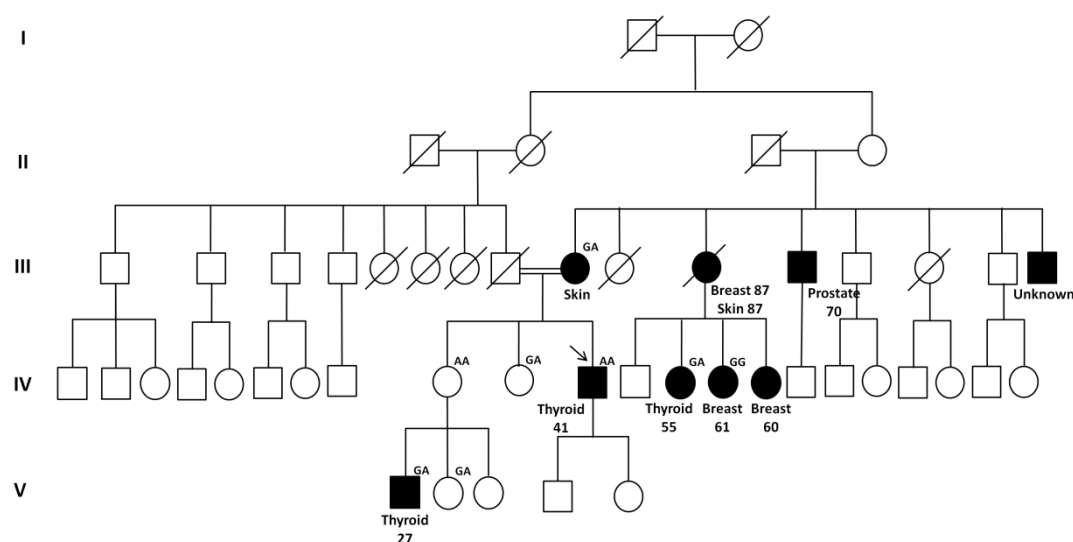
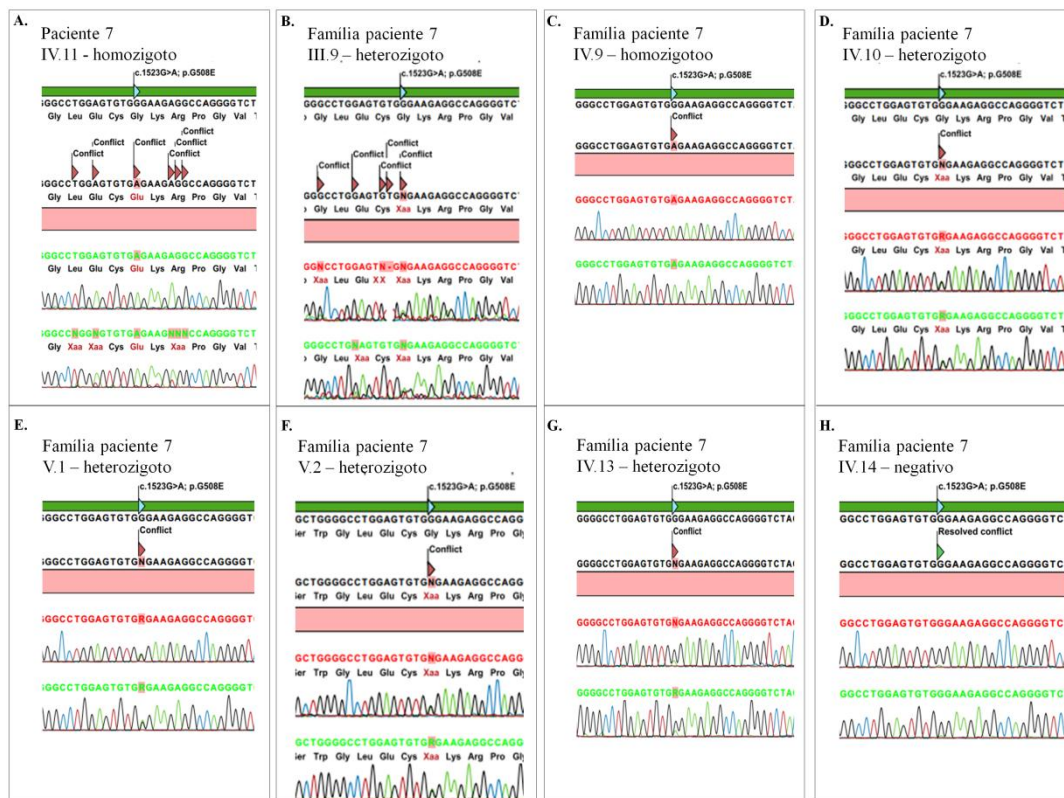


Figura 10. Confirmação e validação da alteração *HABP2* c.1601G>A (rs7080536). Eletroferogramas demonstrando a presença da mutação em homozigose e heterozigose nos membros da família do paciente 7 no gene *HABP2* (NM_004132:exon13).



A paciente 9 também é portadora da mesma alteração *HABP2* c.1601G>A, porém em heterozigose. A mesma foi diagnosticada com carcinoma de mama aos 45 anos e não apresentou CPT. Apresenta histórico familiar de câncer incluindo irmã com CPT aos 68 anos, irmã com CM aos 61 anos, sobrinha com CM e carcinoma de ovário em idade desconhecida, tia materna com CM sem idade definida e irmãos com lesões tireoidianas. A paciente 9 (III.2) também era positiva para a mutação no gene *MUS81* c.1292G>A (rs61754785), e por este motivo foi testada por sequenciamento de Sanger sendo confirmada. Da mesma forma, outros indivíduos da família III.4, III.6 e III.7, também foram testados para ambas as alterações *HABP2* c.1601G>A e *MUS81* c.1292G>A, conforme ilustrado nas Figuras 11 e 12. Como resultado para a alteração *HABP2* c.1601G>A os indivíduos III-2, III-4 e III-7 foram heterozigotos, e o indivíduo III-6 foi negativo para esta alteração. Em contrapartida, para a alteração *MUS81* c.1292G>A todos os indivíduos testados nesta família (III-2, III-4, III-6 e III-7) foram heterozigotos confirmados. Os indivíduos III.4 e III.7, apesar de não terem sido diagnosticados com a forma maligna da doença da tireoide, apresentaram adenoma folicular da tireoide e bócio colóide, respectivamente.

Figura 11. Família da paciente 9, portadora heterozigota da alteração no gene *HABP2* e no *MUS81*. **GG:** Homozigoto para o alelo normal; **GA:** Heterozigoto.

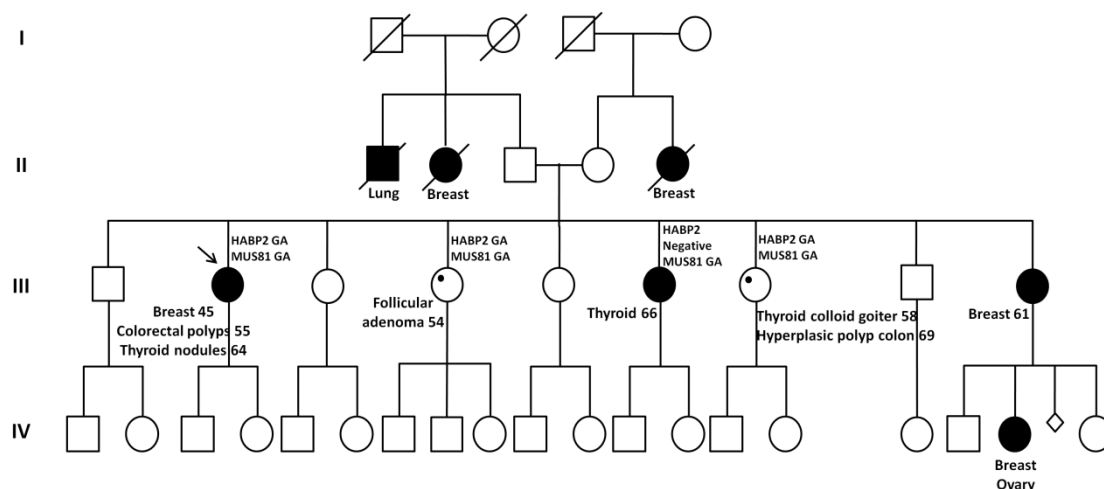
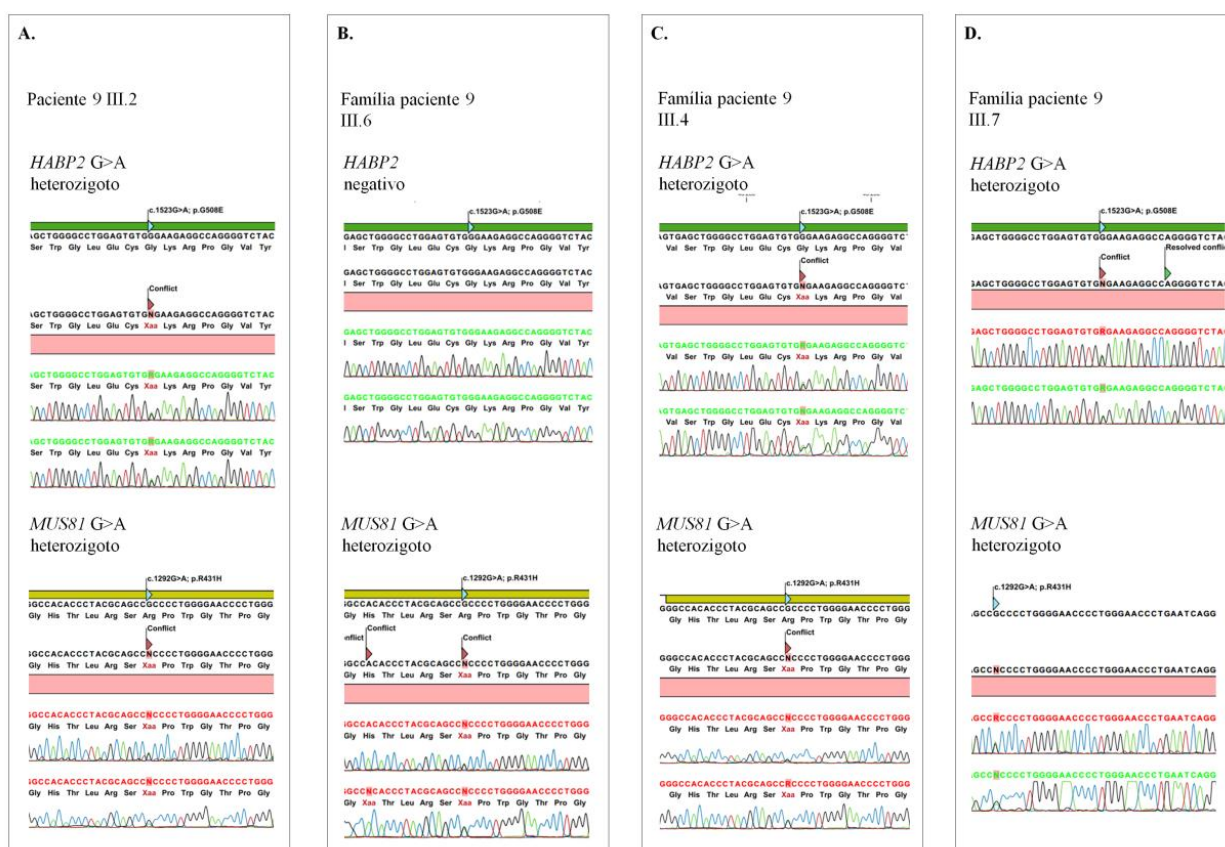


Figura 12. Confirmação e validação da alteração *HABP2* c.1601G>A e *MUS81* c.1292G>A. Eletroferogramas obtidos da análise de sequenciamento de Sanger nos genes *HABP2* (NM_004132:exon13) e *MUS81* (NM_025128:exon13).



6. DISCUSSÃO

O câncer é um problema de saúde pública mundial e por ser uma doença complexa o tratamento envolve um alto custo. Uma pesquisa realizada nos Estados Unidos pelo *National Cancer Institute* (NIH), em 2011, estimou que os gastos no país podem atingir U\$158 bilhões de dólares em 2020, se os custos para o tratamento, idade da população e predição da incidência da doença continuarem estáveis (Mariotto *et al.*, 2011). Em 2015 foi discutida a necessidade de implementação de um programa de prevenção ao câncer em nível mundial, assim como as organizações de prevenção a AIDS/HIV, com o intuito de diminuir em 25% os casos de doenças não reportadas para doenças passíveis de prevenção (Cavalli and Atun, 2015). Existem medidas básicas de prevenção à doença como alimentação saudável e prática de atividade física regular, evitar ingestão de álcool e tabaco e prevenção contra infecções causadas por vírus (HPV, HBV, HCV e *H. Pylori*). A triagem para a detecção precoce da doença também é válida e permite a detecção das fases iniciais da doença com maior taxa de sucesso ao tratamento (World Cancer Research Fund and Institute for Cancer Research, 2009; American Cancer Society, 2015).

A identificação de tumores em fase precoce é fundamental para aumentar as chances de sobrevivência do paciente. A busca de biomarcadores diagnósticos que auxiliem na identificação de pacientes de alto risco para o desenvolvimento de tumores é complexa, pois um marcador ideal deve ser facilmente mensurável, capaz de identificar a doença em estágio pré-sintomático e ser específico (Brennan and Wild, 2015). As alterações genéticas associadas à síndromes de predisposição ao câncer são utilizadas na prática clínica como marcadores e permitem identificar pacientes de alto risco, auxiliando na prevenção dos tumores hereditários não só do paciente índice, mas também de membros destas famílias.

A utilização destes tipos de alterações para o diagnóstico de várias síndromes hereditárias de predisposição ao câncer em pacientes de alto risco já é bem sucedida atualmente, principalmente em países desenvolvidos onde os inúmeros testes são disponíveis (Lammens *et al.*, 2010; Pacini *et al.*, 2010; Imyanitov and Byrski, 2012; Correa *et al.*, 2015;). Uma parcela de indivíduos pré-sintomáticos com alto risco de desenvolvimento de tumores possui interesse pelo teste genético, tanto quando há opções de tratamento, quanto em situações em que é limitado o número de opções de

conduta clínica. Saber sobre seu próprio risco e de seus familiares acaba trazendo segurança e, por sua vez, benefícios psicológicos (Lammens *et al.*, 2010; Taber *et al.*, 2015). Outros benefícios da aplicação destes testes na prática clínica é o entendimento da doença e dos riscos que esta traz para a vida destes pacientes.

Considerando a importância do diagnóstico precoce e o acompanhamento de pacientes com alto risco para desenvolver tumores, faz-se necessária a identificação de alterações com potencial para conferir predisposição hereditária aos carcinomas de mama e tireoide. A utilização de técnicas como análise de exoma vem demonstrando eficiência e eficácia na identificação desses novos marcadores (Roberts *et al.*, 2012, Gylfe *et al.*, 2013; Gara *et al.*, 2015). No entanto o reconhecimento de novas mutações causais em doenças hereditárias é complexo pela relativa alta taxa de mutação germinativa (em torno de $1,0 \times 10^{-8}$), dificuldade de compreender seu efeito molecular e sua contribuição em relação ao risco de desenvolver determinada doença (Shendure and Akey, 2015).

Parece existir uma associação molecular entre os tecidos mamário e tireoidiano que resultam no desenvolvimento de tumores. Pacientes que desenvolveram CT e CM tem um risco aumentado para o desenvolvimento de CM e CT como segundo tumor primário, respectivamente (An *et al.*, 2015; Joseph *et al.*, 2015; Kuo *et al.*, 2016). Outro dado interessante já descrito em literatura é que pacientes que desenvolveram CT seguido de CM e pacientes com ambos os tumores sincrônicos normalmente são positivos para os receptores de estrógeno e progesterona quando comparados à população em geral, evidenciando uma possível relação hormonal entre os tecidos glandulares (An *et al.*, 2015; Kuo *et al.*, 2016). Em nosso estudo, 12 pacientes desenvolveram CM, seis delas com CM (3, 9, 17, 19, 20 e 22) tinham história familiar de CT, uma paciente com CT teve como segundo tumor primário o CM (4), três pacientes apresentaram CM e CT sincrônicos (15, 18 e 21) e duas pacientes com CM desenvolveram CT como segundo tumor primário (14 e 23). O notável é que sete pacientes eram ER+PR+HER- corroborando com os resultados de An *et al.* (2016) e Kuo *et al.* (2016), dando indícios adicionais de que os tumores destas pacientes parecem sofrer uma influência destes hormônios em nível molecular, além de possuírem um melhor prognóstico. A paciente 23 era positiva apenas para o receptor de estrógeno, os dados de outras quatro pacientes não estavam disponíveis nos prontuários.

No presente estudo, foi avaliado o sequenciamento de exoma de 24 pacientes que apresentavam câncer de mama e ou tireoide e história familiar destes tumores. A análise de exomas trouxe inovação para a pesquisa de síndromes hereditárias. Em casos que não há evidência de mutações em genes clássicos, por exemplo, é possível investigar outro novo gene causal de alta penetrância ou alterações em novos genes de baixa penetrância que quando combinados contribuem para o fenótipo da doença de maneira cumulativa (Pinheiro *et al.*, 2015). Em nosso estudo, o principal objetivo foi identificar alterações que pudessem contribuir para o desenvolvimento destes tumores nestas pacientes e em seus familiares.

A identificação de novas alterações candidatas associadas ao risco da doença identificadas pela análise de sequenciamento do exoma passa por uma série de processos e etapas básicas incluindo limpeza, mapeamento, retirada dos *reads* duplicados e, por fim, a identificação e a estratificação das variantes. O processo de limpeza é essencial para a seleção das sequências que possuem uma alta qualidade e são passíveis de serem analisadas. O mapeamento é resultado da comparação das sequências geradas e o genoma de referência, de forma que os *reads* são alocados nas regiões do genoma de maior similaridade. Quanto maior a similaridade, maior será o *Score* de mapeamento (MapQ), onde há penalidade para cada diferença, *mismatch* ou *gaps*, ao passo que pontos positivos são atribuídos para as bases iguais ao genoma de referência. Deste modo, quando um *read* é comparado a uma região genômica é calculada uma pontuação, consequentemente será considerado como melhor alinhamento o que possuir mais alto MapQ, posicionando e conferindo à sequência alinhada a localização mais apropriada nos cromossomos (Langmead and Salzberg *et al.*, 2012).

O genoma utilizado como referência no presente estudo foi o hg19 (<http://hgdownload.soe.ucsc.edu/downloads.html#human>). Muitas vezes o pareamento das sequências pode não ser correto, o que pode ocorrer por vários fatores, incluindo contaminações, sequências que não tiveram sobreposição na formação dos *contigs*, ou até mesmo sequências que possuem elevado número de sobreposições, dificultando seu alinhamento, contudo é interessante mostrar que a qualidade de mapeamento das amostras aqui avaliadas foi alta e não foram observadas perdas significativas de sequências neste processo, contribuindo para uma alta cobertura das regiões genômicas (dados apresentados na Tabela 4).

Em seguida, foi realizada a retirada dos *reads* duplicados. Os *reads* duplicados são compostos por uma série de fragmentos que se iniciam e terminam exatamente na mesma base, podendo ser consequência das PCRs realizadas durante o preparo das bibliotecas, fragmentação do DNA no mesmo local, artefatos devido a regiões repetitivas e *background* na detecção de *clusters* na fase do sequenciamento. Esta limpeza é necessária para que a cobertura das regiões genômicas não seja superestimada e não haja impacto na acurácia de detecção das variantes.

Após a fase de duplicação, os *reads* foram realinhados com o genoma de referência pelos programas de chamada de variantes, ou *SNP calling*. Este realinhamento se faz necessário para a geração dos arquivos VCF. A partir desta comparação é possível identificar as bases ou sequências que diferem do genoma de referência. Estes arquivos compõem uma série de informações relevantes para a filtragem de qualidade das variantes, incluindo qualidade da base referente à alteração (PhredScore), qualidade de mapeamento da base (MapQ), cobertura do SNP, cobertura do sítio do SNP que corresponde à cobertura da região onde a alteração está localizada, informação sobre os alelos (homozigotos ou heterozigotos). Os dados obtidos nesta etapa demonstraram alta qualidade dos *reads* gerados (dados detalhados na Tabela 4).

Atualmente, há uma série de ferramentas disponíveis em literatura para auxiliar no estabelecimento de *pipelines* de análise de bioinformática aplicáveis à dados provenientes de sequenciamento de alto desempenho. Um estudo comparou 13 *pipelines* diferentes utilizando dois métodos de sequenciamento, Illumina e IonProton, quatro alinhadores, Novoalign, Bowtie2, BWA-MEM e Tmap (apenas para IonProton), e quatro programas de chamada de variantes, Freebayes, Samtools, GATK e TVC (apenas para IonProton). No resultado desta comparação, mesmo utilizando *pipelines* diferentes, os dados provenientes da metodologia Illumina demonstraram 91% de concordância, contra apenas 15% do IonProton, considerando apenas variantes confiáveis, de qualidade PhredScore >20. Outro dado interessante foi que a identificação de SNPs e *Indels* não pareceu ser influenciada com uma cobertura de sequências variando de 43,6x a 298,5x (Hwang *et al.*, 2015). Um outro estudo fez uma comparação entre a melhor metodologia para identificação de SNPs, considerando dados provenientes do sequenciamento de genoma total a uma cobertura média de 13x, e dados provenientes de sequenciamento baseado em captura de regiões específicas a uma cobertura média de 87x (Roux *et al.*, 2015). Os resultados indicaram que não há diferença significativa

entre o número de SNPs identificados em ambas as análises. Contudo os autores recomendam que a cobertura mínima para a identificação de variantes com confiabilidade seja de 30x (Roux *et al.*, 2015). Recentemente, Cherukuri *et al.* (2015), avaliando métodos de análise de exoma concluíram que uma cobertura mínima de 20x, com a média variando de 40 a 50x, é suficiente para geração de dados fidedignos e chamada de variantes confiáveis. Em nosso estudo, foi possível identificar variantes genéticas de qualidade, apropriadas para investigação proposta sendo obtidas a partir de uma cobertura média de 95x, variando entre 72x e 135x.

A partir da chamada das variantes, foram avaliadas as alterações detectadas em genes de alta e baixa penetrância, as quais já foram associados à síndromes clássicas de predisposição hereditária ao câncer. Pelo fato de apenas um subgrupo de pacientes terem sido investigadas para mutações no genes *BRCA1*, *BRCA2* e *TP53*, estes genes foram investigados em todos os casos. Entre os 63 genes investigados, foram detectados como alterados sete genes de alta penetrância (Tabela 5) e 17 genes de baixa penetrância (Tabela 6), consideradas comuns de acordo com os bancos de dados 1000 Genomas e 6500 Exomas e classificadas como benignas ou possivelmente deletérias. Foram também encontradas alterações raras de acordo com os bancos de dados 1000 Genomas e 6500 Exomas (frequência alélica menos que 0,05), em 10 genes de alta penetrância, consideradas patogênicas em pelo menos quatro bancos de dados de predição de patogenicidade *in silico*, presentes em pelo menos um ou dois indivíduos da nossa casuística (Tabela 7). Entretanto, as variantes detectadas não são utilizadas na prática clínica para o diagnóstico de síndromes hereditárias.

A variante rs16941 do gene *BRCA1* foi identificada em seis dos 24 casos avaliados, incluindo três pacientes com CT papilífero, dois com CM ductal invasivo e uma paciente com CT papilífero e CM sincrônicos. Recentemente a alteração rs16941 do gene *BRCA1* foi associada ao CT papilífero (Wojcicka *et al.*, 2014) e também descrita em um relato de caso envolvendo pacientes com forte histórico familiar para CM (Tedaldi *et al.*, 2014). Tedaldi *et al.* (2014) avaliaram uma probanda que aos 36 anos ainda não havia desenvolvido nenhum tumor. A análise de um painel de genes envolvendo *BRCA1*, *BRCA2* e *CHEK2* da paciente revelou uma duplicação no gene *CHEK2*, além de alterações em ponto no gene *BRCA1* (rs1799949, rs16940, rs799917, rs1060915, rs1799966) incluindo o SNP rs16941. Sua mãe, que havia sido diagnosticada com tumor de mama bilateral metacrônico aos 46 e 56 anos de idade,

também era portadora das mesmas alterações nos genes *CHEK2* e *BRCA1*. Outros membros da família que haviam desenvolvido tumores de mama, ovário e próstata foram testados e confirmados como sendo portadores da duplicação em *CHEK2* e não foram testados para a variante rs16941 do gene *BRCA1*. A variante rs16941 do gene *BRCA1* é considerada um polimorfismo comum no banco de dados do IARC (<http://brca.iarc.fr/>), além de possuir uma frequência alélica de 0,3 que é considerada comum quando comparado a outros SNPs detectados em nosso estudo. Além disto, o alelo C é considerado como patogênico apenas na predição *in silico* do programa SIFT. Os dados atuais não excluem a necessidade de estudos funcionais para comprovar o real impacto da mutação, contudo outras variantes identificadas aqui parecem ser mais promissoras em relação ao risco para CT e CM por serem raras e preditas como deletérias por mais de um banco de dados.

Entre os nossos achados, 12 das 24 pacientes testadas apresentaram o SNP rs799917 no gene *BRCA1* (Tabela 6). Há uma grande discussão em literatura com relação ao papel desta variante do gene *BRCA1* no risco de desenvolvimento do câncer de mama. Esta alteração é consequência da substituição de uma guanina por uma adenina, que resulta no códon alterado c.2612C>T levando a substituição do aminoácido prolina pela leucina, sendo considerada deletéria e benigna em dois bancos de dados diferentes. Os dados em literatura são controversos, incluindo um efeito moderado na atividade da proteína (Huo *et al.*, 2009; Hasan *et al.*, 2013). Em uma meta-análise, Qin *et al.* (2014) avaliaram oito estudos que investigaram a associação entre CM e o SNP rs799917, incluindo 19.878 indivíduos entre casos e controles. Após rigorosa análise estatística comparando a alteração em homozigose e em heterozigose, além de realizar comparações entre indivíduos caucasianos e asiáticos, os autores inferiram que a variante rs799917 do gene *BRCA1*, não está associada com o risco de câncer de mama nestas populações, uma vez que a correlação entre CM e a alteração não foi significativa em nenhuma das comparações realizadas. Contudo, os autores relataram que apenas 8 dos 47 trabalhos elegíveis preenchiam os critérios de seleção para a inclusão no estudo, os subgrupos populacionais utilizados não eram suficientemente grandes para gerar resultados estatisticamente significativos e foram considerados apenas artigos em inglês. Os autores concluíram que são necessários estudos funcionais para elucidar o efeito desta mutação (Qin *et al.*, 2014).

Em adição, a variante rs799917 do gene *BRCA1* também foi estudada em outros em 660 pacientes com câncer gástrico e 800 controles (Wang *et al.*, 2014). Indivíduos com o genótipo homozigoto rs799917 tinham um risco significativo para desenvolver o câncer gástrico (*odds ratio* de 1,40; 95 % CI = 1,17-1,68; $P = 2,2 \times 10^{-4}$). Os achados encontrados denotam a necessidade de estudos adicionais para melhor compreender o papel desta variante nos mecanismos celulares associados ao desenvolvimento tumoral.

Em 2014, Silva *et al.* investigaram 120 pacientes brasileiras não aparentadas, atendidas pelo departamento de Oncogenética do A.C. Camargo Cancer Center, mesma instituição que este estudo é desenvolvido. As pacientes com CM preenchiam os critérios da Síndrome de Câncer de Mama e Ovário Hereditários (HBOC). Entre as 120 pacientes avaliadas, 31 foram identificadas com alterações nos genes *BRCA1*, *BRCA2*, *TP53* (R337H) e *CHEK2* (1100delC), e em outras duas pacientes foram detectadas duas alterações estruturais nos genes *PTEN* e *ATM*. Uma nova alteração patogênica no gene *BRCA1* (c.560+2T>A) mapeada em um sítio de *splicing* foi detectada em sete pacientes. Um achado interessante foi em relação à taxa de detecção de mutações nos genes *BRCA1* e *BRCA2* em pacientes jovens: 35% em pacientes diagnosticadas com < 35 anos de idade, e entre 23% e 18%, em pacientes diagnosticadas entre 36 a 40 e >45 anos, respectivamente. Ou seja, quanto mais jovem, maior foi a taxa de mutação detectada (Silva *et al.*, 2014). Uma observação interessante é que nenhuma das alterações mapeadas nos genes *BRCA1* e *BRCA2* descritas nas Tabelas 5 e 7 do nosso estudo foram identificadas na casuística brasileira avaliada por Silva *et al.* (2014). Embora a casuística seja pequena, estes dados reforçam a hipótese que as pacientes com câncer de mama e tireoide fazem parte de um espectro diferente de mutações.

Assim como os genes de alta penetrância, todas as alterações encontradas em genes descritos como apresentando penetrância baixa foram categorizados como variantes comuns (frequência dos 1000G > 0,05), sendo a maioria delas benigna segundo os bancos de dados SIFT e Polyphen. Alterações possivelmente deletérias foram encontradas nos genes *TNS1* e *BARD1* em 63% e 46% das amostras, respectivamente. Para o nosso conhecimento, não há relatos de associação do SNP rs3796033 mapeado em *TNS1* com doenças. O gene *TNS1*, que codifica uma proteína localizada na membrana plasmática responsável por adesão à matriz extracelular, foi relatado como envolvido em mecanismos de adesão e migração celular, vias comumente desreguladas em células tumorais (Chen *et al.*, 2002).

O SNP rs1048108, mapeado no gene *BARD1* foi detectado em três pacientes com CT, em quatro com CM e em quatro com ambos os tumores. O SNP rs2229571, localizado no mesmo gene foi detectado em seis pacientes com CT, quatro pacientes com CM e dois com CT e CM. O SNP rs1048108 no gene *BARD1* foi identificado como um dos principais alterados em um estudo de GWAS (*Genome Wide Association*) em pacientes com neuroblastomas (Lee *et al.*, 2014). O SNP rs1048108 foi encontrado em desequilíbrio de ligação com o SNP rs6744811 e, segundo os autores, ambos podem interferir na interação entre *BARD1* e *BRCA1*, ação essencial para a manutenção do genoma que é desempenhada por *BRCA1* (Lee *et al.*, 2014). Os autores ressaltam a dificuldade em associar uma determinada alteração à um fenótipo específico e inferem que a análise de vias envolvendo alterações em ponto pode ser a chave para interpretar o real efeito de determinada alteração. Em nosso estudo, os indivíduos não possuíam história familiar de neuroblastomas. Estes achados não descartam a possibilidade que estas variantes contribuam para o fenótipo em estudo.

Capasso *et al.* (2009) relataram os mesmos SNPs rs2229571 e rs1048108 no gene *BARD1* como associados ao risco de desenvolvimento de neuroblastomas. Usando GWS em mais 189 casos e 1178 controles, os autores relataram seis SNPs localizados em regiões intrônicas como fortemente correlacionados com a doença, e estes encontravam-se em desequilíbrio de ligação com os SNPs rs2229571 e rs1048108. Estes dois SNPs também foram detectados como alterados em nosso estudo (casos 12 e 11, respectivamente). Assim, os autores ressaltam que alterações intrônicas são importantes na busca de genes de predisposição e que o gene *BARD1* e sua interação com *BRCA1* podem ter relevância no processo de supressão tumoral em CM hereditário. Em resumo, estes achados sugerem que as variantes rs1048108 e rs2229571 em *BARD1* são de relevância e importantes alvos para investigações futuras.

Alterações raras e preditas como patogênicas também foram encontradas em genes descritos como clássicos associados às principais síndromes de predisposição ao câncer (Tabela 8), sete não possuem informações quanto sua associação a qualquer fenótipo de doenças. Três destes SNPs já foram associados com risco aumentado de desenvolvimento de carcinomas de mama incluindo rs4987188 (*MSH2*), encontrado nas pacientes 9 e 17; bem como as variantes rs143638171 (*APC*) e rs45551636 (*PALB2*), detectados na paciente 20 (Deb *et al.*, 2014, Leyton *et al.*, 2015, Smolarz *et al.*, 2015). Um dado interessante é que as três pacientes (9, 17, 20) foram diagnosticadas com

carcinomas de mama e não de tireoide, sendo CT presente apenas na história familiar, o que sugere que estes SNPs podem contribuir para o risco aumentado de CM nestas pacientes. Além destes, a análise de busca de novos genes revelou outras alterações raras e preditas como patogênicas em pelo menos três bancos de dados para a amostra 9 nos genes *MUS81* (rs61754785), *SIGIRR* (rs117739035), *L3HYPDH* (rs34741399), *OR51I2* (rs75620804), *ZNF804A* (rs112183442), *MC2R* (rs28926178), *SAMD9L* (rs151304501), *OR51L1* (rs61729748), *TCN2* (rs1131603), *PRKG1* (rs34997494), *SLC12A8* (rs2993631); para a amostra 17 nos genes *SIGIRR* (rs117739035), *MAPT* (rs63750072), *PLCB3* (rs35169799), *S1PR3* (rs34075341), *SULT6B1* (rs45552433), *KCNJ10* (rs1130183), *SLC7A6OS* (rs11548855), *RLTPR* (rs117556162), *ISX* (rs8140287), *DSC3* (rs35630063), *SVEP1* (rs111245230), *TRERF1* (rs11751765); e a amostra 20 com alterações nos genes *VWF* (rs7962217), *IPPK* (rs2277170), *PLEKHA4* (rs62127976), *ABCB8* (rs17545756), *MUS81* (rs61754785), *RNASEH2A* (rs62619782). Em função da patogenicidade predita, todas estas alterações podem agir em conjunto funcionando como alelos de baixa penetrância e contribuindo para o aumento de risco de câncer nestes pacientes. Em especial as alterações rs61754785 (*MUS81*) que está presente nas amostras 9 e 20, e rs35169799 (*PLCB3*) presente na amostra 17, foram selecionadas como candidatas mais promissoras a partir dos resultados obtidos neste estudo. Sem dúvida, é necessária a avaliação funcional destas alterações assim como a análise em outros indivíduos para o cálculo de risco destas variantes e, desta forma ter subsídios para sugerir sua relação com a predisposição ao desenvolvimento de tumores.

Além das alterações descritas acima rs4987188 (*MSH2*), associada a CM (Smolarz *et al.*, 2015) e a variante rs35169799 (*PLCB3*), a paciente 17 ainda apresentou outras alterações nos genes clássicos, consideradas raras. Na Tabela 7 pode-se observar estas alterações: rs28909982 no gene *CHEK2* e rs28897745 no gene *BRCA2*, ambas raras e preditas como patogênicas em 8 bancos de dados. A alteração em *CHEK2* ainda não possui nenhuma associação fenotípica específica, entretanto o SNP em *BRCA2* já foi previamente associado com risco de desenvolvimento de pólipos adenomatosos e carcinoma colorretal em pacientes com idade jovem (Lefevre *et al.*, 2012). O caso 17 desenvolveu CM aos 41 anos e não possuía histórico pessoal ou familiar de carcinoma colorretal ou pólipos adenomatosos. Em conjunto, estes achados parecem indicar que as variantes encontradas na paciente 17 contribuam para o risco de desenvolvimento do CM.

A variante rs730881383 mapeada em *ATM* (Tabela 7) é responsável pela troca de uma arginina por uma glicina no códon 2459 (p.R2459G), sendo consequência da troca de base 7375C>G. Esse tipo de troca de aminoácido ocorre em 30% das mutações germinativas devido a alta taxa de mutação neste códon, cuja consequência pode ser interromper a função normal da estrutura da proteína, contribuindo para um fenótipo deletério (Shendure e Akey, 2015). Além disso, alterações na expressão da proteína codificada pelo gene *ATM* já foi associada com carcinomas de mama esporádicos que apresentam pior prognóstico (Bueno *et al.*, 2014). Especificamente o SNP rs730881383 foi relatado como uma alteração somática e germinativa em carcinomas de ovário (Kachi *et al.*, 2014), contudo ainda não foi relatado em síndromes hereditárias apesar de ser um forte candidato a validação pelo fato de o gene *ATM* ser considerado um supressor tumoral importante (Stagni *et al.*, 2014). Em nossa casuística, a paciente 24 possui esta alteração, bem como outras preditas como patogênicas em pelo menos três bancos de dados descritas na Tabela 7 (rs17158558-*RET*) e na Tabela 10 (*ABCB8*-rs17545756, *RNASEH2A*-rs62619782, *PAX8*-rs3188996, *ADCK4*-rs11538384, *SLC39A12*-rs72778328, *PCSK1*-rs6232, *DSC3*-rs35630063, *TRERF1*-rs11751765). A paciente 24 foi diagnosticada com CPT aos 40 anos, possuindo histórico familiar de irmã com CPT aos 34 anos, prima materna com CPT aos 32 anos, prima paterna com CPT aos 37 anos, tia materna com CM aos 58 anos, tia materna com carcinoma uterino aos 55 anos, pai com carcinoma de estômago aos 52 anos, primo paterno com carcinoma de cólon aos 60 anos, primo paterno com carcinoma de próstata aos 58 anos e prima paterna de 2º grau com linfoma aos 31 anos. Esta família tem uma grande heterogeneidade em relação ao tipo tumoral acometido, contudo o carcinoma de ovário não foi relatado até o momento. É de grande importância a investigação de outros membros da família da paciente 24 para a seleção das variantes associadas com o risco de desenvolvimento de tumores.

A variante rs750746034 (c.80C>T) do gene *MSH2* promove a substituição da prolina pela leucina (p.P27L). Esta alteração foi detectada na paciente 13 com CT aos 36 anos de idade e com um histórico familiar de carcinoma colorretal. Interessantemente, CT e CM já foram reportados em famílias positivas para mutação nos genes de reparo por mau pareamento e com características clínicas para a síndrome de Lynch (Buerki *et al.*, 2012; Pelizzo *et al.*, 2015). Assim como em nossa paciente (caso 13), Pelizzo *et al.* (2015) relataram dois indivíduos diagnosticados com CT

positivos para síndrome de Lynch, incluindo uma mulher aos 47 anos e um homem aos 34 anos. Ambos possuíam alteração em *MLH1* (c.1858G>T) e *MSH2*, (c.1704-1705delAG) respectivamente, diferentes das encontradas em nossa casuística, e o primeiro caso também tinha histórico de CM na família, assim como outros tumores do espectro clássico de Lynch (Pelizzo *et al.*, 2015). O tumor tireoidiano também foi reportado em pacientes com síndrome de Lynch em outros estudos (Broaddus *et al.*, 2004; Stulp *et al.*, 2008). O CM tem sido descrito como um novo tumor do espectro da Síndrome de Lynch. Buerki *et al.* (2012) avaliaram 70 famílias suíças positivas para mutação nos genes *MLH1* e *MSH2* cujo segundo tumor extracolônico mais frequente era o tumor mamário, o que também foi previamente relatado em um grupo de indivíduos brasileiros com diagnóstico de síndrome de Lynch (Silva *et al.*, 2010). Contudo, esta questão ainda é controversa e estudos se fazem necessários para elucidar quais os mecanismos genéticos estão envolvidos na associação entre CM e síndrome de Lynch (Win *et al.*, 2013). Por sua predição de patogenicidade ser alta, a variante rs750746034 em *MSH2* (p.P27L) aqui identificada é candidata a predisposição na Síndrome de Lynch, evidenciando o seu potencial efeito no desenvolvimento de CM e CT. Dentro desta perspectiva, é possível que a paciente 13 seja de fato portadora da Síndrome de Lynch, tendo em seu espectro câncer de mama e tireoide além de história familiar de carcinoma colorretal. Em futuro próximo, planejamos investigar outros membros da família para a presença desta alteração.

Todas as alterações discutidas até este ponto estão mapeadas em genes frequentemente mutados em síndromes de predisposição hereditária a cânceres e foram consideradas raras de acordo com os bancos de dados 1000 genomas e 6500 exomas, tendo sido também preditas como patogênicas por análises *in silico*. Entretanto, estas alterações foram consideradas não patogênicas e sem associação fenotípica pelos bancos de dados BIC, LOVD, CLINVAR e IARC. Estes bancos são utilizados como fonte de referência para a prática clínica, pois combinam as variantes submetidas ao fenótipo de determinada doença humana. As variantes detectadas em nosso estudo encontram-se em um ou dois pacientes, e algumas vezes compartilhadas entre vários casos (Tabela 8). São necessárias investigações funcionais, em um outro subconjunto de indivíduos para verificar se estas alterações podem conferir um risco aumentado para o desenvolvimento de tumores.

Apesar da identificação de alterações em genes altamente penetrantes, nenhuma das alterações detectadas, com exceção do SNP rs121909219 (p.R233X) mapeado no gene *PTEN*, é utilizado ao diagnóstico clínico das síndromes associadas ao CM e CT. A paciente 14 apresentou a mutação no *PTEN*, característica da Síndrome de Cowden, quando da análise de alterações nos genes clássicos associados com predisposição ao desenvolvimento de tumores. No entanto, a paciente 14 também apresentou outras alterações categorizadas como raras. Estes achados foram utilizados como uma referência possibilitando a diminuição do número de genes candidatos como associados com o fenótipo estudado. A detecção desta mutação no gene *PTEN* foi um importante achado que beneficia a família, pois com o diagnóstico confirmado será possível oferecer um aconselhamento genético adequado.

Foram também investigadas alterações em novos genes que pudessem conferir um risco aumentado na predisposição ao desenvolvimento dos tumores de mama e tireoide. Após a aplicação de uma série de filtros, foram identificadas 911 SNVs consideradas raras, com frequência alélica abaixo de 0,05 nos bancos de dados 1000 Genomas e 6500 Exomas, compartilhadas em pelo menos quatro indivíduos de nossa casuística. Atualmente não há um procedimento padrão para seleção de novas variantes com potencial de estarem associada a um fenótipo de doença mendeliana. Muitas vezes esta seleção é realizada com base nas informações de heredogramas, por exemplo, ou até mesmo relacionadas à função dos genes alterados. A seleção das alterações em pelo menos quatro indivíduos foi a abordagem escolhida para uma primeira análise, uma vez que era nosso intuito identificar o maior número de alterações possíveis compartilhadas entre os indivíduos. Foram detectadas variantes interessantes, mas isso não exclui a presença outras alterações que não se encaixem nestes critérios, as quais serão analisadas em um segundo momento. Após esta primeira análise outras alterações foram investigadas em genes recentemente descritos em literatura como associados ao fenótipo aqui estudado, e alterações interessantes foram encontradas (item 5.3 de Resultados), o que conota o dinamismo e versatilidade da análise de exomas.

Baseados nos critérios acima descritos, foram detectadas 287 alterações não-sinônimas, ou seja, alteram o aminoácido na cadeia polipeptídica, dentre as quais 50 são consideradas patogênicas em pelo menos quatro bancos de dados de predição *in silico*. Entre estas 50 alterações encontram-se genes com funções determinantes como organização celular, sinalização e transporte e manutenção do genoma. Dentre estes,

também há produtos proteicos de 13 genes mutados que interagem com genes de alta e baixa penetrância previamente associados com tumores hereditários (Figura 5).

Entre os genes envolvidos na organização celular (Figura 4), foram detectadas variantes nos genes *MAPT* (rs63750072) e *COL5A2* (rs35852101), ambas nos pacientes da família II (Tabela 10). O gene *MAPT* é responsável por codificar a proteína TAU a qual atua na polimerização dos microtúbulos do citoesqueleto e promove a estabilidade celular. Esta proteína também tem sido relatada como associada ao aumento da resposta ao tratamento quimioterápico em câncer (He *et al.*, 2014). Em 2013, uma meta-análise descreveu o gene *MAPT* como um importante candidato a biomarcador de resposta ao tratamento com tamoxifeno em câncer de mama (Mihály *et al.*, 2013). Nesse estudo também foi incluída a investigação dos principais biomarcadores descritos em literatura contidos em três bancos de dados de publicações da ASCO (*Journal of Clinical Oncology*), SABCS (*Cancer Research*) e PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). Esta busca resultou em 68 biomarcadores candidatos. Após esta seleção, os biomarcadores foram avaliados em dados de amostras provenientes em três bancos de dados públicos como GEO (Gene expression. omnibus, www.ncbi.nlm.nih.gov/gds), EGA (Europeangenome–phenomearchive, www.ebi.ac.uk/ega/) e TCGA (The CancerGenome Atlas, <http://cancergenome.nih.gov/>). Foram selecionados 582, 1208 e 23 pacientes, respectivamente, que haviam sido tratadas com tamoxifeno. A comparação e análise dos 68 biomarcadores nas amostras provenientes de bancos de dados resultou em 10 genes candidatos, incluindo *SLC7A5* e *MAPT* como mais promissores. Em conclusão, Mihály *et al.* (2013) sugeriram que alterações na expressão do gene *MAPT* auxiliava na resposta ao tratamento com tamoxifeno nas pacientes avaliadas. Em nossa casuística, duas pacientes (casos 4 e 17) portadoras de alteração rs63750072 realizaram tratamento com tamoxifeno. A primeira no período de 2008 a 2011 e até o momento (Dezembro de 2015) encontra-se sem evidência da doença. Não há informações detalhadas da paciente 17. Em adição, há a necessidade de investigar se essa variante está associada com alteração nos níveis de transcritos e que portanto, poderia dar evidência adicional demonstrando o papel do gene *MAPT* como um biomarcador de resposta ao tratamento com hormonioterapia. Portanto, esta alteração não parece estar associada com o risco aumentado de desenvolvimento de tumores de mama e tireoide em famílias.

A segunda alteração compartilhada na Família II (rs35852101) foi observada gene *COL5A2* (colágeno, tipo V, Alfa 2), o qual é responsável por codificar um tipo de colágeno que compõe o fibrilar, localizado na matriz extracelular e responsável por dar sustentação a vários tecidos. Este gene foi descrito com expressão desregulada em tumores de ovário (Januchowski *et al.*, 2014; Zhang *et al.*, 2014). Alterações neste gene também estão associadas à síndrome rara Ehlers–Danlos, que acomete 1 em cada 20.000 indivíduos e é caracterizada por hiperextensibilidade da pele, cicatrizes atróficas e hipermotilidade articular generalizada (Ritelli *et al.*, 2013). Para o nosso conhecimento, os pacientes portadores desta alteração não possuem características clássicas da síndrome de Ehlers–Danlos. Uma nova avaliação clínica destes pacientes encontra-se em andamento. As demais alterações compartilhadas entre os membros das famílias I e II serão discutidas a seguir. Neste aspecto, esta é uma evidência de outro benefício proporcionado pela análise de SET pois, se confirmada e comprovada a avaliação fenotípica destes indivíduos, a alteração rs35852101 pode agregar informação ao paciente e sua família. O mesmo ponto pode ser abordado em relação rs63750072 do gene *MAPT*, associando este achado à resposta ao tratamento com hormonioterapia.

A atividade de manutenção do genoma, associada a 11% dos genes alterados deste estudo (Figura 4), é um importante processo para o desenvolvimento tumoral em diversos aspectos. Os genes envolvidos na manutenção do genoma são potenciais candidatos a predisposição ao desenvolvimento de tumores. Como citado anteriormente, o produto de 13 dos 50 genes com alterações deletérias foram identificados interagindo diretamente com os genes de alta penetrância associados às síndromes clássicas descritas em literatura (Figura 5), como *BRCA1*, *CHECK2* (Síndrome de Câncer de Mama Hereditário), *TP53* (Síndrome de Li-Fraumeni), *MEN1* (Síndrome das Múltiplas Neoplasias Endócrinas), *PTEN* (Síndrome de Cowden), *PMS1* (Síndrome de Lynch) e *APC* (Polipose Adenomatosa Familiar). Igualmente, interagem com genes de baixa penetrância como *FGFR2*, *ATXN7*, *AKT1*, *IGFBP5*, *CASP8*, *HMMR* e *ESR1* (CM hereditário) (Shuen e Foulkes, 2011; Lalloo e Evans, 2012; OMIM:188550). Estes 13 genes são classificados como de organização celular ou manutenção do genoma (Figura 5). O câncer é uma doença de instabilidade genética, cujo acúmulo de alterações no DNA resulta na desregulação de processos vitais para o bom funcionamento celular. Estão envolvidas nestes mecanismos moleculares as vias de reparo por excisão de bases, recombinação homóloga, mau pareamento e organização dos telômeros (revisado em

Hoeijmakers, 2001). Muitos *drivers* já foram identificados como atuando nestas vias de manutenção do DNA como *TP53*, *BRCA1* e *BRAC2* (revisado em Hoeijmakers, 2001). Em nossa análise foram identificados cinco genes envolvidos com esta função: *MUS81*, *ERCC5*, *DDIAS*, *ZC3H12D* e *RNASEH2A* (Figura 4.B), os quais serão discutidos no decorrer deste texto.

Quatro pacientes (1, 3, 10, 12) foram detectadas com a variante rs1047769 que corresponda à troca de uma adenina por uma guanina na posição 256 (c.A256G), que dará origem a troca de aminoácido p.M86V da proteína codificada pelo gene *ERCC5*. Na família I, duas pacientes eram portadoras da variante rs1047769: mãe (caso 3), diagnosticada com CM aos 49 anos e sua filha (caso 1), diagnosticada com CPT aos 31 anos, contudo a paciente 2 (CPT) também da família I não era portadora desta variante. As pacientes 10 e 12, também portadoras desta variante, desenvolveram CPT aos 37 e 38 anos, respectivamente. O gene *ERCC5* é responsável por codificar uma endonuclease que age na excisão do final 3' no processo de reparo a danos no DNA induzidos por radiação UV. Alterações neste gene estão associadas com a síndrome Xeroderma Pigmentoso (OMIM:278780). Recentemente, Na *et al.* (2015) avaliaram uma possível relação entre alguns SNPs mapeados neste gene com CM esporádico. Os autores relataram uma associação significativa do genótipo GA + GG do SNP rs2094258 do gene *ERCC5* em pacientes portadoras de CM que eram tabagistas. Anteriormente, Rajaraman *et al.* (2008) avaliaram o mesmo SNP encontrado em nossa casuística (rs1047769) em relação à radiação ionizante. Os autores investigaram 1083 controles e 859 pacientes com CM que foram radiologistas por pelo menos dois anos e que foram expostas a baixas doses de radiação. Os autores não encontraram associação significativa entre o SNP rs1047769 e o CM (Rajaraman *et al.*, 2008). A paciente 12 portadora do SNP rs1047769 possui história familiar de câncer de pele em idade tardia, além disso, a alteração não segrega em todos os indivíduos da Família I. Pelos dados obtidos neste estudo e pelos relatos da literatura, esta alteração não parece ter efeito no risco de desenvolvimento de câncer de mama e tireoide.

As pacientes 8, 10, 11 e 18 foram diagnosticadas com CPT aos 26, 38, 40 e 36 anos, respectivamente. Estas pacientes possuem o SNP rs61997220 (c.317A>G;p.K106R) mapeado no gene *ZC3H12D*, e que foi predito como patogênico em cinco bancos de dados. O gene *ZC3H12D* é responsável por codificar uma proteína *zinc finger* que interage com receptores do tipo *Toll-like* e sua resposta aos ligantes NK

e NF- κ B, importantes para o processo inflamatório em macrófagos (Huang *et al.*, 2012). Este gene também foi reportado como putativo supressor tumoral (Minagawa *et al.*, 2007). Especificamente esta alteração c.317A>G;p.K106R foi detectada com perda de heterozigose em tumores de pulmão esporádicos (Wang *et al.*, 2007). Estes autores detectaram a alteração c.317A>G;p.K106R em heterozigose em regiões dos tumores de pulmão, devido a heterogeneidade tumoral, nesta mesma região, o outro alelo funcional era perdido por uma deleção em 6q23-25. Interessantemente, a variante rs61997220 foi encontrada no sangue periférico de um indivíduo com carcinoma de pulmão avaliado na mesma casuística. Estes resultados davam evidências de que o gene *ZC3H12D* atuava como um supressor tumoral. O efeito de supressão do gene foi confirmado por ensaio funcional onde o alelo A (normal) suprimiu a formação de colônias em 30% quando comparado ao alelo G (mutado) (Wang *et al.*, 2007). Em nossa casuística apenas a amostra 10 apresentou histórico familiar de câncer de pulmão, um tio do ramo paterno sem idade definida. No entanto, este gene pode ser um candidato interessante e deve ser explorado em relação à sua predisposição aos carcinomas de mama e tireoide hereditários.

A paciente 15, diagnosticada com tumores sincrônicos de mama e tireoide aos 67 anos, possuía apenas uma alteração (rs61902276) dentre as 50 descritas na Tabela 8. Esta variante está mapeada no gene *DDIAS* cuja troca de base c.G34A está envolvida com a substituição de uma valina por uma isoleucina no códon 12 da proteína. A variante possui frequência alélica de 0,04 no banco de dados 1000G, sendo um pouco mais frequente na população asiática (frequência alélica na população sul-asiática de 0,08; <http://exac.broadinstitute.org/variant/11-82625814-G-A>). Também foi considerada deletéria em três bancos de dados de predição *in silico*. Além da paciente 15, o SNP rs61902276 estava presente nos casos 11, 12 e 23, sendo que esta última paciente também desenvolveu tumores múltiplos. O SNP rs61902276 ainda não foi associado a algum fenótipo específico, embora o gene *DDIAS* (*DNA damage-induced apoptosis suppressor*), também conhecido como *noxin* (Nakaya *et al.*, 2007), seja relatado como sendo estimulado e estimulando a proteína p53, prevenindo a apoptose, agindo na parada do ciclo celular e auxiliando na ação dos mecanismos de reparo. A expressão deste gene foi detectada em células submetidas a estresse como radiação γ , radiação UV, adriamicina (agente utilizado em medicamentos quimioterápicos) e peróxido de hidrogênio. Foi também relatada expressão aumentada do gene *DDIAS* em tecidos que

estão em constante proliferação como medula óssea, baço e testículos (Nakaya *et al.*, 2007). Em carcinoma hepatocelular, este gene foi relatado como potencial oncogene, sendo detectado um aumento do número de cópias (SNParray) e aumento da expressão o que estimulava a proliferação celular e a tumorigênese *in vivo* (Zhang, Z. Z. *et al.*, 2015). Apesar de ter sido detectado na paciente 15 diagnosticada com CT e CM em idade tardia, este é um candidato interessante para futuras validações, pois sua função parece ser determinante para o ciclo celular. Os familiares destes pacientes estão sendo convidados para participarem do estudo para que seja realizada uma análise de segregação da alteração.

Todavia, a maioria dos casos possui mais de uma alteração as quais interagem com genes de alta e baixa penetrância. A alteração mais frequentemente detectada na casuística avaliada é rs35169799 mapeada no *PLCB3*, presente em seis pacientes (1, 2, 3, 10, 17, 22), incluindo três da mesma família (Família I, Tabela 10). Este gene é responsável por codificar uma subunidade da enzima fosfolipase C beta ($PLC\beta$), que atua na ativação de uma cadeia de enzimas que abrem canais de cálcio para o meio intracelular. O $PLC\beta$ atua como sinalizador secundário de vários processos celulares (Adjobo-Hermans *et al.*, 2013). O supressor tumoral *PTEN*, relatado em tumores de mama e associado com 85% dos casos da Síndrome de Cowden, interage diretamente com o complexo proteico PLC, estimulando o descolamento de PLC em direção à membrana plasmática para que o complexo possa atuar como sinalizador. Assim a via de sinalização de fosfatases PLC, envolvida em progressão do ciclo celular, pode ser mais um dos processos regulados pelo gene *PTEN* (Alvarez-Breckenridge *et al.*, 2007). O gene *PLCB3* também já foi descrito como associado com a síndrome MEN1, que envolve tumores da paratireoide, pâncreas, células endócrinas do duodeno e da glândula pituitária (Gobl *et al.*, 1995). Contudo, de Wit *et al.* (1997) excluíram o gene *PLCB3* do espectro da síndrome MEN1, pois os autores não identificaram erros estruturais na proteína ou alterações genômicas no DNA do sangue e tumoral dos pacientes. Outras alterações no gene *PLCB3* foram investigadas e apenas a variante rs12146487 foi detectada como patogênica em um banco de dados e com frequência alélica ≤ 0.1231 . Esta alteração foi detectada em 5 indivíduos (9, 12, 21, 23, e 24). Apesar de ser outra alteração no gene *PLCB3*, o qual é um importante candidato por sua função, o SNP rs12146487 pode ser considerado apenas uma variante comum e não contribuir com o fenótipo tumoral.

Outros três genes identificados com maior frequência entre as amostras foram *F5* (rs6027), *MUS81* (rs61754785) e *RNASEH2A* (rs62619782), sendo cada alteração identificada em pelo menos 5 indivíduos. O alelo rs6027, mapeado no gene *F5* e que possui frequência alélica de 0,04 no banco de dados 1000 Genomas, está envolvido na troca de base c.A6665G. Este SNP é considerado patogênico em oito bancos de dados por conta da troca de aminoácido p.D2222G e estava presente em três pacientes com CT (6, 10, 12), em dois pacientes com CT e CM (4, 18), incluindo apenas um caso da Família II (mas não em seu irmão afetado com CT). O gene *F5* codifica um cofator de coagulação sanguínea e alterações neste gene estão relacionados com patologias de coagulação (Peng *et al.*, 2014). Outras alterações que não as encontradas em nosso estudo (rs12120605, rs6427202, rs9332542 e rs6427199) já foram associadas aos carcinomas de mama negativos para os receptores hormonais. Os autores ressaltam a importância de desenvolver terapias direcionadas a processos de coagulação em pacientes com câncer (Tinholt *et al.*, 2014). Além disso, a paciente 18, que possui apenas esta alteração dentre as 13 descritas na Tabela 10, é ER+ PR+ HER2- e, portanto, não tem o perfil de expressão dos receptores hormonais negativos como descrito por Tinholt *et al.* (2014). As informações sobre os receptores hormonais não estavam disponíveis para a paciente 4. Esses resultados sugerem que a alteração detectada no gene *F5* pode não ter papel no fenótipo estudado.

O gene *MUS81* codifica uma endonuclease que, recentemente, foi descrita por Fu *et al.* (2015) como protetora da integridade do genoma, respondendo a fatores exógenos à replicação do DNA. A ação deste produto é essencial para o sucesso da duplicação do DNA, uma vez que atua em conjunto com outras proteínas na forquilha de replicação de eucariotos, na integridade dos fragmentos de *Okazaki* e no reparo por excisão de bases (Fu *et al.*, 2015). A alteração c.G1292A; p.R431H (rs61754785) em *MUS81*, gene que interage com *BRAC1* e *CHEK2* de acordo com a análise de interação proteica (Figura 5), foi observada em cinco pacientes (casos 6, 9, 11, 20, 22), sendo três diagnosticadas com CM e dois com CT. O gene *MUS81* já foi descrito como supressor tumoral com fenótipo haploinsuficiente (McPherson *et al.*, 2004). A proteína mutada Mus81, parte do complexo Mus81-Eme1, faz com que a célula seja altamente sensível a ação da droga mitomicina C. Esta droga possui ação citotóxica para a interação entre fitas de DNA, processo importante para replicação e reparo por recombinação homóloga. No mesmo estudo, McPherson *et al.* (2004) relataram que camundongos

Mus81 selvagens, Mus81^{+/-} e Mus81^{-/-} possuem instabilidade genômica em níveis diferenciados, além de conferir um fenótipo tumorigênico aos camundongos em maior e menor escala quando mutados em homozigose ou heterozigose, respectivamente. Os autores concluíram que o gene *MUS81* pode conferir uma predisposição ao desenvolvimento de câncer tanto em homozigose quanto em heterozigose, podendo atuar como fator de risco em haploinsuficiência (McPherson *et al.*, 2004). A deficiência da proteína Mus81 também demonstrou uma incapacidade celular de recuperação frente a exposição de agentes que inibem a replicação do DNA (Fu *et al.*, 2015).

Dada a importância deste gene para a replicação e reparo do DNA, mais estudos estão sendo realizados em nosso laboratório na tentativa de elucidar o papel da alteração encontrada em *MUS81* e sua contribuição para a predisposição de risco no desenvolvimento tumoral. A variante rs61754785 no gene *MUS81* foi confirmada por sequenciamento de Sanger nos indivíduos 6, 9, 11, 20 e 22. Os familiares da paciente 9 também foram testados e a alteração foi encontrada segregando em três casos: III.4, III.6 e III.7 (Figuras 11 e 12).

Uma análise de modelagem da proteína Mus81 foi realizada em colaboração com o Dr. Richard Charles Garrat, Instituto de Física, USP, São Carlos, SP. Baseado nos dados de Gwon *et al.*, 2014), foi verificado que a proteína faz parte do complexo Mus81-Eme1, quando este complexo está ligado ao DNA de dupla fita. O complexo Mus81-Eme1 possui dois domínios: domínio nuclease e domínio HhH2 com uma conexão *linker* entre eles, que é flexível em relação ao ligamento e desligamento do complexo à fita dupla de DNA. O resíduo R431, local em que se localiza a mutação p.R431H, em seu estado normal sofre uma mudança conformacional importante em função da ligação entre o complexo e o DNA, pois a estrutura do resíduo assume orientações diferentes nas duas condições: ligada e desligada. A troca p.R431H parece influenciar a estabilidade do complexo, demonstrando interferir fisicamente na região *linker* o que culmina na incompatibilidade da estrutura sem o DNA. Isto parece afetar a transição entre a conformação ligada e não ligada à dupla hélice, consequentemente comprometendo a atividade de Mus81. Estudos funcionais estão em andamento para comprovar o papel do gene *MUS81* e sua variante em processos tumorais.

Outras alterações no gene *MUS81* foram pesquisadas e detectadas, incluindo duas alterações raras: rs115472389 presente no indivíduo 20 ($F \leq 0.0189$, patogênica em dois bancos de dados) e rs34381357 no indivíduo 6 ($F \leq 0.0155$; patogênica em um

banco de dados). A variante comum rs13817 foi detectada em 20 indivíduos ($F \leq 0.5526$, patogênica em dois bancos de dados). Mesmo com outras alterações sendo detectadas no gene *MUS81* a mais promissora parece ser rs61754785, pois além de rara, é predita como patogênica em quatro bancos de dados, está presente em cinco indivíduos de nossa casuística e, ao que indica a análise de modelagem de proteína, confere uma mudança conformacional à proteína Mus81 e sua ligação ao DNA.

A outra variante encontrada em maior frequência entre os casos estudados (pacientes 6, 7, 19, 20) foi a rs62619782 – c.615T>A;p.D205E – no gene *RNASEH2*. A proteína RNaseH2, codificada pelo gene *RNASEH2A*, é uma ribonuclease que age na quebra dos fragmentos de Okazaki ou de ribonucleotídeos ligados ao filamento de DNA (Feng e Cao, 2015). As alterações nos resíduos G37S, R186W, R235Q, T240M, N212I, F230L, R108W e R291H neste gene estão associadas com a síndrome Aicardi-Goutieres, uma doença neurológica rara e grave, causada pela disfunção do complexo RNaseH2, composto pelos genes *RNASEH2A*, *RNASEH2B* e *RNASEH2C* (Coffin *et al.*, 2011). Em nossa análise a variante deste gene (rs62619782) não foi detectada nos indivíduos das Famílias I e II, apenas em indivíduos não aparentados diagnosticados com CT (6, 7, 24) e CM (19, 20) e não coincidem com as descritas por Coffin *et al.* (2011). Segundo o banco de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*), o produto deste gene interage com as enzimas DNA ligase 1 e FLAP endonuclease 1 (FEN1), esta última atua também no reparo a danos do DNA (<http://www.genome.jp/kegg/pathway/ko/ko03030.html>). Para o nosso conhecimento, os indivíduos analisados não eram portadores da síndrome Aicardi-Goutieres. O complexo RNase H2 é um importante fator para a manutenção da estabilidade do genoma pois atua na quebra de ribonucleotídeos de filamento único, deixando-os expostos para a maquinaria de reparo, o que representa um papel de supressão tumoral. Embora a atividade do complexo RNase H2 não esteja esclarecida, o transcrito da subunidade *RNASEH2A* foi encontrado como desregulado com alta expressão em tumores de bexiga, cérebro, mama, carcinomas espinocelulares e leucemias (Flanagan *et al.*, 2009). Outro achado interessante é que em linhagens de câncer de mama (como MCF-7^{kd}), ocorre a perda do fenótipo tumoral quando este gene apresenta expressão aumentada demonstrando seu papel de supressor tumoral (Feng and Cao, 2015). Por seu papel no desenvolvimento tumoral, seria interessante avaliar o potencial da alteração rs62619782 no gene *RNASEH2A*, sendo este outro candidato interessante detectado em

nosso estudo. Adicionalmente, a paciente 6 (CPT aos 53 anos) e a paciente 20, (CM aos 49 ano), compartilham as alterações rs62619782 (*RNASEH2A*) e rs61754785 (*MUS81*) o que pode contribuir para o aparecimento do fenótipo tumoral.

Recentemente, os genes *SRRM2* e *HABP2* foram descritos como associados ao carcinoma de tireoide não medular hereditário e o gene *PARP4* como associado aos carcinomas de tireoide e mama no mesmo paciente (Gara *et al.*, 2015; Ikeda *et al.*, 2015; Tomsic *et al.*, 2015). Estes genes foram investigados com mais detalhes em nossos dados de exoma.

O gene *SRRM2* (*Serine/Arginine Repetitive Matrix 2*) é um ortólogo do gene *Cwc21p* em *C. elegans*, importante fator que atua no mecanismo de *splicing* alternativo (Grainger *et al.*, 2009). Este gene foi reportado como biomarcador molecular da doença de Parkinson uma vez que foi detectado com alta expressão no sangue de pacientes afetados, quando comparados a pacientes com outras doenças neurológicas (Shehadeh *et al.*, 2010). Recentemente, Tomsic *et al.* (2015) investigaram sete pacientes com tumores de tireoide familiar por meio de SET e, utilizando rigoroso critério de seleção de variantes por meio de análise de ligação, identificaram duas alterações candidatas nos genes *SRRM2* e *CHD9*. Após isso, os autores esperavam encontrar uma destas alterações em um pequeno grupo controle de tumores esporádicos de tireoide para confirmar a associação significativa das mutações com CT, demonstrando assim a baixa frequência alélica da mutação. Apenas a variante p.S346F do gene *SRRM2* foi detectada neste pequeno grupo de tumores esporádicos (Tomsic *et al.*, 2016). Em nossa casuística esta alteração não foi encontrada, com tudo outras seis (c.5990G>A, c.6098G>C, c.5801G>A, c.3107G>C, c.3527C>G e c.986A>G) alterações no gene *SRRM2* foram detectadas em seis indivíduos diferentes (casos 24, 4, 7, 17, 16 e 2, respectivamente). Todas estas alterações são raras de acordo com o banco de dados ExAC e 1000 Genomas, além de serem preditas como patogênicas em pelo menos um banco de dados (Tabela 9). Este dado levanta questões sobre a potencial associação deste gene e o desenvolvimento não apenas de CT, mas também de CM na forma de uma nova síndrome hereditária. Estes achados precisam ser validados em parentes dos pacientes testados assim como ensaios funcionais para desvendar o papel destas variantes no risco de desenvolvimento de tumores de mama e tireoide.

Também recentemente, Ikeda *et al.* (2015) investigaram 14 pacientes diagnosticados com tumores primários de CM e CT antes dos 50 anos de idade, sem

exposição à radiação prévia e com forte histórico familiar para ambos os tumores. Para a identificação dos possíveis genes candidatos os autores focaram em genes de vias de reparo a danos e, após análise de estatística, o gene *PARP4* foi o que se mostrou significativamente associado com o fenótipo estudado. Os autores relataram que as alterações p.G496V e p.T1170I foram as mais frequentes estando presente em cinco e um indivíduo, respectivamente, resultando de um total de 14 pacientes investigados na casuística inicial (Ikeda *et al.*, 2015). Este gene codifica a enzima poly(ADP-ribosil)ação (*poly(ADP-ribosyl)ation*) responsável por catalisar a formação da polimerase ADP-ribose que age em resposta a danos no DNA resultante de quebras de fita única ou dupla (Kickhoefer *et al.*, 1999). Em um estudo de GWAs, os SNPs no gene *PARP4* foram significativamente associados com carcinomas de glândula salivar (Xu *et al.*, 2015). Por agir indiretamente na ação de reparo a danos no DNA, este gene parece ser interessante e pode estar associado com o desenvolvimento de tumores de mama e de tireoide. Em nossa casuística, a paciente 20 (Tabela 9) apresentou uma alteração nova e predita como patogênica em nove bancos de dados (c.2078A>G;p.Y693C). Outras três alterações neste gene também foram encontradas. O SNP rs2275660, que apesar de ser considerado patogênico em cinco bancos de dados é uma alteração comum na população em geral e inclusive em nosso estudo, pois estava presente em todas as amostras avaliadas, com exceção da paciente 19. A variante comum rs7571 foi predita como patogênica em apenas um banco e foi detectada nos indivíduos 6 e 20. Além destas, a alteração rs35200240, considerada rara e deletéria em um banco de dados, foi verificada na paciente 9. As variantes raras aqui detectadas em um total de 3 pacientes (a paciente 20 apresentou duas variantes deste gene) não foram as mesmas relatadas por Ikeda *et al.* (2015). Entretanto, isso não invalida a relevância desses dados. Outros estudos estão em andamento para investigar o gene *PARP4* como candidato a predisposição de carcinomas de mama e tireoide.

Recentemente, foi publicado um relato descrevendo a alteração rs7080536, no gene *HABP2* como associada com o risco de desenvolvimento do carcinoma de tireoide hereditário não-medular (Gara *et al.*, 2015). O gene *HABP2* é responsável pela degradação de ácido hialurônico da matriz extracelular e alterações neste gene já foram associadas com doenças tromboembólicas e cardiovasculares (Sedding *et al.*, 2006; Parahuleva *et al.*, 2008; Parahuleva *et al.*, 2013). Gara *et al.* (2015) descreveram a alteração rs7080536 no gene *HABP2*, a c.1601G>A que resulta na substituição do

aminoácido glicina pelo ácido glutâmico p.G534E, em uma probanda diagnosticada com CT papilífero aos 53 anos, assim como seu irmão mais velho diagnosticado aos 62 anos, o qual foi a óbito. A probanda havia tido uma filha portadora da mesma alteração e que havia desenvolvido CT papilífero aos 31 anos. O irmão da probanda teve cinco filhos e destes, quatro possuíam a mesma alteração A>G, bem como CT papilífero, sendo uma mulher e três homens diagnosticados aos 42, 39, 34 e 37 anos, respectivamente. Na última geração desta família, todos os portadores de A>G possuíam nódulos tireoideanos. Os autores avaliaram o potencial papel deste gene como supressor tumoral por meio de ensaios funcionais e concluíram que a elevada expressão do transcrito contendo a alteração, mesmo em heterozigose, estimula a migração e a proliferação celular. A ausência de um segundo evento mutacional nas amostras tumorais sugeriu aos autores que esta mutação poderia ter um efeito dominante-negativo. Além disso, a expressão desta proteína foi avaliada em tecido normal e tumoral, onde uma expressão alterada foi detectada nos tumores, mas não em tecido normal (Gara *et al.*, 2015). Este estudo gerou uma grande discussão em literatura sobre o papel do gene *HABP2* na predisposição ao carcinoma de tireoide não medular.

Em nosso estudo, esta variante foi identificada em duas pacientes (casos 7 e 9). No primeiro caso, a variante rs7080536 do gene *HABP2* foi encontrada em homozigose, e posteriormente testada em outros sete indivíduos da família da paciente 7, sendo um indivíduo homozigoto assim como o probando, dois indivíduos heterozigotos com carcinoma papilífero de tireoide aos 27 e aos 55 anos, um indivíduo heterozigoto com câncer de pele, dois indivíduos heterozigotos sem manifestar a doença e um indivíduo negativo para a variante, este último uma paciente do sexo feminino que havia desenvolvido CM aos 61 anos (Figura 9). O fato de esta mutação ter sido encontrada em homozigose levanta questionamentos em relação a sua ação dominante negativa, descrita por Gara *et al.* (2015). Nessa família, em particular, um número razoável de parentes tem a variante mas não tem o CPT. Na família da paciente 9, três parentes foram testados sendo que dois indivíduos heterozigotos desenvolveram nódulos benignos da tireoide, e o único indivíduo que havia desenvolvido carcinoma de tireoide aos 66 anos era negativo para esta variante (Figura 14). Interessantemente, a família 9 também é portadora da alteração em *MUS81*, discutida acima. Estes achados nos levam a questionar a hipótese levantada por Gara *et al.* (2015), de que esta alteração está de fato associada ao desenvolvimento dos CT hereditário por três motivos: (1) a alteração

não segrega em todos os indivíduos acometidos por câncer; (2) pacientes com tumores de mama, mas não de tireoide, também eram portadores desta alteração e portanto, não parece ser uma alteração exclusiva de CT familiar e (3) não há efeito dominante negativo para esta variante uma vez que foi detectada em homozigose sem qualquer diferença fenotípica. Além disso, outras questões também já foram levantadas neste aspecto por outros autores uma vez que esta alteração não foi detectada em indivíduos chineses com histórico de CT hereditário não medular e pelo conflito entre os dados de frequência alélica nos diferentes bancos de dados (Sponziello *et al.* 2015, Zhao *et al.* 2015, Zhou *et al.*, 2015).

Recentemente, Tomsic *et al.*, (2016) avaliaram a variante p.G534E (rs7080536) do gene *HABP2* em 179 probandos com histórico familiar de CT não medular, 1160 tumores esporádicos e 1395 tecido normais de tireoide. O SNP foi detectado em 6,1% dos probandos, em 8 % dos tumores esporádicos e 8.7% das amostras controle. Uma família com 25 indivíduos avaliados revelou quatro pacientes homozigotos, incluindo um paciente não afetado pela doença. Além disso, a alteração germinativa também foi investigada nos 1160 indivíduos com tumores esporádicos e dois destes também apresentaram a alteração em homozigose. A avaliação da expressão do transcrito em três tumores esporádicos e cinco controles revelou que não foi encontrada expressão do gene *HABP2* (Tomsic *et al.*, 2016). Este achado reforça o fato da alteração p.G534E não estar associada com CT não medular hereditário.

A busca de alterações em novos genes de predisposição tornou possível verificar que a maioria das alterações não está presente em 100% dos casos. Foi então questionado se as alterações observadas poderiam interagir entre si ou com outros genes de relevância em síndromes de predisposição hereditária ao câncer. Para responder essa questão, foi realizada uma análise de vias para reconhecer e entender possíveis mecanismos e vias celulares que pudessem contribuir para o fenótipo em questão. O enriquecimento de vias demonstrou ser uma ótima ferramenta para o cumprimento deste objetivo e ajudou a identificar mais de 90 vias potencialmente desreguladas. Duas vias em especial merecem destaque: a via de sinalização Rap1 e a via de síntese de hormônio tireoidiano (Tabela 11).

Alguns produtos gênicos da via de sinalização Rap1 atuam como supressores tumorais. A perda da regulação do gene *Rap1GAP*, responsável por inibir a molécula GTPase da família RAS denominada Rap1, foi associada com carcinomas papilíferos da

tireoide (Tsygankova *et al.*, 2007; Zuo *et al.*, 2010). Zuo *et al.* (2010) avaliaram 40 nódulos hiperplásicos, 49 adenomas foliculares e 27 carcinomas foliculares, 78 carcinomas papilíferos e 3 anaplásicos da tireoide e relataram que a expressão do gene *Rap1GAP* era variável, sendo mantida nos nódulos hiperplásicos, com expressão diminuída porém estatisticamente não significativa nos tumores benignos, e encontrava-se progressivamente diminuída nos carcinomas. Segundo os autores esta diminuição de expressão foi promovida por mecanismos envolvendo a inativação por mutação, LOH e metilação da região promotora do gene *Rap1GAP*. Os autores relataram também uma forte correlação entre a baixa expressão do gene e extensa invasão tumoral (Zuo *et al.*, 2010). No presente estudo, os genes alterados e identificados na via de sinalização Rap1 (Figura 7) foram *ITGAL*, *PLCB3*, *PDGFRB*, *SIPA1L3*, *ADCY2*, *PDGFD* e *TEK* sendo que os quatro primeiros foram considerados patogênicos em pelo menos um banco de dados. Alterações nestes genes foram observadas em 14 indivíduos, sendo que cada uma estava presente em pelo menos quatro pacientes. Além disto, as alterações nos genes *PLCB3* (já descrito acima), *PDGFRB* e *SIPA1L3* são compartilhadas entre os membros da Família I (Tabela 12), incluindo mãe da probanda com CM aos 49 anos, a probanda e sua irmã com CT aos 24 e 31 anos de idade, respectivamente. A desregulação da via de sinalização Rap1 pode contribuir para o desenvolvimento do câncer, uma vez que os membros desta via estão envolvidos em processos como crescimento, diferenciação, adesão celular e proliferação.

Recentemente, o gene *PDGFRB* (*Platelet-derived growth factor, beta polypeptide*) foi descrito com expressão aumentada em carcinomas de tireoide e hipermetilado em câncer de mama, ambos em tumores esporádicos (Li *et al.*, 2015; Zhang, H. *et al.*, 2015). O *PDGFRB* é considerado um importante componente do processo de inflamatório, por ser um receptor da membrana celular que estimula o crescimento tumoral e a divisão celular. Além disso, o produto deste gene é capaz de induzir a transcrição de *VEGF* o qual também desempenha papel importante na angiogênese e metástase (Zhang *et al.*, 2015). Assim, esta alteração poderia contribuir para o fenótipo tumoral, mas é possível que não tenha um papel relevante na predisposição a estes tumores.

O *SIPA1L3* (*signal-induced proliferation-associated 1 like 3*) não foi explorado em cânceres, sendo que, até o momento, apenas a alteração c.4489C>T (p.(R1497*)) neste gene foi descrita como associada com o desenvolvimento de catarata congênita

(Evers *et al.*, 2015). Não temos conhecimento se os membros com a alteração aqui descrita apresentam história de catarata na família. Aparentemente esta alteração não parece relevante no fenótipo estudado, mas pode ser útil no aconselhamento genético destas famílias.

A análise de enriquecimento revelou também duas alterações consideradas patogênicas na via de síntese de hormônios tireoidianos (Figura 5): *PCLB3* (pacientes 1, 2, 3 da família I e os casos 10, 17 e 22), anteriormente citado como associado à síndrome MEN1 e já discutido acima (De Wit *et al.*, 1997); e *PAX8* (casos 4 e 5 da família II, casos 23 e 24). Ambas as alterações foram identificados em indivíduos diagnosticados com os tumores de mama e/ou tireoide. Rearranjos cromossômicos como *PAX8-PPARc* já são bem estabelecidos em tumores da tireoide e estão envolvidos com a via ERK, uma das principais cascatas de regulação de proliferação celular e sobrevivência (Santarpia *et al.*, 2010). Além disto, a via de hormônios tireoideanos atua diretamente na produção de T3 e T4, já demonstrados desregulados em pacientes com câncer de mama. Uma meta-análise revelou que altos níveis de T3 estão associados significativamente com pacientes com tumores mamários, quando comparados aos controles do estudo (Shi *et al.*, 2014). Desta forma, o *PAX8* tem potencial para contribuir para o risco de desenvolvimento de câncer de mama e tireoide. Estudos futuros focarão na análise destes genes em membros destas famílias, como uma primeira estratégia para investigar seu papel no risco de desenvolvimento da doença.

Os estudos descritos acima bem como os resultados alcançados até o momento evidenciam que existem de fato alterações genéticas que interferem em mecanismos moleculares importantes, e que por sua vez podem contribuir para o desenvolvimento dos carcinomas mamários e tireoidianos em sua forma hereditária. Apesar de a casuística avaliada ser heterogênea, o objetivo inicial deste estudo foi avaliar pacientes que desenvolveram CT e ou CM, com histórico familiar destes tumores. Os critérios de inclusão e exclusão foram estipulados levando em conta as características clínicas dos pacientes, com avaliação minuciosa do histórico familiar com o intuito de diminuir o viés de uma casuística muito complexa. Nenhuma alteração foi identificada em 100% dos casos, contudo os genes alterados parecem interagir de alguma maneira, ou em vias comumente desreguladas ou por desempenharem papel como os de manutenção do genoma, processo fundamental para o bom funcionamento celular. Por fim, a técnica de sequenciamento de exoma demonstrou ser uma metodologia adequada e possibilitou a

identificação de alterações em genes importantes tendo sido confirmada por sequenciamento de Sanger para alguns genes testados. Este estudo identificou novas alterações genéticas em pacientes com CT e ou CM, com histórico destes tumores, revelando potenciais candidatos à predisposição hereditária para o desenvolvimento destes tumores.

7. CONCLUSÃO

- A análise de exoma de pacientes com diagnóstico de CT e/ou CM em idade jovem, com histórico familiar destes tumores, permitiu identificar alterações raras, de acordo com os bancos de dados 1000 Genomas e 6500 Exomas, e consideradas patogênicas por análise de predição *in silico*. Estas alterações estavam presentes tanto em genes clássicos previamente associados com síndromes de predisposição ao câncer, quanto em novos genes candidatos relacionado a funções como de manutenção celular e do genoma. Nenhum das alterações encontradas em genes clássicos, com exceção de p.R233X no gene *PTEN* da paciente 14, é utilizada atualmente na prática clínica para o diagnóstico de uma síndrome hereditária.
- Nos genes de alta e baixa penetrância já descritos em literatura foram detectadas 38 alterações não-sinônimas comuns, sendo apenas cinco consideradas possivelmente deletérias ou deletérias em apenas um banco de dados de predição de patogenicidade. Outras 13 alterações raras, com frequência alélica $< 0,05$, e preditas como patogênicas em pelo menos três bancos de dados, foram detectadas em no máximo dois indivíduos cada uma.
- Na busca de novos genes de predisposição foram selecionadas as alterações presentes em pelo menos quatro indivíduos, reportadas com frequência alélica $< 0,05$ em bancos de dados 1000 Genomas e 6500 Exomas, o que resultou em 911 SNPs. Destas, 287 eram exônicas e não sinônimas. O número de alterações foi reduzido para 50 quando consideradas apenas as alterações patogênicas em pelo menos três bancos de dados de patogenicidade.
- As 50 variantes patogênicas identificadas estavam mapeadas em 50 genes distintos, que potencialmente podem estar associados com a predisposição aos tumores de mama e tireoide. Destes 50 genes, 46 codificavam proteínas que apresentam interação com outras proteínas. Quarenta e um por cento destes genes foram classificados de acordo com sua ontologia gênica em Organização Celular, 11% em Sinalização e Transporte, 7% em Transcrição, 11% em Manutenção do Genoma, 9% em Metabolismo de Proteínas, 4% em Outros Metabolismos, 2% em Tradução e 15% não possuíam classificação. Interessantemente, 13 genes alterados interagiam com genes de alta e baixa

penetrância previamente descritos como associados às síndromes de predisposição ao câncer.

- A busca de vias potencialmente desreguladas revelou a via de sinalização Rap1 e via de síntese de hormônio tireoidiano. A via de sinalização Rap1 confirma resultados prévios descritos em carcinomas papilíferos de tireoide. A segunda via revelou o envolvimento dos genes *PLCB3* e *PAX8*, cujas alterações estão presentes em 10 pacientes diferentes, inclusive segregando nos pacientes das famílias I e II. As alterações nestes dois genes são mutuamente exclusivas entre os pacientes. Em adição, 10 indivíduos não aparentados apresentam alterações em genes desta via.
- Os membros das famílias I e II possuem alterações que segregam entre os indivíduos avaliados, contudo não compartilham nenhuma variante detectada nesta primeira etapa de análise de dados.
- Foram encontradas mutações em genes recentemente descritos em literatura como associados com CT e ou CM hereditários, incluindo *HABP2*, *SRRM2* e *PARP4*.
- A variante *HABP2* c.1601G>A foi selecionada para confirmação dos achados e análise de segregação por estar presente em homozigose em um indivíduo de nossa casuística, resultante de casamento consanguíneo. A alteração foi encontrada segregando na família do paciente 7 a da paciente 9, tanto em indivíduos com câncer como em indivíduos que não manifestaram a doença até o presente momento. No entanto, em ambas as famílias dos pacientes 7 e 9, dois familiares com CM e CT, respectivamente, não eram portadores de *HABP2* c.1601G>A.
- O SNP *MUS81* c.1292G>A foi encontrado segregando na paciente 9 e em todos seus os familiares avaliados. Levando em conta esta análise de segregação, a função do gene *MUS81* e a análise de modelagem de proteína, a variante *MUS81* c.1292G>A parece contribuir com a predisposição dos tumores de mama e tireoide em sua forma hereditária.
- O *pipeline* de análise desenvolvido para este estudo revelou seu potencial para os objetivos propostos, viabilizando a identificação de alterações de base única em genes de alta e baixa penetrância assim como em novos genes candidatos. Também mostrou-se confiável com a confirmação de algumas variantes testadas.

8. REFERÊNCIAS

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012 Nov 1;491(7422):56-65.

A, Ma YP, Dobbins SE, Houlston RS. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. *J Clin Oncol*. 2015 Feb 10;33(5):426-32.

Achatz MI, Olivier M, Le Calvez F, Martel-Planche G, Lopes A, Rossi BM, Ashton-Prolla P, Giugliani R, Palmero EI, Vargas FR, Da Rocha JC, Vettore AL, Hainaut P. The TP53 mutation, R337H, is associated with Li-Fraumeni and Li-Fraumeni-like syndromes in Brazilian families. *Cancer Lett*. 2007 Jan 8;245(1-2):96-102.

Adjobo-Hermans MJ, Crosby KC, Putyrski M, Bhageloe A, van Weeren L, Schultz C, Goedhart J, Gadella TW Jr. PLC β isoforms differ in their subcellular location and their CT-domain dependent interaction with G α_q . *Cell Signal*. 2013 Jan;25(1):255-63.

Al-Baimani K, Bazzarelli A, Clemons M, Robertson SJ, Addison C, Arnaout A. Invasive Pleomorphic Lobular Carcinoma of the Breast: Pathologic, Clinical, and Therapeutic Considerations. *Clin Breast Cancer*. 2015 Dec;15(6):421-5.

Alvarez-Breckenridge CA, Waite KA, Eng C. PTEN regulates phospholipase D and phospholipase C. *Hum Mol Genet*. 2007 May 15;16(10):1157-63

American Cancer Society. Cancer Facts & Figures 2015. Atlanta: American Cancer Society; 2015.

An JH, Hwangbo Y, Ahn HY, Keam B, Lee KE, Han W, Park do J, Park IA, Noh DY, Youn YK, Cho BY, Im SA, Park YJ. A Possible Association Between Thyroid Cancer and Breast Cancer. *Thyroid*. 2015 Dec;25(12):1330-8.

Aoude LG, Heitzer E, Johansson P, Gartside M, Wadt K, Pritchard AL, Palmer JM, Symmons J, Gerdes AM, Montgomery GW, Martin NG, Tomlinson I, Kearsley S, Hayward NK. POLE mutations in families predisposed to cutaneous melanoma. *Fam Cancer*. 2015 Dec;14(4):621-8.

Aoude LG, Heitzer E, Johansson P, Gartside M, Wadt K, Pritchard AL, Palmer JM, Symmons J, Gerdes AM, Montgomery GW, Martin NG, Tomlinson I, Kearsley S, Hayward NK. POLE mutations in families predisposed to cutaneous melanoma. *Fam Cancer*. 2015 Dec;14(4):621-8.

Auvinen A, Curtis RE, Ron E. Risk of subsequent cancer following breast cancer in men. *J Natl Cancer Inst*. 2002 Sep 4;94(17):1330-2.

Aza-Carmona M, Barreda-Bonis AC, Guerrero-Fernández J, González-Casado I, Gracia R, Heath KE. Familial glucocorticoid deficiency due to compound heterozygosity of two novel MC2R mutations. *J Pediatr Endocrinol Metab*. 2011;24(5-6):395-7.

Backes C, Harz C, Fischer U, Schmitt J, Ludwig N, Petersen BS, Mueller SC, Kim YJ, Wolf NM, Katus HA, Meder B, Furtwängler R, Franke A, Bohle R, Henn W, Graf N, Keller A, Meese E. New insights into the genetics of glioblastoma multiforme by familial exome sequencing. *Oncotarget*. 2015 Mar 20;6(8):5918-31.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011 Sep 27;12(11):745-55.

Bardhan P, Bui MM, Minton S, Loftus L, Carter WB, Laronga C, Ismail-Khan R. HER2-positive male breast cancer with thyroid cancer: an institutional report and review of literature. *Ann Clin Lab Sci*. 2012 Spring;42(2):135-9.

Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med*. 2014 Sep 18;371(12):1170.

BREAST CANCER SCREENING AND DIAGNOSIS Version 1.2015. 7/15/2015 National Comprehensive Cancer Network, Inc. 2014, All Rights Reserved.

Brennan P, Wild CP. Genomics of Cancer and a New Era for Cancer Prevention. *PLoS Genet*. 2015 Nov 5;11(11):e1005522.

Broaddus RR, Lynch PM, Lu KH, Luthra R, Michelson SJ. Unusual tumors associated with the hereditary nonpolyposis colorectal cancer syndrome. *Mod Pathol*. 2004 Aug;17(8):981-9.

Brown KR, Otasek D, Ali M, McGuffin MJ, Xie W, Devani B, Toch IL, Jurisica I. NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. *Bioinformatics*. 2009 Dec 15;25(24):3327-9.

Bueno RC, Canevari RA, Villacis RA, Domingues MA, Caldeira JR, Rocha RM, Drigo SA, Rogatto SR. ATM down-regulation is associated with poor prognosis in sporadic breast carcinomas. *Ann Oncol*. 2014 Jan;25(1):69-75.

Buerki N, Gautier L, Kovac M, Marra G, Buser M, Mueller H, Heinimann K. Evidence for breast cancer as an integral part of Lynch syndrome. *Genes Chromosomes Cancer*. 2012 Jan;51(1):83-91.

Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM. Ductal carcinoma in situ of the breast. *N Engl J Med*. 2004 Apr 1;350(14):1430-41.

Capasso M, Devoto M, Hou C, Asgharzadeh S, Glessner JT, Attiyeh EF, Mosse YP, Kim C, Diskin SJ, Cole KA, Bosse K, Diamond M, Laudenslager M, Winter C, Bradfield JP, Scott RH, Jagannathan J, Garriss M, McConville C, London WB, Seeger RC, Grant SF, Li H, Rahman N, Rappaport E, Hakonarson H, Maris JM. Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. *Nat Genet*. 2009 Jun;41(6):718-23.

Cascón A, Comino-Méndez I, Currás-Freixes M, de Cubas AA, Contreras L, Richter S, Peitzsch M, Mancikova V, Inglada-Pérez L, Pérez-Barrios A, Calatayud M, Azriel S, Villar-Vicente R, Aller J, Setién F, Moran S, Garcia JF, Río-Machín A, Letón R, Gómez-Graña Á, Apellániz-Ruiz M, Roncador G, Esteller M, Rodríguez-Antona C, Satrústegui J, Eisenhofer G, Urioste M, Robledo M. Whole-exome sequencing identifies MDH2 as a new familial paraganglioma gene. *J Natl Cancer Inst*. 2015 Mar 11;107(5).

Cavalli F, Atun R. Towards a global cancer fund. *Lancet Oncol*. 2015 Feb;16(2):133-4.
Chen H, Duncan IC, Bozorgchami H, Lo SH. Tensin1 and a previously undocumented family member, tensin2, positively regulate cell migration. *Proc Natl Acad Sci US A*. 2002 Jan 22;99(2):733-8.

Chen H, Duncan IC, Bozorgchami H, Lo SH. Tensin1 and a previously undocumented family member, tensin2, positively regulate cell migration. *Proc Natl Acad Sci US A*. 2002 Jan 22;99(2):733-8.

Cherukuri PF, Maduro V, Fuentes-Fajardo KV, Lam K; NISC Comparative Sequencing Program, Adams DR, Tifft CJ, Mullikin JC, Gahl WA, Boerkoel CF. Replicate exome-

sequencing in a multiple-generation family: improved interpretation of next-generation sequencing data. *BMC Genomics*. 2015 Nov 25;16(1):998.

Chubb D, Broderick P, Frampton M, Kinnersley B, Sherborne A, Penegar S, Lloyd A, Ma YP, Dobbins SE, Houlston RS. Genetic diagnosis of high-penetrance susceptibility for colorectal cancer (CRC) is achievable for a high proportion of familial CRC by exome sequencing. *J Clin Oncol*. 2015 Feb 10;33(5):426-32.

Coffin SR, Hollis T, Perrino FW. Functional consequences of the RNase H2A subunit mutations that cause Aicardi-Goutieres syndrome. *J Biol Chem*. 2011 May 13;286(19):16984-91.

Comino-Méndez I, Gracia-Aznárez FJ, Schiavi F, Landa I, Leandro-García LJ, Letón R, Honrado E, Ramos-Medina R, Caronia D, Pita G, Gómez-Graña A, de Cubas AA, Inglada-Pérez L, Maliszewska A, Taschin E, Bobisse S, Pica G, Loli P, Hernández-Lavado R, Díaz JA, Gómez-Morales M, González-Neira A, Roncador G, Rodríguez-Antona C, Benítez J, Mannelli M, Opocher G, Robledo M, Cascón A. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet*. 2011 Jun 19;43(7):663-7.

Consorti F, Di Tanna G, Milazzo F, Antonaci A. Nulliparity enhances the risk of second primary malignancy of the breast in a cohort of women treated for thyroid cancer. *World J Surg Oncol*. 2011 Aug 12;9:88.

Correa R, Salpea P, Stratakis CA. Carney complex: an update. *Eur J Endocrinol*. 2015 Oct;173(4):M85-97.

Cox DG, et al, Consortium of Investigators of Modifiers of BRCA1/2. Common variants of the BRCA1 wild-type allele modify the risk of breast cancer in BRCA1 mutation carriers. *Hum Mol Genet*. 2011 Dec 1;20(23):4732-47. doi: 10.1093/hmg/ddr388. Epub 2011 Sep 2. PubMed PMID: 21890493; PubMed Central PMCID: PMC3733139.

Cybulski C, Carrot-Zhang J, Kluźniak W, Rivera B, Kashyap A, Wokołorczyk D, Giroux S, Nadaf J, Hamel N, Zhang S, Huzarski T, Gronwald J, Byrski T, Szwiec M, Jakubowska A, Rudnicka H, Lener M, Masojć B, Tonin PN, Rousseau F, Górski B, Dębniak T, Majewski J, Lubiński J, Foulkes WD, Narod SA, Akbari MR. Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet*. 2015 Jun;47(6):643-6.

da Silva FC, de Oliveira LP, Santos EM, Nakagawa WT, Aguiar Junior S, Valentin MD, Rossi BM, de Oliveira Ferreira F. Frequency of extracolonic tumors in Brazilian families with Lynch syndrome: analysis of a hereditary colorectal cancer institutional registry. *Fam Cancer*. 2010 Dec;9(4):563-70.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011 Aug 1;27(15):2156-8.

Datta SR, McQuillin A, Rizig M, Blaveri E, Thirumalai S, Kalsi G, Lawrence J, Bass NJ, Puri V, Choudhury K, Pimm J, Crombie C, Fraser G, Walker N, Curtis D, Zvelebil M, Pereira A, Kandaswamy R, St Clair D, Gurling HM. A threonine to isoleucine missense mutation in the pericentriolar material 1 gene is strongly associated with schizophrenia. *Mol Psychiatry*. 2010 Jun;15(6):615-28.

Davies MJ. De-mystifying genes: identifying and managing hereditary cancer. *ONS Connect*. 2007;22(8 Suppl):43-4. Review.

Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J*. 2013 Mar 6;32(5):617-28.

de Kock L, Sabbaghian N, Soglio DB, Guillerman RP, Park BK, Chami R, Deal CL, Priest JR, Foulkes WD. Exploring the association Between DICER1 mutations and differentiated thyroid carcinoma. *J Clin Endocrinol Metab*. 2014 Jun;99(6):E1072-7.

de Wit MJ, Landsvater RM, Sinke RJ, Geurts van Kessel A, Lips CJ, Höppener JW. Exclusion of the phosphatidylinositol-specific phospholipase C beta 3 (PLC beta 3) gene as candidate for the multiple endocrine neoplasia type 1 (MEN 1) gene. *Hum Genet*. 1997 Jan;99(1):133-7.

Deb S, Wong SQ, Li J, Do H, Weiss J, Byrne D, Chakrabarti A, Bosma T; kConFab Investigators, Fellowes A, Dobrovic A, Fox SB. Mutational profiling of familial male breast cancers reveals similarities with luminal A female breast cancer with rare TP53 mutations. *Br J Cancer*. 2014 Dec 9;111(12):2351-60.

DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA Cancer J Clin*. 2016 Jan;66(1):31-42.

Donner I, Kiviluoto T, Ristimäki A, Aaltonen LA, Vahteristo P. Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. *Fam Cancer*. 2015 Jun;14(2):241-6.

Dox DG, et al. Consortium of Investigators of Modifiers of BRCA1/2. Common variants of the BRCA1 wild-type allele modify the risk of breast cancer in BRCA1 mutation carriers. *Hum Mol Genet*. 2011 Dec 1;20(23):4732-47.

Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DG, Chenevix-Trench G, Rahman N, Robson M, Domchek SM, Foulkes WD. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med*. 2015 Jun 4;372(23):2243-57.

Esteban-Jurado C, Vila-Casadesús M, Garre P, Lozano JJ, Pristoupilova A, Beltran S, Muñoz J, Ocaña T, Balaguer F, López-Cerón M, Cuatrecasas M, Franch-Expósito S, Piqué JM, Castells A, Carracedo A, Ruiz-Ponte C, Abulí A, Bessa X, Andreu M, Bujanda L, Caldés T, Castellví-Bel S. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet Med*. 2015 Feb;17(2):131-42.

Evers C, Paramasivam N, Hinderhofer K, Fischer C, Granzow M, Schmidt-Bacher A, Eils R, Steinbeisser H, Schlesner M, Moog U. SIPA1L3 identified by linkage analysis and whole-exome sequencing as a novel gene for autosomal recessive congenital cataract. *Eur J Hum Genet*. 2015 Dec;23(12):1627-33.

Fan C, Wei J, Yuan C, Wang X, Jiang C, Zhou C, Yang M. The functional TP53 rs1042522 and MDM4 rs4245739 genetic variants contribute to Non-Hodgkin lymphoma risk. *PLoS One*. 2014 Sep 9;9(9):e107047.

Feng S, Cao Z. Is the role of human RNase H2 restricted to its enzyme activity? *Prog Biophys Mol Biol*. 2015 Nov 19. pii: S0079-6107(15)00183-2.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015 Mar 1;136(5):E359-86.

Flanagan JM, Funes JM, Henderson S, Wild L, Carey N, Boshoff C. Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADARB1 as putative anticancer drug targets. *Mol Cancer Ther*. 2009 Jan;8(1):249-60.

- Fu H, Martin MM, Regairaz M, Huang L, You Y, Lin CM, Ryan M, Kim R, Shimura T, Pommier Y, Aladjem MI. The DNA repair endonuclease Mus81 facilitates fast DNA replication in the absence of exogenous damage. *Nat Commun*. 2015 Apr 16;6:6746.
- Galanti MR, Ekbom A, Grimelius L, Yuen J. Parental cancer and risk of papillary and follicular thyroid carcinoma. *Br J Cancer*. 1997;75(3):451-6.
- Gara SK, Jia L, Merino MJ, Agarwal SK, Zhang L, Cam M, Patel D, Kebebew E. Germline HAP2 Mutation Causing Familial Nonmedullary Thyroid Cancer. *N Engl J Med*. 2015 Jul 30;373(5):448-55.
- Garber JE, Offit K. Hereditary cancer predisposition syndromes. *J Clin Oncol*. 2005 Jan 10;23(2):276-92.
- Gómez García EB, Knoers NV. Gardner's syndrome (familial adenomatous polyposis): a cilia-related disorder. *Lancet Oncol*. 2009 Jul;10(7):727-35.
- Garner CN, Ganetzky R, Brainard J, Hammel JP, Berber E, Siperstein AE, Milas M. Increased prevalence of breast cancer among patients with thyroid and parathyroid disease. *Surgery*. 2007 Dec;142(6):806-13; discussion 813.e1-3. Epub 2007 Nov 5. Erratum in: *Surgery*. 2008 Feb;143(2):302.
- Gibson BA, Kraus WL. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat Rev Mol Cell Biol*. 2012 Jun 20;13(7):411-24.
- Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet*. 2012 May;20(5):490-7.
- Gobl AE, Chowdhary BP, Shu W, Eriksson L, Larsson C, Weber G, Oberg K, Skogseid B. Assignment of the mouse homologue of a human MEN1 candidate gene, phospholipase C-beta 3 (Plcb3), to chromosome region 19B by FISH. *Cytogenet Cell Genet*. 1995;71(3):257-9.
- Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Senn HJ; Panel members. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol*. 2013 Sep;24(9):2206-23.
- Goto M, Miller RW, Ishikawa Y, Sugano H. Excess of rare cancers in Werner syndrome (adult progeria). *Cancer Epidemiol Biomarkers Prev*. 1996 Apr;5(4):239-46.

Gracia-Aznarez FJ, Fernandez V, Pita G, Peterlongo P, Dominguez O, de la Hoya M, Duran M, Osorio A, Moreno L, Gonzalez-Neira A, Rosa-Rosa JM, Sinilnikova O, Mazoyer S, Hopper J, Lazaro C, Southey M, Odefrey F, Manoukian S, Catucci I, Caldes T, Lynch HT, Hilbers FS, van Asperen CJ, Vasen HF, Goldgar D, Radice P, Devilee P, Benitez J. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PLoS One*. 2013;8(2):e55681.

Grainger RJ, Barrass JD, Jacquier A, Rain JC, Beggs JD. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA*. 2009 Dec;15(12):2161-73.

Gustavsson EK, Trinh J, Guella I, Vilariño-Güell C, Appel-Cresswell S, Stoessl AJ, Tsui JK, McKeown M, Rajput A, Rajput AH, Aasly JO, Farrer MJ. DNAJC13 genetic variants in parkinsonism. *Mov Disord*. 2015 Feb;30(2):273-8.

Gwon GH, Jo A, Baek K, Jin KS, Fu Y, Lee JB, Kim Y, Cho Y. Crystal structures of the structure-selective nuclease Mus81-Eme1 bound to flap DNA substrates. *EMBO J*. 2014 May 2;33(9):1061-72.

Gylfe AE, Katainen R, Kondelin J, Tanskanen T, Cajuso T, Hänninen U, Taipale J, Taipale M, Renkonen-Sinisalo L, Järvinen H, Mecklin JP, Kilpivaara O, Pitkänen E, Vahteristo P, Tuupanen S, Karhu A, Aaltonen LA. Eleven candidate susceptibility genes for common familial colorectal cancer. *PLoS Genet*. 2013;9(10):e1003876.

Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. *Orphanet J Rare Dis*. 2009 Oct 12;4:22.

Hampel H, Sweet K, Westman JA, Offit K, Eng C. Referral for cancer genetics consultation: a review and compilation of risk assessment criteria. *J Med Genet*. 2004 Feb;41(2):81-91.

Hancock L, Beckly J, Geremia A, Cooney R, Cummings F, Pathan S, Guo C, Warren BF, Mortensen N, Ahmad T, Jewell D. Clinical and molecular characteristics of isolated colonic Crohn's disease. *Inflamm Bowel Dis*. 2008 Dec;14(12):1667-77.

Hasan TN, Shafi G, Syed NA, Alsaif MA, Alsaif AA, Alshatwi AA. Lack of association of BRCA1 and BRCA2 variants with breast cancer in an ethnic population of Saudi Arabia, an emerging high-risk area. *Asian Pac J Cancer Prev*. 2013;14(10):5671-4.

He W, Zhang D, Jiang J, Liu P, Wu C. The relationships between the chemosensitivity of human gastric cancer to paclitaxel and the expressions of class III β -tubulin, MAPT, and survivin. *Med Oncol*. 2014 May;31(5):950.

Hedley AJ, Jones SJ, Spiegelhalter DJ, Clements P, Bewsher PD, Simpson JG, Weir RD. Breast cancer in thyroid disease: fact or fallacy? *Lancet*. 1981 Jan 17;1(8212):131-3.

Hemminki K, Scélo G, Boffetta P, Møller H, Tracey E, Andersen A, Brewster DH, Pukkala E, McBride M, Kliever EV, Chia KS, Pompe-Kirn V, Martos C, Jonasson JG, Li X, Brennan P. Second primary malignancies in patients with male breast cancer. *Br J Cancer*. 2005 Apr 11;92(7):1288-92.

Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature*. 2001 May 17;411(6835):366-74.

Huang S, Qi D, Liang J, Miao R, Minagawa K, Quinn T, Matsui T, Fan D, Liu J, Fu M. The putative tumor suppressor Zc3h12d modulates toll-like receptor signaling in macrophages. *Cell Signal*. 2012 Feb;24(2):569-76.

Huo X, Lu C, Huang X, Hu Z, Jin G, Ma H, Wang X, Qin J, Wang X, Shen H, Tang J. Polymorphisms in BRCA1, BRCA1-interacting genes and susceptibility of breast cancer in Chinese women. *J Cancer Res Clin Oncol*. 2009 Nov;135(11):1569-75.

Hwang S, Kim E, Lee I, Marcotte EM. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep*. 2015 Dec 7;5:17875.

Ikeda Y, Kiyotani K, Yew PY, Kato T, Tamura K, Yap KL, Nielsen S, Mester J, Eng C, Nakamura Y, Grogan RH. Germline PARP4 mutations in patients with primary thyroid and breast cancers. *Endocr Relat Cancer*. 2015 Dec 23.

Imyanitov EN, Byrski T. Systemic treatment for hereditary cancers: a 2012 update. *Hered Cancer Clin Pract*. 2013 Apr 1;11(1):2.

INCA. Estimativa 2014: Incidência de Câncer no Brasil. Rio de Janeiro, Brasil.: 2014.

Jamieson D, Cresti N, Bray J, Sludden J, Griffin MJ, Hawsawi NM, Famie E, Mould EV, Verrill MW, May FE, Boddy AV. Two minor NQO1 and NQO2 alleles predict poor response of breast cancer patients to adjuvant doxorubicin and cyclophosphamide therapy. *Pharmacogenet Genomics*. 2011 Dec;21(12):808-19.

Januchowski R, Zawierucha P, Ruciński M, Zabel M. Microarray-based detection and expression analysis of extracellular matrix proteins in drug-resistant ovarian cancer cell lines. *Oncol Rep.* 2014 Nov;32(5):1981-90.

Joseph KR, Edirimanne S, Eslick GD. The association between breast cancer and thyroid cancer: a meta-analysis. *Breast Cancer Res Treat.* 2015 Jul;152(1):173-81.

Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, Zhang Q, Koboldt DC, Xie M, Kandoth C, McMichael JF, Wyczalkowski MA, Larson DE, Schmidt HK, Miller CA, Fulton RS, Spellman PT, Mardis ER, Druley TE, Graubert TA, Goodfellow PJ, Raphael BJ, Wilson RK, Ding L. Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun.* 2014;5:3156.

Kazakov VS, Demidchik EP, Astakhova LN. Thyroid cancer after Chernobyl. *Nature.* 1992 Sep 3;359(6390):21.

Kickhoefer VA, Siva AC, Kedersha NL, Inman EM, Ruland C, Streuli M, Rome LH. The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase. *J Cell Biol.* 1999 Sep 6;146(5):917-28.

Klaassen I, Özgören B, Sadowski CE, Möller K, van Husen M, Lehnhardt A, Timmermann K, Freudenberg F, Helmchen U, Oh J, Kemper MJ. Response to cyclosporine in steroid-resistant nephrotic syndrome: discontinuation is possible. *Pediatr Nephrol.* 2015 Sep;30(9):1477-83.

Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat Rev Cancer.* 2006 Apr;6(4):292-306.

Ku CS, Cooper DN, Wu M, Roukos DH, Pawitan Y, Soong R, Iacopetta B. Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. *Mod Pathol.* 2012 Aug;25(8):1055-68.

Kuo JH, Chabot JA, Lee JA. Breast cancer in thyroid cancer survivors: An analysis of the Surveillance, Epidemiology, and End Results-9 database. *Surgery.* 2016 Jan;159(1):23-30.

La Vecchia C, Malvezzi M, Bosetti C, Garavello W, Bertuccio P, Levi F, Negri E. Thyroid cancer mortality and incidence: a global overview. *Int J Cancer.* 2015 May 1;136(9):2187-95.

Lalloo F, Evans DG. Familial breast cancer. *Clin Genet.* 2012 Aug;82(2):105-14.

Lammens CR, Aaronson NK, Wagner A, Sijmons RH, Ausems MG, Vriends AH, Ruijs MW, van Os TA, Spruijt L, Gómez García EB, Kluijdt I, Nagtegaal T, Verhoef S, Bleiker EM. Genetic testing in Li-Fraumeni syndrome: uptake and psychosocial consequences. *J Clin Oncol*. 2010 Jun 20;28(18):3008-14.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014 Jan;42(Database issue):D980-5.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012 Mar 4;9(4):357-9.

Larsen MJ, Thomassen M, Gerdes AM, Kruse TA. Hereditary breast cancer: clinical, pathological and molecular characteristics. *Breast Cancer (Auckl)*. 2014 Oct 15;8:145-55.

Lee YH, Kim JH, Song GG. Genome-wide pathway analysis in neuroblastoma. *Tumour Biol*. 2014 Apr;35(4):3471-85.

Lefevre JH, Bonilla C, Colas C, Winney B, Johnstone E, Tonks S, Day T, Hutnik K, Boumertit A, Soubrier F, Midgley R, Kerr D, Parc Y, Bodmer WF. Role of rare variants in undetermined multiple adenomatous polyposis and early-onset colorectal cancer. *J Hum Genet*. 2012 Nov 26;57(11):709-16.

Leyton Y, Gonzalez-Hormazabal P, Blanco R, Bravo T, Fernandez-Ramires R, Morales S, Landeros N, Reyes JM, Peralta O, Tapia JC, Gomez F, Waugh E, Ibañez G, Pakomio J, Grau G, Jara L. Association of PALB2 sequence variants with the risk of familial and early-onset breast cancer in a South-American population. *BMC Cancer*. 2015 Jan 31;15:30.

Li FP, Fraumeni JF Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med*. 1969 Oct;71(4):747-52.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9.

Li Z, Guo X, Wu Y, Li S, Yan J, Peng L, Xiao Z, Wang S, Deng Z, Dai L, Yi W, Xia K, Tang L, Wang J. Methylation profiling of 48 candidate genes in tumor and matched normal tissues from breast cancer patients. *Breast Cancer Res Treat*. 2015 Feb;149(3):767-79.

Li FP, Fraumeni JF Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med.* 1969 Oct;71(4):747-52.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009 Aug 15;25(16):2078-9.

Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, Eng C, Parsons R. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat Genet.* 1997 May;16(1):64-7.

LiVolsi VA. Papillary thyroid carcinoma: an update. *Mod Pathol.* 2011 Apr;24 Suppl 2:S1-9.

Liyanarachchi S, Wojcicka A, Li W, Czetwertynska M, Stachlewska E, Nagy R, Hoag K, Wen B, Ploski R, Ringel MD, Kozłowicz-Gudzinska I, Gierlikowski W, Jazdzewski K, He H, de la Chapelle A. Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. *Thyroid.* 2013 Dec;23(12):1532-40.

LLOYD KM 2nd, DENNIS M. Cowden's disease. A possible new symptom complex with multiple system involvement. *Ann Intern Med.* 1963 Jan;58:136-42.

Luque RM, Ibáñez-Costa A, López-Sánchez LM, Jiménez-Reina L, Venegas-Moreno E, Gálvez MA, Villa-Osaba A, Madrazo-Atutxa AM, Japón MA, de la Riva A, Cano DA, Benito-López P, Soto-Moreno A, Gahete MD, Leal-Cerro A, Castaño JP. A cellular and molecular basis for the selective desmopressin-induced ACTH release in Cushing disease patients: key role of AVPR1b receptor and potential therapeutic implications. *J Clin Endocrinol Metab.* 2013 Oct;98(10):4160-9.

Ma XD, Cai GQ, Zou W, Huang YH, Zhang JR, Wang DT, Chen BL. First evidence for the contribution of the genetic variations of BRCA1-interacting protein 1 (BRIP1) to the genetic susceptibility of cervical cancer. *Gene.* 2013 Jul 25;524(2):208-13.

Ma XD, Cai GQ, Zou W, Huang YH, Zhang JR, Wang DT, Chen BL. First evidence for the contribution of the genetic variations of BRCA1-interacting protein 1 (BRIP1) to the genetic susceptibility of cervical cancer. *Gene.* 2013 Jul 25;524(2):208-13.

Majdak-Paredes EJ, Fatah F. Hereditary breast cancer syndromes and clinical implications. *J Plast Reconstr Aesthet Surg.* 2009 Feb;62(2):181-9.

Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet.* 2011 Sep;48(9):580-9.

Malkin D, Li FP, Strong LC, Fraumeni JF Jr, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science.* 1990 Nov 30;250(4985):1233-8. Erratum in: *Science.* 1993 Feb 12;259(5097):878.

Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst.* 2011 Jan 19;103(2):117-28. doi: 10.1093/jnci/djq495. Epub 2011 Jan 12. Erratum in: *J Natl Cancer Inst.* 2011 Apr 20;103(8):699.

Masciari S, Van den Abbeele AD, Diller LR, Rastarhuyeva I, Yap J, Schneider K, Digianni L, Li FP, Fraumeni JF Jr, Syngal S, Garber JE. F18-fluorodeoxyglucose-positron emission tomography/computed tomography screening in Li-Fraumeni syndrome. *JAMA.* 2008 Mar 19;299(11):1315-9.

Mavaddat N, et al. Consortium of Investigators of Modifiers of BRCA1/2. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *Cancer Epidemiol Biomarkers Prev.* 2012 Jan;21(1):134-47.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010 Sep;20(9):1297-303.

McPherson JP, Lemmers B, Chahwan R, Pamidi A, Migon E, Matysiak-Zablocki E, Moynahan ME, Essers J, Hanada K, Poonepalli A, Sanchez-Sweatman O, Khokha R, Kanaar R, Jasin M, Hande MP, Hakem R. Involvement of mammalian Mus81 in genome integrity and tumor suppression. *Science.* 2004 Jun 18;304(5678):1822-6.

Mihály Z, Kormos M, Lánckzy A, Dank M, Budczies J, Szász MA, Györfly B. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. *Breast Cancer Res Treat.* 2013 Jul;140(2):219-32.

Mihály Z, Kormos M, Lánckzy A, Dank M, Budczies J, Szász MA, Györfly B. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. *Breast Cancer Res Treat.* 2013 Jul;140(2):219-32.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science.* 1994 Oct 7;266(5182):66-71.

Milne RL, et al. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Hum Mol Genet.* 2014 Nov 15;23(22):6096-111.

Minagawa K, Yamamoto K, Nishikawa S, Ito M, Sada A, Yakushijin K, Okamura A, Shimoyama M, Katayama Y, Matsui T. Dereglulation of a possible tumour suppressor

gene, ZC3H12D, by translocation of IGK@ in transformed follicular lymphoma with t(2;6)(p12;q25). *Br J Haematol.* 2007 Oct;139(1):161-3.

Morrow M, Schnitt SJ, Norton L. Current management of lesions associated with an increased risk of breast cancer. *Nat Rev Clin Oncol.* 2015 Apr;12(4):227-38.

Na N, Dun E, Ren L, Li G. Association between ERCC5 gene polymorphisms and breast cancer risk. *Int J Clin Exp Pathol.* 2015 Mar 1;8(3):3192-7.

Nachalon Y, Katz O, Alkan U, Shvero J, Popovtzer A. Radiation-Induced Thyroid Cancer: Gender-Related Disease Characteristics and Survival. *Ann Otol Rhinol Laryngol.* 2015 Oct 5. pii: 0003489415608200.

Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. *Oncogene.* 2004 Aug 23;23(38):6445-70.

Nakaya N, Hemish J, Krasnov P, Kim SY, Stasiv Y, Michurina T, Herman D, Davidoff MS, Middendorff R, Enikolopov G. noxin, a novel stress-induced gene involved in cell cycle and apoptosis. *Mol Cell Biol.* 2007 Aug;27(15):5430-44. Epub 2007 May 21.

Narod S. The prevention of hereditary breast and ovarian cancer: a personal view. *Hered Cancer Clin Pract.* 2004 Feb 15;2(1):5-10.

Nejda N, Iglesias D, Moreno Azcoita M, Medina Arana V, González-Aguilera JJ, Fernández-Peralta AM. A MLH1 polymorphism that increases cancer risk is associated with better outcome in sporadic colorectal cancer. *Cancer Genet Cytogenet.* 2009 Sep;193(2):71-7.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet.* 2010 Sep;42(9):790-3.

Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol.* 2011 Aug 30;7(10):569-80.

Nosé V. Familial thyroid cancer: a review. *Mod Pathol.* 2011 Apr;24 Suppl 2:S19-33.

Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University, Baltimore, MD. MIM Number: MIM 188550: 8/31/2015:. World Wide Web URL: <http://omim.org/>

Pacini F, Castagna MG, Cipri C, Schlumberger M. Medullary thyroid carcinoma. *Clin Oncol (R Coll Radiol).* 2010 Aug;22(6):475-85.

Palles C, et al. Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Guarino E, Salguero I, Sherborne A, Chubb D, Carvajal-Carmona LG, Ma Y,

Kaur K, Dobbins S, Barclay E, Gorman M, Martin L, Kovac MB, Humphray S; CORGI Consortium; WGS500 Consortium, Lucassen A, Holmes CC, Bentley D, Donnelly P, Taylor J, Petridis C, Roylance R, Sawyer EJ, Kerr DJ, Clark S, Grimes J, Kearsey SE, Thomas HJ, McVean G, Houlston RS, Tomlinson I. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet.* 2013 Feb;45(2):136-44. doi: 10.1038/ng.2503. Epub 2012 Dec 23. Erratum in: *Nat Genet.* 2013 Jun;45(6):713. Guarino Almeida, Estrella [corrected to Guarino, Estrella].

Paradiso A, Formenti S. Hereditary breast cancer: clinical features and risk reduction strategies. *Ann Oncol.* 2011 Jan;22 Suppl 1:i31-6.

Parahuleva MS, Kanse SM, Parviz B, Barth A, Tillmanns H, Bohle RM, Sedding DG, Hölschermann H. Factor Seven Activating Protease (FSAP) expression in human monocytes and accumulation in unstable coronary atherosclerotic plaques. *Atherosclerosis.* 2008 Jan;196(1):164-71.

Parahuleva MS, Maj R, Hölschermann H, Parviz B, Abdallah Y, Erdogan A, Tillmanns H, Kanse SM. Regulation of monocyte/macrophage function by factor VII activating protease (FSAP). *Atherosclerosis.* 2013 Oct;230(2):365-72.

Park DJ, Odefrey FA, Hammet F, Giles GG, Baglietto L, ABCFS, MCCC, Hopper JL, Schmidt DF, Makalic E, Sinilnikova OM, Goldgar DE, Southey MC. FAN1 variants identified in multiple-case early-onset breast cancer families via exome sequencing: no evidence for association with risk for breast cancer. *Breast Cancer Res Treat.* 2011 Dec;130(3):1043-9.

Pelizzo MR, Pennelli G, Zane M, Galuppini F, Colletti PM, Merante Boschini I, Rubello D. Papillary thyroid carcinoma (PTC) in Lynch syndrome: Report of two cases and discussion on Lynch syndrome behaviour and genetics. *Biomed Pharmacother.* 2015 Aug;74:9-16.

Peng KT, Huang KC, Huang TW, Lee YS, Hsu WH, Hsu RW, Ueng SW, Lee MS. Single nucleotide polymorphisms other than factor V Leiden are associated with coagulopathy and osteonecrosis of the femoral head in Chinese patients. *PLoS One.* 2014 Aug 13;9(8):e104461.

Pinheiro H, Oliveira P, Oliveira C. Hereditary cancer risk assessment: challenges for the next-gen sequencing era. *Front Oncol.* 2015 Mar 23;5:62.

Ponti A, Lynge E, James T, Májek O, von Euler-Chelpin M, Anttila A, Fitzpatrick P, Mano MP, Kawai M, Scharpantgen A, Fracheboud J, Hofvind S, Vidal C, Ascunce N, Salas D, Bulliard JL, Segnan N, Kerlikowske K, Taplin S; ICSN DCIS Working group. International variation in management of screen-detected ductal carcinoma in situ of the breast. *Eur J Cancer.* 2014 Oct;50(15):2695-704.

Qi XP, Chen XL, Ma JM, Du ZF, Fei J, Yang CP, Cheng J, Song QZ, Han JS, Jin HY, Chen ZG, Wang JQ, Yang YP, Ying RB, Liu WT, Zhao Y, Chen CY, Jiang HL, Ke HP, Zhang XN. RET proto-oncogene genetic screening of families with multiple endocrine neoplasia type 2 optimizes diagnostic and clinical management in China. *Thyroid*. 2012 Dec;22(12):1257-65.

Qin TT, Chen T, Zhang Q, Du HN, Shu YQ, Luo K, Zhu LJ. Association between BRCA1 rs799917 polymorphism and breast cancer risk: A meta-analysis of 19,878 subjects. *Biomed Pharmacother*. 2014 Sep;68(7):905-10.

Rajaraman P, Bhatti P, Doody MM, Simon SL, Weinstock RM, Linet MS, Rosenstein M, Stovall M, Alexander BH, Preston DL, Sigurdson AJ. Nucleotide excision repair polymorphisms may modify ionizing radiation-related breast cancer risk in US radiologic technologists. *Int J Cancer*. 2008 Dec 1;123(11):2713-6.

Ramus SJ, et al. Ovarian Cancer Association Consortium. Consortium analysis of 7 candidate SNPs for ovarian cancer. *Int J Cancer*. 2008 Jul 15;123(2):380-8.

Rebbeck TR, et al. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. *JAMA*. 2015 Apr 7;313(13):1347-61. doi: 10.1001/jama.2014.5985. Erratum in: *JAMA*. 2015 Aug 11;314(6):628.

Repapi E, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet*. 2010 Jan;42(1):36-44.

Ritelli M, Dordoni C, Venturini M, Chiarelli N, Quinzani S, Traversa M, Zoppi N, Vascellaro A, Wischmeijer A, Manfredini E, Garavelli L, Calzavara-Pinton P, Colombi M. Clinical and molecular characterization of 40 patients with classic Ehlers-Danlos syndrome: identification of 18 COL5A1 and 2 COL5A2 novel mutations. *Orphanet J Rare Dis*. 2013 Apr 12;8:58.

Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, Gallinger S, Schwartz AG, Syngal S, Cote ML, Axilbund J, Schulick R, Ali SZ, Eshleman JR, Velculescu VE, Goggins M, Vogelstein B, Papadopoulos N, Hruban RH, Kinzler KW, Klein AP. ATM mutations in patients with hereditary pancreatic cancer. *Cancer Discov*. 2012 Jan;2(1):41-6.

Robson M, Offit K. Clinical practice. Management of an inherited predisposition to breast cancer. *N Engl J Med*. 2007 Jul 12;357(2):154-62.

Roux PF, Marthey S, Djari A, Moroldo M, Esquerré D, Estellé J, Klopp C, Lagarrigue S, Demeure O. Comparison of whole-genome (13X) and capture (87X) resequencing methods for SNP and genotype callings. *Anim Genet*. 2015

Santagata S, Thakkar A, Ergonul A, Wang B, Woo T, Hu R, Harrell JC, McNamara G, Schwede M, Culhane AC, Kindelberger D, Rodig S, Richardson A, Schnitt SJ, Tamimi

RM, Ince TA. Taxonomy of breast cancer based on normal cell phenotype predicts outcome. *J Clin Invest*. 2014 Feb;124(2):859-70.

Santarpia L, Myers JN, Sherman SI, Trimarchi F, Clayman GL, El-Naggar AK. Genetic alterations in the RAS/RAF/mitogen-activated protein kinase and phosphatidylinositol 3-kinase/Akt signaling pathways in the follicular variant of papillary thyroid carcinoma. *Cancer*. 2010 Jun 15;116(12):2974-83.

Santos M, Azevedo T, Martins T, Rodrigues FJ, Lemos MC. Association of RET genetic polymorphisms and haplotypes with papillary thyroid carcinoma in the Portuguese population: a case-control study. *PLoS One*. 2014 Oct 17;9(10):e109822.

Savas S, Geraci J, Jurisica I, Liu G. A comprehensive catalogue of functional genetic variations in the EGFR pathway: protein-protein interaction analysis reveals novel genes and polymorphisms important for cancer research. *Int J Cancer*. 2009 Sep 15;125(6):1257-65.

Sedding D, Daniel JM, Muhl L, Hersemeyer K, Brunsch H, Kemkes-Matthes B, Braun-Dullaeus RC, Tillmanns H, Weimer T, Preissner KT, Kanse SM. The G534E polymorphism of the gene encoding the factor VII-activating protease is associated with cardiovascular risk due to increased neointima formation. *J Exp Med*. 2006 Dec 25;203(13):2801-7.

Shehadeh LA, Yu K, Wang L, Guevara A, Singer C, Vance J, Papapetropoulos S. SRRM2, a potential blood biomarker revealing high alternative splicing in Parkinson's disease. *PLoS One*. 2010 Feb 8;5(2):e9104.

Shendure J, Akey JM. The origins, determinants, and consequences of human mutations. *Science*. 2015 Sep 25;349(6255):1478-83.

Shi XZ, Jin X, Xu P, Shen HM. Relationship between breast cancer and levels of serum thyroid hormones and antibodies: a meta-analysis. *Asian Pac J Cancer Prev*. 2014;15(16):6643-7.

Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. *Ann Oncol*. 2015 Jul;26(7):1291-9.

Shuen AY, Foulkes WD. Inherited mutations in breast cancer genes--risk and response. *J Mammary Gland Biol Neoplasia*. 2011 Apr;16(1):3-15.

Silva FC, Lisboa BC, Figueiredo MC, Torrezan GT, Santos EM, Krepschi AC, Rossi BM, Achatz MI, Carraro DM. Hereditary breast and ovarian cancer: assessment of point mutations and copy number variations in Brazilian patients. *BMC Med Genet*. 2014 May 15;15:55.

Smolarz B, Makowska M, Samulak D, Michalska MM, Mojs E, Wilczak M, Romanowicz H. Association between single nucleotide polymorphisms (SNPs) of XRCC2 and XRCC3 homologous recombination repair genes and triple-negative breast cancer in Polish women. *Clin Exp Med*. 2015 May;15(2):151-7.

Snape K, Ruark E, Tarpey P, Renwick A, Turnbull C, Seal S, Murray A, Hanks S, Douglas J, Stratton MR, Rahman N. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. *Breast Cancer Res Treat*. 2012 Jul;134(1):429-33.

Sobreira N, Modaff P, Steel G, You J, Nanda S, Hoover-Fong J, Valle D, Pauli RM. An anadysplasia-like, spontaneously remitting spondylometaphyseal dysplasia secondary to lamin B receptor (LBR) gene mutations: further definition of the phenotypic heterogeneity of LBR-bone dysplasias. *Am J Med Genet A*. 2015 Jan;167A(1):159-63.

Sonnenblick A, Fumagalli D, Sotiriou C, Piccart M. Is the differentiation into molecular subtypes of breast cancer important for staging, local and systemic therapy, and follow up? *Cancer Treat Rev*. 2014 Oct;40(9):1089-95.

Sonnenblick A, Fumagalli D, Sotiriou C, Piccart M. Is the differentiation into molecular subtypes of breast cancer important for staging, local and systemic therapy, and follow up? *Cancer Treat Rev*. 2014 Oct;40(9):1089-95.

Stagni V, Oropallo V, Fianco G, Antonelli M, Cinà I, Barilà D. Tug of war between survival and death: exploring ATM function in cancer. *Int J Mol Sci*. 2014 Mar 27;15(4):5388-409.

Stranneheim H, Wedell A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *J Intern Med*. 2016 Jan;279(1):3-15.

Stulp RP, Herkert JC, Karrenbeld A, Mol B, Vos YJ, Sijmons RH. Thyroid cancer in a patient with a germline MSH2 mutation. Case report and review of the Lynch syndrome expanding tumour spectrum. *Hered Cancer Clin Pract*. 2008 Feb 15;6(1):15-21.

Taber JM, Aspinwall LG, Stump TK, Kohlmann W, Champine M, Leachman SA. Genetic test reporting enhances understanding of risk information and acceptance of prevention recommendations compared to family history-based counseling alone. *J Behav Med*. 2015 Oct;38(5):740-53.

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S, Janecki T, Jiang P, Kehrler R,

Leblanc JF, Mitchell JT, McArthur-Morrison J, Nguyen K, Peng Y, Samson C, Schroeder M, Snyder SC, Steele L, Stringfellow M, Stroup C, Swedlund B, Swense J, Teng D, Thomas A, Tran T, Tranchant M, Weaver-Feldhaus J, Wong AK, Shizuya H, Eyfjord JE, Cannon-Albright L, Tranchant M, Labrie F, Skolnick MH, Weber B, Kamb A, Goldgar DE. The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat Genet.* 1996 Mar;12(3):333-7.

Tedaldi G, Danesi R, Zampiga V, Tebaldi M, Bedei L, Zoli W, Amadori D, Falcini F, Calistri D. First evidence of a large CHEK2 duplication involved in cancer predisposition in an Italian family with hereditary breast cancer. *BMC Cancer.* 2014 Jul 1;14:478.

Tinholt M, Viken MK, Dahm AE, Vollan HK, Sahlberg KK, Garred O, Børresen-Dale AL, Jacobsen AF, Kristensen V, Bukholm I, Kåresen R, Schlichting E, Skretting G, Lie BA, Sandset PM, Iversen N. Increased coagulation activity and genetic polymorphisms in the F5, F10 and EPCR genes are associated with breast cancer: a case-control study. *BMC Cancer.* 2014 Nov 19;14:845.

Tomsic J, Fultz R, Liyanarachchi S, He H, Senter L, de la Chapelle A. HABP2 G534E Variant in Papillary Thyroid Carcinoma. *PLoS One.* 2016 Jan 8;11(1):e0146315.

Tomsic J, He H, Akagi K, Liyanarachchi S, Pan Q, Bertani B, Nagy R, Symer DE, Blencowe BJ, de la Chapelle A. A germline mutation in SRRM2, a splicing factor gene, is implicated in papillary thyroid carcinoma predisposition. *Sci Rep.* 2015 Jul 2;5:10566.

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015 Mar;65(2):87-108.

Tsygankova OM, Prendergast GV, Puttaswamy K, Wang Y, Feldman MD, Wang H, Brose MS, Meinkoth JL. Downregulation of Rap1GAP contributes to Ras transformation. *Mol Cell Biol.* 2007 Oct;27(19):6647-58.

Virnig BA, Tuttle TM, Shamliyan T, Kane RL. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. *J Natl Cancer Inst.* 2010 Feb 3;102(3):170-8.

Visser WF, Verhoeven-Duif NM, de Koning TJ. Identification of a human trans-3-hydroxy-L-proline dehydratase, the first characterized member of a novel family of proline racemase-like enzymes. *J Biol Chem.* 2012 Jun 22;287(26):21654-62.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010 Sep;38(16):e164.

Wang K, Xu L, Pan L, Xu K, Li G. The functional BRCA1 rs799917 genetic polymorphism is associated with gastric cancer risk in a Chinese Han population. *Tumour Biol.* 2015 Jan;36(1):393-7.

Wang M, Vikis HG, Wang Y, Jia D, Wang D, Bierut LJ, Bailey-Wilson JE, Amos CI, Pinney SM, Petersen GM, de Andrade M, Yang P, Wiest JS, Fain PR, Schwartz AG, Gazdar A, Minna J, Gaba C, Rothschild H, Mandal D, Kupert E, Seminara D, Liu Y, Viswanathan A, Govindan R, Anderson MW, You M. Identification of a novel tumor suppressor gene p34 on human chromosome 6q25.1. *Cancer Res.* 2007 Jan 1;67(1):93-9.

Wang Y, Ray AM, Johnson EK, Zuhlke KA, Cooney KA, Lange EM. Evidence for an association between prostate cancer and chromosome 8q24 and 10q11 genetic variants in African American men: the Flint Men's Health Study. *Prostate.* 2011 Feb 15;71(3):225-31.

Win AK, Lindor NM, Jenkins MA. Risk of breast cancer in Lynch syndrome: a systematic review. *Breast Cancer Res.* 2013 Mar 19;15(2):R27.

Wójcicka A, Czetwertyńska M, Świerniak M, Długosińska J, Maciąg M, Czajka A, Dymecka K, Kubiak A, Kot A, Płoski R, de la Chapelle A, Jazdzewski K. Variants in the ATM-CHEK2-BRCA1 axis determine genetic predisposition and clinical presentation of papillary thyroid carcinoma. *Genes Chromosomes Cancer.* 2014 Jun;53(6):516-23.

World Cancer Research Fund / American Institute for Cancer Research. Policy and Action for Cancer Prevention. Food, Nutrition, and Physical Activity: a Global Perspective Washington DC: AICR, 2009

Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer.* 2013 Mar;13(3):184-99.

Xu L, Tang H, Chen DW, El-Naggar AK, Wei P, Sturgis EM. Genome-wide association study identifies common genetic variants associated with salivary gland carcinoma and its subtypes. *Cancer.* 2015 Jul 15;121(14):2367-74.

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-

exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013 Oct 17;369(16):1502-11.

Yeager N, Klein-Szanto A, Kimura S, Di Cristofano A. Pten loss in the mouse thyroid causes goiter and follicular adenomas: insights into thyroid function and Cowden disease pathogenesis. *Cancer Res*. 2007 Feb 1;67(3):959-66.

Yerushalmi R, Hayes MM, Gelmon KA. Breast carcinoma--rare types: review of the literature. *Ann Oncol*. 2009 Nov;20(11):1763-70.

Yip L, Carty SE. Systematic screening after Chernobyl: insights on radiation-induced thyroid cancer. *Cancer*. 2015 Feb 1;121(3):339-40.

Yu GP, Li JC, Branovan D, McCormick S, Schantz SP. Thyroid cancer incidence and survival in the national cancer institute surveillance, epidemiology, and end results race/ethnicity groups. *Thyroid*. 2010 May;20(5):465-73.

Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical management of breast cancer heterogeneity. *Nat Rev Clin Oncol*. 2015 Jul;12(7):381-94.

Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol*. 2011 May;12(5):477-88.

Zhang H, Teng X, Liu Z, Zhang L, Liu Z. Gene expression profile analyze the molecular mechanism of CXCR7 regulating papillary thyroid carcinoma growth and metastasis. *J Exp Clin Cancer Res*. 2015 Feb 12;34:16.

Zhang Q, Burdette JE, Wang JP. Integrative network analysis of TCGA data for ovarian cancer. *BMC Syst Biol*. 2014 Dec 31;8:1338.

Zhang ZZ, Huang J, Wang YP, Cai B, Han ZG. NOXIN as a cofactor of DNA polymerase-primase complex could promote hepatocellular carcinoma. *Int J Cancer*. 2015 Aug 15;137(4):765-75.

Zhao H, Wang D, Fu D, Xue L. Predicting the potential ankylosing spondylitis-related genes utilizing bioinformatics approaches. *Rheumatol Int*. 2015 Jun;35(6):973-9.

Zhao X, Li X, Zhang X. HABP2 Mutation and Nonmedullary Thyroid Cancer. *N Engl J Med*. 2015 Nov 19;373(21):2084.

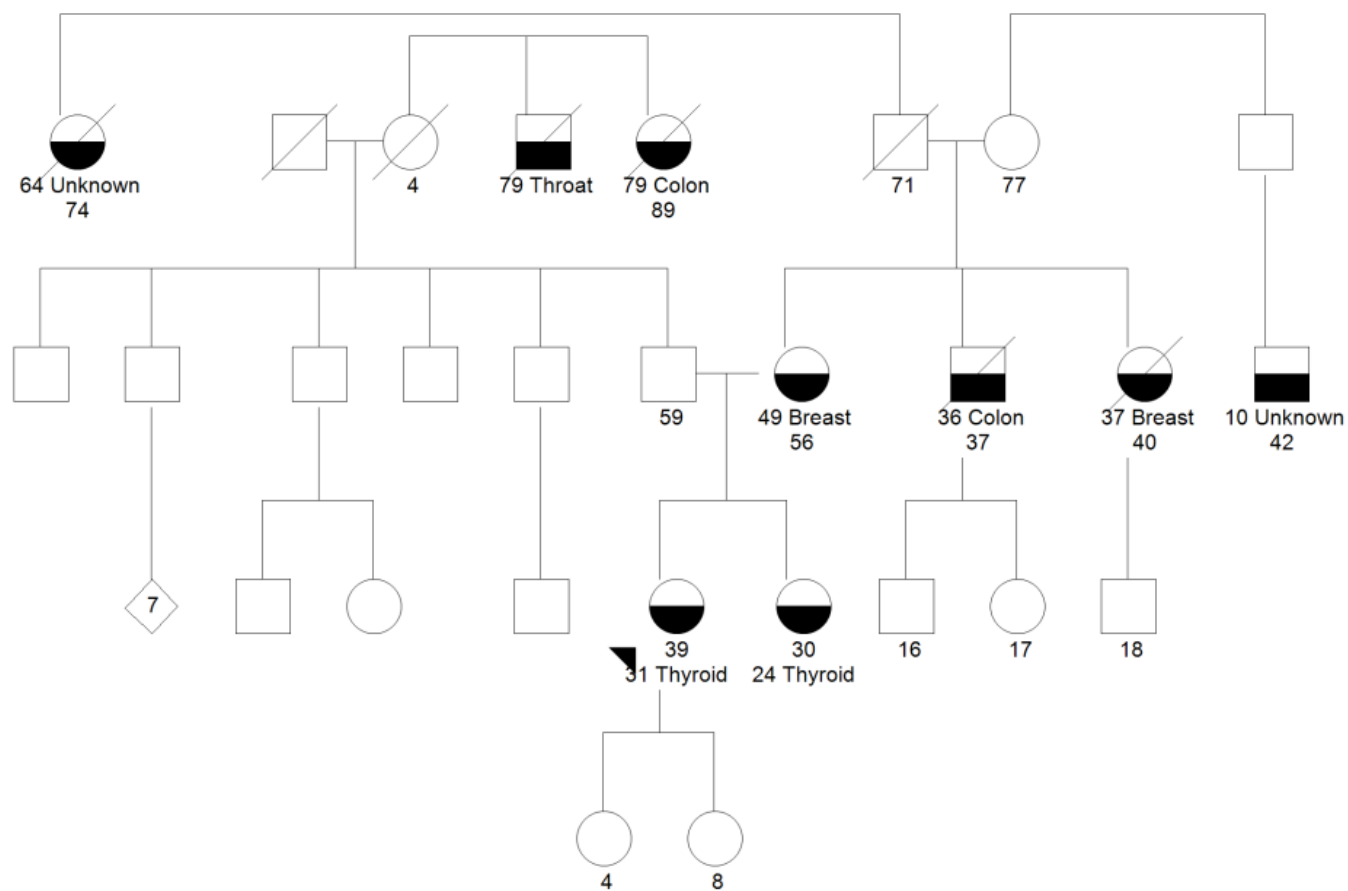
Zhou EY, Lin Z, Yang Y. HABP2 Mutation and Nonmedullary Thyroid Cancer. *N Engl J Med*. 2015 Nov 19;373(21):2084-5.

Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol*. 2015 Apr;3(4):286-95.

Zuo H, Gandhi M, Edreira MM, Hochbaum D, Nimgaonkar VL, Zhang P, Dipaola J, Evdokimova V, Altschuler DL, Nikiforov YE. Downregulation of Rap1GAP through epigenetic silencing and loss of heterozygosity promotes invasion and progression of thyroid tumors. *Cancer Res*. 2010 Feb 15;70(4):1389-97.

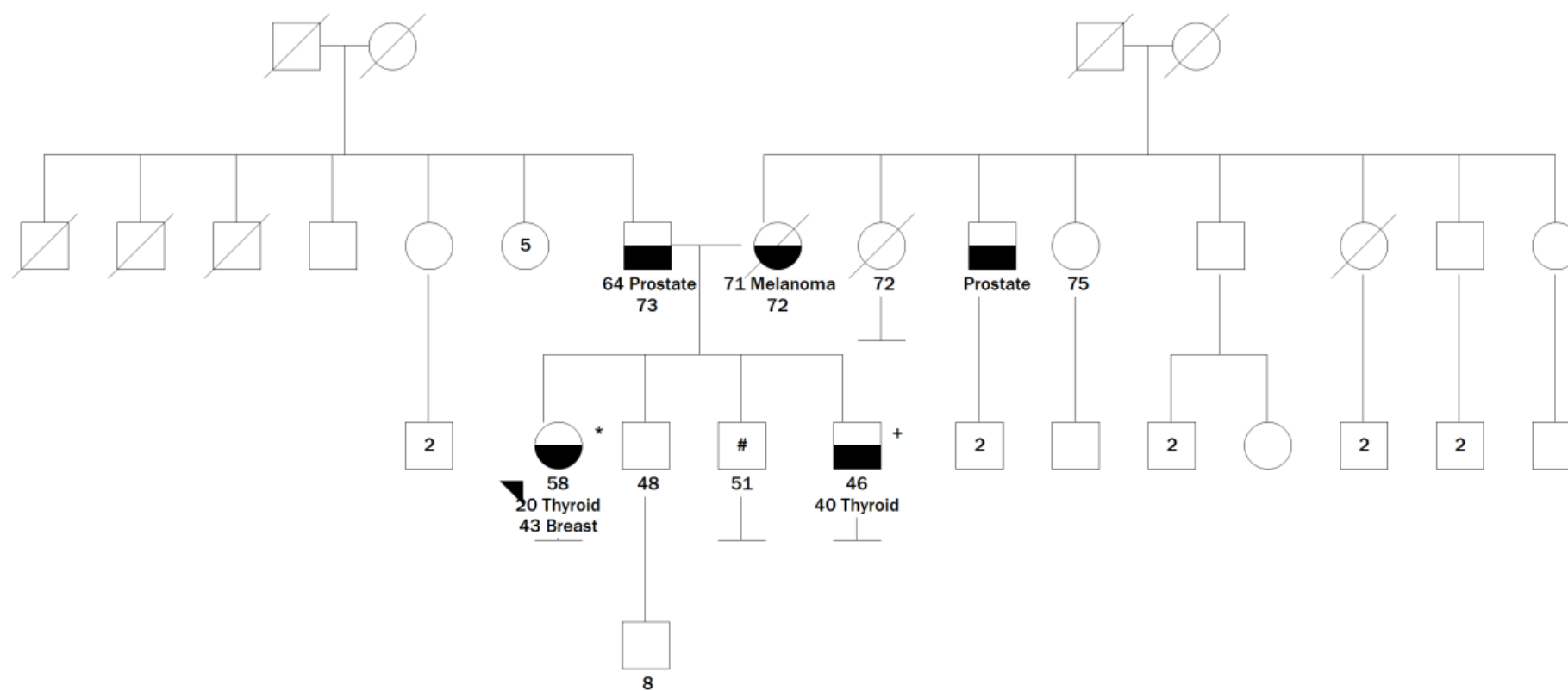
SUPLEMENTO 1
HEREDOGRAMAS

Família I
Pacientes 1,2,3

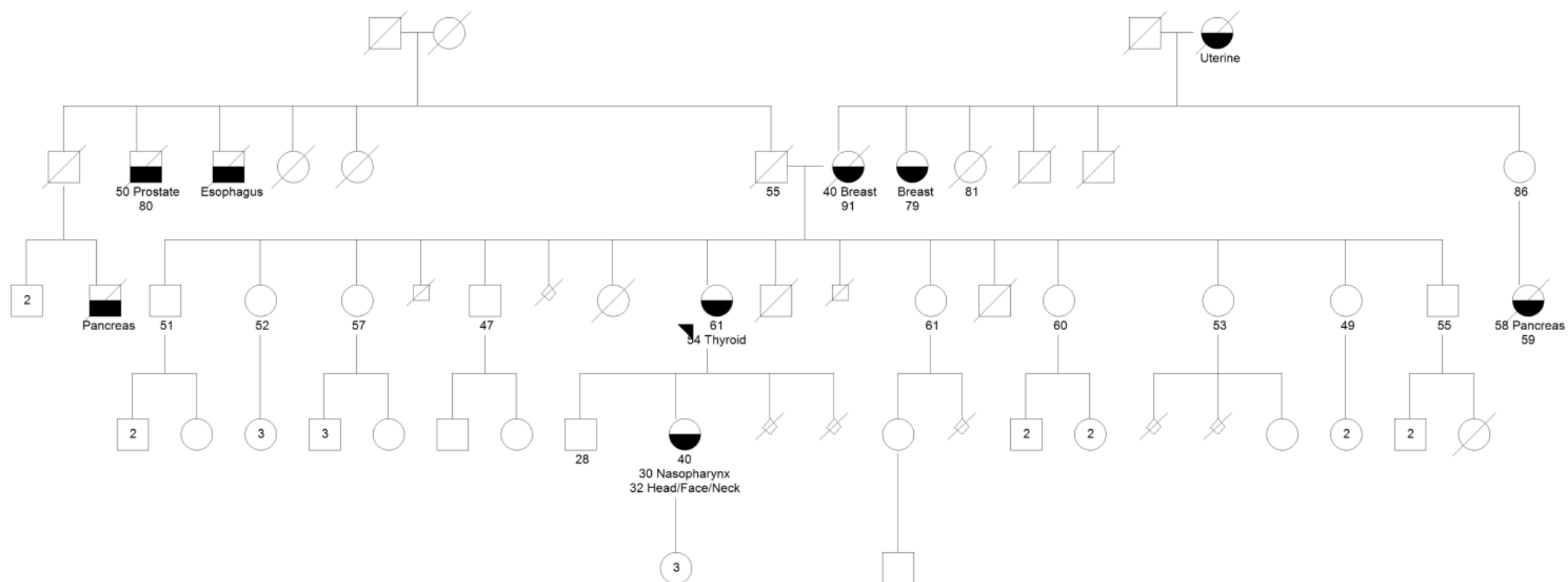


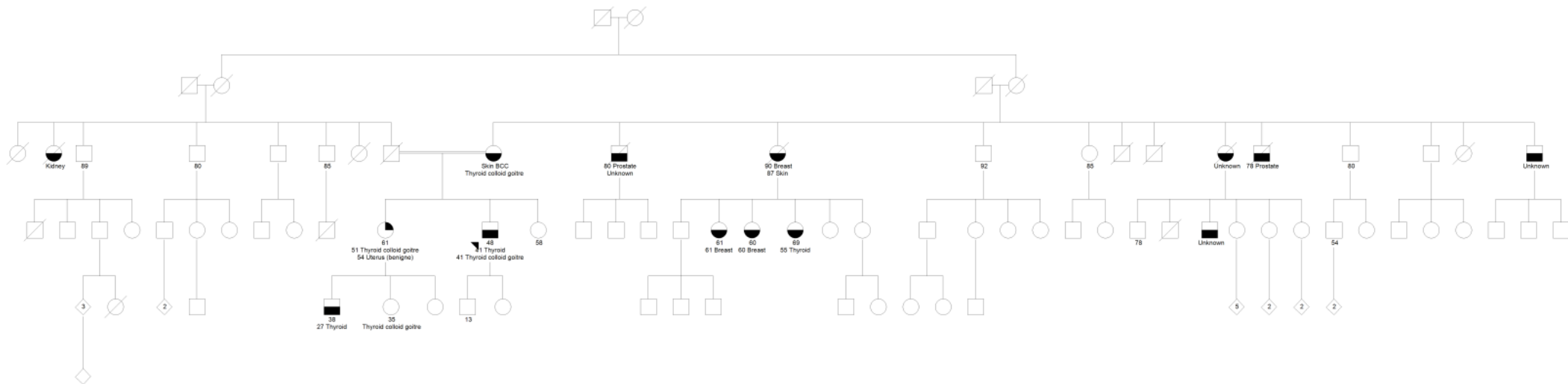
Família II
Pacientes 4,5

HEMANGIOMA CAPILAR / nevos melanocíticos
 + papilífero
 * folicular

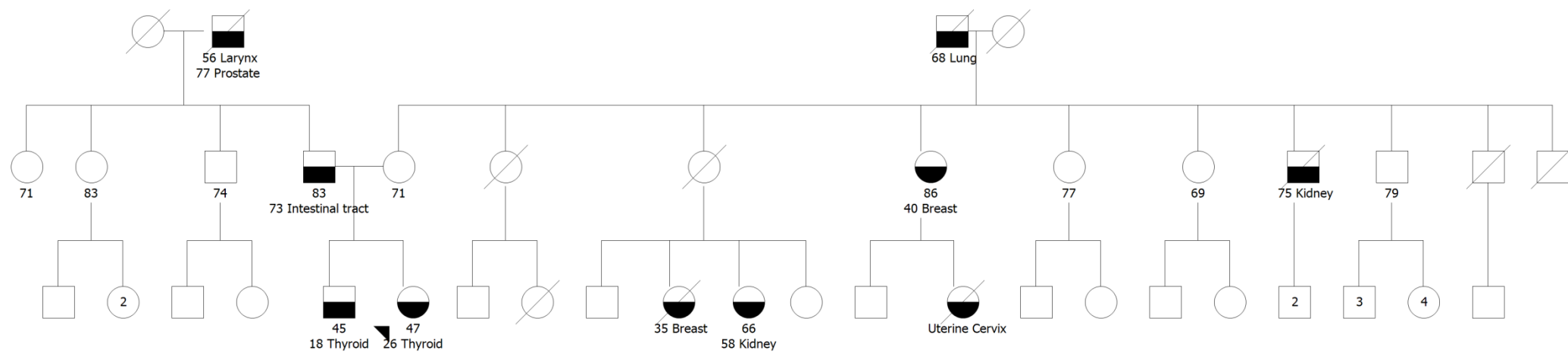


Paciente 6

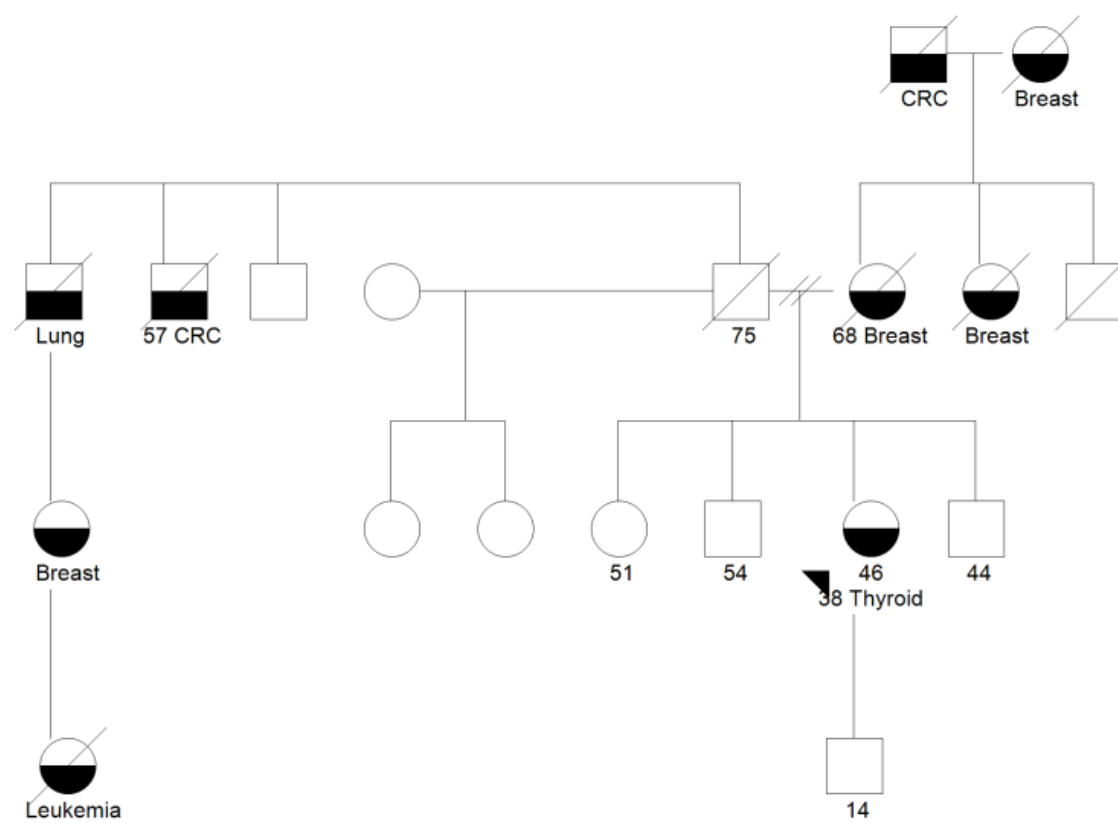




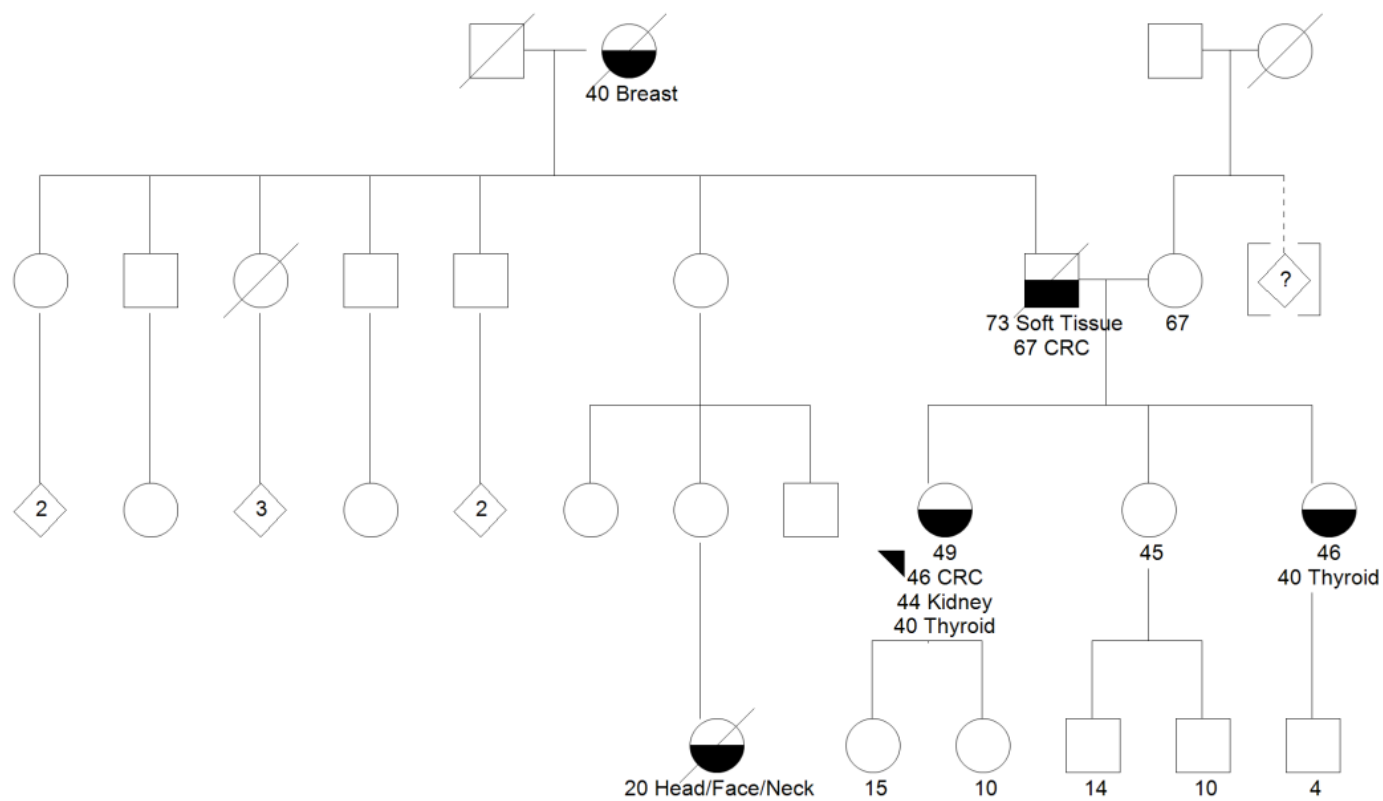
Paciente 8



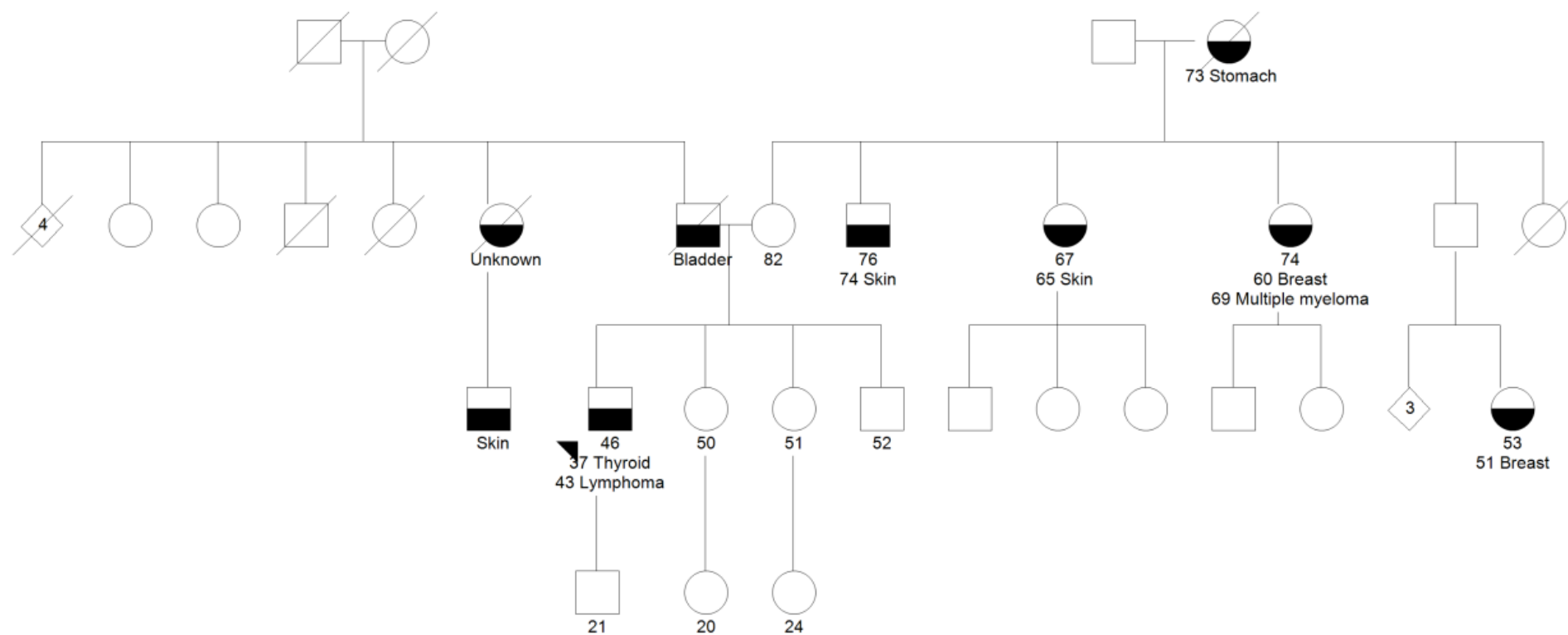
Paciente 10



Paciente 11



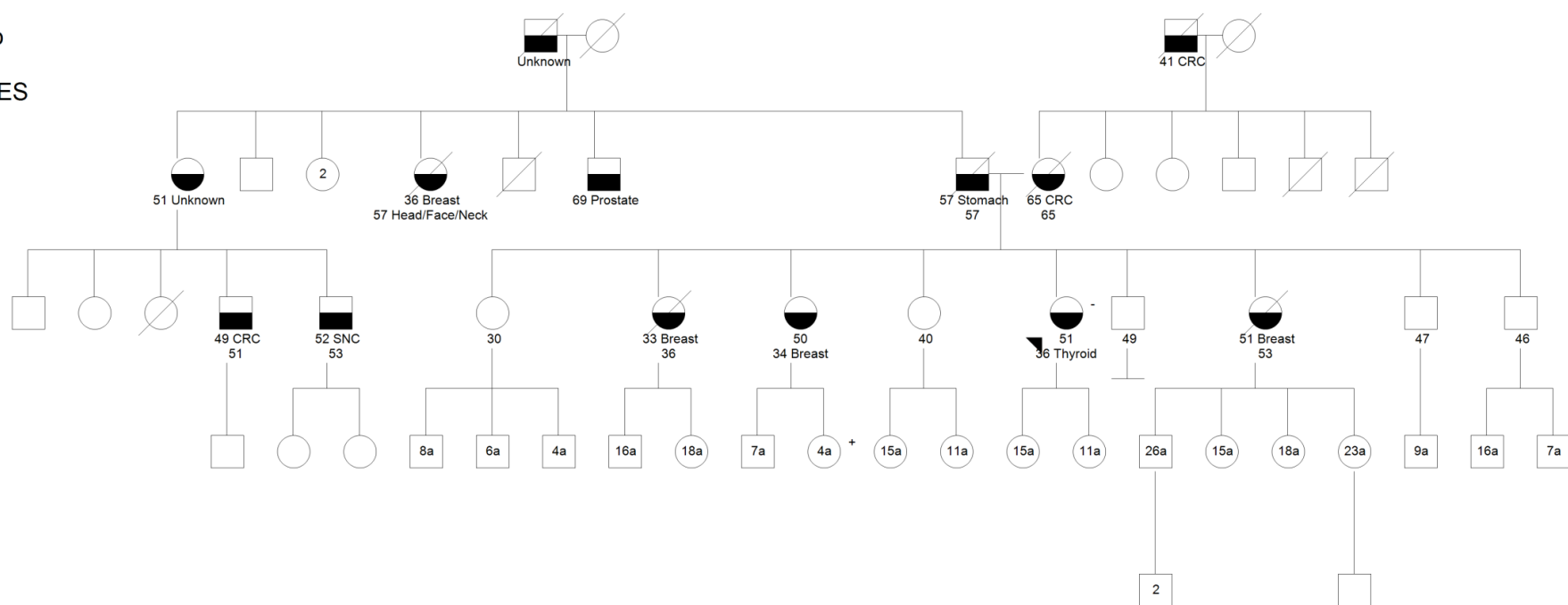
Paciente 12

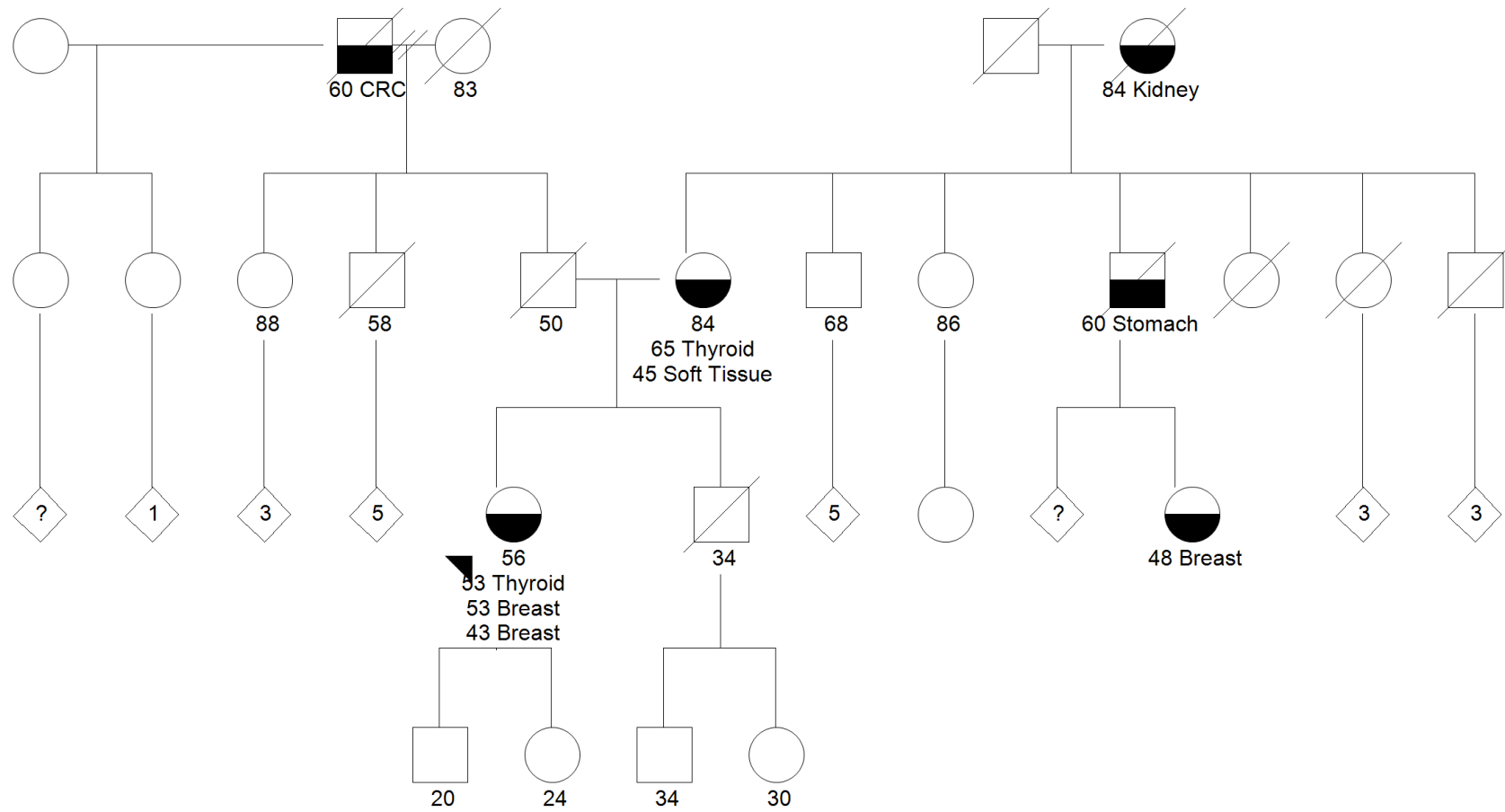


Paciente 13

+ hipotireoidismo

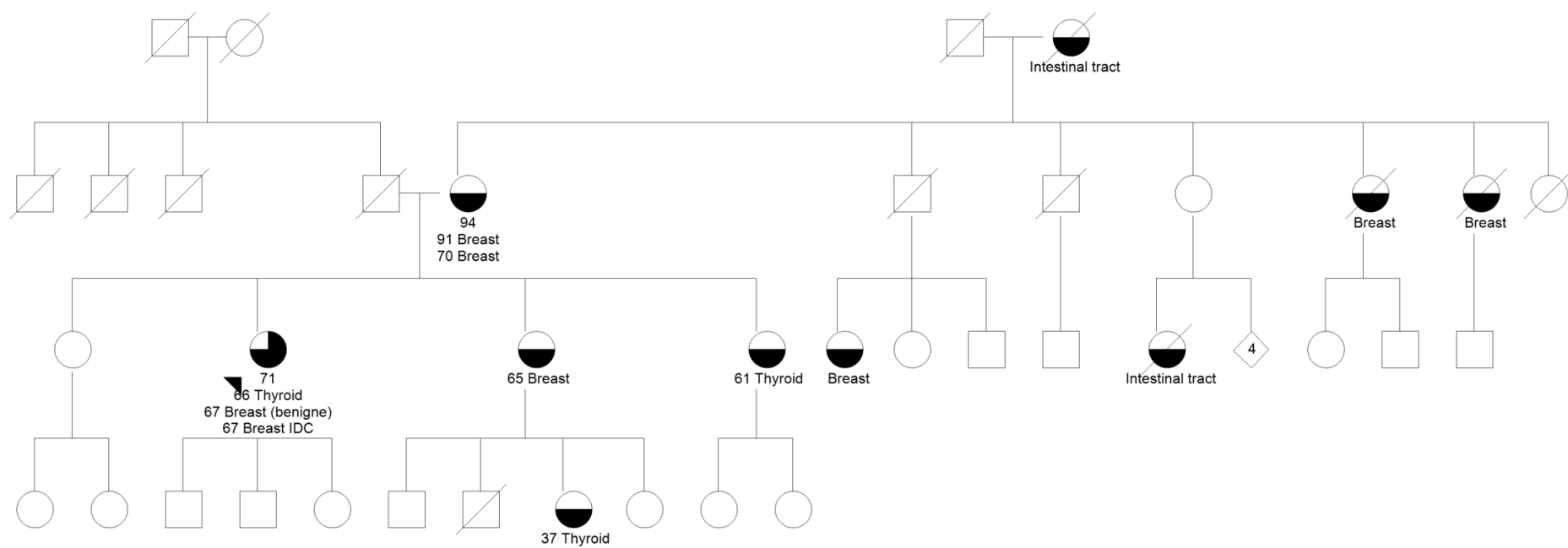
- PITS PALMARES





Paciente 15

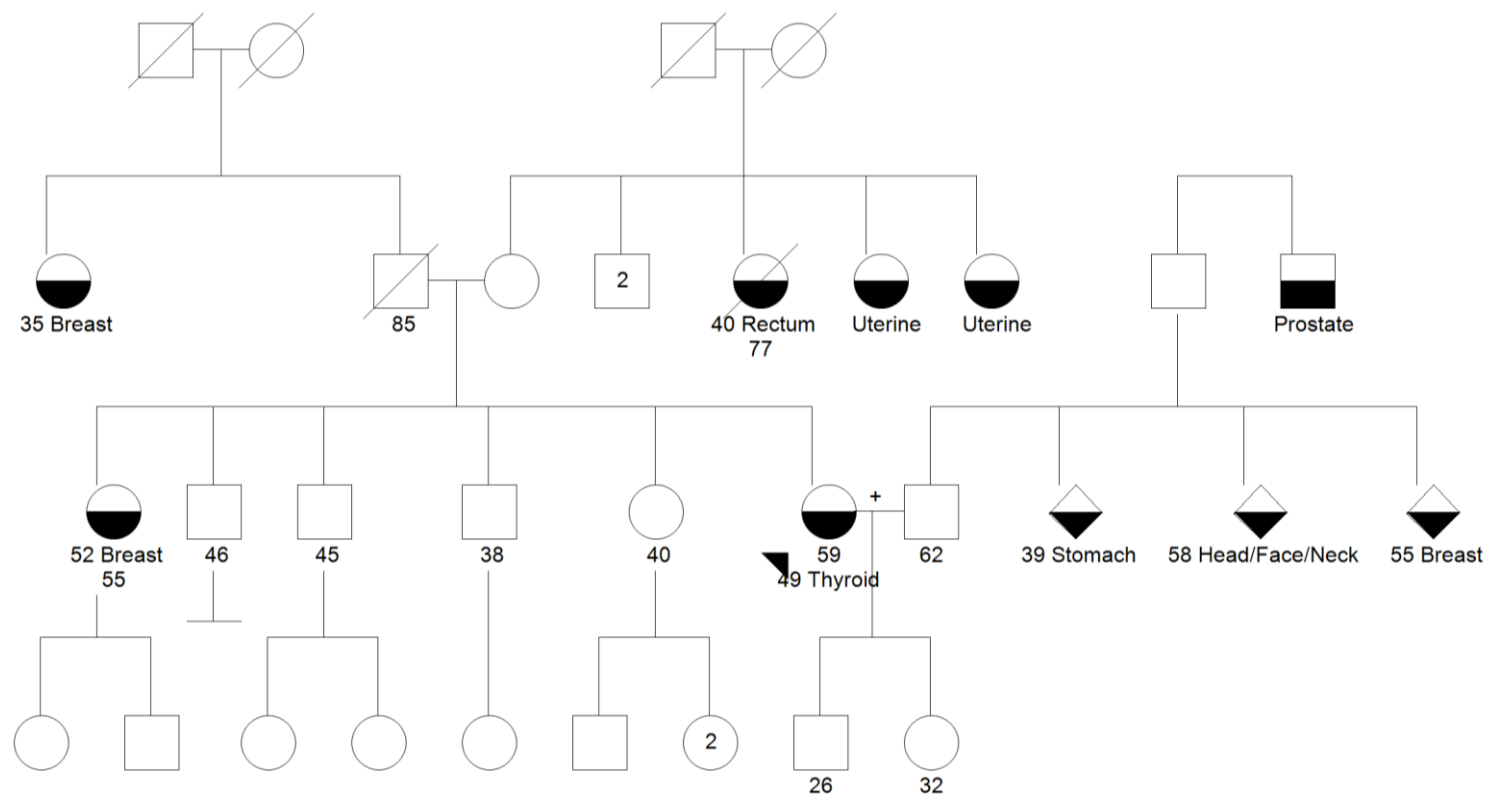
* Ca. mama bilateral



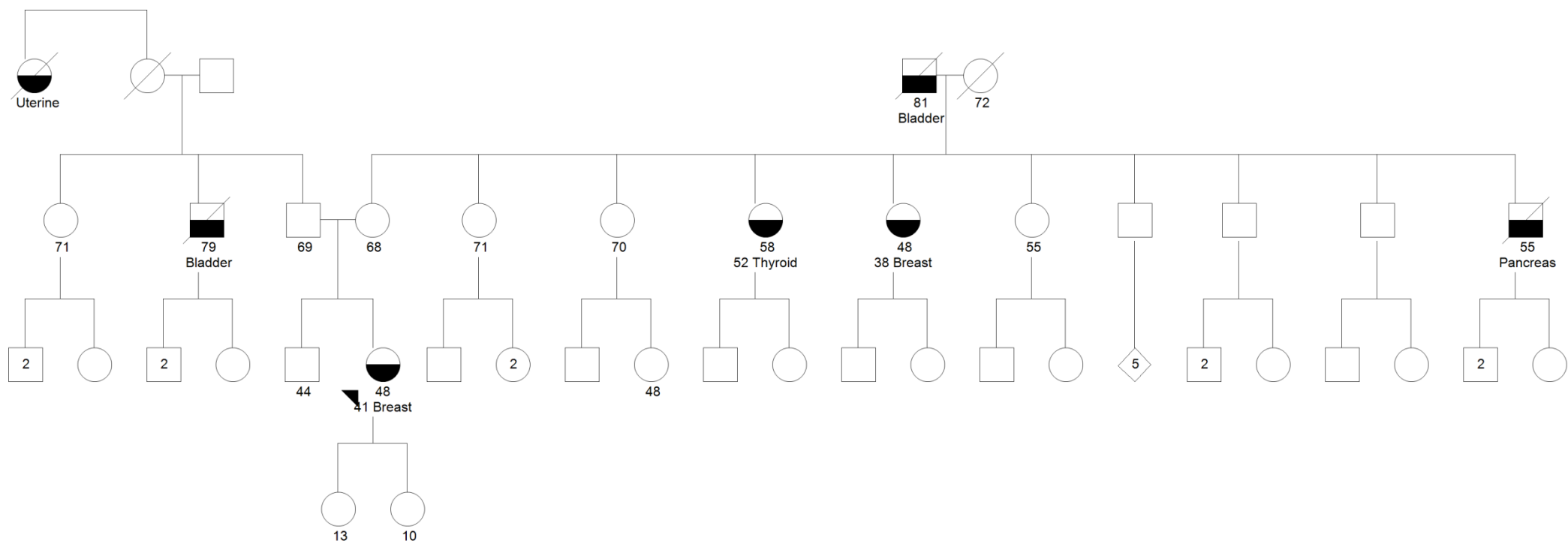
Paciente 16

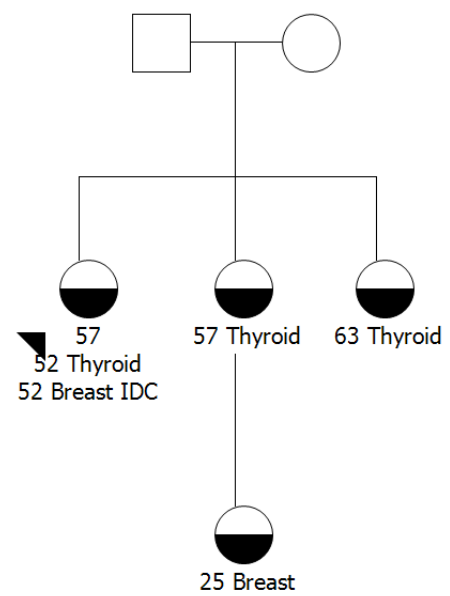
+ FIBROADENOMA e PÓLIPOS ENDOMETRIAIS

* NÓDULOS RETROAURICULARES e CISTO EM PUNHO (retirados s/ AP).

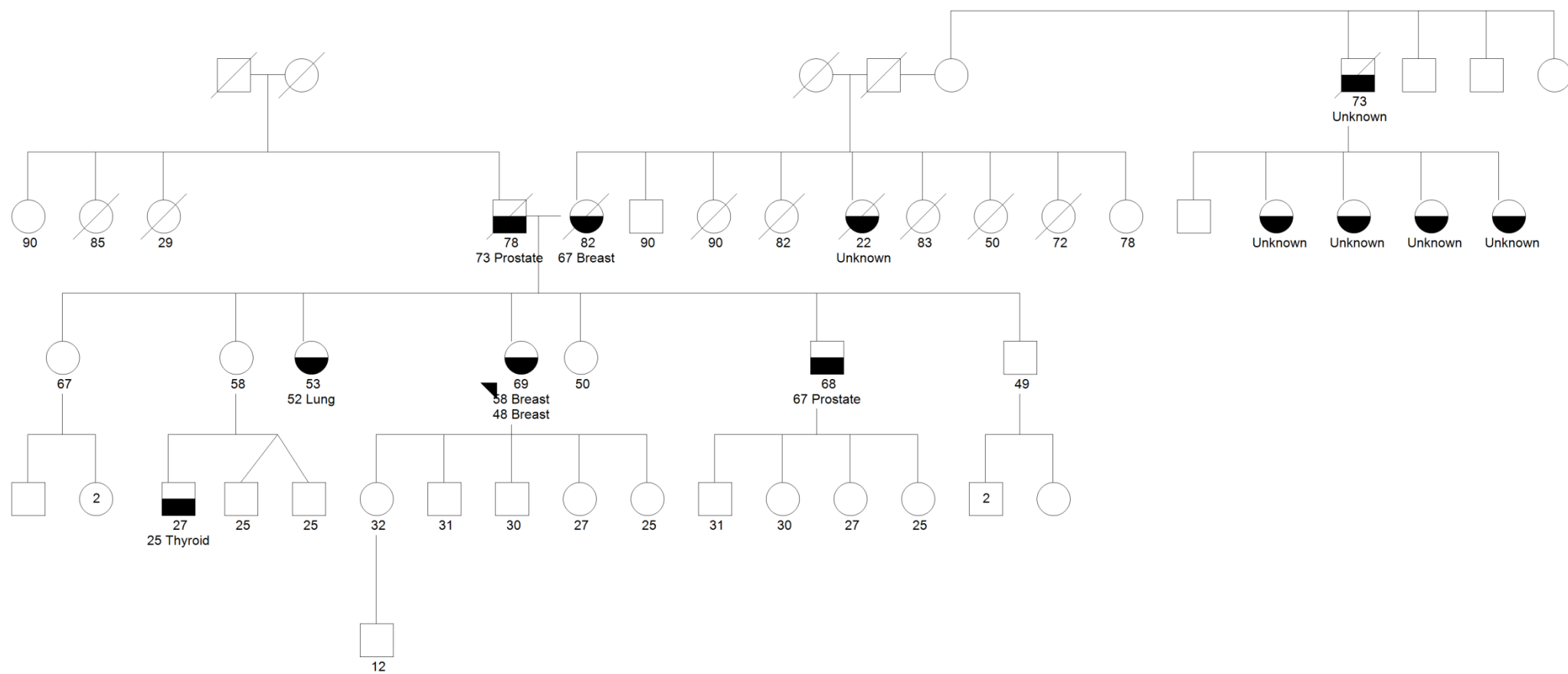


Paciente 17

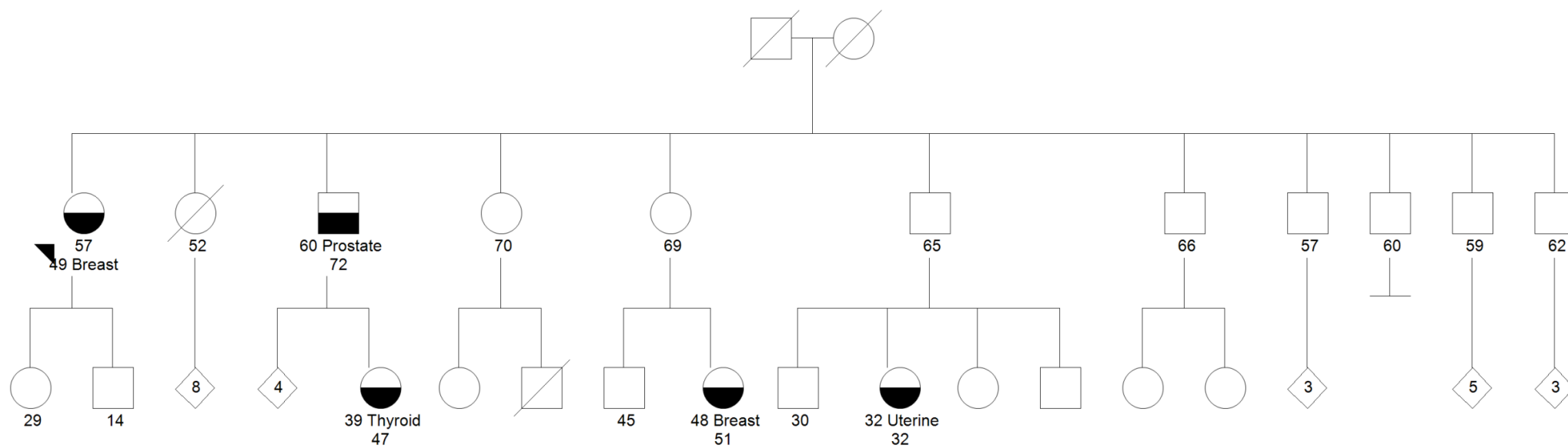


Paciente 18

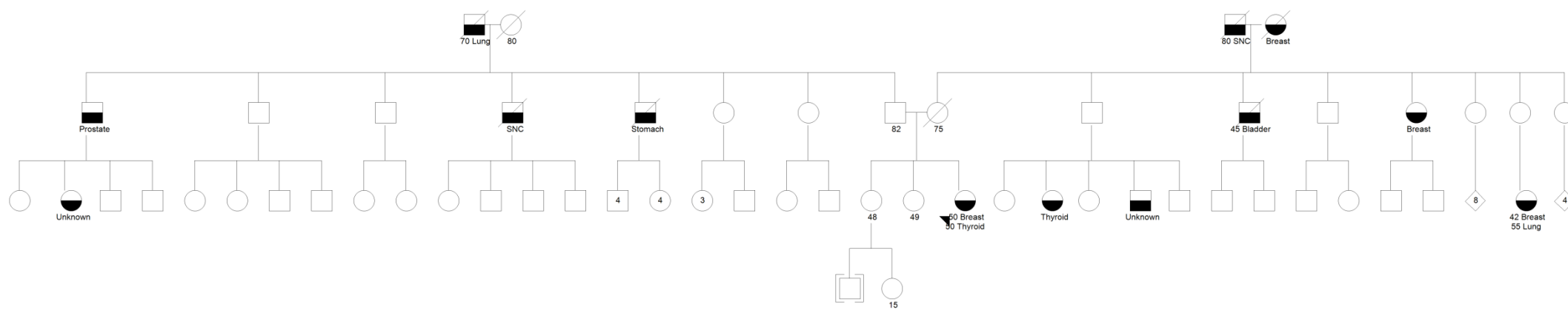
Paciente 19



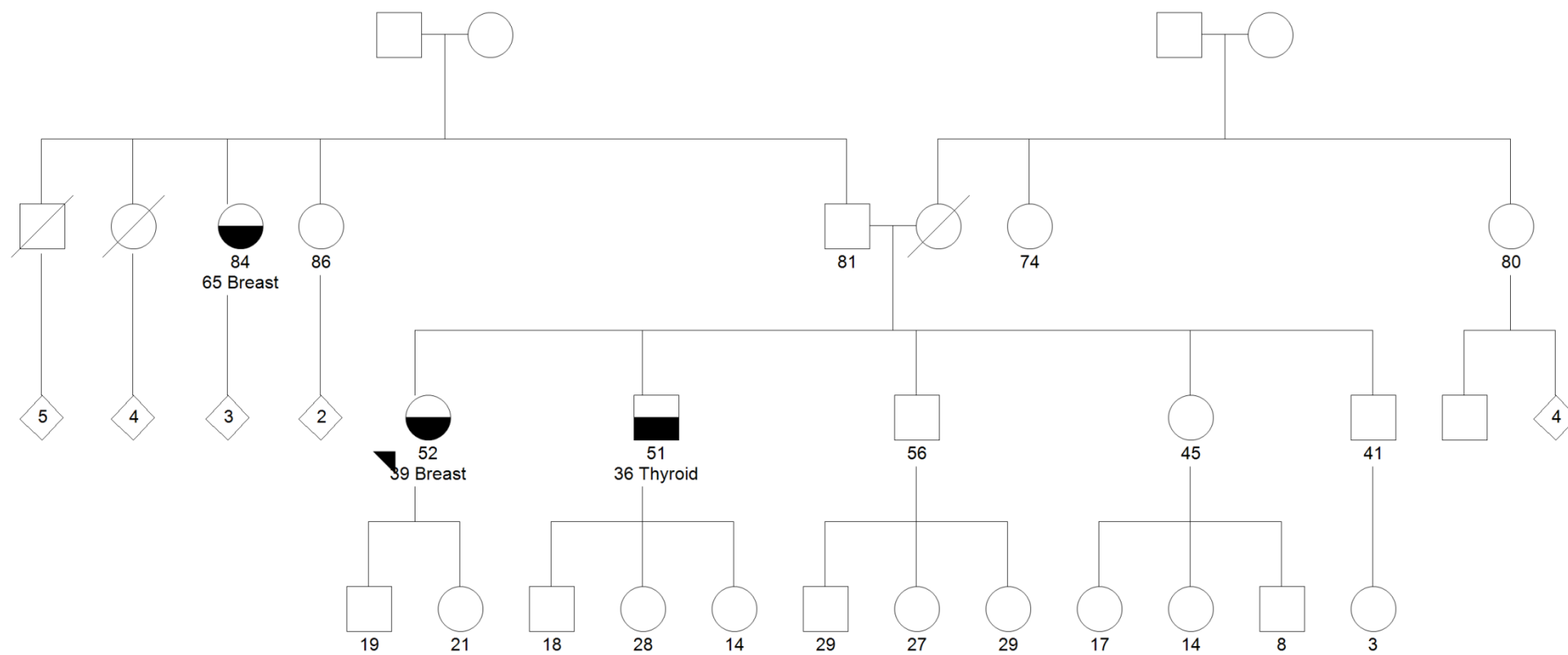
Paciente 20



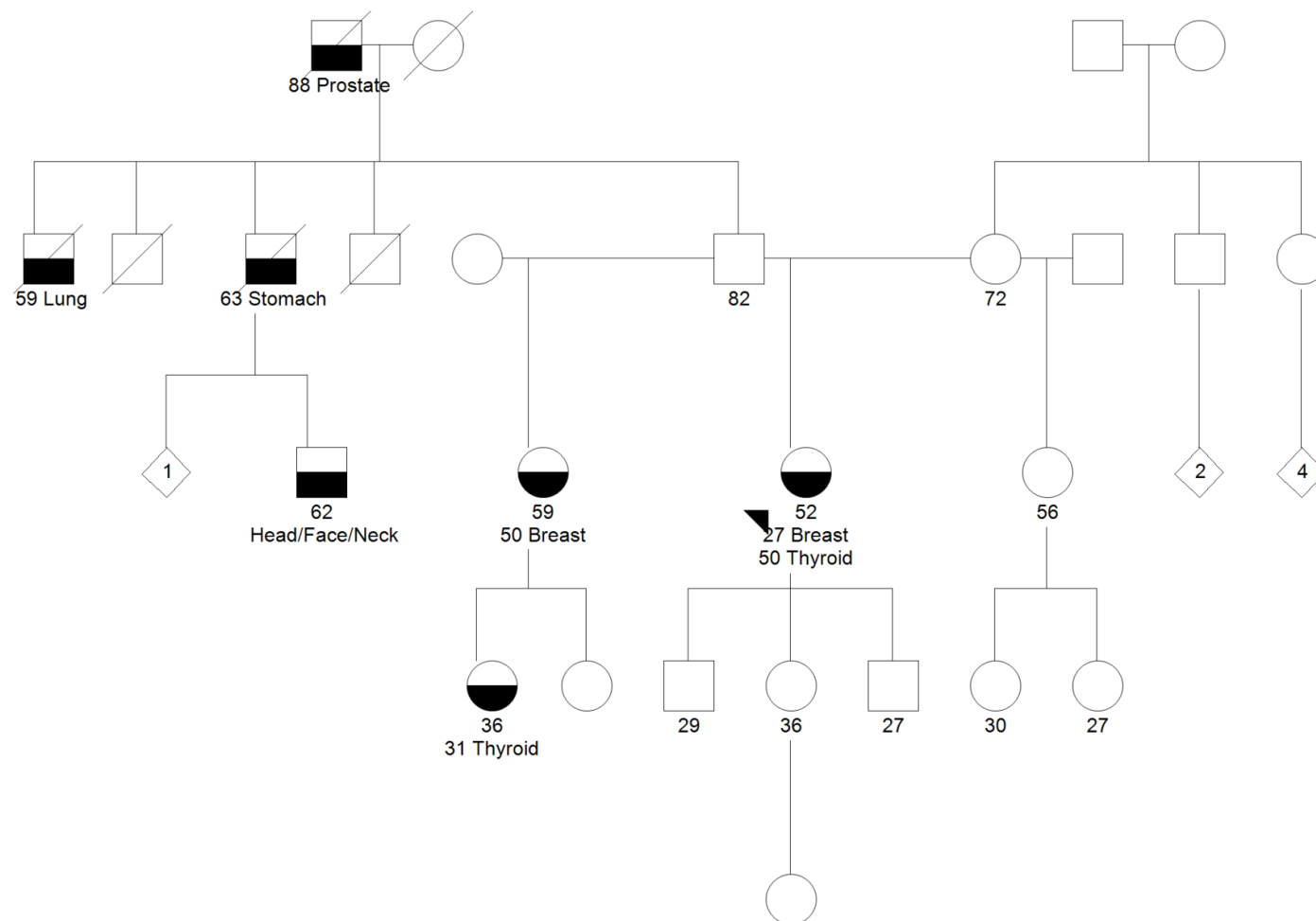
Paciente 21



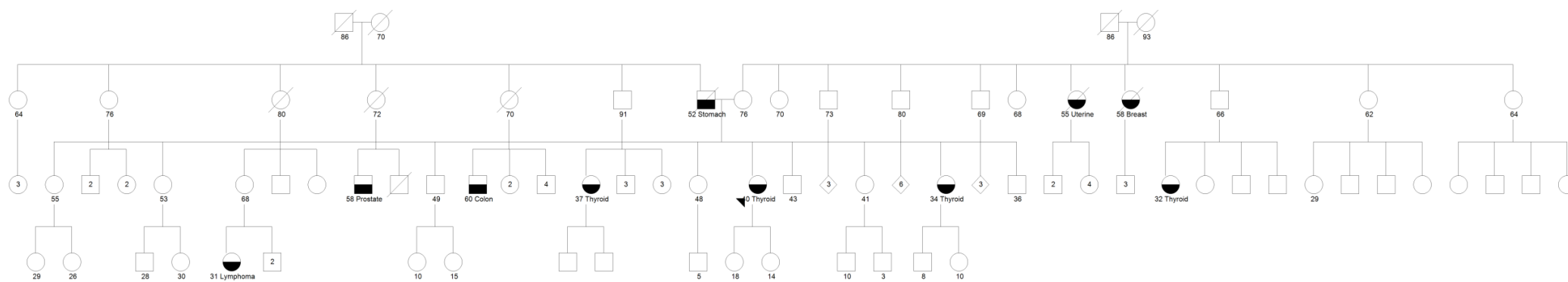
Paciente 22



Paciente 23



Paciente 24



SUPLEMENTO 2 Vias enriquecidas resultantes da inclusão de todas as alterações não sinônimas. Nesta análise foram incluídas as 287 alterações não-sinônimas contendo as 50 consideradas patogênicas (em pelo menos 3 bancos de dados de patogenidade). Todas as alterações não-sinônimas foram utilizadas para evidenciar o maior número de vias possível e pelo fato de terem potencial para atuar no desenvolvimento da doença em estudo, uma vez que a não-patogenicidade destas alterações é apenas predita por análises *in silico*.

Vias identificadas com PathDIP (ferramenta construída pelo nosso colaborador Dr. Igor Jurisica)

Via alterada enriquecida	p-value	Nº de alterações patogênicas por via	Nº de genes com alteração não-patogênica (não-sinônimas)	Total de genes na via	Taxa de alteração patogênica por via
KEGG Vascular smooth muscle contraction Homo sapiens human	0,0012	5	2	7	71%
KEGG Gap junction Homo sapiens human	0,0015	3	3	6	50%
Reactome Semaphorin interactions	0,0023	3	2	5	60%
Reactome Other semaphorin interactions	0,0023	2	1	3	67%
KEGG Calcium signaling pathway Homo sapiens human	0,0031	7	1	8	88%
PID PDGF receptor signaling network	0,0041	1	1	2	50%
SMPDB Corticotropin Activation of Cortisol Production	0,0041	1	1	2	50%
Reactome Lysine catabolism	0,0054	2	0	2	100%
KEGG Axon guidance Homo sapiens human	0,0081	5	1	6	83%
IPAVS Glutaric aciduria_type3 metabolite pathway	0,0085	2	0	2	100%
Reactome Biotin transport and metabolism	0,0085	2	0	2	100%
Wikipathways Angiogenesis overview	0,0100	1	3	4	25%
PID LPA receptor mediated events	0,0111	2	2	4	50%
Reactome Caspase mediated cleavage of cytoskeletal proteins	0,0121	2	0	2	100%
BioCarta extrinsic prothrombin activation pathway	0,0141	2	0	2	100%
SMPDB Glutaric Aciduria Type III	0,0141	2	0	2	100%
SMPDB Hyperlysinemia I Familial	0,0141	2	0	2	100%
SMPDB Hyperlysinemia II or Saccharopinuria	0,0141	2	0	2	100%
SMPDB Lysine Degradation	0,0141	2	0	2	100%
SMPDB Pyridoxine dependency with seizures	0,0141	2	0	2	100%
SMPDB Saccharopinuria Hyperlysinemia II	0,0141	2	0	2	100%
UniProt Pathways Amino acid metabolism	0,0148	1	0	1	100%
UniProt Pathways acetoacetate and fumarate from L phenylalanine step 3 6	0,0148	1	0	1	100%
UniProt Pathways tryptophan metabolism	0,0148	1	0	1	100%
Reactome Activation and oligomerization of BAK protein	0,0148	0	1	1	0%
systems biology org adipocyte_v2 1	0,0150	9	2	11	82%
KEGG Adipocytokine signaling pathway Homo sapiens human	0,0154	4	0	4	100%
KEGG Thyroid hormone synthesis Homo sapiens human	0,0168	2	2	4	50%
HumanCyc 1D i myo i inositol hexakisphosphate biosynthesis II mammalian	0,0184	2	0	2	100%

KEGG Glycosaminoglycan biosynthesis keratan sulfate Homo sapiens human	0,0184	1	1	2	50%
Reactome Signal Transduction	0,0198	23	11	34	68%
EHMN De novo fatty acid biosynthesis	0,0211	1	2	3	33%
KEGG Rap1 signaling pathway Homo sapiens human	0,0218	4	3	7	57%
KEGG cGMP PKG signaling pathway Homo sapiens human	0,0221	5	1	6	83%
PID Urokinase type plasminogen activator uPA and uPAR mediated signaling	0,0223	1	2	3	33%
Reactome Inositol phosphate metabolism	0,0235	3	0	3	100%
Reactome cGMP effects	0,0256	2	0	2	100%
KEGG Salivary secretion Homo sapiens human	0,0257	3	1	4	75%
Reactome Gene Expression	0,0279	7	2	9	78%
Wikipathways Activation of Gene Expression by SREBP SREBF	0,0282	2	0	2	100%
Wikipathways Angiogenesis	0,0282	0	2	2	0%
KEGG Platelet activation Homo sapiens human	0,0288	4	1	5	80%
Reactome Cytosolic iron sulfur cluster assembly	0,0289	1	0	1	100%
UniProt Pathways malonyl CoA biosynthesis	0,0289	1	0	1	100%
UniProt Pathways malonyl CoA from acetyl CoA step 1 1	0,0289	1	0	1	100%
Reactome Activation translocation and oligomerization of BAX	0,0289	0	1	1	0%
Reactome Transfer of LPS from LBP carrier to CD14	0,0289	0	1	1	0%
systems biology org Skeletal_muscle_v2 9	0,0298	9	2	11	82%
KEGG Prostate cancer Homo sapiens human	0,0313	2	2	4	50%
KEGG Lysine degradation Homo sapiens human	0,0316	3	0	3	100%
OntoCanro Caspase Cascade in Apoptosis	0,0316	2	1	3	67%
Wikipathways Nicotine Activity on Dopaminergic Neurons	0,0337	0	2	2	0%
PID Validated transcriptional targets of TAp63 isoforms	0,0361	3	0	3	100%
Wikipathways Blood Clotting Cascade	0,0365	2	0	2	100%
Reactome Termination of O glycan biosynthesis	0,0365	0	2	2	0%
Reactome GPCR downstream signaling	0,0384	13	5	18	72%
PID Caspase Cascade in Apoptosis	0,0392	2	1	3	67%
BioCarta intrinsic prothrombin activation pathway	0,0394	2	0	2	100%
KEGG Proximal tubule bicarbonate reclamation Homo sapiens human	0,0394	2	0	2	100%
KEGG Vitamin digestion and absorption Homo sapiens human	0,0394	2	0	2	100%
PharmGKB Sympathetic Nerve Pathway Neuroeffector Junction	0,0394	2	0	2	100%
Reactome Transmission across Chemical Synapses	0,0423	4	2	6	67%
KEGG Viral myocarditis Homo sapiens human	0,0424	2	1	3	67%
Reactome Nitric oxide stimulates guanylate cyclase	0,0424	2	0	2	100%
Wikipathways Signal Transduction of S1P Receptor	0,0424	2	0	2	100%

Reactome Sema4D induced cell migration and growth cone collapse	0,0424	1	1	2	50%
HumanCyc 1D i myo i inositol hexakisphosphate biosynthesis V from Ins 1 3 4 P3	0,0427	1	0	1	100%
HumanCyc 2 oxoglutarate decarboxylation to succinyl CoA	0,0427	1	0	1	100%
HumanCyc biotin carboxyl carrier protein assembly	0,0427	1	0	1	100%
IPAVS Biotinidase_Deficiency_metabolite_pathway	0,0427	1	0	1	100%
KEGG Biotin metabolism Homo sapiens human	0,0427	1	0	1	100%
Reactome Potassium transport channels	0,0427	1	0	1	100%
SMPDB Biotin Metabolism	0,0427	1	0	1	100%
SMPDB Biotinidase Deficiency	0,0427	1	0	1	100%
SMPDB Multiple carboxylase deficiency neonatal or early onset form	0,0427	1	0	1	100%
HumanCyc acyl CoA hydrolysis	0,0427	0	1	1	0%
PharmGKB Anti diabetic Drug Nateglinide Pathway Pharmacokinetics	0,0427	0	1	1	0%
PharmGKB Rosiglitazone Pharmacokinetic Pathway	0,0427	0	1	1	0%
Reactome Transport of fatty acids	0,0427	0	1	1	0%
SMPDB Rosiglitazone Metabolism Pathway	0,0427	0	1	1	0%
UniProt Pathways thyroid hormone biosynthesis	0,0427	0	1	1	0%
systems biology org hepatocyte_v2 9	0,0427	8	2	10	80%
Wikipathways Calcium Regulation in the Cardiac Cell	0,0432	4	1	5	80%
EHMN Lysine metabolism	0,0440	3	1	4	75%
IPAVS Calcium_Signaling_Pathway_Humanv3	0,0440	4	1	5	80%
Wikipathways GPCRs Class A Rhodopsin like	0,0444	6	1	7	86%
Reactome Activation of gene expression by SREBF SREBP	0,0454	2	0	2	100%
EHMN Saturated fatty acids beta oxidation	0,0454	1	1	2	50%
IPAVS Alkaptonuria_metabolite_pathway	0,0454	1	1	2	50%
KEGG Inositol phosphate metabolism Homo sapiens human	0,0456	3	0	3	100%
PID S1P3 pathway	0,0485	2	0	2	100%
Reactome Synthesis of IP3 and IP4 in the cytosol	0,0485	2	0	2	100%
Reactome Signaling by GPCR	0,0485	13	6	19	68%
systems biology org beta_cell_v2 3	0,0499	9	2	11	82%
Reactome Sema4D in semaphorin signaling	0,0516	1	1	2	50%
Wikipathways Spinal Cord Injury	0,0546	3	1	4	75%
RB Pathways CycD1_CDK4 6_module	0,0547	2	0	2	100%
PharmGKB Selective Serotonin Reuptake Inhibitor Pathway Pharmacodynamics	0,0547	1	1	2	50%
Reactome Keratan sulfate biosynthesis	0,0547	1	1	2	50%

Vias identificadas com IPA – Ingenuity Pathway Analysis

Via alterada/enriquecida	Banco de dados	ID da via	Nº de genes alterados	Background	P-Valor
--------------------------	----------------	-----------	-----------------------	------------	---------

			radosnavia	avia	
<i>Othersemaphorininteractions</i>	Reactome	REACT_19200	3	19	0,00437
<i>Gapjunction</i>	KEGGPATHWAY	hsa04540	6	89	0,00453
<i>Semaphorininteractions</i>	Reactome	REACT_19271	5	67	0,00489
<i>Vascularsmoothmusclecontraction</i>	KEGGPATHWAY	hsa04270	7	121	0,00495
<i>PDGFreceptorsignalingnetwork</i>	PID	pdgf_pathway	2	6	0,00575
<i>extrinsicprothrombinactivationpathway</i>	BioCarta	100168	2	12	0,0073
<i>Lysinecatabolism</i>	Reactome	REACT_1298	2	8	0,00961
<i>Calciumsignalingpathway</i>	KEGGPATHWAY	hsa04020	8	180	0,01192
<i>Biotintransportandmetabolism</i>	Reactome	REACT_11153	2	11	0,01616
<i>1D-myo- inositolhexakisphosphatebiosynthesisII (mammalian)</i>	BioCyc	PWY-6362	2	15	0,01665
<i>Caspase- mediatedcleavageofcytoskeletalprotein s</i>	Reactome	REACT_13541	2	12	0,01866
<i>LPAreceptormediatedevents</i>	PID	lysophospholipid_pathway	4	65	0,01892
<i>intrinsicprothrombinactivationpathway</i>	BioCarta	100122	2	21	0,01923
<i>ActivationofgeneexpressionbySREBF(SREBP)</i>	Reactome	REACT_147904	3	38	0,02443
<i>Recyclingofbileacidsandsalts</i>	Reactome	REACT_11042	2	15	0,02707
<i>Urokinase- typeplasminogenactivator(uPA)anduP AR-mediatedsignaling</i>	PID	upa_upar_pathway	3	42	0,02866
<i>Glycosaminoglycanbiosynthesis- keratansulfate</i>	KEGGPATHWAY	hsa00533	2	15	0,03038
<i>Adipocytokinesignalingpathway</i>	KEGGPATHWAY	hsa04920	4	70	0,03226
<i>Thyroidhormonesynthesis</i>	KEGGPATHWAY	hsa04918	4	72	0,03509
<i>Inositolphosphatemetabolism</i>	Reactome	REACT_150154	3	46	0,03864
<i>Integrincellsurfaceinteractions</i>	Reactome	REACT_13552	4	84	0,04572
<i>superpathwayofinositolphosphatecomp ounds</i>	BioCyc	PWY-6371	3	66	0,04679
<i>Regulationofcholesterolbiosynthesisby SREBP(SREBF)</i>	Reactome	REACT_147797	3	51	0,04924
<i>TerminationofO-glycanbiosynthesis</i>	Reactome	REACT_115835	2	22	0,05124
<i>ValidatedtranscriptionaltargetsofTAp6 3isoforms</i>	PID	tap63pathway	3	54	0,05207
<i>fosgeneexpressionanddrugabuse</i>	BioCarta	100161	1	5	0,05507
<i>Synthesisofbileacidsandbilesaltsvia7al pha-hydroxycholesterol</i>	Reactome	REACT_11041	2	23	0,05514
<i>CaspaseCascadeinApoptosis</i>	PID	caspase_pathway	3	56	0,05665
<i>Sema4Dinducedcellmigrationandgrow th-conecollapse</i>	Reactome	REACT_19277	2	24	0,05915
<i>RecruitmentofNuMAto mitotic centroso mes</i>	Reactome	REACT_15510	2	24	0,05915
<i>Alphaadrenergicreceptorsignalingpath way</i>	PANTHER	P00002	2	23	0,05961
<i>Lysinedegradation</i>	KEGGPATHWAY	hsa00310	3	52	0,05984
<i>Proximal tubule bicarbonate reclamation</i>	KEGGPATHWAY	hsa04964	2	23	0,06159
<i>cGMP-PKG signaling pathway</i>	KEGGPATHWAY	hsa04022	6	167	0,06333
<i>Axonguidance</i>	KEGGPATHWAY	hsa04360	5	127	0,06414
<i>Prostatecancer</i>	KEGGPATHWAY	hsa05215	4	89	0,06475
<i>Rap1 signaling pathway</i>	KEGGPATHWAY	hsa04015	7	211	0,06557