



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

**IDENTIFICAÇÃO DE POLIMORFISMOS DE NUCLEOTÍDEO ÚNICO COM
EFEITOS DELETÉRIOS EM TRANSCRIPTOMAS DE CÂNCER GÁSTRICO**

VIVIANE SANTOS DA SILVA

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Biotecnologia, da Universidade Federal do Pará, como requisito parcial para obtenção do grau de Mestre em Biotecnologia.

Orientador: Prof. Dr. Sylvain Henri Darnet.

Belém-Pará
Setembro/2015

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Silva, Viviane Santos da, 1985-

Identificação de polimorfismos de nucleotídeo único com efeitos deletérios em transcriptomas de câncer gástrico / Viviane Santos da Silva. - 2015.

Orientador: Sylvain Darnet.

Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Ciências Biológicas, Programa de Pós-Graduação em Biotecnologia, Belém, 2015.

1. Adenocarcinoma. 2. Estômago - Câncer - Aspectos genéticos. I. Título.

CDD 22. ed. 616.99433

INSTITUIÇÃO E FONTE FINANCIADORA

Este trabalho foi realizado no Laboratório de Biologia Computacional, do Instituto de Ciências Biológicas, da Universidade Federal do Pará.

Teve como fonte financiadora a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

AGRADECIMENTOS

Ao Professor Dr. Sylvain Darnet pela orientação, apoio e confiança.

A CAPES pelo apoio financeiro.

Ao Programa de Pós-Graduação em Biotecnologia pela oportunidade de ingressar no curso de mestrado.

À Universidade Federal do Pará pela infraestrutura disponível.

Aos amigos do Laboratório de Biologia Computacional pelo companheirismo e aprendizado.

À minha família pelo apoio incondicional e incentivo.

SUMÁRIO

1. INTRODUÇÃO	9
1.1. O QUE É CÂNCER?	10
1.2. CÂNCER GÁSTRICO	11
1.2.1. O que é Câncer gástrico?	11
1.2.2. Fatores epidemiológicos e etiológicos do câncer gástrico.....	14
1.3. HORMÔNIOS ESTEROIDES E O DESENVOLVIMENTO DO CÂNCER GÁSTRICO.	14
1.4. ALTERAÇÕES GENÉTICAS	15
1.4.1. Alterações genéticas e sua influência no desenvolvimento do câncer.	15
1.5. IDENTIFICAÇÃO DE ALTERAÇÕES GENÉTICAS NO GENOMA.	17
2. OBJETIVOS.....	19
2.1. OBJETIVO GERAL	19
2.2. OBJETIVOS ESPECÍFICOS	19
3. MATERIAIS E MÉTODOS.....	20
3.1. OBJETO DE ESTUDO	20
3.2. GENES CANDIDATOS	20
3.3. ANÁLISE DOS TRANSCRIPTOMAS	21
3.3.1. Pré-processamento do conjunto de dados.....	22
3.3.2. Alinhamento das leituras	23
3.3.3. Localização e identificação de SNP	23
3.4. ANÁLISE FUNCIONAL DOS SNPs.....	24
4. RESULTADOS	25
4.1. CONJUNTO DE DADOS	25
4.2. IDENTIFICAÇÃO DE SNPS.....	26
5. DISCUSSÃO	33
6. CONCLUSÃO	36
REFERÊNCIAS BIBLIOGRÁFICAS.....	37
APÊNDICES	42
ANEXOS.....	53

LISTA DE ABREVIATURAS

CG - Câncer Gástrico

dbSNP - *Single nucleotide polymorphism database*

DEG - *Differential expressed genes*

EBI - *European Bioinformatics Institute*

ENA - *European Nucleotide Archive*

GO - *Gene Ontology*

HP - *Helicobacter pylori*

KEGG - *Kyoto Encyclopedia of Genes and Genomes*

MNNG - Metilnitronitrosoguanidina

NGS - *Next-Generation Sequencing*

PROVEAN - *Protein Variation Effect Analyzer*

RNA-seq - *RNA sequencing*

SNP - *Single Nucleotide Polymorphism*

LISTA DE ILUSTRAÇÕES

FIGURA 1- Diagrama das características importantes da célula cancerígena.	10
FIGURA 2- Principais etapas envolvidas no processo de carcinogênese gástrica, segundo tipo histológico de tumor.	13
FIGURA 3- Principais classes de Polimorfismo de nucleotídeo único (SNP).	17
FIGURA 4- <i>Pipeline</i> bioinformático elaborado para identificar SNPs em genes relacionados à biossíntese dos hormônios esteroides, colesterol e receptor de progesterona em transcriptomas.	21
GRÁFICO 1- Distribuição dos transcriptomas segundo idade dos indivíduos e tipo histológico dos tumores.	25
GRÁFICO 2- Distribuição dos SNPs identificados nos transcriptomas analisados segundo via biossintética, idade, tipo histológico do tumor e transcriptoma com e sem câncer gástrico.	27
GRÁFICO 3- SNPs identificados em transcriptomas indivíduos com câncer gástrico e sem câncer gástrico. A: Distribuição dos SNPs encontrados nos genes relacionados às vias de biossíntese dos hormônios esteroides e do colesterol por tipo de substituição de base nucleotídica. B: Distribuição dos SNPs do tipo transição segundo substituição de nucleotídeo. C: Distribuição dos SNPs do tipo transversão segundo substituição de nucleotídeo.	28
GRÁFICO 4- Distribuição dos SNPs preditos com efeitos deletérios.	29

RESUMO

O câncer gástrico é uma das principais causas de mortalidade por câncer no mundo. Seu desenvolvimento está associado a fatores relacionados ao estilo de vida e a alterações genéticas que podem modificar genes e/ou vias biossintéticas e metabólicas importantes para a manutenção da integridade celular e tecidual. Pesquisas envolvendo câncer têm sido realizadas com o intuito de identificar genes e mutações que possam ser utilizados como marcadores genéticos e possíveis alvos terapêuticos. Estudos realizados com genes codificantes de proteínas da via dos hormônios esteroides já identificaram polimorfismos que podem estar relacionadas ao risco de câncer. Por tratar-se de uma via importante para processos fisiológicos e patológicos, estudos de identificação de polimorfismos genéticos nesta via, e na via de biossíntese do colesterol poderão contribuir com o entendimento da carcinogênese gástrica. Diante disso, foi realizada uma análise bioinformática em transcriptomas de tecidos gástricos com câncer e sem câncer, com o objetivo de identificar SNPs em genes codificantes de enzimas das vias de biossíntese dos hormônios esteroides, do colesterol e do receptor de progesterona que possam estar relacionadas ao desenvolvimento do câncer gástrico. A análise foi realizada por meio da Plataforma Galaxy, da ferramenta de alinhamento *BOWTIE* e do software de análise funcional de variações *PROVEAN Genome Variants*. SNPs deletérios foram identificadas nos genes *CYP51A1*, *DHCR24* e *SQLE* da via de biossíntese de hormônios esteroides e nos genes *CYP3A5*, *HSD17B12*, *UGT1A1*, *UGT1A5*, *UGT1A6* e *AKRIC3* da via biossintética do colesterol. O gene *CYP3A5* apresentou o maior número de SNPs deletérios. Estes resultados indicam uma possível participação dos genes analisados nos mecanismos moleculares envolvido no desenvolvimento do câncer gástrico.

Palavras-chave: Sequenciadores de nova geração, câncer gástrico, transcriptoma, polimorfismos genético, SNP.

ABSTRACT

Gastric cancer is one of the leading causes of cancer mortality in the world. Its development is associated with factors related to lifestyle and genetic alterations that can modify genes and important biosynthetic and metabolic pathways that help maintaining cellular and tissue integrity. One of the main cancer research objectives is identify genes and mutations that may be used as genetic markers and potential therapeutic targets. Studies with genes related to the pathway of steroid hormones proteins have already identified polymorphisms that may be related to the risk of cancer. Because it is an important route for physiological and pathological processes identification of genetic polymorphisms in this and cholesterol biosynthesis pathway may contribute to the understanding of gastric carcinogenesis. Therefore, we performed a comprehensive bioinformatics analysis of transcriptome datasets from gastric tissues with and without cancer, in order to identify SNPs in genes associated to enzymes of biosynthetic pathways of steroid hormones, cholesterol and progesterone receptor that may be related to gastric cancer. The analysis was performed using the Galaxy Platform, the Bowtie alignment tool and the Provean software for functional analysis of variations. Deleterious mutations have been identified in *CYP51A1*, *DHCR24* and *SQLE* genes in the steroid hormone biosynthesis pathway, and in *CYP3A5*, *HSD17B12*, *UGT1A1*, *UGT1A5*, *UGT1A6* and *AKR1C3* genes in the cholesterol biosynthetic pathway. The *CYP3A5* gene had the highest number of deleterious SNPs. These results indicate a possible participation of genes analyzed in the molecular mechanisms involved in the development of gastric cancer.

Key-words: Next-generation sequencing, gastric cancer, transcriptome, genetic polymorphisms, SNP.

1. INTRODUÇÃO

A publicação do genoma humano, em 2001, representou um grande marco na história da biologia. A partir de então, muito esforço tem sido empregado para compreender as variações genéticas humanas (CASALS *et al.*, 2012; DOSS *et al.*, 2010). Além disso, avanços tecnológicos computacionais e experimentais em biologia molecular têm revolucionado campos inteiros da medicina e biologia (DOSS *et al.*, 2010).

O uso de sequenciadores de nova geração (NGS – *Next - Generation Sequencing*) tem gerado uma grande quantidade de dados que estão disponíveis para a comunidade científica, podendo ser utilizado nas mais diversificadas análises e aplicações, como descoberta de alterações genéticas em partes ou na totalidade do genoma, estudos de transcriptomas de células, tecido ou de organismos, perfis epigenéticos, classificação de espécies, identificação de novos genes, entre outros (METZKER, 2010). Nesse contexto, a técnica de RNA-Seq (*RNA Sequencing*), vem emergindo como um dos principais sistemas de obtenção de perfis quantitativos de transcriptomas (MUTZ *et al.*, 2013).

Na perspectiva clínica existe um grande potencial para a utilização de tecnologias NGS no manejo e na descoberta de novos tratamentos na área da saúde (MELDRUM *et al.*, 2011).

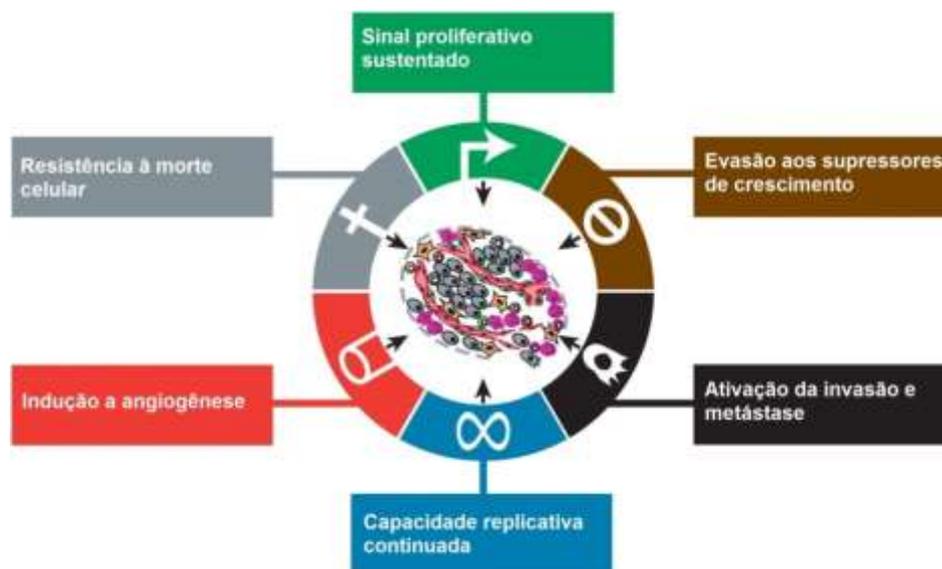
Consórcios científicos, formados por grupos de pesquisa de diversas partes do mundo, como o *Cancer Genome Atlas - TCGA* (<http://cancergenome.nih.gov/>), o *International Cancer Genome Consortium - ICGC* (<https://icgc.org/icgc>), o *International Human Genome Sequencing Consortium* (<https://www.genome.gov/11006939>), têm unido esforços para elucidar os aspectos genéticos, epigenéticos e moleculares do câncer, uma das patologias que mais afetam a população mundial.

Diante disso, estudos envolvendo novas tecnologias de sequenciamento e análises de transcriptomas de câncer poderão contribuir significativamente no conhecimento sobre a patogênese da doença, gerando tratamentos mais específicos e formas de diagnósticos mais precoces com base no perfil genético dos pacientes (MELDRUM *et al.*, 2011).

1.1. O QUE É CÂNCER?

O termo ‘câncer’ engloba mais de uma centena de doenças que apresentam fatores de risco e epidemiológicos distintos. Caracteriza-se pela presença de células com capacidade proliferativa descontrolada e metastática (STRATTON *et al.*, 2009). Para HANAHAN e WEINBERG (2011), a capacidade de manutenção do sinal proliferativo, a insensibilidade aos fatores inibitórios de crescimento celular, resistência à morte celular, capacidade replicativa continuada, indução à angiogênese, atividade invasora e metastática são características importantes da célula do câncer e estão fortemente associadas ao crescimento e disseminação do tumor (Figura 1).

FIGURA 1- Diagrama das características importantes da célula cancerígena.



Fonte: Traduzida de HANAHAN e WEINBERG (2011).

O desenvolvimento de células de câncer envolve modificações que levam ao aumento da atividade de proteínas codificadas pelos chamados oncogenes e/ou modificações que ocasionam a inativação dos genes supressores de tumor (BERTRAM, 2001; SARKAR *et al.*, 2013).

Grande parte das mutações que estão envolvidas no processo carcinogênico corresponde a mutações somáticas, adquiridas ao longo da vida, que surgem a partir de danos químicos ao DNA e que implicam no funcionamento inadequado de genes importantes

(BERTRAM, 2001). No quadro 1, são apresentados os principais tipos de mutação envolvidos no processo de desenvolvimento do câncer e seus efeitos no produto gênico.

QUADRO 1- Principais mutações/modificações que implicam no desenvolvimento de células de câncer.

Tipo de mutação	Efeito
Mutações de ponto	Mudança de aminoácido na sequência proteica.
Alteração na fase de leitura do DNA, aparecimento de códon de parada	Proteínas truncadas.
Instabilidade cromossômica	Amplificação gênica, hiperexpressão ou expressão inadequada de genes importantes.
Rearranjo ou quebra cromossômica	Perda ou fusão de genes (proteínas quiméricas ou com função alterada).
Modificações epigenéticas (metilação de citosinas, etc)	Silenciamento gênico

Fonte: Adaptado de BERTRAM (2001).

1.2. CÂNCER GÁSTRICO

1.2.1. O que é Câncer gástrico?

O Câncer gástrico (CG) representa um grupo heterogêneo de tumores com morfologia, aspectos moleculares e histológicos diferentes, originando-se, na maior parte das vezes, a partir do epitélio glandular, recebendo a denominação de adenocarcinoma (YAKIREVICH; RESNICK, 2013).

Estágios iniciais da doença caracterizam-se por uma fase assintomática ou por sintomas não específicos. Em estágios avançados da doença, sintomas como dores abdominais persistentes, anorexia e perda de peso, podem ser comumente observados (CORREA, 2013).

Recentemente, foi proposta uma classificação baseada em características moleculares dos tumores gástricos, nessa classificação quatro tipos de adenocarcinoma gástrico foram estabelecidos: tumores Epstein-Barr vírus (EBV) - positivo, tumores com

instabilidade em microssatélite, tumores estáveis genomicamente e tumores com instabilidade cromossômica (CANCER GENOME ATLAS RESEARCH NETWORK, 2014).

Outros sistemas de classificação já foram propostos, entretanto, os comumente utilizados são de Lauren e da Organização Mundial de Saúde (WHO - 2010) (YAKIREVICH; RESNICK, 2013).

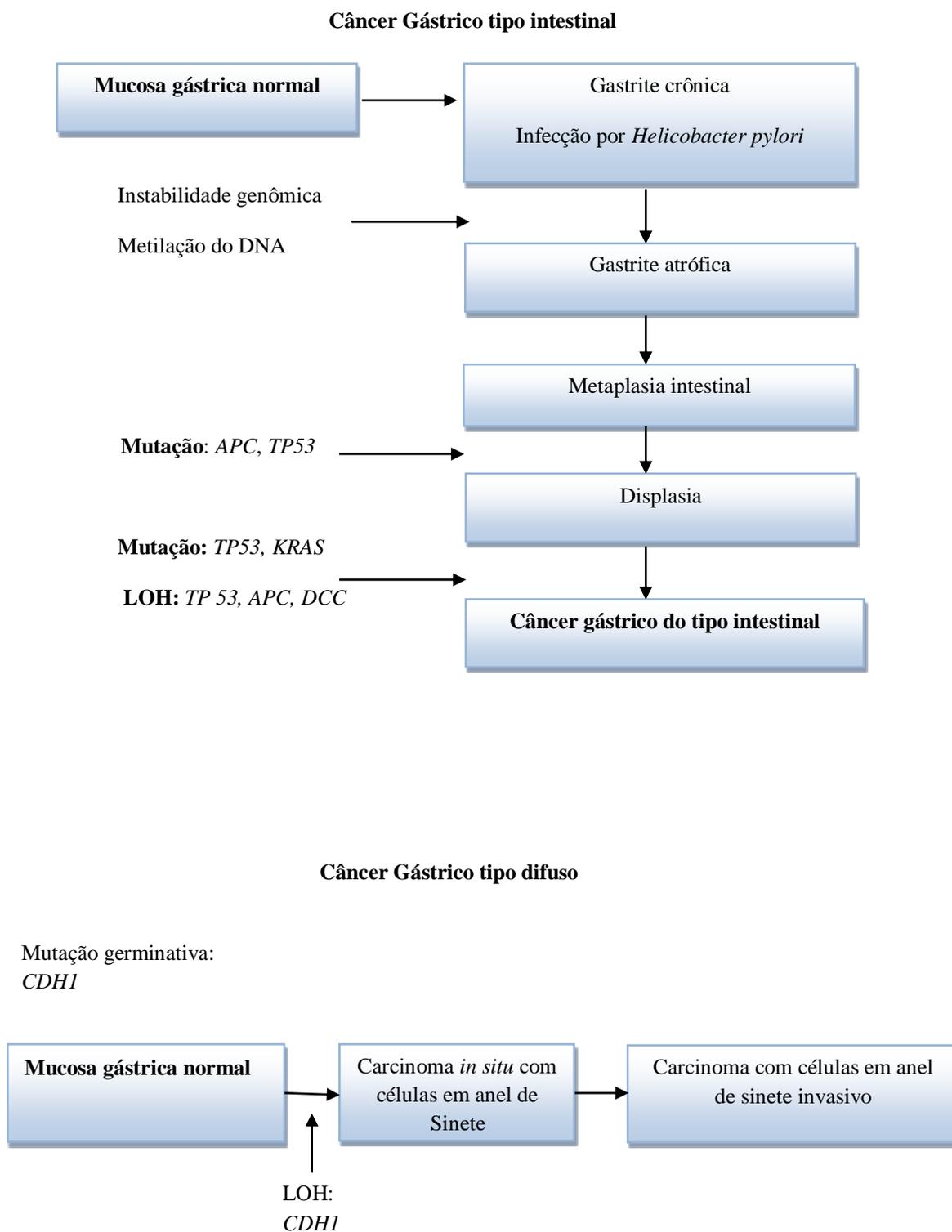
Histologicamente, o CG pode ser classificado em tipo intestinal e difuso. O CG do tipo intestinal caracteriza-se pela formação de glândulas com graus distintos de diferenciação e pela produção de mucina extracelular, o tipo difuso, é composto por células pouco coesas sem a formação de glândulas (YAKIREVICH; RESNICK, 2013).

Múltiplos mecanismos genéticos e epigenéticos levam a modificações da mucosa gástrica e conseqüentemente ao desenvolvimento do cancer gástrico do tipo intestinal. Inflamação crônica, atrofia, metaplasia e displasia gástrica são eventos importantes e consecutivos envolvidos neste tipo de câncer e englobam diversas alterações genéticas (YAKIREVICH; RESNICK, 2013). Variações genéticas ocasionam alterações em integrantes importantes dos mecanismos proliferativos, regulatórios e transcricionais das células, além de outros componentes celulares (ALTIERI *et al*, 2012).

O CG difuso apresenta alterações em moléculas envolvidas na interação e adesão célula-célula, resultando na invasão de células isoladas ou pequenos grupos de células. Mutação no gene *CDH1* configura-se como a base do processo de desenvolvimento deste tipo de CG (YAKIREVICH; RESNICK, 2013).

Na figura abaixo, são mostrados, resumidamente, os principais eventos ocorridos durante a oncogênese gástrica nos tipos intestinal e difuso.

FIGURA 2- Principais etapas envolvidas no processo de carcinogênese gástrica, segundo tipo histológico de tumor.



Fonte: Adaptado de YAKIREVICH; RESNICK (2013).

1.2.2. Fatores epidemiológicos e etiológicos do câncer gástrico

Segundo estatísticas mundiais do câncer, em 2012 o CG foi o 5º tipo de câncer maligno mais comum e a 3ª maior causa de morte por câncer no mundo, sendo diagnosticados mais de 900 mil novos casos, destes, mais de 70% ocorreram em regiões em desenvolvimento, como América do Sul, Ásia Ocidental, América Central e Europa Meridional (FERLAY *et al.*, 2015).

Na América do Sul, neste mesmo período, o CG foi o 6º tipo de câncer mais incidente, com 45.945 casos diagnosticados, e o 3º em mortalidade, com 37.221 mortes. No Brasil foram diagnosticados 19.690 casos e 16.077 mortes provocadas pela doença, a taxa de incidência foi cerca de 2 vezes maior entre os homens quando comparado as mulheres (WHO, 2012).

A etiologia do CG é multifatorial, envolvendo fatores exógenos e endógenos. Entre os fatores endógenos temos o histórico familiar de CG e alterações genéticas; e entre os exógenos ou fatores ambientais temos o consumo elevado de sódio e de alimentos com grande concentração de compostos nitrosos, o baixo consumo de legumes e verduras, infecção por *Helicobacter pylori*, além do uso de cigarro e consumo de bebidas alcoólicas (ZILBERSTEIN *et al.*, 2012). Vale resaltar que fatores endógenos e exógenos atuam conjuntamente no desenvolvimento do câncer.

1.3. HORMÔNIOS ESTERÓIDES E O DESENVOLVIMENTO DO CÂNCER GÁSTRICO.

Hormônios esteróides são os maiores determinantes de processos fisiológicos e patológicos (IGAZ *et al.*, 2002). Os hormônios esteróides têm como precursor principal a molécula de colesterol. A via do mevalonato, importante para a biossíntese do colesterol, e de prenilação de proteínas, tem sido implicada em diversos aspectos do desenvolvimento e progressão do câncer (GRUENBACHER; THURNHER, 2015).

A via do mevalonato converte o mevalonato em esteróis, dentre esses esteróis temos a molécula de colesterol, seus intermediários estão envolvidos em diversas funções, como modificações pós-traducionais de proteínas envolvidas na sinalização intracelular, crescimento celular, expressão gênica, glicosilação de proteínas e montagem do citoesqueleto (BUHAESCU; IZZEDINE, 2007).

Alterações na biossíntese e metabolismo desses hormônios parecem estar envolvidos na patogênese de diversas doenças, incluindo síndromes endócrinas e diferentes tipos de câncer (IGAZ *et al.*, 2002). Sequências das enzimas envolvidas na biossíntese e metabolismo dos hormônios esteróides já foram bem descritas na literatura e diversas regiões de polimorfismo tem sido encontradas, sendo que algumas parecem estar associadas com certas patologias (IGAZ *et al.*, 2002).

Tecidos gástricos normais e com câncer expressam receptores de hormônios esteróides, tais receptores reconhecem e permitem a ligação de estrogênios e androgênios (KARAT *et al.*, 1999). FREEDMAN e colaboradores (2009) sugerem que polimorfismos genéticos em genes da via biossintética e metabólica dos hormônios sexuais possam desempenhar algum papel na etiologia do câncer gástrico.

Estudos como o de HYLAND *et al.* (2013) e de ASHTON *et al.* (2010) já demonstraram uma associação entre polimorfismos em genes das vias metabólica e biossintética e o risco de câncer. HYLAND e colaboradores identificaram alterações genéticas presentes nos genes *SULT2B1*, *CYP1B1*, *CYP3A5*, *SHBG* e *CYP11A1* como associados ao risco de desenvolvimento de carcinoma epidermóide de esôfago. Já ASHTON e colaboradores (2010), em um estudo realizado com mulheres caucasianas, reportaram que polimorfismos nos genes *AR*, *CYP11A1*, *CYP1B1*, *ESR1* e *GSTM1* estão associados ao risco de câncer uterino.

Além desses estudos, SUN *et al.* (2011) identificaram três regiões de polimorfismos em genes da via metabólica dos hormônios sexuais relacionados à agressividade do câncer de próstata.

1.4. ALTERAÇÕES GENÉTICAS

1.4.1. Alterações genéticas e sua influência no desenvolvimento do câncer.

Mais de 90% dos casos de câncer são desencadeados por mutações somáticas acumuladas no genoma ao longo da vida (VANDIN *et al.*, 2012; TORKAMANI; SCHORK, 2008). Essas mutações incluem mudanças de nucleotídeos únicos e alteração no número de cópias e na estrutura dos genes (VANDIN *et al.*, 2012). Tais mutações resultam de erros ocorridos durante o processo replicação ou por exposição à carcinógenos (TORKAMANI; SCHORK, 2008).

O maior desafio encontrado em estudos sobre essa temática é a distinção funcional entre mutações *drivers* e mutações *passengers*. Mutações do tipo *driver* são responsáveis pelo desenvolvimento do câncer e as *passenger* são mutações acumuladas em células somáticas, mas que não estão envolvidas no desenvolvimento da doença (VANDIN *et al.*, 2012; TORKAMANI; SCHORK, 2008). Uma regra para predição de mutações *drivers* é a identificação recorrente de mutações (ou genes mutados recorrentemente) em um grande número de pacientes com câncer (VANDIN *et al.*, 2012). Estudos de associação de polimorfismos têm desempenhado papéis cruciais na descoberta de correlações genéticas entre variantes genômicas e doenças complexas (DOSS *et al.*, 2010).

1.4.2. Impactos das alterações genéticas no fenótipo.

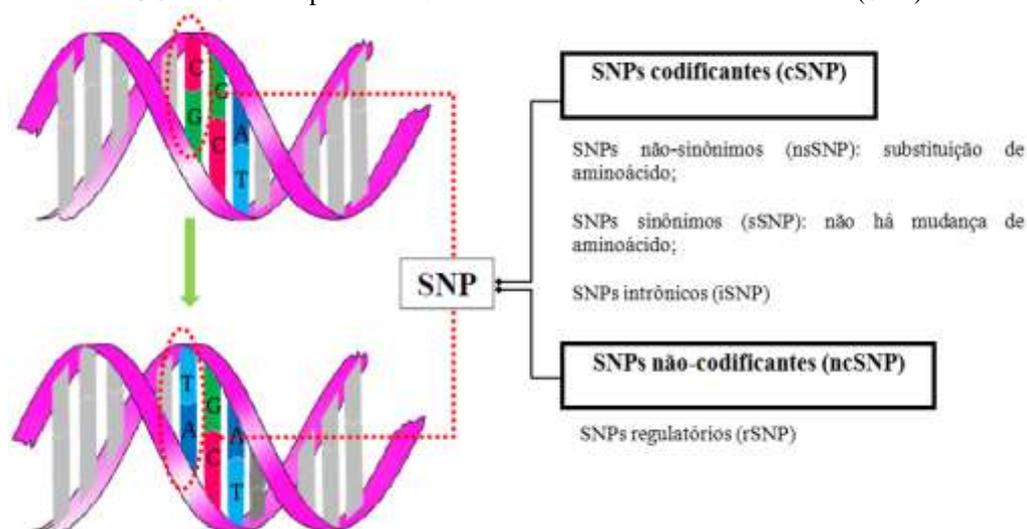
Mutações somáticas em células de câncer podem englobar diversas classes de alterações na sequência de DNA, como substituições de uma única base nucleotídica por outra; inserções ou deleções de pequenos ou grandes segmentos de DNA; rearranjos; aumento ou diminuição do número de cópias de genes (STRATTON *et al.*, 2009).

O principal tipo de variação genética humana é o chamado polimorfismo de nucleotídeo único ou *Single Nucleotide Polymorphism* (SNP), esse tipo de variação caracteriza-se pela alteração de um único nucleotídeo na sequência do genoma (LI *et al.*, 2012).

SNPs podem ser encontrados em regiões codificantes, não codificantes e intergênicas, sendo classificados em sinônimos, quando a sequência polipeptídica permanece a mesma, apesar da alteração de nucleotídeo, e em não sinônimos, quando este resulta em mudança de aminoácido na sequência protéica, podendo também gerar um códon de parada (*nonsense*) (LI *et al.*, 2012).

Além disso, SNPs presentes em regiões não codificantes podem alterar regiões de *splicing*, regiões de ligação de fatores de transcrição e sequências de RNAs não codificantes (LI *et al.*, 2012). As várias classes de SNPs podem ser visualizadas na figura 3.

FIGURA 3- Principais classes de Polimorfismo de nucleotídeo único (SNP).



Fonte: Traduzida de LI *et al.* (2012).

Variações não sinônimas constituem 50% das mutações conhecidas envolvidas com doenças hereditárias humanas, demonstrando assim o papel importante dessas variações na saúde humana e seus fortes efeitos no fenótipo (DOSS *et al.*, 2010).

1.5. IDENTIFICAÇÃO DE ALTERAÇÕES GENÉTICAS NO GENOMA.

O desenvolvimento da tecnologia NGS tem permitido a análise do conjunto de éxons (exomas) ou do genoma total de tumores malignos. O sequenciamento de exomas tem gerado a descoberta da maioria das substituições somáticas condutoras do processo de carcinogênese, de inserções, deleções e de alterações no número de cópias de genes codificadores de proteínas, portanto, a análise do conjunto de éxons pode ser considerada uma metodologia vantajosa e efetiva na descoberta de genes envolvidos no desenvolvimento do câncer, por permitir a análise do padrão de mutação presente em amostras de câncer (ALEXANDROV *et al.*, 2014).

A identificação de SNPs pode ser uma ferramenta importante na prática clínica e na descoberta de novas formas de diagnóstico e de terapias para diversas patologias, entretanto, o grande número de SNPs identificados até o momento inviabiliza a realização de experimentos que demonstrem o impacto de tais alterações genéticas na funcionalidade do produto a ser codificado. Ferramentas bioinformáticas vêm sendo desenvolvidas com o intuito selecionar SNPs considerados prejudiciais ao funcionamento de moléculas

importantes à célula (MAH *et al*, 2011), portanto, a predição das repercussões de mutações nas proteínas pode ser uma ótima opção para a análise de SNPs.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Investigar a possível influência dos hormônios sexuais no processo de carcinogênese gástrica, através da busca de SNPs com efeitos deletérios em genes relacionados às vias biossintéticas dos hormônios esteróides e do colesterol e no gene do receptor de progesterona em transcriptomas de tecidos gástricos com câncer.

2.2. OBJETIVOS ESPECÍFICOS

- Elaborar um *pipeline* bioinformático para identificação, em transcriptomas, de polimorfismos de nucleotídeo único (SNP) em genes codificantes das proteínas das vias de biossíntese do colesterol e dos hormônios esteroides e do receptor de progesterona;
- Reprocessar dados de transcriptomas de tecidos gástricos com câncer, disponíveis em bancos de dados, em busca de SNPs nos genes relacionados às vias de biossíntese do colesterol e dos hormônios esteroides e do receptor de progesterona;
- Analisar os transcriptomas das linhagens celulares de câncer gástrico ACP02 e ACP03 para identificar SNPs nos genes relacionados às vias de biossíntese do colesterol e dos hormônios esteroides e do receptor de progesterona.
- Identificar SNPs com efeitos deletérios nos genes estudados através da análise funcional dos SNPs identificados.

3. MATERIAIS E MÉTODOS

3.1. OBJETO DE ESTUDO

Nesta pesquisa foram reprocessados os dados de transcriptomas pertencentes ao estudo de KIM *et al.* (2012), com número de submissão SRP012016 no *European Nucleotide Archive* (ENA) do *European Bioinformatics Institute* (EMBL-EBI) e os transcriptomas das linhagens celulares de câncer gástrico ACP02 e ACP03, isoladas de indivíduos brasileiros, do sexo masculino, com idade de 66 e 63 anos e CG tipo difuso e intestinal, respectivamente. ACP02 e ACP03 já foram caracterizadas citogeneticamente por um grupo de pesquisa da Universidade Federal do Pará e configuram-se como as primeiras linhagens celulares de CG isoladas a partir de indivíduos sul-americanos (LEAL *et al.*, 2009).

Todos os transcriptomas analisados são resultado de experimentos de RNA-seq, sequenciados com Plataforma ABI SOLiD. O número de submissão de cada transcriptoma do estudo SRP012016 pode ser visualizado no APÊNDICE A, assim como informações sobre estágio da doença, total de leituras e tamanho dos arquivos.

3.2. GENES CANDIDATOS

Foram analisados os genes codificantes (RNAs mensageiros) das enzimas que participam das vias de biossíntese dos hormônios esteroides, biossíntese do colesterol e o gene do receptor de progesterona. A listagem completa dos genes referentes à via de biossíntese dos hormônios esteroides foi obtida diretamente no banco de dados *KEGG PATHWAY database* (<http://www.kegg.jp/kegg/pathway.html>), selecionando *hsa* (*Homo sapiens*) no parâmetro organismo e utilizando como termo de busca '*steroid hormone biosynthesis*'. Os genes relacionados à via de biossíntese de colesterol foram selecionados a partir dos estudos de BUHAESCU; IZZEDINE (2007) e MARIJANOVIC *et al.* (2003). As vias de biossíntese dos hormônios esteróides e do colesterol podem ser visualizadas nos anexos A e B, respectivamente.

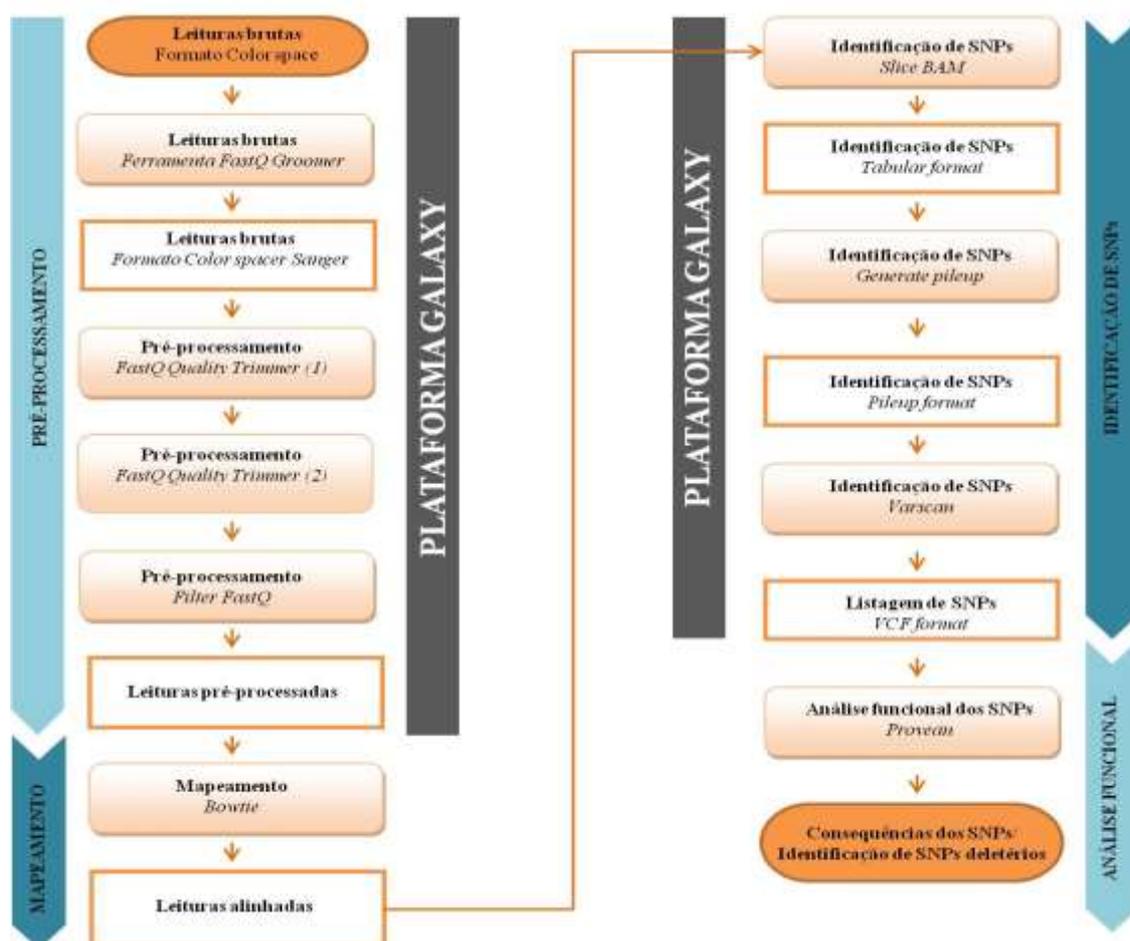
A partir da listagem dos genes, foi obtida a localização, identificada pelas posições inicial e final de cada gene no genoma humano (GRCh37/hg19) através da página do *UCSC*

Genome Browser (<https://genome.ucsc.edu/cgi-bin/hgGateway>). Às posições inicial e final dos genes foram acrescidos 5 kilo bases (Kb) para que possíveis SNPs presentes em regiões regulatórias fossem inseridos no conjunto final de SNPs.

3.3. ANÁLISE DOS TRANSCRIPTOMAS

A análise dos transcriptomas envolveu as etapas de (i) pré-processamento, (ii) mapeamento/alinhamento das leituras no genoma humano, (iii) identificação de polimorfismos e (iv) análise funcional dos polimorfismos identificados, conforme pode ser visualizado na figura 1. As etapas de pré-processamento e de identificação de polimorfismos foram realizadas na Plataforma Galaxy (GIARDINE *et al.*, 2005), disponível no endereço eletrônico <https://usegalaxy.org/>.

FIGURA 4- Pipeline bioinformático elaborado para identificar SNPs em genes relacionados à biossíntese dos hormônios esteroides, colesterol e receptor de progesterona em transcriptomas.



Fonte: Elaborado pelo autor.

Os transcriptomas provenientes do ENA foram diretamente importados na Plataforma Galaxy utilizando-se o número de submissão no ENA, optando-se pelo arquivo no formato *Fastq file (galaxy)*. Os transcriptomas das linhagens celulares de CG ACP02 e ACP03 foram importados do computador para a Plataforma Galaxy.

3.3.1. Pré-processamento do conjunto de dados

O pré-processamento dos dados envolveu os processos de conversão de formato dos dados, “trimagem” e filtragem das leituras. Após inserção do conjunto de dados dos transcriptomas na plataforma Galaxy, arquivos com formato *Fastq file (galaxy)* foram convertidos para o formato *Color Space Sanger* por meio da ferramenta *FASTQ Groomer*, disponível no conjunto de ferramentas *NGS: QC and manipulation* da Plataforma Galaxy.

Com a finalidade de melhorar a qualidade do conjunto de dados, bases de baixa qualidade e leituras de tamanho reduzido foram eliminadas, para isso, os arquivos passaram por duas etapas de trimagem e uma etapa de filtragem. A ferramenta *FastQC Read Quality reports (NGS: QC and manipulation)* foi utilizada para monitorar o padrão de qualidade das leituras antes e depois da etapa de pré-processamento.

3.3.1.1. Trimagem das leituras

Na etapa de ‘trimagem’ foi utilizada a ferramenta *FASTQ Quality Trimmer*, presente no kit de ferramentas *NGS: QC and manipulation*. O primeiro processo de trimagem teve como parâmetros: *Trimer ends = 3' only, Window size = 3, Aggregate action for window: means of score e Quality score = 20*. E o segundo processo, parâmetros: *Trimer ends = 3' only, Window size = 1, Aggregate action for window: means of score e Quality score = 18*. Estes parâmetros foram utilizados em todas as análises realizadas.

3.3.1.2. Filtragem das leituras

A filtragem dos dados foi realizada através da ferramenta *Filter FASTQ*, também presente em *NGS: QC and manipulation*. Nesta fase, sequências com tamanho inferior a 30 pares de base (*Minimum size = 30*) foram eliminadas do conjunto de dados.

3.3.2. Alinhamento das leituras

O alinhamento das leituras foi feito com a ferramenta *BOWTIE*, utilizando parâmetros *default* (número de *mismatches* (-n) = 2, número de bases (-l) = 28 e soma dos valores de qualidades dos *mismatches* (-e) = 70), tendo como genoma de referência o genoma humano (hg19).

3.3.3. Localização e identificação de SNP

A localização dos SNPs nos genes analisados foi feita com a ferramenta *Slice BAM by provided regions*. Nesta etapa, foram utilizados os arquivos resultantes do mapeamento contra o genoma humano e o arquivo com todas as localizações de SNPs presentes nos genes estudados disponíveis nos bancos de dados.

O arquivo com as localizações dos SNPs disponíveis nos bancos de dados foi obtido diretamente na plataforma Galaxy por meio de *UCSC Main table browser*, presente em *Get Data*, selecionando a opção *variation* no parâmetro *group*; *snp138Common* em *table* e *position* em *region*. No parâmetro *region*, a listagem com a localização (número do cromossomo, posição inicial e final) dos genes estudados foi inserida clicando-se no ícone *define regions*.

Como resultado da ferramenta *Slice BAM by provided regions*, foi gerado um arquivo no formato BAM, este arquivo necessitava ser convertido para o formato *pileup*, para que as etapas restantes pudessem ser realizadas. Esta conversão foi realizada por meio da ferramenta *Generate pileup format from BAM dataset*, presente em *NGS: SAM Tools*.

Após a conversão dos formatos, o arquivo final foi processado na ferramenta *Varscan for variant detection*, presente em *NGS: Variant analysis*, para gerar uma listagem com a localização, o nucleotídeo alterado e o nucleotídeo da referência de todos os polimorfismos encontrados nos transcriptomas analisados.

Na ferramenta *Varscan* foram utilizados os parâmetros: *Minimum read depth*=8, *Minimum supporting reads*=2, *Minimum base quality at a position to count a read* =15, *Minimum variant allele frequency threshold*=0,01 e *p-value threshold for calling variants* = 0,99.

3.4. ANÁLISE FUNCIONAL DOS SNPs

A listagem com as coordenadas (número do cromossomo, posição no genoma, base nucleotídica da referência e base nucleotídica encontrada) de todos os SNPs identificados no conjunto de dados, foi submetida a uma análise funcional dos possíveis efeitos destas mutações na funcionalidade proteica. Para isso, foi utilizada a ferramenta *PROVEAN Genome Variants*, disponível no servidor web PROVEAN (*Protein Variation Effect Analyzer*), disponível na página <http://provean.jcvi.org/index.php>.

A ferramenta *PROVEAN Genome Variants* fornece uma abordagem generalizada para prever efeitos funcionais de variações na sequência de aminoácidos das proteínas ocasionadas por substituições únicas ou múltiplas de nucleotídeos, inserções e deleções. SNPs que ocasionam mudança de aminoácido na sequência proteica são classificados pelo *PROVEAN* em neutros e deletérios, sendo preditos como deletérios SNPs com score $\leq -2,5$ (CHOI *et al.*, 2015).

4. RESULTADOS

4.1. CONJUNTO DE DADOS

O conjunto de dados com todos os transcriptomas totalizou aproximadamente 2 bilhões leituras ou cerca de 94 Giga bases. Após as etapas de ‘trimagem’ e filtragem, as leituras apresentaram tamanho variando de 30 a 50 pares de base (pb). O percentual médio de mapeamento contra o genoma de referência foi de 44%.

Dos 32 transcriptomas obtidos por sequenciamento em Plataforma SOLiD que foram analisados, 24 eram provenientes de indivíduos com câncer gástrico, 6 sem câncer gástrico e 2 das linhagens celulares ACP02 e ACP03 de CG. A análise de transcriptomas sequenciados por um mesmo tipo de sequenciador permite a uniformização dos parâmetros e de todo o plano de análise, sem a necessidade de adaptações ou criação de *pipelines* adicionais.

Mais de 80% dos transcriptomas era procedente de indivíduos do sexo masculino, 50 % dos tumores eram classificados histologicamente como tipo difuso e 31% como tipo intestinal (Gráfico 1). A média de idade dos indivíduos foi de 62,5 anos (DP= 10,77), entre os sexos a média de idade foi de 62,8 anos (DP=9,84) para homens e de 61,3 anos (DP= 15,01) para mulheres (Tabela 1).

GRÁFICO 1- Distribuição dos transcriptomas segundo idade dos indivíduos e tipo histológico dos tumores.

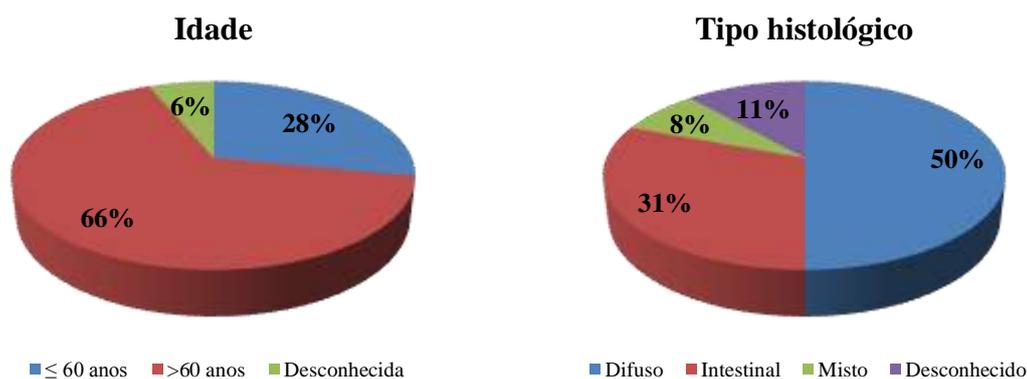


TABELA 1- Distribuição dos SNPs identificados nos genes relacionados às vias de biossíntese de hormônios esteróides e de colesterol segundo idade, sexo, tipo histológico do tumor e transcriptoma de indivíduo com e sem câncer gástrico.

Variáveis	Transcriptomas analisados		SNPs							
			Via de Biossíntese dos Hormônios Esteróides				Via de Biossíntese do Colesterol			
	N	%	N	%	M	DP	N	%	M	DP
<i>Idade</i>										
≤ 60 anos	9	28	7	12	0,77	1,09	8	22	0,88	1,16
>60 anos	21	66	45	72	2,09	3,33	29	78	1,38	1,90
Desconhecida	2	6	10	16	5,00	7,07	0	0	0,00	0,00
<i>Sexo</i>										
Masculino	26	81	54	87	2,07	3,50	28	73	1,07	1,83
Feminino	6	19	8	13	1,33	2,45	9	27	1,50	0,83
<i>Tipo histológico</i>										
Difuso	13	50	13	24	1,00	1,15	13	42	1,00	1,47
Intestinal	8	30	25	46	3,15	5,13	12	39	1,5	2,39
Misto	2	8	6	11	3,00	1,41	6	19	3,00	1,41
Desconhecido	3	12	10	19	3,33	5,77	0	0	0,00	0,00
<i>Linhagem celular</i>										
ACP02	1	50	-	-	-	-	-	-	-	-
ACP03	1	50	-	-	-	-	-	-	-	-
<i>Transcriptoma</i>										
Indivíduo com câncer gástrico	26	81	54	87	2,07	3,46	31	84	1,19	1,78
Indivíduo sem câncer gástrico	6	19	8	13	1,33	1,50	6	16	1,00	2,06
Total	32	100	62	100	1,93	3,18	37	100	1,15	1,68

N: quantidade; M: Média; DP: Desvio padrão.

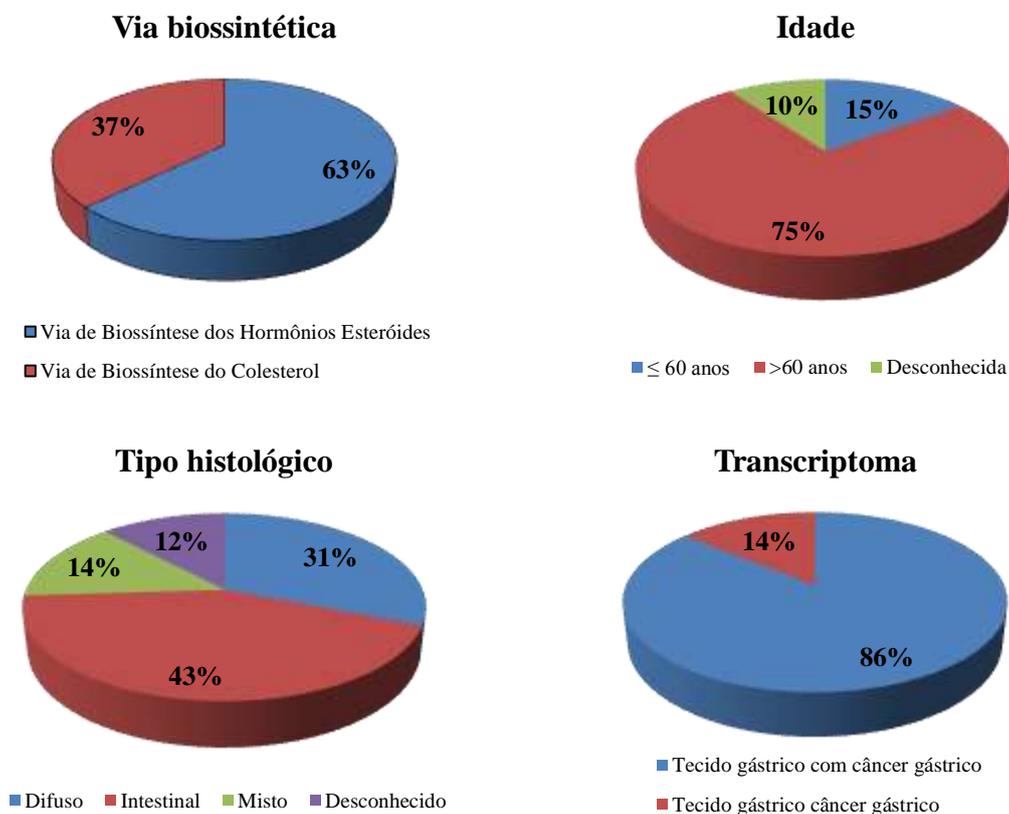
4.2. IDENTIFICAÇÃO DE SNPS

A presença de SNPs foi investigada em 56 genes codificantes de enzimas envolvidas na via de biossíntese de hormônios esteróides (APÊNDICE B), 17 genes da via de biossíntese do colesterol (APÊNDICE C) e o gene do receptor de progesterona (gene *PGR*).

Dos 99 SNPs identificados nos transcriptomas analisados, 74% foram identificados em transcriptomas de indivíduos com idade igual ou superior a 60 anos; 86% estavam presentes em transcriptomas de CG; 63 % eram em genes relacionados aos hormônios esteróides e 37% em genes da via de biossíntese do colesterol. Com os parâmetros adotados

nesta análise, não foi identificado nenhum SNP nas linhagens celulares ACP02 e ACP03 de CG.

GRÁFICO 2 - Distribuição dos SNPs identificados nos transcriptomas analisados segundo via biossintética, idade, tipo histológico do tumor e transcriptoma com e sem câncer gástrico.

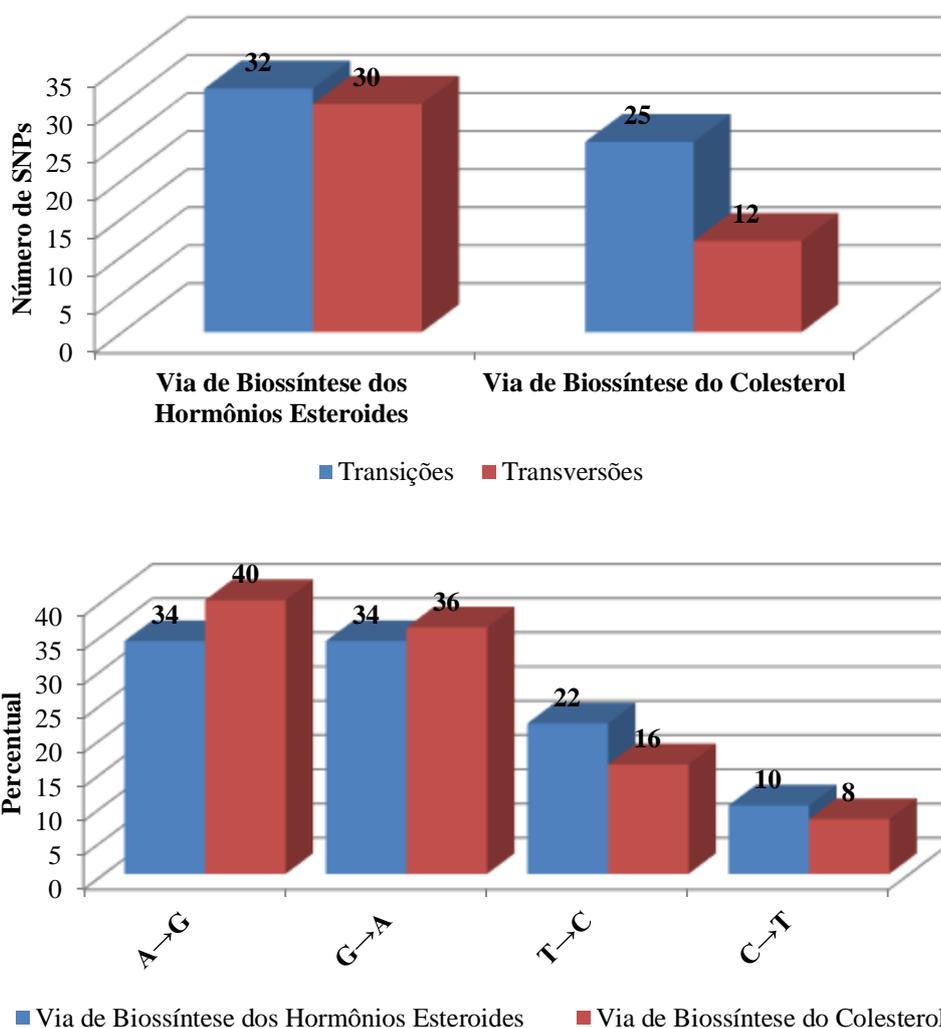


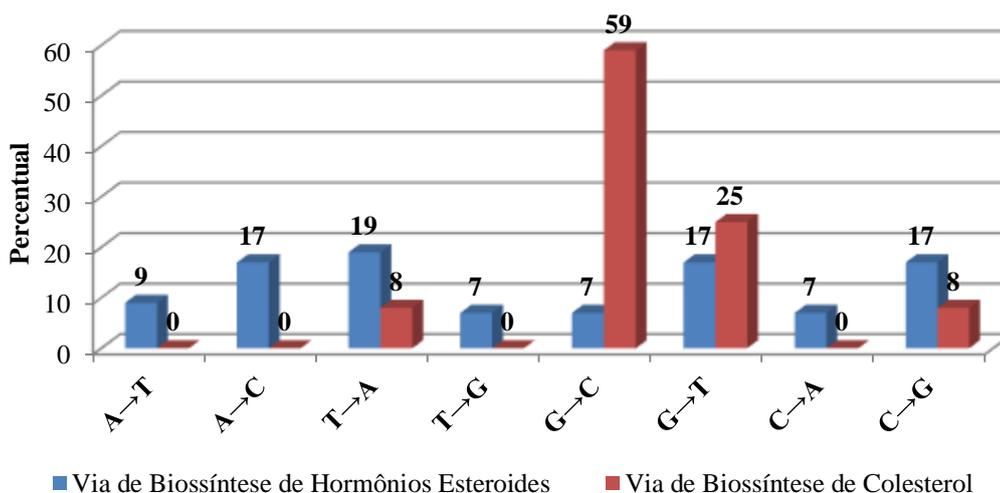
Nos genes relacionados à via de biossíntese do colesterol, no total 37 SNPs foram encontrados, destes 84% localizavam-se em transcriptomas de CG, 42% em transcriptomas de CG do tipo difuso e 78% em transcriptomas de indivíduos com mais de 60 anos. Entre os genes relacionados à via de biossíntese dos hormônios esteroides, 62 SNPs foram localizados, onde 87% foram identificados em transcriptomas de CG, 46% em transcriptomas de CG do tipo intestinal e 72% em transcriptomas de indivíduos com idade superior a 60 anos (Tabela 1).

Em 50% dos SNPs identificados nos genes da via dos hormônios esteroides e em 65% dos SNPs da via de biossíntese do colesterol, as substituições das bases nucleotídicas foram transições. Nos genes de ambas as vias de biossíntese, mais de 60% das transições eram substituições de bases púricas (A↔G), como mostra a figura 5.

Entre as transversões, nos genes via dos hormônios esteroides, a maior parte (19%) era substituição de uma guanina por uma citosina (G→C) e, nos genes da via de biossíntese do colesterol, 16% era substituição de uma citosina por uma adenina (C→A) (Figura 5).

GRÁFICO 3- SNPs identificados em transcriptomas indivíduos com câncer gástrico e sem câncer gástrico. A: Distribuição dos SNPs encontrados nos genes relacionados às vias de biossíntese dos hormônios esteroides e do colesterol por tipo de substituição de base nucleotídica. **B:** Distribuição dos SNPs do tipo transição segundo substituição de nucleotídeo. **C:** Distribuição dos SNPs do tipo transversão segundo substituição de nucleotídeo.

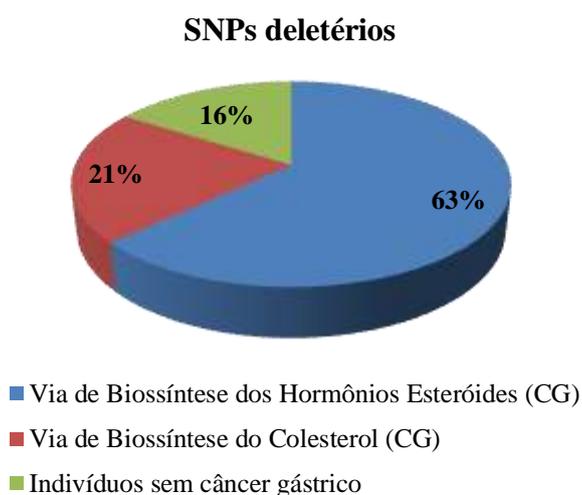




Fonte: Elaborada pelo autor.

No total, 19 SNPs preditos com efeitos deletérios foram identificados nos transcriptomas analisados, destes, 63% estavam localizados em genes relacionados à via de biossíntese dos hormônios esteróides, em transcriptomas de câncer gástrico e 21% em transcriptomas de tecidos gástricos sem câncer. O restante (16%) foi localizado entre os genes da biossíntese de colesterol, em transcriptomas de câncer gástrico. Em 52,6% dos SNPs preditos como deletérios, ocorreu mudança da primeira base no códon alterado e 68,4% corresponderam a transições nucleotídicas (Tabela 2).

Gráfico 4- Distribuição dos SNPs preditos com efeitos deletérios.



Em transcriptomas de câncer gástrico, SNPs deletérios foram localizados nos genes *CYP3A5*, *HSD17B12*, *UGT1A1*, *UGT1A5*, *UGT1A6* e *AKR1C3* da via de biossíntese dos

hormônios esteróides. Os genes *CYP3A5* e *AKRIC3* apresentaram SNPs com os menores valores de score, o que revela o forte de impacto desses SNPs no produto proteico desses genes. Entre os genes da via de biossíntese de colesterol, SNPs deletérios foram identificados nos genes *CYP51A1*, *DHCR24* e *SQLE*, sendo que SNPs nos genes *DHCR24* e *CYP51A1* apresentaram os menores valores de score (Tabela 2). Em transcriptomas de indivíduos sem CG foram identificados SNPs deletérios nos genes *AKRIC1* e *AKRIC3*, relacionados à via de biossíntese de hormônios esteroides (Tabela 2).

TABELA 2- SNPs com efeitos deletérios identificados nos genes codificantes de proteínas envolvidas nas vias de biossíntese de hormônios esteróides e de colesterol e receptor de progesterona em transcriptomas de câncer gástrico.

Gene	Localização do SNP		Códon alterado	Mudança de aminoácido	Score (PROVEAN)
	Cromossomo	Posição			
Via de Biossíntese dos Hormônios Esteroides					
Câncer gástrico					
<i>CYP3A5</i>	7	99250300	ACT [T/G]GC AAG	p.C377G	-9,06
	7	99264585	GGA A[A/G]A CTC	p.K141R	-2,53
	7	99258223	GAA [A/T]CC ACC	p.T309S	-3,05
	7	99258162	CAG C[A/T]G AAA	p.Q329L	-4,84
	7	99270240	GTG C[T/G]A GTG	p.L94R	-5,22
<i>HSD17B12</i>	11	43876380	CAA T[C/T]C CGA	p.S267F	-3,84
<i>UGT1A1</i>	2	234676982	AAT [G/A]CA AAG	p.A401T	-2,52
<i>UGT1A5</i>	2	234676982	AAT [G/A]CA AAG	p.A402T	-2,56
<i>UGT1A6</i>	2	234676982	AAT [G/A]CA AAG	p.A402T	-2,88
<i>AKRIC3</i>	10	5138738	AGT G[T/G]G AAG	p.V74G	-6,47
	10	5141634	GTG T[G/A]C AAC	p.C188Y	-10,78
	10	5144341	TGC [A/G]AG TCG	p.K88E	-3,49
Sem câncer gástrico					
<i>AKRIC1</i>	10	5010537	GGA [A/G]AA ATA	p.K138E	-3,03
	10	5011114	CTC A[A/G]G TAC	p.K183R	-2,52
<i>AKRIC3</i>	10	5141616	GGA C[T/C]C AAG	p.L182P	-6,84
	10	5141618	CTC [A/G]AG TAC	p.K183E	-3,61
Via de Biossíntese do Colesterol					
Câncer gástrico					
<i>CYP51A1</i>	7	91761164	TTC T[C/A]C CCA	p.S72Y	-6,00
<i>DHCR7</i>	11	71150074	CCT [C/T]GG ATC	p.R228W	-7,29
<i>SQLE</i>	8	126015517	GGA [G/A]CT GGC	p.A131T	-2,89

Entre os SNPs identificados nos transcriptomas analisados, 13 estão cadastrados no banco de dados dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), a maior parte foi identificada

em transcriptomas de indivíduos com câncer gástrico, 69% têm sua localização em regiões intrônicas e 3'UTR, entretanto, dois SNPs ocasionavam mudança de aminoácido. Os SNPs rs7374 (*DHCR24*), rs9205 (*FDFT1*), rs162562 (*CYP1B1*), rs15524 (*CYP3A5*) e rs12429 (*AKRIC3*) foram identificados em mais de um transcriptoma, conforme pode ser observado na Tabela 3.

Tabela 3- SNPs cadastrados no banco de dados dbSNP identificados nos genes relacionados às vias de biossíntese dos hormônios esteróides, do colesterol e do receptor de progesterona presentes nos transcriptomas analisados.

SNP ID	MAF	Gene	Função	Transcriptoma (n)			
				Câncer gástrico*			Sem câncer gástrico
				Intestinal	Difuso	Misto	
Via de Biossíntese do Colesterol							
rs7374	G=0,4359/2183	<i>DHCR24</i>	3'UTR	-	-	1	1
rs1138030	A=0,2260/1132	<i>DHCR24</i>	3'UTR	-	1	-	-
rs9205	C=1,3069/1537	<i>FDFT1</i>	Sinônimo	1	-	2	2
rs12673910	A=0,1697/850	<i>CYP51A1</i>	3'UTR	1	-	-	-
rs188890480	G=0,0020/10	<i>MSMO1</i>	Missense (P→A)	1	-	-	-
Via de Biossíntese dos Hormônios Esteróides							
rs4775934	C=0,2041/1022	<i>CYP19A1</i>	Íntron	-	-	1	-
rs162562	G=0,2668/1336	<i>CYP1B1</i>	3'UTR	2	2	-	-
rs15524	G=0,3480/1743	<i>CYP3A5</i>	ncRNA , 3'UTR	2	2	-	1
rs2040992	G=0,0056/28	<i>CYP3A5</i>	Íntron	1	-	-	-
rs11037683	C=0,053/252	<i>HSD17B12</i>	3'UTR	-	1	-	-
rs12529	C=0,4203/2105	<i>AKRIC3</i>	Missense (H→Q)	1	1	1	-
rs12387	G=0,1518/760	<i>AKRIC3</i>	Sinônimo (K→K)	1	1	-	-
rs1937840	C=0,4171/2089	<i>AKRIC3</i>	Íntron	-	1	-	-

*Classificação histológica de Lauren.

SNPs sinônimos foram identificados nos genes *FDPS*, *CYP51A1*, *DHCR7*, *FDFT1*, *SQLE*, *DHCR24* e *MSMO1* da via de biossíntese de colesterol e nos genes *CYP3A5*, *HSD17B12*, *AKRIC3*, *UGT1A6*, *UGT1A10*, *UGT1A9*, *UGT1A4*, *UGT1A5*, *UGT1A7*, *UGT1A8* e *UGT1A3*, da via de biossíntese de hormônios esteróides.

O resultado completo da análise dos transcriptomas, com todos os SNPs identificados, pode ser visualizado no APÊNDICE D. Informações adicionais sobre a linhagem celular ACP03 de câncer gástrico, podem ser obtidas no artigo que está sendo elaborado sobre expressão diferencial e identificação de SNPs desta linhagem (APÊNDICE E).

5. DISCUSSÃO

Mundialmente, o CG apresenta diferenças nas taxas de incidência entre as várias regiões do mundo, sendo mais prevalente em países menos desenvolvidos e em homens (FERLAY, 2015). Neste trabalho, a prevalência do sexo masculino no conjunto de dados analisado pode ser explicada pela diferença na incidência de CG na população de origem dos indivíduos, na qual, segundo estatísticas mundiais do câncer, a taxa de incidência de CG em homens é mais que o dobro encontrado entre as mulheres, sendo o CG, o tipo de câncer que mais acomete a população masculina.

Diferenças nas taxas de incidência de CG entre homens e mulheres podem sugerir um potencial papel dos hormônios sexuais nos mecanismos de desenvolvimento do câncer gástrico, entretanto, devido ao reduzido quantitativo de indivíduos do sexo feminino com CG no presente estudo, não é possível afirmar esta informação com exatidão. No entanto, estudos de caráter experimental realizados em ratos, já demonstraram uma possível influência dos hormônios sexuais no desenvolvimento do câncer gástrico (FURUKAWA *et al.* (1982); WAKUI *et al.* (2011). Nesses estudos, a incidência de CG, induzido pelo agente mutagênico metilnitrosoguanidina (MNNG), foi consideravelmente superior em ratos machos que em ratos fêmeas. FURUKAWA *et al.* (1982) identificaram uma menor incidência de CG em ratos machos castrados comparado aos não castrados, e uma incidência maior em fêmeas castradas com idade mais avançada, evidenciando assim uma possível participação de hormônios sexuais no desenvolvimento deste tipo de câncer.

Neste trabalho, a busca por SNPs foi realizada em um conjunto de genes codificantes de enzimas integrantes das vias de biossíntese dos hormônios esteroides e do colesterol, assim como o gene do receptor de progesterona. Dessa maneira, SNPs que possam ocasionar alterações nos mecanismos normais dessas enzimas puderam ser identificados.

Até o momento, poucos genes envolvidos na via de biossíntese dos hormônios esteroides têm sido foco de estudos envolvendo câncer, principalmente, câncer gástrico. CHO *et al.* (2012), em um estudo realizado com indivíduos coreanos, analisaram 5 genes relacionados à via dos hormônios esteroides e identificaram os polimorfismos rs16964228, rs1902580 e rs1004982, presentes no gene *CYP19A1*, como alterações associadas significativamente ao risco de câncer gástrico naquela população.

No presente estudo, todos os genes relacionados à via dos hormônios esteroides foram analisados, o que permitiu a realização de uma busca exaustiva por SNPs nos genes

desta via. Em transcriptomas de indivíduos com CG, polimorfismos com efeitos deletérios foram identificados nos genes *CYP3A5*, *HSD17B12*, *UGT1A1*, *UGT1A5*, *UGT1A6* e *AKR1C3* da via dos hormônios esteroides.

O gene *CYP3A5* codifica uma enzima pertencente à superfamília de citocromos P450, responsáveis pela metabolização de xenobióticos, e de compostos endógenos como ácido araquidônico e hormônios esteroides, além de participar da sinalização e regulação celular. *CYP3A5* é uma das principais proteínas envolvidas no metabolismo de carcinógenos, junto com *CYP3A4*, *CYP1A1*, *CYP2A6* e *CYP2E1* (MURRAY, 2000). Uma mutação deletéria nessa proteína pode ocasionar danos nos mecanismos enzimáticos e consequentemente, danos a vias celulares importantes. SNPs no gene *CYP3A5* já foram associados ao risco de desenvolvimento do câncer de esôfago, assim como SNPs presentes em outros genes da via de biossíntese dos hormônios sexuais (HYLAND *et al.*, 2013), entretanto, GERVASINI *et al.* (2007), investigando polimorfismos nos genes *CYP3A4* e *CYP3A5* em indivíduos com cânceres gastrointestinais, não conseguiram identificar nenhum alelo nesses genes associados ao desenvolvimento do câncer.

O gene *HSD17B12* é responsável por uma das enzimas da família das 17 β -hidroxiesteroide desidrogenase. Esta família de enzimas é responsável por regular a transformação de estrogênios e androgênios em sua forma ativa (MINDNICH *et al.*, 2004), desenvolvendo, portanto, uma atividade essencial na biossíntese dos hormônios esteroides (SUN *et al.*, 2011). A enzima *HSD17B12* desempenha um papel multifuncional, estando envolvida na conversão de estrona em estradiol, alongação da cadeia de ácidos graxos e conversão do ácido palmítico a ácido araquidônico – precursor dos esteróis e do mediador inflamatório prostaglandina E (SUN *et al.*, 2011). Segundo FRYCZ *et al.* (2015), níveis reduzidos de *HSD17B2* estão associados a características clínico patológicas do CG.

Os genes *UGT1A1*, *UGT1A5* e *UGT1A6* fazem parte do grupo de enzimas UDP-glucuronil transferases que catalisam as reações de conversão de substratos hidrofóbicos a glicuronídeos hidrofílicos para a eliminação via vesícula biliar e rins (STRASSBURG *et al.*, 2000). Genes desse grupo de enzimas são diferencialmente regulados em tecidos hepáticos e extra-hepáticos do trato gastrointestinal e atuam de forma tecido-específico (STRASSBURG *et al.*, 2000). SNPs que causam modificações na estrutura dessas enzimas poderiam interferir na conversão de fármacos, carcinógenos, entre outros compostos,

fazendo com que tais substâncias se acumulem no organismo, provocando efeitos prejudiciais à saúde.

SNPs em genes relacionados à via do mevalonato, ligada a biossíntese do colesterol, podem influenciar no mecanismo normal de genes importantes como *p53* (GRUENBACHER; THURNER, 2015), por exemplo, assim como na produção normal das substâncias que dependem de precursores presentes nesta via. Neste trabalho, SNPs com efeitos deletérios foram identificados nos genes *CYP51A1*, *DHCR24* e *SQLE* em indivíduos com CG. Essas enzimas atuam em importantes reações dentro da via do mevalonato, como por exemplo, o gene *SQLE* que é responsável pela enzima esqualeno monooxigenase, que atua na primeira reação da via, convertendo o esqualeno a epóxido de esqualeno. Certamente, alterações genéticas em *SQLE* poderiam comprometer as etapas subsequentes desta via.

Os genes *CYP51A1* e *DHCR24* codificam enzimas que atuam em reações intermediárias da via do mevalonato (MARIJANOVIC *et al.*, 2003). Modificações no funcionamento adequado destas enzimas poderiam fazer com que compostos precursores não fossem produzidos, inviabilizando a biossíntese de compostos importantes, como os hormônios esteroides.

A criação de novos tratamentos contra câncer tem sido possível graças aos avanços ocorridos nas áreas de genômica e ao uso da bioinformática em estudos que buscam elucidar os mecanismos de oncogênese (ARABUTAMI, 2005). Nesse contexto, novas tecnologias têm sido aplicadas na identificação de SNPs em genes e regiões genômicas que possam estar envolvidas no desenvolvimento de doenças (ENGLE *et al.*, 2006). Novos estudos têm identificado genes adicionais relacionados ao desenvolvimento do câncer (GREENMAN *et al.*, 2007; KAN *et al.*, 2007). Essas informações podem servir de base para a elaboração de novas formas de diagnóstico ou de prognóstico da doença, como a utilização de genes associados ao câncer como biomarcadores.

6. CONCLUSÃO

O *pipeline* bioinformático utilizado no presente estudo permitiu a identificação de 99 SNPs nos genes analisados. A análise funcional dos SNPs resultou na identificação de 15 SNPs preditos com efeitos deletérios nos genes *CYP3A5*, *HSD17B12*, *UGT1A1*, *UGT1A5*, *UGT1A6* e *AKR1C3* da via de biossíntese de hormônios esteroides e nos genes *CYP51A1*, *DHCR24* e *SQLE* relacionados à via de biossíntese do colesterol em transcriptomas de CG. Entre os genes utilizados na análise, o gene *CYP3A5* destacou-se por apresentar o maior número de SNPs, principalmente, em transcriptomas de indivíduos com CG, além de possuir a maior parte dos SNPs preditos como deletérios. Entretanto, análises mais precisas necessitam ser realizadas para comprovar possíveis influências do gene *CYP3A5* no desenvolvimento do CG.

Apesar de pouco evidentes, os demais genes que apresentaram SNPs deletérios podem ser analisados em um conjunto maior de transcriptomas de câncer gástrico, pois desempenham papéis importantes nas vias das quais fazem parte e alterações em seu funcionamento poderiam ocasionar modificações nos mecanismos moleculares celulares.

Estudos adicionais são necessários para que os papéis dos SNPs identificados em genes das vias de biossíntese dos hormônios esteroides e do colesterol no desenvolvimento do CG sejam consolidados. A partir disso, tais genes poderão ser empregados na elaboração de diagnósticos baseados em alterações genéticas ou na expressão gênica.

Análises de transcriptomas de CG, com foco em genes relacionados às vias de biossíntese importantes, podem ser utilizados para que novos aspectos moleculares do câncer sejam explorados, de forma a acrescentar informações significativas e pertinentes sobre os mecanismos envolvidos na oncogênese gástrica.

REFERÊNCIAS BIBLIOGRÁFICAS

ALEXANDROV, L. B.; STRATTON, M. R. Mutational signatures: the pattern of somatic mutations hidden in cancer genomes. **Current Opinion in Genetic and development**, 24: 52-60, 2014.

ALTIERI, F.; ARCARI, P.; RIPPA, E. Gastric cancer: molecular pathology state. **Current Topics in Gastritis**, 241 - 250, 2012.

ARABUTAMI, H. Discovery of a new biomarker for gastroenterological cancers. **J Gastroenterol**, 40(Suppl XVI): I-6, 2005.

ASHTON, K. A. *et al.* Polymorphisms in genes of the steroid hormone biosynthesis and metabolism pathways and endometrial cancer risk. **Cancer Epidemiology**, 34: 328-337, 2010.

BERTRAM, J. S. The molecular biology of cancer. **Molecular Aspects of Medicine**, 21: 167-223, 2001.

BUHAESCU, I.; IZZEDINE, H. Mevalonate pathway: A clinical and therapeutical implications. **Clinical Biochemistry**, 40: 575-584, 2007.

CANCER GENOME ATLAS RESEARCH NETWORK. Comprehensive molecular characterization of gastric adenocarcinoma. **Nature**, 513: 202 – 209, 2014.

CASALS, F. *et al.* Next-generation sequencing approaches for genetic mapping of complex diseases. **Journal of Neuroimmunology**, 248: 10-22, 2012.

CHO, L. Y. *et al.* Genetic susceptibility factors on genes involved in the steroid hormone biosynthesis pathway and progesterone receptor for gastric cancer risk. **Plos One**, 7(10): e47603, 2012.

CHOI, Y.; CHAN, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**, 1-3, 2015.

CORREA, P. Gastric cancer: Overview. **Gastroenterol Clin N Am**, 42: 211- 217, 2013.

DOSS, C. G. P. *et al.* Prioritization of candidate SNPs in colon cancer using bioinformatics tools: an alternative approach for a cancer biologist. **Interdiscip Sci Comput Life Sci**, 2: 320-346, 2010.

ENGLE, L.J.; SIMPSON, C.L.; LANDERS, J.E. Using high-throughput SNP technologies to study cancer. **Oncogene**, 25:1594–1601, 2006.

FERLAY, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major pattern in GLOBOCAN 2012. **Int. J. Cancer**, 136: E359-E386, 2015.

FREEDMAN, N. D. *et al.* Polymorphisms in estrogen- and androgen-metabolizing genes and the risk of gastric cancer. **Carcinogenesis**, 30(1): 71-77, 2009.

FRYCZ, B. A. *et al.* Expression of 17 β -hydroxysteroid dehydrogenase type 2 is associated with some clinicopathological features in gastric cancer. **Biomedicine & Pharmacotherapy**, 70: 24–27, 2015.

FURUKAWA, H. *et al.* Effect of sex hormones on carcinogenesis in the stomach of rats. **Cancer research**, 42: 5181-5182, 1982.

GERVASINI, G. *et al.* Genetic variability in CYP3A4 and CYP3A5 in primary liver, gastric and colorectal cancer patients. **BMC Cancer**, 7:118, 2007.

GIARDINE, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. **Genome Res**, 15: 1451-1455, 2005.

GREENMAN, C. *et al.* Patterns of somatic mutation in human cancer genomes. **Nature**, 448: 153-158, 2007.

GRUENBACHER, G.; THURNHER, M. Mevalonate metabolism in cancer. **Cancer Letters**, 356: 192-196, 2015.

WORLD HEALTH ORGANIZATION (WHO). GLOBOCAN 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012. Disponível no site: <<http://globocan.iarc.fr/Default.aspx>>. Acesso em: maio/2015.

HANAHAN, D.; WEINBERG, R. Hallmarks of Cancer: The Next Generation. **Cell**, 144: 646-674, 2011.

HYLAND, P. L. *et al.* Genetics variants in sex hormone pathway genes and risk of esophageal squamous cell carcinoma. **Carcinogenesis**, 34(5): 1062-1068, 2013.

IGAZ, P. *et al.* Genomics of steroid hormones: in silico analysis of nucleotide sequence variants (polymorphisms) of the enzymes involved in the biosynthesis and metabolism of steroid hormones. **Journal of Steroid Biochemistry & Molecular Biology**, 82: 359-367, 2002.

KAN, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. **Nature**, 466: 869-875, 2010.

KARAT, D. *et al.* Expression of oestrogen and progesterone receptors in gastric cancer: a flow cytometric study. **British Journal of Cancer**, 80(8): 1271-1274, 1999.

KIM, Y. H.; LIANG, H.; LIU, X. AMPK α modulation in cancer progression: Multilayer integrative analysis of the whole transcriptome in Asian gastric cancer. **Cancer Res**, 72: 2512-2521, 2012.

LEAL, M. F. *et al.* Establishment and conventional cytogenetic characterization of three gastric cancer cell lines. **Cancer Genetics and Cytogenetics**, 195: 85-91, 2009.

LI, L.; CHEN, Q.; WEI, D-Q. Prediction and Functional Analysis of Single Nucleotide Polymorphisms. **Current Drug Metabolism**, 13: 1012-1023, 2012.

MAH, J.T.L.; LOW, E.S.H.; LEE, E. In silico SNP analysis and bioinformatics tools: a review of state of the art to aid drug discovery. **Drug Discovery Today**, 16: 800-809, 2011.

MARIJANOVIC, Z. *et al.* Closing the gap: Identification of human 3-ketosteroid reductase, the last unknown enzyme of mammalian cholesterol biosynthesis. **Molecular Endocrinology**, 17(9): 1715-1725, 2003.

MELDRUM, C.; DOYLE, M.A.; TOTHILL, R.W. Next-generation sequencing for cancer diagnostics: a practical perspective. **Clin Biochem Rev**, 32: 177-195, 2011.

METZKER, M. L. Sequencing technologies - the next generation. **Nature Reviews/Genetic**, 11:31-46, 2010.

MINDNICH, R.; Möller, G.; Adamski, J. The role of 17 beta-hydroxysteroid dehydrogenases. **Molecular and Cellular Endocrinology**, 218: 7–20, 2004.

MURRAY, G. I. The role cytochrome P450 in tumor development and progression and its potential in therapy. **J Pathol**, 192: 419-426, 2000.

MUTZ, K. *et al.* Transcriptome analysis using next-generation sequencing. **Current Opinion in Biotechnology**, 24: 22-30, 2013.

SARKAR, S. *et al.* Cancer Development, Progression, and Therapy: An Epigenetic Overview. **Int. J. Mol. Sci.**, 14: 21087-21113, 2013.

STRASSBURG, C. P. *et al.* Polymorphic Gene Regulation and Interindividual Variation of UDP-glucuronosyltransferase Activity in Human Small Intestine. **J. Biol. Chem**, 275 (46): 36164-36171, 2000.

STRATTON, M. R.; CAMPBELL, P. J.; FUTREAL, P. A. The cancer genome. **Nature**, 458: 719-724, 2009.

SUN, T. *et al.* The impact of common genetic variations in genes of the sex hormones metabolic pathways on steroid hormone levels and prostate cancer aggressiveness. **Cancer Prev Res**, 4: 2044-2050, 2011.

TORKAMANI, A.; SCHORK, N. J. Prediction of cancer driver mutations in protein kinases. **Cancer Res**, 68(6): 1675-1682, 2008.

VANDIN, F.; UPFAL, E.; RAPHAEL, B.J. De novo discovery of mutated driver pathways in cancer. **Genome Res**, 22: 375-385, 2012.

WAKUI, S. *et al.* Sex-Associated Difference in Estrogen Receptor β Expression in N-Methyl-N'-Nitro-N-Nitrosoguanidine-Induced Gastric Cancers in Rats. **Comparative Medicine**, 61(5):412-418, 2011.

YAKIREVICH, E.; RESNICK, M. B. Pathology of gastric cancer and its precursor lesions. **Gastroenterol Clin N Am**, 42: 261 - 284, 2013.

ZILBERSTEIN, B.; JACOB, C. E.; CECCONELLO, I. Gastric cancer trends in epidemiology. **Arq Gastroenterol**, 49(3): 177- 178, 2012.

APÊNDICES

APÊNDICE A- Número de acesso dos transcriptomas disponíveis no SRA, características clínicas dos pacientes e quantidade de leituras por transcriptoma.

Número de acesso no ENA	Transcriptoma	Sexo	Idade	Estágio da doença*	Classificação histológica**	Total de leituras (milhões)	Tamanho do arquivo (Gb)
SRR452335	1	M	83	I	Misto	59,9	2
SRR452336	2	M	74	IV	Difuso	60,0	3
SRR452337	3	M	63	III	Intestinal	95,9	4
SRR452338	4	M	58	III	Intestinal	70,7	3
SRR452339	5	F	70	II	Difuso	82,7	4
SRR452340	6	M	64	IV	Difuso	81,7	4
SRR452341	7	M	76	I	Difuso	53,0	2
SRR452342	8	F	67	Normal	Normal	67,1	3
SRR452343	9	M	61	I	Intestinal	72,1	3
SRR452344	10	M	69	Normal	Normal	63,4	3
SRR452345	11	M	69	II	Difuso	74,7	3
SRR452346	12	M	37	Normal	Normal	89,5	4
SRR452347	13	M	56	Normal	Normal	36,2	4
SRR452348	14	F	60	IV	Misto	64,6	3
SRR452349	15	M	69	IV	Difuso	59,0	2
SRR452350	16	M	52	II	Desconhecido	53,2	2
SRR452351	17	F	32	Normal	Normal	68,8	3
SRR452352	18	F	73	III	Intestinal	91,3	4
SRR452353	19	F	66	I	Difuso	59,7	3
SRR452354	20	M	66	II	Intestinal	73,8	3
SRR452355	21	M	64	IV	Difuso	58,0	2
SRR452356	22	M	66	Normal	Normal	55,7	2
SRR452357	23	M	52	III	Difuso	78,7	4
SRR452358	24	M	47	III	Intestinal	62,6	3
SRR452359	25	M	61	III	Intestinal	73,7	3
SRR452360	26	M	64	I	Difuso	96,0	4
SRR452361	27	M	55	III	Difuso	80,7	4
SRR452362	28	M	72	III	Difuso	89,3	4
SRR452363	29	M	–	IV	Desconhecido	77,4	3
SRR452364	30	M	–	II	Desconhecido	73,9	3

*American Joint Commission on Cancer.

**Classificação histológica de Lauren.

Fonte: Adaptada de KIM *et al.* (2012).

APÊNDICE B (Continua)- Lista dos genes codificantes das enzimas participantes da via de biossíntese dos hormônios esteróides e do receptor de progesterona analisados no estudo.

Símbolo do gene	Nome do gene	Localização do transcrito no genoma (incluindo UTRs)*	Tamanho do transcrito (pares de base)
AKR1C1	Aldo-keto reductase family 1, member C1	10: 5.005.454-5.010.035	4.582
AKR1C2	Aldo-keto reductase family 1, member C2	10 :5.042.103-5.046.206	4.104
AKR1C3	Aldo-keto reductase family 1, member C3	10: 5.005.454-5.149.878	144.425
AKR1C4	Aldo-keto reductase family 1, member C4	10: 5.238.798-5.260.910	22.113
AKR1D1	Aldo-keto reductase family 1, member D1	7: 137.761.178-137.803.050	41.873
COMT	Catechol-O-methyltransferase	7: 137.761.178-137.803.050	41.873
CYP11A1	Cytoome P450, family 11, subfamily A, polypeptide 1	15: 74.631.928-74.660.081	28.154
CYP11B1	Cytoome P450, family 11, subfamily B, polypeptide 1	8: 143.953.773-143.961.236	7.464
CYP11B2	Cytoome P450, family 11, subfamily B, polypeptide 2	8: 143.991.975-143.999.259	7.285
CYP17A1	Cytoome P450, family 17, subfamily A, polypeptide 1	10: 104.590.288-104.597.290	7.003
CYP19A1	Cytoome P450, family 19, subfamily A, polypeptide 1	15: 51.514.546-51.630.795	116.25
CYP1A1	Cytoome P450, family 1, subfamily A, polypeptide 1	15: 75.011.883-75.017.877	5.995
CYP1A2	Cytoome P450, family 1, subfamily A, polypeptide 2	15: 75.041.184-75.048.941	7.758
CYP1B1	Cytoome P450, family 1, subfamily B, polypeptide 1	2: 38.294.746-38.303.323	8.578
CYP21A2	Cytoome P450, family 21, subfamily A, polypeptide 1	6: 32.006.093-32.009.447	3.355
CYP2E1	Cytoome P450, family 2, subfamily E, polypeptide 1	10: 135.340.867-135.352.620	11.754
CYP3A4	Cytoome P450, family 3, subfamily A, polypeptide 4	7: 99.354.583-99.381.811	27.229
CYP3A5	Cytoome P450, family 3, subfamily A, polypeptide 5	7: 99.268.845-99.277.621	8.777
CYP3A7	Cytoome P450, family 3, subfamily A, polypeptide 7	7: 99.302.660-99.332.819	30.160
CYP7A1	Cytoome P450, family 7, subfamily A, polypeptide 1	8: 59.402.737-59.412.720	9.984

*Cromossomo: posição inicial-posição final.

Fonte: KEGG Pathway database.

APÊNDICE B (Continua) - Lista dos genes codificantes das enzimas participantes da via de biossíntese dos hormônios esteróides e do receptor de progesterona analisados no estudo.

Símbolo do gene	Nome do gene	Localização do transcrito no genoma (incluindo UTRs)*	Tamanho do transcrito (pares de base)
CYP7B1	Cytochrome P450, family 7, subfamily B, polypeptide 1	8: 65.508.529-65.711.348	202.82
HSD11B1	Hydroxysteroid (11-beta) dehydrogenase 1	1: 209.859.525-209.908.295	48.771
HSD11B2	Hydroxysteroid (11-beta) dehydrogenase 2	16: 67.465.036-67.471.454	6.419
HSD17B1	Hydroxysteroid (17-beta) dehydrogenase 1	17: 40.703.984-40.707.232	3.249
HSD17B12	Hydroxysteroid (17-beta) dehydrogenase 12	11: 43.702.143-43.878.169	176.027
HSD17B2	Hydroxysteroid (17-beta) dehydrogenase 2	16: 82.068.858-82.132.139	63.282
HSD17B3	Hydroxysteroid (17-beta) dehydrogenase 3	9: 98.997.589-99.064.434	66.846
HSD17B6	Hydroxysteroid (17-beta) dehydrogenase 6	12: 57.157.108-57.181.574	24.467
HSD17B7	Hydroxysteroid (17-beta) dehydrogenase 7	1: 162.760.496-162.782.608	22.113
HSD17B8	Hydroxysteroid (17-beta) dehydrogenase 8	6: 33.172.414-33.174.608	2.195
HSD3B1	Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1	1: 120.049.941-120.057.681	7.741
HSD3B2	Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2	1: 119.957.270-119.965.662	8.393
SRD5A1	Steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)	5: 6.633.500-6.669.675	36.176
SRD5A2	Steroid-5-alpha-reductase, alpha polypeptide 2	2: 31.749.656-31.806.040	56.385
STS	Steroid sulfatase (microsomal), isozyme S	X: 7.137.472-7.272.682	135.211
SULT1E1	Sulfotransferase family 1E, estrogen-preferring, member 1	4: 70.714.844-70.725.870	11.027
SULT2B1	Sulfotransferase family, cytosolic, 2B, member 1	19: 49.078.993-49.102.684	23.692
UGT1A1	UDP glucuronosyltransferase 1 family, polypeptide A1	2: 234.668.919-234.680.261	11.343
UGT1A10	UDP glucuronosyltransferase 1 family, polypeptide A10	2: 234.545.123-234.681.951	136.829
UGT1A3	UDP glucuronosyltransferase 1 family, polypeptide A3	2: 234.637.773-234.678.310	40.538

*Cromossomo: posição inicial-posição final.

Fonte: KEGG Pathway database.

APÊNDICE B (Conclusão) - Lista dos genes codificantes das enzimas participantes da via de biossíntese dos hormônios esteróides e do receptor de progesterona analisados no estudo.

Símbolo do gene	Nome do gene	Localização do transcrito no genoma (incluindo UTRs)*	Tamanho do transcrito (pares de base)
UGT1A4	UDP glucuronosyltransferase 1 family, polypeptide A4	2: 234.637.773-234.681.951	44.179
UGT1A5	UDP glucuronosyltransferase 1 family, polypeptide A5	2: 234.621.638-234.678.310	56.673
UGT1A6	UDP glucuronosyltransferase 1 family, polypeptide A6	2: 234.601.512-234.680.261	78.75
UGT1A7	UDP glucuronosyltransferase 1 family, polypeptide A7	2: 234.590.584-234.678.310	87.727
UGT1A8	UDP glucuronosyltransferase 1 family, polypeptide A8	2: 234.526.291-234.678.310	152.02
UGT1A9	UDP glucuronosyltransferase 1 family, polypeptide A9	2: 234.580.544-234.678.310	97.767
UGT2A1	UDP glucuronosyltransferase 2 family, polypeptide A1	4: 70.454.135-70.518.967	64.833
UGT2A2	UDP glucuronosyltransferase 2 family, polypeptide A2	4: 70.454.135-70.505.360	51.226
UGT2A3	UDP glucuronosyltransferase 2 family, polypeptide A3	4: 69.794.177-69.817.509	23.333
UGT2B10	UDP glucuronosyltransferase 2 family, polypeptide B10	4: 69.870.295-69.886.113	15.819
UGT2B11	UDP glucuronosyltransferase 2 family, polypeptide B11	4: 69.681.713-69.697.741	16.029
UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15	4: 69.512.315-69.536.494	24.180
UGT2B17	UDP glucuronosyltransferase 2 family, polypeptide B17	4: 69.402.903-69.434.245	31.343
UGT2B28	UDP glucuronosyltransferase 2 family, polypeptide B28	4: 70.146.217-70.160.768	14.552
UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	4: 70.345.883-70.391.732	45.850
UGT2B7	UDP glucuronosyltransferase 2 family, polypeptide B7	4: 69.962.193-69.978.705	16.513
PGR	Progesterone receptor	11: 100.900.355-101.000.458	100.104

*Cromossomo: posição inicial-posição final.

Fonte: KEGG Pathway database.

APÊNDICE C - Lista dos genes codificantes das enzimas que fazem parte da via de biossíntese do colesterol analisados no estudo.

Símbolo do gene	Nome do gene	Posição do transcrito no genoma (incluindo UTRs)*	Tamanho do transcrito (pares de base)
<i>Via do Mevalonato ao Esqualeno*</i>			
<i>CYP51A1</i>	Cytochrome P450, family 51, subfamily A, polypeptide 1	7: 91.741.463-91.763.840	22.378
<i>DHCR24</i>	24-dehydrocholesterol reductase	1: 55.315.300-55.331.353	16.054
<i>DHCR7</i>	7-dehydrocholesterol reductase	11: 71.145.457-71.159.477	14.021
<i>EBP</i>	Emopamil binding protein (sterol isomerase)	X: 48.380.164-48.387.104	6.941
<i>HSD17B7</i>	Hydroxysteroid (17-beta) dehydrogenase 7	1: 162.760.496-162.782.608	22.113
<i>LSS</i>	Lanosterol Synthase (2,3-Oxidosqualene-Lanosterol Cyclase)	21: 47.608.360-47.648.738	40.379
<i>NSDHL</i>	NAD(P) dependent sterol dehydrogenase-like	X:151.999.511-152.037.907	38.397
<i>SC4MOL</i>	Methylsterol monooxygenase 1	4:166.248.818-166.264.314	15.497
<i>SC5D</i>	Sterol-C5-desaturase	11:121.163.388-121.184.119	20.732
<i>SQLE</i>	Squalene epoxidase	8:126.010.720-126.034.525	23.806
<i>TM7SF2</i>	Transmembrane 7 superfamily member 2	11:64.879.326-64.883.707	4.382
<i>Via do Esqualeno ao Colesterol**</i>			
<i>FDFT1</i>	Farnesyl-diphosphate farnesyltransferase 1	8:11.653.082-11.696.818	43.737
<i>FDPS</i>	Farnesyl diphosphate synthase	1:155.278.539-155.290.457	11.919
<i>HMGCR</i>	3-Hydroxy-3-Methylglutaryl-CoA Reductase	5:74.632.993-74.657.926	24.934
<i>MVD</i>	Mevalonate (Diphospho) Decarboxylase	16:88.718.348-88.723.674	5.327
<i>MVK</i>	Mevalonate kinase	12:110.011.500-110.035.071	23.572
<i>PMVK</i>	Phosphomevalonate kinase	1:154.897.208-154.909.484	12.277

*Cromossomo: posição inicial-posição final.

Fonte:* BUHAESCU; IZZEDINE (2007); ** MARIJANOVIC *et al.* (2003).

APÊNDICE D (Continua) – Quadro com todos os SNPs identificados nos genes codificantes de proteínas envolvidas nas vias de biossíntese dos hormônios esteróides, do colesterol e do receptor de progesterona nos transcriptomas analisados.

Transcriptoma	SNPs					
	Via de biossíntese dos hormônios esteróides*	Símbolo do gene	Tipo de mutação	Via de biossíntese do colesterol*	Símbolo do gene	Tipo de mutação
Tecido gástrico normal						
8	-	-	-	-	-	-
10	7, 99260766, G, C 10, 5010537, A, G 10, 5014447, A, T 10, 5141616, T, C	<i>CYP3A5</i> <i>AKRIC1</i> <i>AKRIC1</i> <i>AKRIC3</i>	- Não sinônima** - Não sinônima**	-	-	-
12	-	-	-	-	-	-
13	10, 5141618, A, G	<i>AKRIC3</i>	Não sinônima	1, 55316322, A, G 1, 55331122, G, T 8, 11689119, G, C	<i>DHCR24</i> <i>DHCR24</i> <i>FDFTI</i>	rs7374 - Sinônima
17	11, 43877934, C, A	<i>HSD17B12</i>	rs1061810	8, 11689119, G, C	<i>FDFTI</i>	Sinônima
22	7, 99245914, A, G 10, 5011114, A, G	<i>CYP3A5</i> <i>AKRIC1</i>	rs15524 Não sinônima**	1, 55317248, C, T 5, 74638539, T, C	<i>DHCR24</i> <i>HMGCR</i>	- Não sinônima
Câncer gástrico						
1	7, 99250300, A, C 11, 43785541, C, T 11, 43876380, C, T 15, 51518769, C, T	<i>CYP3A5</i> <i>HSD17B12</i> <i>HSD17B12</i> <i>CYP19A1</i>	Não sinônima** - Não sinônima** rs4775934	1, 55316322, A, G 1, 155279969, A, G 7, 91758233, A, G 8, 11689119, G, C	<i>DHCR24</i> <i>FDPS</i> <i>CYP51A1</i> <i>FDFTI</i>	rs7374 Não sinônima Sinônima Sinônima
2	2, 38297515, G, T	<i>CYP11B1</i>	rs16262	-	-	-
3	-	-	-	-	-	-
4	-	-	-	-	-	-
5	7, 99245914, A, G	<i>CYP3A5</i>	rs15524	1, 55317685, G, A 7, 91761164, G, T	<i>DHCR24</i> <i>CYP51A1</i>	rs1138030 Não sinônima**
6	2, 38297515, G, T 7, 99245914, A, G	<i>CYP11B1</i> <i>CYP3A5</i>	rs16262 rs15524	1, 55315445, A, G 7, 91761130, A, G 8, 11683545, G, A 11, 71150074, G, A 11, 71156007, G, C	<i>DHCR24</i> <i>CYP51A1</i> <i>FDFTI</i> <i>DHCR7</i> <i>DHCR7</i>	- Sinônima Não sinônima Não sinônima** -
7	11, 43876813, A, C	<i>HSD17B12</i>	rs11037683	8, 11667246, G, A	<i>FDFTI</i>	Não sinônima
9	-	-	-	-	-	-
11	7, 99264585, T, C	<i>CYP3A5</i>	Não sinônima**	-	-	-
14	10, 5136651, C, G 10, 5139685, G, A	<i>AKRIC3</i> <i>AKRIC3</i>	Não sinônima Não sinônima	1, 55315380, A, G 8, 11689119, G, C	<i>DHCR24</i> <i>FDFTI</i>	- Sinônima
15	-	-	-	8, 126015517, G, A	<i>SQLE</i>	Não sinônima**
16	-	-	-	-	-	-

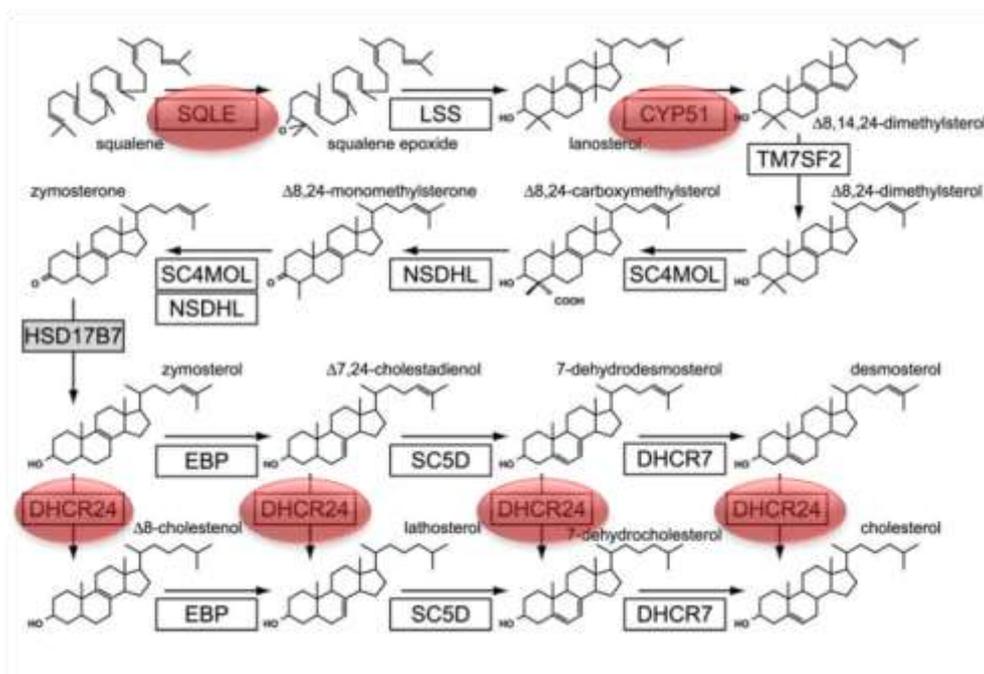
*Cromossomo, posição, base no genoma de referência, base encontrada. **Mutações não sinônimas com efeitos deletérios.

APÊNDICE D (Conclusão) - Listagem dos SNPs identificados nos genes codificantes de proteínas envolvidas nas vias de biossíntese dos hormônios esteróides, do colesterol e do receptor de progesterona nos transcriptomas analisados.

Transcriptoma	SNPs					
	Via de biossíntese dos hormônios esteróides*	Símbolo do gene	Tipo de mutação	Via de biossíntese do colesterol*	Símbolo do gene	Tipo de mutação
Câncer gástrico						
27	-	-	-	-	-	-
28	7, 99261592, T, C	<i>CYP3A5</i>	Não sinônima	-	-	-
29	-	-	-	-	-	-
30	2, 234675778, A, T	<i>UGT1A8+</i>	Sinônima	-	-	-
	7, 99250173, T, A	<i>CYP3A5</i>	-			
	7, 99258413, G, A	<i>CYP3A5</i>	-			
	7, 99258415, G, A	<i>CYP3A5</i>	-			
	7, 99258419, T, C	<i>CYP3A5</i>	-			
	10, 5138738, T, G	<i>AKR1C3</i>	Não sinônima**			
	10, 5141634, G, A	<i>AKR1C3</i>	Não sinônima**			
	10, 5144341, A, G	<i>AKR1C3</i>	Não sinônima**			
10, 5248353, G, A	<i>AKR1C4</i>	-				
11, 43819925 T A	<i>HSD17B12</i>	Não sinônima				
Linhagens celulares do câncer gástrico						
ACP02	-	-	-	-	-	-
ACP03	-	-	-	-	-	-

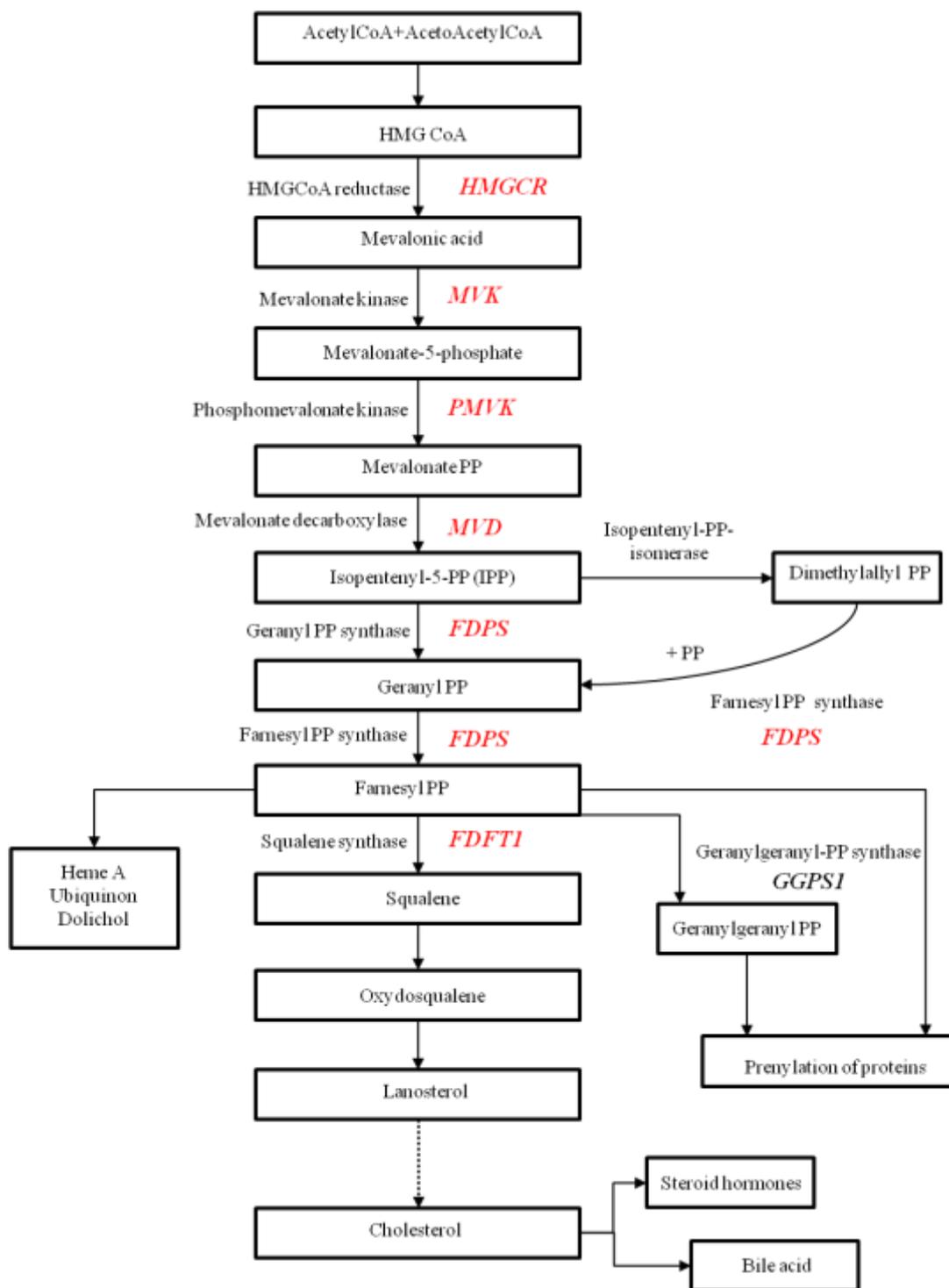
*Cromossomo, posição, base no genoma de referência, base encontrada. **Mutações não sinônimas com efeitos deletérios.

APÊNDICE F– Localização dos SNPs com efeitos deletérios na Via de Biossíntese do Colesterol identificados nos transcriptomas.



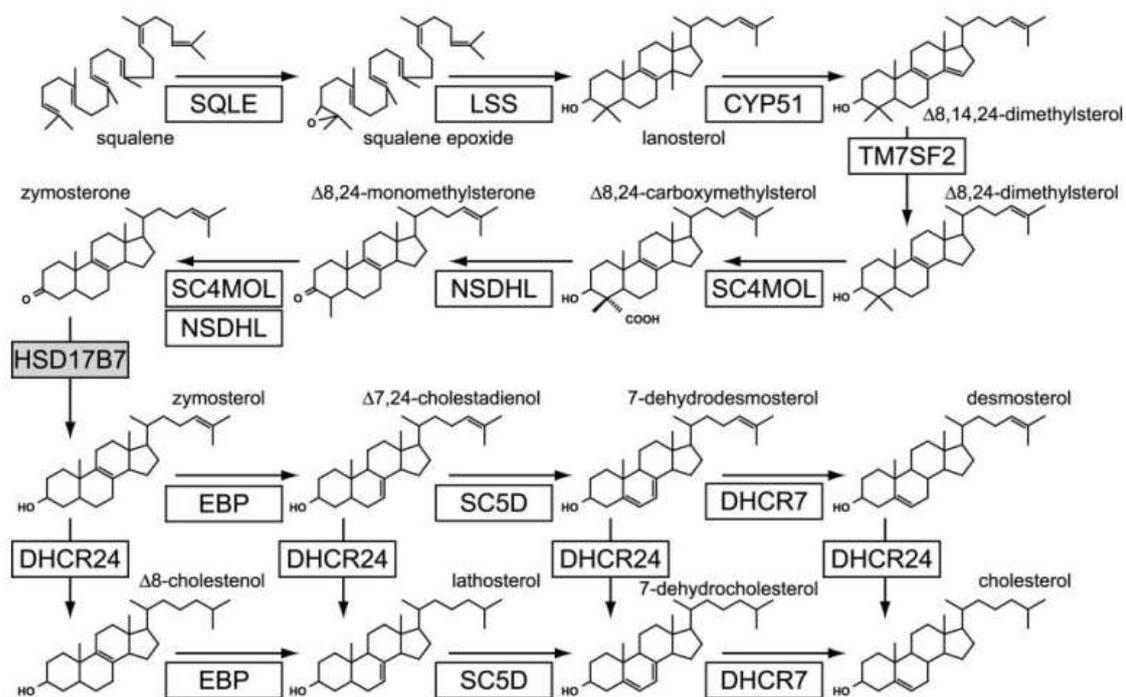
ANEXOS

ANEXO B - Via de biossíntese do colesterol: Do mevalonato ao esqualeno.



Fonte: Adaptado de BUHAESCU; IZZEDINE (2007).

ANEXO B (Continuação) - Via de biossíntese do colesterol: Do escaleno ao colesterol.



Fonte: MARIJANOVIC *et al.* (2003).