

## **MAPEAMENTO DOS GENES DIFERENCIALMENTE EXPRESSOS EM VIAS DE ESTABILIDADE GENÔMICA<sup>1</sup>**

### *MAPPING OF DIFFERENTIALLY EXPRESSED GENES IN GENOMIC STABILITY WAYS*

Sylvio André Garcia Vieira<sup>2</sup> e Eder Maiquel Simão<sup>3</sup>

#### **RESUMO**

Devido às alterações em genes, uma série de processos de defesa é ativada pelos mecanismos de manutenção do genoma (GMM). Os mecanismos incluem as vias de ciclo celular, resposta ao dano do DNA e apoptose, que fazem parte de uma barreira a qual tenta impedir a evolução das mutações em tecidos saudáveis. Essa barreira é observada essencialmente em tecidos pré-cancerosos, como o adenoma de cólon. Suas características envolvem a desativação das vias de ciclo celular e a ativação das vias de apoptose e resposta aos danos do DNA. Tendo em vista estas características, o objetivo deste trabalho foi analisar os genes diferencialmente expressos e estatisticamente significantes em adenoma do cólon e ranqueá-los entre as vias GMM. Por meio da linguagem Java, foram gerados grafos que permitiram a visualização destas relações. Constatou-se que a grande maioria dos genes identificados está ligada diretamente aos processos de oncogênese. Verificou-se também que, na presença de adenoma no organismo, as vias do ciclo celular apresentam maior atividade do que as demais estudadas.

**Palavras-chave:** apoptose, biologia computacional, genoma.

#### **ABSTRACT**

*Due to some changes in genes, a series of protection processes is activated by genome maintenance mechanisms (GMM). The mechanisms include cell cycle pathways, response to DNA damage and apoptosis, which are part of a barrier which tries to prevent the development of mutations in healthy tissues. This barrier is mainly observed in pre-cancerous tissues such as colon adenoma. Its features involve disabling the cell cycle pathways and the activation of the process of apoptosis and response to DNA damage. Given these characteristics, the objective of this study is to analyze the genes that are differentially expressed and statistically significant in colonic adenoma in order to rank them among the GMM ways. Some graphs were generated, via Java language, in order to allowed the visualization of these relationships. It was found out that the great majority of the identified genes is directly linked to oncogenesis processes. It was also found out that in the presence of adenoma in the body, the pathways of the cell cycle exhibit higher activity than the others studied.*

**Keywords:** apoptosis, computational biology, genome.

---

<sup>1</sup> Trabalho resultante de Tese de Doutorado.

<sup>2</sup> Aluno de doutorado do Programa de Pós-graduação em Nanociências - Centro Universitário Francisco. E-mail: sylvio@unifra.br

<sup>3</sup> Orientador - Centro Universitário Francisco. E-mail: edersimao@gmail.com

## INTRODUÇÃO

Com o propósito de encontrar as interações entre proteínas, foram desenvolvidas ferramentas capazes de identificar perfis metabólicos, tanto no que tange às interações proteicas quanto à criação de redes complexas que descrevem essas interações e suas vias, como descrito no artigo de Goh et al. (2007). As vias determinam uma ligação funcional entre proteínas que apresentam funções semelhantes. Se a função de uma proteína é conhecida, então se pode inferir que as proteínas ligadas a ela agem na mesma subvia. As subvias são conjuntos mais específicos do que as vias, como as de reparo por duplas quebras do DNA (*DSBR*) ou as de reparo de bases mal pareadas (*MMR*) (SIMÃO, 2012).

Com o aumento das pesquisas na área da genética, surgiu a necessidade de criação e de expansão de bancos de dados como o *NCI Pathway*, que armazenam e descrevem grupos de proteínas, chamadas de vias de ativação de doenças, que, alteradas em conjunto, venham a desencadear doenças (GOH et al., 2007). A análise dessas vias pode ser realizada por meio de ferramentas computacionais, as quais são desenvolvidas com o intuito de contribuir para a modelagem das redes complexas e para a sua visualização (BASTIAN et al., 2009).

A união da tecnologia, em especial das ferramentas computacionais, com o conhecimento sobre as ligações proteicas pode contribuir para a identificação e a avaliação das interações envolvidas no surgimento das doenças genéticas, bem como para a identificação de processos de cura. No caso do câncer, por exemplo, as proteínas se ligam para iniciar um processo de reparo aos danos do DNA (DDR, do inglês, *DNA damage repair*) em uma série de genes envolvidos na resposta ao dano oxidativo, identificado em um estudo com expressão gênica transcricional (OLIVEIRA, 2007). Estudos realizados por Oliveira (2007) e Kessler et al. (2013) evidenciaram alterações semelhantes em proteínas específicas associadas a vias. Algumas dessas alterações foram encontradas em vias de manutenção do genoma (GMM, do inglês, *Genome Maintenance Mechanisms*), que são formadas por subvias especializadas em realizar a manutenção do ácido desoxirribonucléico (DNA). Essas vias apresentam funções como regulação do ciclo celular, DDR, apoptose e senescência, que se mostram envolvidas em algumas doenças genéticas. É relevante identificar os pontos de verificação do ciclo celular, pois é por meio deles que se torna possível avaliar a existência e o grau dos danos causados ao DNA (FOLGER et al., 2011).

Na ocorrência de danos ao DNA e de acordo com a avaliação destes, inicia-se o processo de DDR. O processo ocorre com uma parada temporária no ciclo celular e é seguido pela ativação de vias especializadas na reparação dos danos. Caso o reparo não seja realizado, é ativado um mecanismo de destruição chamado de morte celular programada ou apoptose (GRIVICICH et al., 2007).

A ativação desses conjuntos de genes tem o propósito de garantir a proliferação e a manutenção celular. Porém, em decorrência de uma inconformidade com a manutenção do DNA, pode ocorrer o surgimento de alterações e, conseqüentemente, o desenvolvimento de doenças genéticas.

O mapeamento das interações entre proteínas e subvias em diferentes tipos de doenças genéticas e suas correlações, juntamente com a identificação de suas semelhanças e diferenças genéticas, poderá tornar-se um diferencial e, até mesmo, uma estratégia para melhorar a qualidade do tratamento dessas doenças (WOOD et al., 2005).

Dentre as doenças genéticas conhecidas, há o câncer de cólon, e, anterior ao câncer, há o adenoma de cólon, que é uma espécie de tumor benigno, conhecido por lesão pré-cancerosa, com grande potencial para o desenvolvimento de câncer (SALCES et al., 2004).

No processo de desenvolvimento e evolução do câncer, apenas 1% do genoma humano está envolvido e ainda precisa estar combinado com o ambiente em que o indivíduo está inserido, bem como com seus hábitos alimentares e com a qualidade do ar que respira (JEFFORD; IRMINGER-FINGER, 2006). O trabalho de Goh et al. (2007) concentrou o resultado de algumas pesquisas e elaborou uma rede de doenças humanas conhecidas, na qual todas as doenças foram relacionadas aos seus genes promotores.

Os objetivos, no presente trabalho, foram identificar os genes mais expressos pelo cálculo de diferenciação de expressão e construir uma rede de interação entre as vias GMM e uma doença genética conhecida como adenoma de cólon.

## **METODOLOGIA**

Foram utilizadas as subvias apresentadas na tabela 1, que mostra as três vias de manutenção do genoma estudadas (ciclo celular (CC), resposta ao dano do DNA (DDR) e apoptose (APO)), com suas respectivas subvias, obtidas da ontologia Ontocancro<sup>4</sup>.

Os dados referentes ao adenoma de cólon foram extraídos do banco de dados (GEO), do qual foram selecionados transcritos de 32 adenomas de cólon e 32 amostras de controle (tecidos normais), retiradas de 28 pacientes, dos quais quatro forneceram duas amostras cada em adenomas diferentes. As amostras são pareadas adenoma/normal oriundas de um mesmo paciente. A plataforma GPL570 (*Affymetrix Human Genome U133 Plus*) (GEO, 2007) foi utilizada para identificação dos nomes dos genes. Os pacientes tinham idades entre 38 e 79 anos, com média de 59,5 anos, sendo vinte do sexo masculino, incluindo os quatro que forneceram duas amostras, e oito do sexo feminino.

Um *software* foi desenvolvido em linguagem Java e MySQL. Inicialmente, as amostras dos 32 casos de adenoma de cólon e dos 32 de controle foram inseridas no banco de dados MySQL separadamente. Houve, então, uma seleção por meio do cálculo de diferenciação de expressão, que consiste em equacionar os valores dos transcritos, dividindo o valor de expressão do gene em adenoma pelo valor de expressão do seu gene controle. O resultado desta equação caracterizou a diferenciação de expressão. Por fim, foram então selecionados os cinquenta genes que possuíam a maior diferenciação.

<sup>4</sup> Ontocancro (ontologia do câncer). Disponível em: <[www.ontocancro.org](http://www.ontocancro.org)>.

A quantidade de genes que interagem com as subvias e que também interagem no processo de adenoma foi o fator determinante no ranqueamento de quais vias estão mais envolvidas com o adenoma. Quanto mais genes estavam presentes, maior foi a classificação da via.

O software aplicou uma simplificação do algoritmo PageRank<sup>5</sup>, que comandou uma busca no banco de dados, relacionando cada uma das vias em estudo com os genes mais expressos em adenoma de cólon. Dessa maneira, foi elaborada uma matriz no formato de arquivo computacional “xml”, na qual foram definidos os nós que fazem parte do grafo (Figura 1a) e as ligações destes (Figura 1b). O sistema criou um grafo que ajustou a posição de cada via de acordo com a classificação recebida no cruzamento com o adenoma.

**Tabela 1** - Relação das subvias utilizadas neste trabalho, agrupadas pela via de manutenção do genoma.

Via do ciclo celular			
Sigla	Nome da subvia	Sigla	Nome da subvia
CCC	Cell Cycle Checkpoints	MMG1	Mitotic M-M/G1 Phases
CCM	Cell Cycle, Mitotic	MSC	Mitotic Spindle Checkpoint
CCCR	Cyclins and Cell Cycle Regulation	RBTSC	Rb Tumor supressor/Check. P. Sign. In Response to Damage
G1SC	G1/S Checkpoint	RDNR	Regulation of DNA Replication
G1SDNA	G1/S DNA Damage Checkpoints	RMCC	Regulation of Mitotic Cell Cycle
G2/MC	G2/M Checkpoint	SP	S Phase
Via de resposta ao dano do DNA			
Sigla	Nome da subvia	Sigla	Nome da subvia
ATM	ATM signalling	HRDSB	He Repair of Replication-Independent DSB
ATR	ATR signalling	MR	Mismatch Repair
BER	Base Excision Repair	NHEJ	Non-Homologous and Joining
DSBR	Double-Strand Break Repair	NER	Nucleotide Excision Repair
FAP	Fanconi Anemia Pathway	DNADSB	Processing of DNA DSB ends Recruitment of Repair and sig. Proteins
HR	Homologous Recombination		
Via da apoptose			
Sigla	Nome da subvia	Sigla	Nome da subvia
AHS	Apoptosis – Homo Sapiens	IATR345	Induction of Apoptosis Through DR3 and DR4/5 death receptors
AEP	Apoptotic Execution Phase	IPA	Intrinsic Pathway for Apoptosis
ASRDDR	Apoptotic Signalling in response to DNA damage	RA	Regulation of Apoptosis
CCA	Caspase Cascade in Apoptosis	TNFR	TNF Receptor Signalling Pathway
DRS	Death Receptor Signalling	TNFR1	TNFR1 Signalling Pathway
EPA	Extrinsic Pathway for Apoptosis	TNFR2	TNFR2 Signalling Pathway
GMAP	Granzyme a Mediated Apoptosis Pathway		

<sup>5</sup> Algoritmo de busca desenvolvido pela empresa Google Inc. utilizado na classificação das páginas da internet atribuindo relevância a cada uma delas em função do assunto pesquisado.

**Figura 1** - Fragmento do arquivo “xml” montado pelo software e aplicado na montagem dos grafos. Em (a) pode-se perceber a identificação dos nós, chamados de “node id=1”, que recebe o nome de “ADENOMA é também é identificado com o gênero “D” que neste caso significa ser a doença, “node id=2” recebe o nome de “BER” que é a via de reparo por excisão de bases, pertencente à via de resposta ao dano de DNA e é identificada como gênero “V” de via. Em (b) pode-se identificar os relacionamentos em que o nó com identificação “1” é ligado ao de identificação “2”, depois ao “3” e assim sucessivamente até o término do arquivo, indicando todas as ligações que foram identificadas pelo *software*.

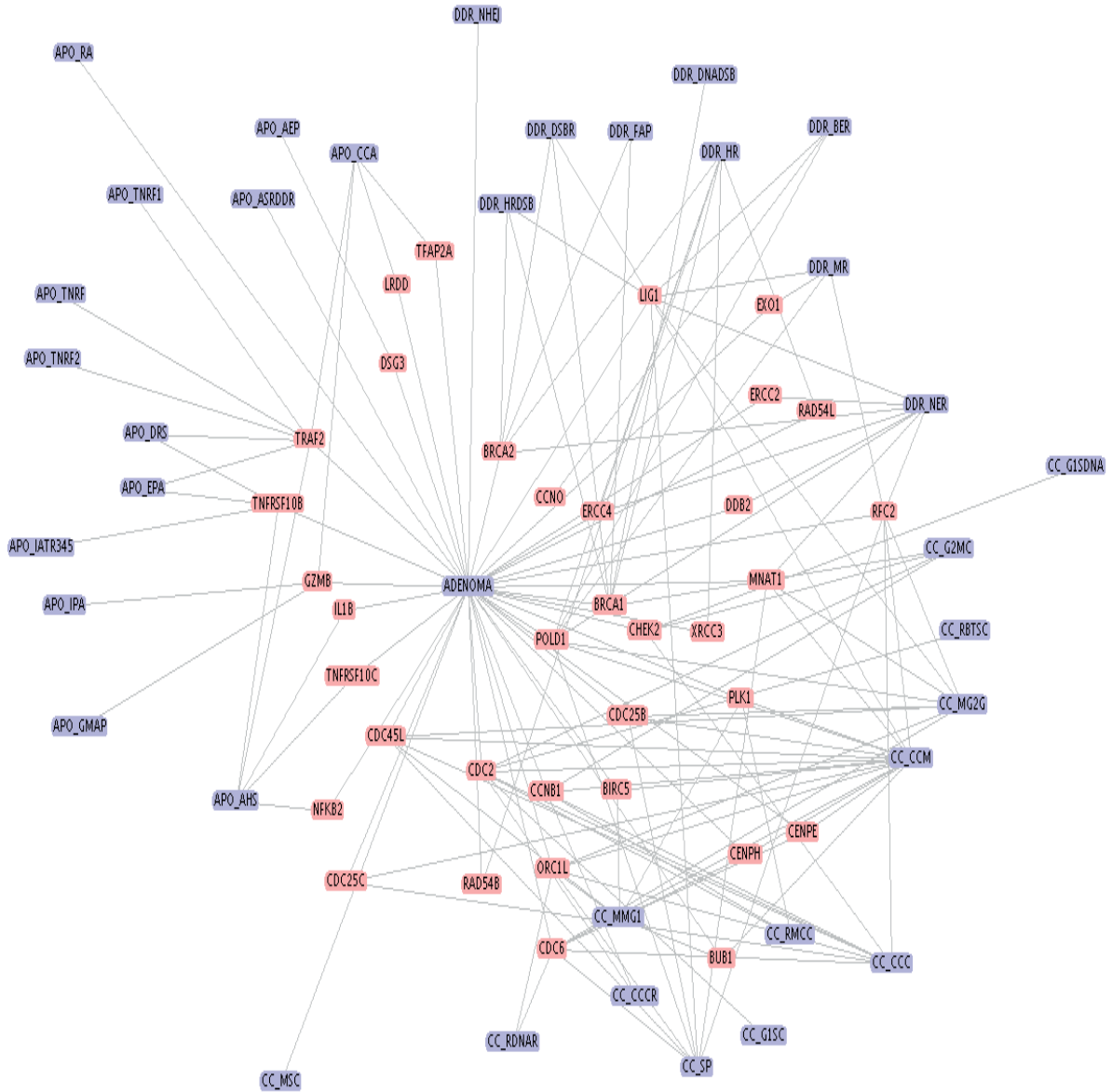
<pre> &lt;node id="1"&gt; &lt;data key="name"&gt;ADENOMA&lt;/data&gt; &lt;data key="gender"&gt;D&lt;/data&gt; &lt;/node&gt; &lt;node id="2"&gt; &lt;data key="name"&gt;BER&lt;/data&gt; &lt;data key="gender"&gt;V&lt;/data&gt; &lt;/node&gt; &lt;node id="3"&gt; &lt;data key="name"&gt;DNADSB&lt;/data&gt; &lt;data key="gender"&gt;V&lt;/data&gt; &lt;/node&gt; </pre> <p style="text-align: center;">(a)</p>	<pre> &lt;edge source="1" target="2"&gt;&lt;/edge&gt; &lt;edge source="1" target="3"&gt;&lt;/edge&gt; &lt;edge source="1" target="4"&gt;&lt;/edge&gt; </pre> <p style="text-align: center;">(b)</p>
--	---

## RESULTADOS

A construção das bases de dados e a aplicação do algoritmo de ranqueamento possibilitou a análise das vias de manutenção do genoma com o adenoma de cólon. Com a elaboração do grafo (Figura 2), foi possível verificar a interação destes genes com os genes presentes nos adenomas. O arquivo “xml” montado pelo sistema, que pode ser observado na figura 1, é interpretado pelo software que avalia as ligações e efetua a construção do grafo, agrupando, sempre que possível, os genes responsáveis pelas conexões das vias e aproximando as vias que estão mais conectadas.

Os grafos foram gerados inicialmente abrangendo todas as subvias e seus respectivos genes, disponíveis na ontologia Ontocancro, e envolvidos com ciclo celular, resposta ao dano do DNA e apoptose. O grafo gerado pela interação de uma grande quantidade de subvias e proteínas correspondentes resulta em um ambiente visual bastante confuso, com muitos nós e arestas, o que prejudica bastante a sua compreensão. O filtro e a classificação proporcionados pelo algoritmo permitem uma compreensão mais rápida das inter-relações entre a doença e as vias GMM, como ilustrado na figura 2.

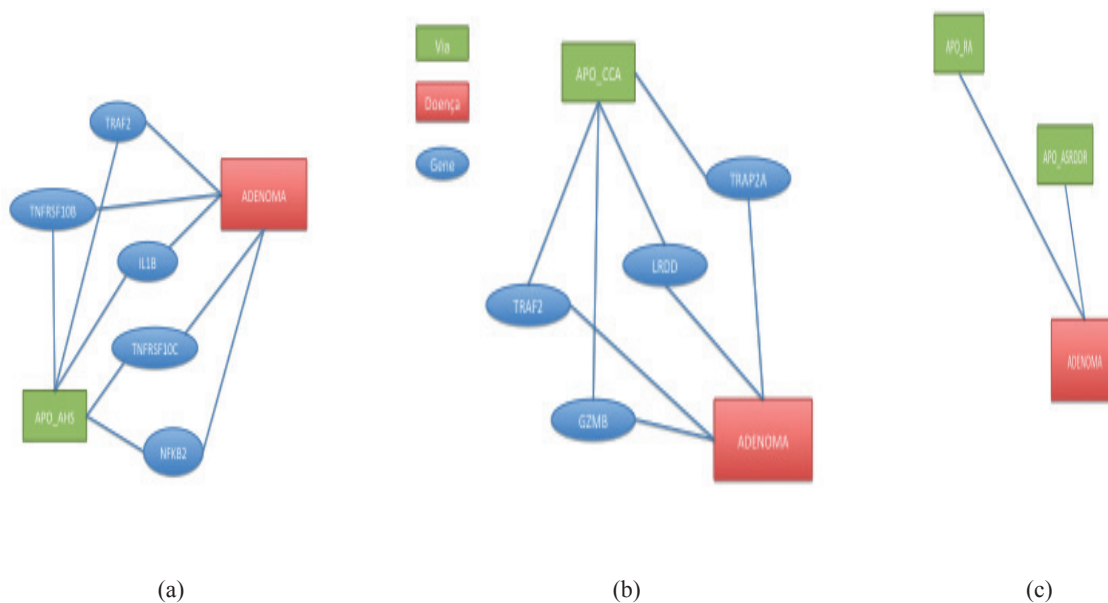
**Figura 2** - Inter-relação entre o adenoma e as subvias das vias GMM. Os pontos em azul representam as subvias e os pontos em salmão representam os genes. Dessa forma, é possível identificar quais são os genes mais envolvidos com as subvias e o adenoma. As subvias da apoptose ASRDDDR e RA, bem como a subvia MSC do ciclo celular e a subvia NHEJ, não se relacionam com o adenoma por nenhum gene ranqueado.



A técnica de visualização por grafos permite observação e compreensão bem mais detalhadas, não somente das vias, como também dos genes envolvidos e de suas interações. Pôde ser identificado, por exemplo, que a subvia conhecida como apoptose do *homo sapiens* (AHS) possui cinco genes expressos que estão presentes nos cinquenta genes com maior alteração na presença

do adenoma, que são os genes supressores de tumor (*TRAF2* e *TNFRSF10C*), o gene *IL1B*, que é promotor de inflamação, e o *NFKB2*, que é vinculado à deficiência imunológica (CHEN et al., 2013). Esta configuração pode indicar a presença de atividade anormal no organismo, que pode ser identificada na figura 3a. A subvia de cascata de caspases da apoptose (CCA) possui quatro genes interligados ao adenoma (Figura 3b), o *TRAF2*, que é um gene receptor de TNF; o *TFAP2A*, que é um gene ativador/inibidor de transcrição; o *LRDD*, que é uma leucina induzida pelo *p53* com domínio de morte; e o *GZMB*, que codifica a granzima B, pró-apoptótica (GARCIA, 2010). Por outro lado, identificou-se que a subvia da regulação da apoptose (RA) e a subvia de sinalização de apoptose em resposta ao dano de DNA (ASRDDR) não possuem nenhum dos genes ligados aos cinquenta genes diferencialmente expressos da doença pesquisada (Figura 3c). Essas ligações realizadas pelo *software* vêm ao encontro ao que a literatura citada já apresenta, o que corrobora os resultados alcançados.

**Figura 3** - Visualização parcial específica do grafo apresentado na figura 3, demonstrando a localização da Apoptose do *Homo Sapiens* (parcial) (a) e Cascata de caspases (parcial) (b) e a subvia RA, que não possui genes associados ao adenoma (c).



Os genes que promoveram as maiores relações, de acordo com a via envolvida, e que estiveram presentes em pelo menos duas subvias estão apresentados na tabela 2.



**Tabela 2** - Relação dos genes envolvidos na doença e a quantidade de subvias em que apareceram agrupados por via de manutenção do genoma.

Via do ciclo celular			
Gene	Quantidade de subvias	Gene	Quantidade de subvias
CDC2	8	CHEK2	3
CCNB1	6	MNAT1	3
CDC6	5	RFC2	3
ORC1L	5	BIRC5	2
CDC45L	4	BUB1	2
PLK1	4	CENPE	2
CDC25B	3	CENPH (2)	2
CDC25C	3	POLD1	2
Via de resposta ao dano do DNA			
Gene	Quantidade de subvias	Gene	Quantidade de subvias
BRCA1	5	POLD1	4
BRCA2	5	ERCC4	2
LIG1	5	RFC2	2
Via da apoptose			
Gene	Quantidade de subvias	Gene	Quantidade de subvias
TRAF2	7	GZMB	3
TNRSF10B	4		

## CONCLUSÃO

A utilização de grafos para registrar as interações entre os genes e as vias permite uma visualização rápida e clara sobre quais subvias estão envolvidas com a doença e quais os genes que os correlacionam. Permite também avaliar a ocorrência e a frequência das interações nas diferentes subvias, bem como identificar por agrupamento de via. Assim, tornou-se possível identificar que alguns genes são relacionados a várias subvias e, quando afetados por uma doença, alteram-se para, em conjunto, poderem realizar a função desta, mas que, quando em vias diferentes, eles não se relacionam. Dessa forma, existem diferentes redes de interação de genes no processo de formação de adenoma com as vias, distintos em ciclo celular, apoptose e de resposta aos danos do DNA.

Foi possível identificar por meio deste trabalho que, na presença de adenoma, há uma forte atividade nas subvias de ciclo celular, uma atividade menos evidenciada das subvias de resposta ao dano do DNA e uma baixa atividade das subvias da apoptose. Disso é possível inferir que, no adenoma, há uma grande produção de células e uma tentativa mediana do organismo em tentar o reparo em células alteradas. Porém, também se pôde inferir que há uma baixa tentativa por parte do organismo em eliminar células que não podem ser reparadas.

Analisando as características apresentadas, verifica-se que os genes podem se relacionar uns com os outros e com suas vias, bem como com processos de formação de doenças. Estas interações permitiram identificar a relação biológica de uma doença genética com suas possíveis rotas de forma-



ção dentro dos organismos. Tais rotas de doenças são explicitadas quando se determina toda a sequência de genes que sofrem mutações para sua formação (GOH et al., 2007). Há também uma sequência de mutações no processo de cura do organismo, em que uma ou várias vias são ativadas, a partir das mutações de um ou mais genes. Sequencialmente, outros genes vão sofrendo estas mutações na tentativa de combater doenças (KUEHN, 2010).

Os grafos, associados às ferramentas computacionais e ao algoritmo destas, permitiram agilizar o processo de identificação das correlações entre as vias de manutenção do genoma e o processo de adenoma. Alguns bancos de dados já estão adaptados a padrões de nomenclatura para genes e proteínas, o que permite ir adiante com este trabalho para realizar o mapeamento das relações entre as vias de manutenção do genoma e outras doenças genéticas.

Com a análise dos genes que estão entre os mais alterados no processo de adenoma e das subvias de manutenção do genoma, verificou-se uma relação dos genes com a ativação do sistema imunológico. Um exemplo é o *NFKB*, que desempenha um papel fundamental na regulação da resposta imunitária à infecção (GILMORE, 2006). Outros genes que fazem parte da formação do câncer, como o *BRCA1* e o *BRCA2* (YAO et al., 2014), estão bastante expressos no adenoma de cólon e presentes em cinco subvias de resposta ao dano do DNA.

A identificação das subvias envolvidas nos processos de ativação de doenças pode indicar um novo caminho para investigação de processos de tratamento.

## REFERÊNCIAS

BASTIAN, M.; HEYMANN, S.; JACOMY, M. **Gephi**: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. North America, mar. 2009 (ICWSM09).

CHEN, K. et al. Germline Mutations in *NFKB2* Implicate the Noncanonical NF- $\kappa$ B Pathway in the Pathogenesis of Common Variable Immunodeficiency. **The American Journal of Human Genetics**, v. 93, p. 812-824, 2013.

FOLGER, O. et al. Predicting selective drug targets in cancer through metabolic networks. **Molecular systems biology**, v. 7, n. 7, p. 501, 2011.

GARCIA, F. B. **Análise de polimorfismos no promotor do gene da perforina e nos éxons 2, 3 e 5 do gene da granzima b em grupos étnicos brasileiros**. 2010. 103f. Tese (Doutorado em Patologia Clínica) - Programa de Pós-graduação em Patologia, Universidade Federal do Triângulo Mineiro, Uberaba - MG, 2010.

GEO. **Site oficial**. 2007. Disponível em: <<http://1.usa.gov/112231H>>. Acesso em: 20 jun. 2014.

GILMORE, T. D. Introduction to NF-kappaB: players, pathways, perspectives. **Oncogene**, v. 25, n. 51, p. 6680-4, 2006.

GOH, K. et al. The human disease network. **PNAS**, v. 104, n. 21, p. 8685-8690, 2007.

GRIVICICH, I.; REGNER, A.; BRONDANI, A. R. Morte celular por Apoptose. **Revista Brasileira de Cancerologia**, v. 53, n. 3, p. 335-343. 2007.

JEFFORD, C. E.; IRMINGER-FINGER, I. Mechanisms of chromosome instability in cancers. **Critical reviews in oncology/hematology**, v. 59, p. 1-14, 2006.

KESSLER, T.; HACHE, H.; WIERLING, C. Integrative analysis of cancer-related signaling pathways. **Frontiers in Physiology**, v. 4, Article 124, 2013.

KUEHN, B. M. Genomics Illuminates a deadly Brain Cancer. **American Medical Association. JAMA**, v. 303, n. 10, p. 925-927, 2010.

OLIVEIRA, D. V. P. de. **Avaliação da expressão dos genes processadores de danos oxidativo em pacientes Alzheimer**. 2007. 73f. Dissertação (Mestrado em Genética) – Programa de Pós-graduação em Ciências, Faculdade de Medicina de Ribeirão Preto - Departamento de Genética, Ribeirão Preto, 2007.

SALCES, I. et al. Tissue ca-19.9 content in colorectal adenomas and its value in the assesment of dysplasia. **Revista Española De Enfermedades Digestivas (rev. esp. enferm. dig.)**, Madrid, v. 96, n. 49, p. 246-254, 2004.

SIMÃO, E. M. **Dinâmica de transição pré-câncer para câncer: estudo da expressão de vias de manutenção do genoma**. 2012. 101f. Tese (Doutorado em Física) – Programa de Pós-graduação em Física, Centro de Ciências Naturais e Exatas, Universidade Federal de Santa Maria, Santa Maria – RS, 2012.

WOOD, R. D.; MITCHELL, M.; LINDAHL, T. Human DNA repair genes. Fundamental and Molecular Mechanisms of mutagenesis. **Science Direct, ELSEVIER**, v. 577, Issues 1-2,4, p. 275-283, 2005.

YAO, K. et al. Nipple-Sparing Mastectomy in BRCA1/2 Mutation Carriers: An Interim Analysis and Review of the Literature. **Annals of Surgical Oncology**, v. 22, Issue 2, p. 370-376, 2015.