

Uso da ontologia Ontocancro para comprovar o envolvimento de genes com a barreira de progressão do câncer

Laís Falcade, Karlise Soares, Karine de Souza, Eder Simão,
José Carlos Marino Mombach, Giovani Rubert Librelotto

Programa de Pós-Graduação em Informática – PPGI
Universidade Federal de Santa Maria – UFSM
Av. Roraima, 1000, Camobi, 97105-900
Santa Maria, RS

librelotto@inf.ufsm.br

Abstract. *An important question involving the study of the human genome refers to research into the expression of genes involved in cancer formation. The Ontocancro 2.0 was developed to assist with this task and provide data integration related to pathways of genome maintenance mechanisms (GMM). Allied to it is an algorithm for statistical analysis of the values of diversity and relative activity. The purpose of this article is to demonstrate how Ontocancro can demonstrate quantitatively the results obtained by studies that describe a barrier to progress of cancer in precancerous lesions.*

Resumo. *Um dos maiores questionamentos envolvendo o estudo do genoma humano refere-se à investigação sobre a expressão de genes envolvidos com a formação do câncer. A Ontocancro 2.0 foi desenvolvida para auxiliar nesta tarefa e prover a integração de dados relacionados às vias dos mecanismos de manutenção do genoma (GMM). Aliado a ela há um algoritmo para análise estatística dos valores de atividade e diversidade relativa. O objetivo deste artigo é demonstrar como a Ontocancro pode comprovar quantitativamente os resultados obtidos por estudos que descrevem uma barreira de evolução do câncer em lesões pré-cancerosas.*

1. Introdução

Uma grande quantidade de informações tem sido produzida, nos últimos anos, através do estudo do genoma humano. Diversas pesquisas estão sendo realizadas com o objetivo de interpretar o funcionamento dos componentes que fazem parte desse sistema. Um dos maiores questionamentos refere-se à interação entre proteínas e qual a importância da expressão de suas funcionalidades dentro do organismo humano.

A integração das informações passou a ser essencial para a quantificação dos elementos moleculares de um sistema biológico e também para a formação de modelos a serem utilizados na construção de hipóteses de previsão para explicar as propriedades emergentes, que caracterizam as interações que ocorrem com os interatomos [Hood *et al.*, 2004].

Uma das maiores dificuldades na integração dos dados encontra-se no tratamento das informações, uma vez que encontram-se dispersos em diversos bancos de dados públicos. A ontologia Ontocancro permite a integração de dados relacionados

às vias dos Mecanismos de Manutenção do Genoma (GMM), auxiliando na investigação do funcionamento das redes metabólicas de genes envolvidos com o câncer [Librelotto et al. 2009]. As informações extraídas a partir da ontologia estão disponíveis¹ através de uma aplicação *Web* denominada Ontocancro 2.0. Nesta aplicação, também há um algoritmo para análise estatística dos valores de atividade e diversidade relativa, que são métodos quantitativos de expressão de genes usados na biologia sistêmica [Pereira, 2012].

Halazonetis (2008) propôs um modelo que descreve uma barreira de evolução tumoral, em tecidos pré-cancerosos, responsável pela ativação das vias metabólicas de Apoptose e de Resposta aos Danos, mas quando rompida dá origem a um tecido maligno. Este estudo motivou o grupo de pesquisadores da Ontocancro a elaborar um modelo quantitativo que serve de apoio computacional ao modelo inicialmente proposto.

O objetivo deste artigo é demonstrar como a Ontocancro 2.0 executa a análise quantitativa através do cálculo da atividade e diversidade relativa proposta por Castro e colaboradores [Castro, M. A. A. et al. 2007] e testar os resultados obtidos com o estudo proposto por [Halazonetis et al. 2008]. Para isso, o artigo aborda na seção 2, o trabalho realizado por Halazonetis e colaboradores e uma breve explicação sobre os mecanismos de manutenção do genoma. A seção 3 contém informações sobre a ontologia Ontocancro e os dados armazenados em seu banco de dados. A seção 4 apresenta os cálculos disponibilizados na interface da Ontocancro 2.0 para encontrar os valores de atividade relativa e diversidade relativa. A seção 5 descreve um estudo de caso comparando as abordagens da Ontocancro e da ferramenta ViaComplex.

2. Investigação da formação do pré-câncer

Dentre as pesquisas relacionadas às doenças humanas, o câncer tem se destacado como um importante campo de estudo, uma vez que existem diversas variações desta doença que evoluem de acordo com fatores genéticos e ambientais [Simão, 2012].

Sabe-se que o surgimento de carcinomas (tumores malignos relacionados aos tecidos moles) decorre da formação desordenada das células devido à perda da estabilidade do genoma. Esse crescimento anormal pode evoluir para o câncer passando por três estágios pré-cancerosos, representados na Figura 1. A progressão tumoral ocorre devido a alterações em genes ligados aos GMM.

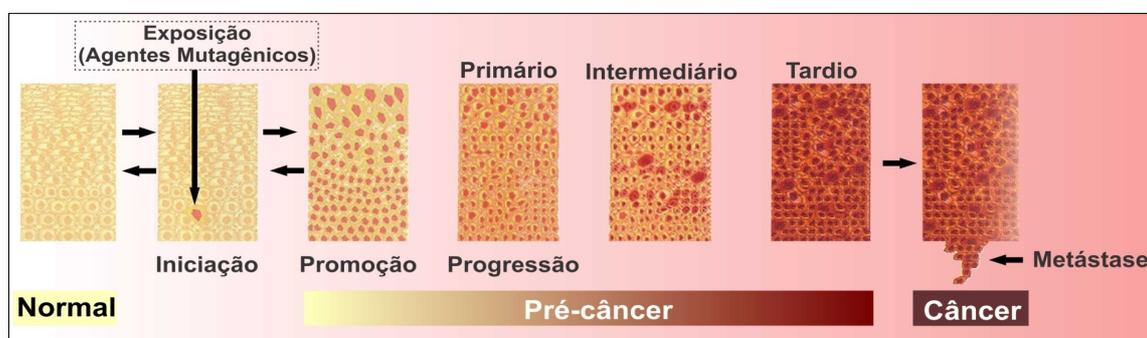


Figura 1 - Progressão tumoral. Adaptada de [Simão 2012].

¹A Ontocancro 2.0 encontra-se disponível no endereço www.ontocancro.org

Os GMMs são responsáveis por corrigir falhas e garantir a integridade do DNA e conseqüentemente, a sobrevivência da célula. Estes mecanismos envolvem uma série de proteínas e vias metabólicas que interagem para responder a danos no DNA e agir na supressão de tumores. A apoptose, por exemplo, é uma conhecida via metabólica responsável por ativar a morte celular em caso de danos irreversíveis.

Com base no estudo qualitativo realizado por Halazonetis et al (2008), os tecidos pré-cancerosos, em geral, apresentam uma diminuição da expressão em vias de manutenção do genoma relacionada ao índice de proliferação celular, seguida por um aumento da atividade da via de apoptose e reparo de danos no DNA (DDR), surgindo, assim, uma barreira de progressão anti-câncer que resiste a transformação tumoral. Uma vez que esta barreira é rompida, origina-se então, o tecido maligno.

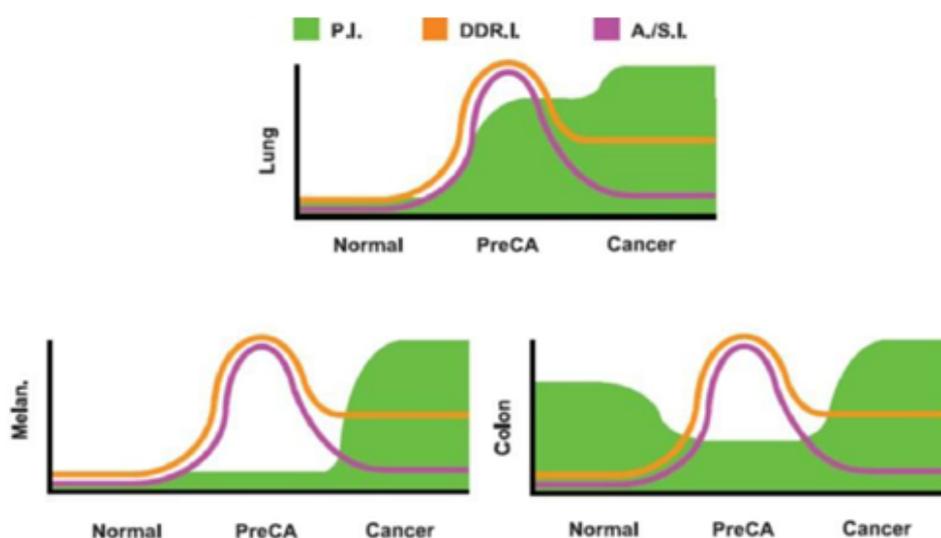


Figura 2 - Barreira de evolução do câncer em lesões pré-cancerosas (preCA). O índice de proliferação (P.I.), o índice de resposta ao dano do DNA (DDR.I.) e o índice de apoptose e senescência (A./S.I.) são mostrados para os tecidos normais, pré-câncer (preCA) e câncer de pulmão, melanoma e colorretal. Adaptada de Halazonetis et al. 2008.

Segundo Halazonetis, a ativação de oncogenes causa as alterações dos tecidos pré-cancerosos, desencadeando uma proliferação de mutações. Em resposta, o tecido inicia o mecanismo de Reparo de Danos do DNA ativando as vias especializadas que criam a barreira anti-câncer.

A transformação do fenótipo de uma célula normal em um fenótipo alterado é observada em tumores que surgem de duas formas: a primeira delas é baseada no acúmulo de anomalias genéticas observadas ao longo do período de vida, chamadas de mutações somáticas e que aparecem na maioria dos tumores sólidos, e em segundo lugar aparecem às mutações germinativas que são passadas aos descendentes de forma hereditária. [Castro, M. et al. 2007]

A próxima seção apresenta as informações obtidas para o armazenamento no banco de dados da Ontocancro 2.0

3. Armazenamento de dados biológicos e a Ontocancro 2.0

O gerenciamento da grande demanda de informação produzida nas pesquisas exige a utilização de bancos de dados para o seu armazenamento. A maior dificuldade dos cientistas é a padronização de termos o que causa hoje uma enorme dificuldade em integrar diferentes bancos de dados e informações. Essa dificuldade foi solucionada com a criação de alguns bancos públicos usados para padronizar as descobertas provenientes do Projeto Genoma Humano [Watson 1990].

O *Gene Nomenclature Committee* (HUGO) é um banco de dados de genes, utilizado para atribuir um único nome e símbolo para mais de 32 mil genes, dos quais mais de 19 mil codificam proteínas. Além do HUGO outros bancos de dados se destacam como: *Biocarta*, *Reactome*, *Kegg*, *Gene Ontology (GO)* e *National Center for Biotechnology Information (NCBI)* que são grandes repositórios de informações moleculares provenientes de vários organismos. Os genes encontrados no banco de dados *Gene Ontology* apresentam evidências físicas inferidas por curadores que extraem informações relevantes que apoiam a participação dos genes em determinadas funções.

Existem também bancos de dados que são usados para armazenar uma grande quantidade de estudos gerados a partir de microarranjos. O banco de dados *Gene Expression Omnibus* (GEO) é um repositório público funcional de dados usado para armazenar transcriptomas. O banco de dados é composto por 866.597 amostras (*samples*) e 10.987 plataformas (*platforms*) divididas em mais de 29 mil estudos (microarranjos) diferentes [GEO 2013]. O GEO é considerado o maior banco de dados de microarranjos e nele estão depositados transcriptomas fabricados por diversas empresas em destaque a *Affymetrix Gene Chip*. A *Affymetrix* é uma empresa fundada em 1992 por Stephen Fodor e tem como objetivo fabricar microarranjos de DNA [Gohlmann and Talloen, 2009].

Entretanto, a falta de um vocabulário unificado para genes, subvias e vias de manutenção do genoma e estabilidade genômica fez com que a ontologia Ontocancro fosse desenvolvida [Librelotto, et al. 2009]. A ontologia é um modelo de dados que representa um conjunto de conceitos dentro de um determinado domínio e os relaciona entre si [Gruber 1993]. O banco de dados da Ontocancro é baseado em uma ontologia que busca organizar e integrar as informações de interatomos e transcriptomas disponíveis a partir dos bancos de dados já citados. Todas as vias implementadas na ontologia Ontocancro são curadas e suas informações mais relevantes são extraídas de diferentes repositórios. Entre estas informações buscadas nos bancos destaca-se a descrição do gene extraída do Banco de dados *String*, o símbolo aprovado pelo HUGO e o EntrezGene, que é um identificador numérico do gene [Simão, et al. 2010].

A Ontocancro 2.0 é uma aplicação *web* que, além de integrar informações de vias e subvias de manutenção do genoma envolvidas no ciclo celular, resposta ao dano do DNA, apoptose e senescência, integra também estudos de microarranjos envolvidos em câncer extraídos do banco de dados GEO. Em sua versão atual, a Ontocancro disponibiliza quatro estudos:

- **GSE10927** (*Human adrenocortical carcinomas, adenomas, and normal*): Contém 33 amostras de carcinomas, 22 amostras de adenomas (tecidos pré-cancerosos), e 10 amostras de tecido normal, de diferentes pacientes. Contendo ensaios de mRNA utilizando o microarranjo *Affymetrix*

HG_U133_plus_2 arrays, formado por 54.675 conjuntos de sondas [Giordano et al. 2009]

- **GSE27155** (*Human thyroid carcinomas, adenomas and normal*): Contém 4 amostras de tecido normal, 17 de tecido folicular de adenoma da tireoide e 13 de tecido folicular de carcinoma. Estas amostras contêm ensaios de mRNA utilizando o microarranjo *Affymetrix* HG_U133A, formado por 22.283 conjuntos de sondas [Giordano et al. 2006].
- **GSE19650** (*Human pancreatic carcinomas, adenomas, and normal*): Contém 6 amostras de carcinomas, 6 adenomas, e 7 de tecido normal, de diferentes pacientes. Contendo ensaios de mRNA utilizando o microarranjo *Affymetrix* HG_U133_plus_2 arrays, formado por 54.675 conjuntos de sondas [Hiraoka et al. 2011]
- **GSE4183** (*Human colon rectal carcinomas, adenomas, inflammations and normal*): Contém 8 amostras de tecido normal, 15 adenomas, 15 carcinoma e 15 com inflamação. Estas amostras contêm ensaios de mRNA utilizando o microarranjo *Affymetrix* HG_U133A, formado por 54.675 conjuntos de sondas [Galamb et al. 2010].

Para que a Ontocancro 2.0 pudesse servir de apoio computacional para a comprovação dos resultados obtidos pelos estudos de Halazonetis, foi desenvolvido um algoritmo para traçar o perfil das vias de estabilidade genômica com base no cálculo da atividade relativa e diversidade relativa, que serão abordados na seção 4 deste artigo.

4. Cálculos de Atividade e Diversidade Relativas

Castro e colaboradores (2007) introduziram um método de análise direto de expressão de vias e subvias. Os cálculos da atividade relativa e da diversidade relativa são usados para avaliar os níveis de expressão de conjuntos de genes e definir os padrões de expressão entre os genes das vias.

Para calcular a atividade relativa de uma dada via α com um número de genes M_α , deve-se somar a expressão dos genes em dois grupos de vias: o primeiro grupo representa as amostras de tecidos alterados ou experimentais N_α^e e o segundo grupo é composto pelas amostras de tecidos normais ou controle N_α^y . Então, a atividade relativa n_α da via α será dada por:

$$n_\alpha = \frac{N_\alpha^e}{N_\alpha^e + N_\alpha^y} \quad (1)$$

O valor de n_α varia entre $0 \leq n_\alpha \leq 1$, se $n_\alpha < 0,5$ a atividade da via com amostra alterada é menor que a atividade do controle, enquanto que $n_\alpha > 0,5$ representa o caso inverso [Castro, M. et al. 2007]. Para caracterizar de forma quantitativa a diversidade para uma via α é utilizada a entropia de Shannon que é descrita como:

$$H_\alpha = - \frac{1}{\ln(M_\alpha)} \sum_i^{M_\alpha} p(i, \alpha) \ln p(i, \alpha) \quad (2)$$

onde M_α é o número de genes na via e $p(i, \alpha)$ é a frequência da diversidade do gene i , dada por:

$$p(i, \alpha) = \frac{s(i, \alpha)}{N_\alpha} \quad (3)$$

com $s(i,\alpha)$ sendo a atividade do gene (i) e N_α a soma da expressão dos genes na via (α). O termo $\ln(M_\alpha)$ é um fator de normalização que garante $0 \leq H_\alpha \leq 1$, desta forma, pode-se comparar as vias com diferentes quantidades de genes. Tendo como referência o sinal controle da amostra, pode-se definir a diversidade relativa (h_α) como:

$$h_\alpha = \frac{H_\alpha^e}{H_\alpha^e + H_\alpha^y} \quad (4)$$

onde H_α^e e H_α^y são as diversidades das amostras com alteração (experimento) e o controle, respectivamente. O valor de h_α varia entre $0 \leq h_\alpha \leq 1$, se $h_\alpha < 0,5$ implica que $H_\alpha^e < H_\alpha^y$, isto é, a diversidade dos valores de expressão dos genes na via é menor para a amostra alterada do que para o controle, enquanto que $h_\alpha > 0,5$ representa o caso inverso. Na Figura 3 os genes das amostras de controle apresentam níveis de expressão com valores mais próximos do que os níveis de expressão dos genes do experimento, isso indica que $H_\alpha^e < H_\alpha^y$ e $h_\alpha < 0,5$. Ou seja, há uma diminuição da diversidade nos níveis de expressão dos genes de experimento em relação ao controle. Para $H_\alpha^e > H_\alpha^y$ tem-se o caso contrário e haverá aumento da diversidade relativa $h_\alpha > 0,5$ do experimento *versus* o controle. Na reprogramação da expressão dos genes poderão ocorrer níveis de aumento ou diminuição da diversidade relativa entre os genes de uma via.

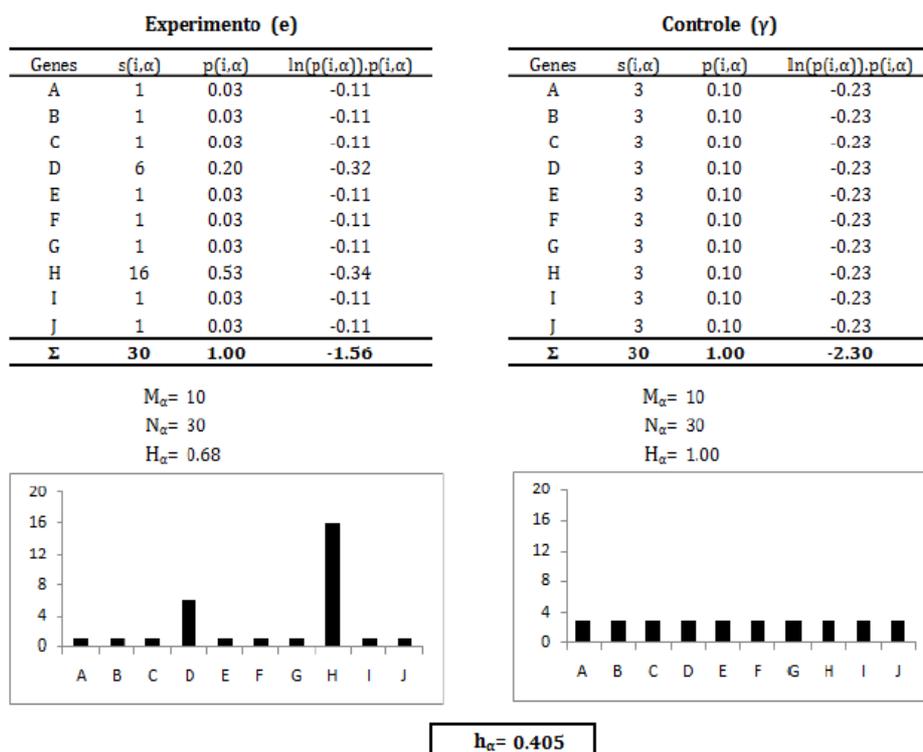


Figura 3 - Exemplo mostrando os níveis de expressão de 10 genes quaisquer, com diminuição da diversidade do experimento em relação ao controle ($H_\alpha^e < H_\alpha^y$). Para este caso a diversidade dos genes de controle é maior do que o experimento, desta forma a diversidade relativa é menor que 0,5 deste modo $h_\alpha = 0,405$. [Simão, 2012]

Para determinar se uma alteração em uma via (atividade) ou em um conjunto de genes (diversidade) é estatisticamente significativa em um determinado estudo (microarranjo), aplica-se o método de *bootstrap*. Esse método é usado para calcular a distribuição amostral de h_α e n_α através de uma análise de reamostragem aleatória que

cofre todos os genes do estudo com repetições que variam de 100 a 100.000 com as mesmas quantidades de genes das vias de interesse para investigar a convergência do valor de p (p -value) da amostra sobre uma distribuição de probabilidades normal.

O nível de significância em uma distribuição de probabilidades é comparado com o valor de p e serve para delimitar se os resultados da atividade ou diversidade relativa de um determinado conjunto de genes são ou não significativos a um determinado nível. Desta forma, se o nível de significância for fixado em 0,05 os valores de p menores que 0,05 ($p < 0,05$) apresentarão aumento ou diminuição significativa de expressão da via em relação ao estudo. Dependendo em qual dos lados da distribuição de probabilidade unicaudal cair o valor de p poderá ter aumento ou diminuição de expressão.

As análises mostradas acima podem ser calculadas usando o *software ViaComplex*. O *software* é um aplicativo usado para construir mapas funcionais de expressão gênica e utiliza a entropia de Shannon para obter um parâmetro quantitativo usado para caracterizar a atividade e a diversidade relativa de vias. O *software* é utilizado para calcular significâncias de vias pelo método de *bootstrap* e pela correção de falsos positivos. Os falsos positivos são usados como método de controle estatístico quando se faz os testes de múltiplas hipóteses (*bootstrap*) para efetuar uma correção nas múltiplas comparações (entre as vias). O código fonte do *software ViaComplex* contendo todas as equações apresentadas acima está disponível em ambiente virtual [Castro, M. A. A. et al. 2009].

5. Estudo de Caso comparando a Ontocancro com o Software ViaComplex

Para este estudo de caso, foram selecionadas amostras de tecidos da tireoide, da série GSE27155, que foram usados para determinar a presença da barreira anti-câncer presente em tecidos pré-cancerosos (adenomas). O conjunto de vias GMM extraídas da base de dados da Ontocancro é formado por duas vias de apoptose (*Death Receptor Signalling; Apoptotic Signaling in Response to DNA Damage*) e duas vias de ciclo celular (*RB Tumor Suppressor/Checkpoint Signaling in Response to DNA Damage; G2/M Checkpoints*).

Nos estudos realizados por Halazonetis e seus colaboradores é possível constatar em tecidos pré-cancerosos, um aumento na atividade das vias de apoptose e uma diminuição da atividade das vias de ciclo celular. Aplicando a abordagem da Ontocancro às amostras de tecidos pré-cancerosos (adenomas) das vias mencionadas, a mesma constatação foi verificada, como indicado no gráfico da Figura 4.

Halazonetis também propõe que em tecidos cancerosos (carcinomas) há um aumento da atividade em vias de ciclo celular e uma diminuição da atividade em vias de apoptose. Este aumento de atividade em vias de ciclo celular dá-se a partir da não ocorrência de apoptose em células alteradas. O organismo então começa a proliferar estas células modificadas aumentando gradativamente o grau de câncer. O que também foi constatado na análise com a ontologia.

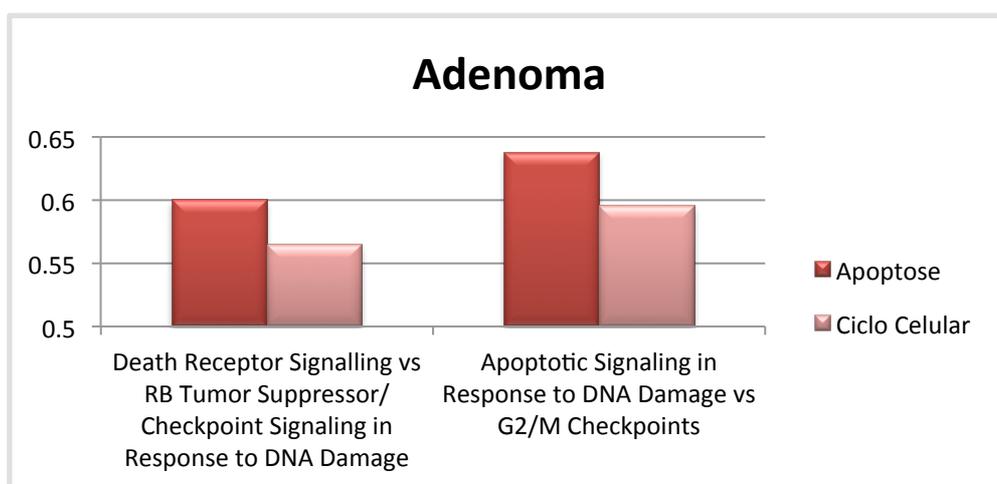


Figura 4 - Gráfico resultante da análise de tecidos pré-cancerosos.

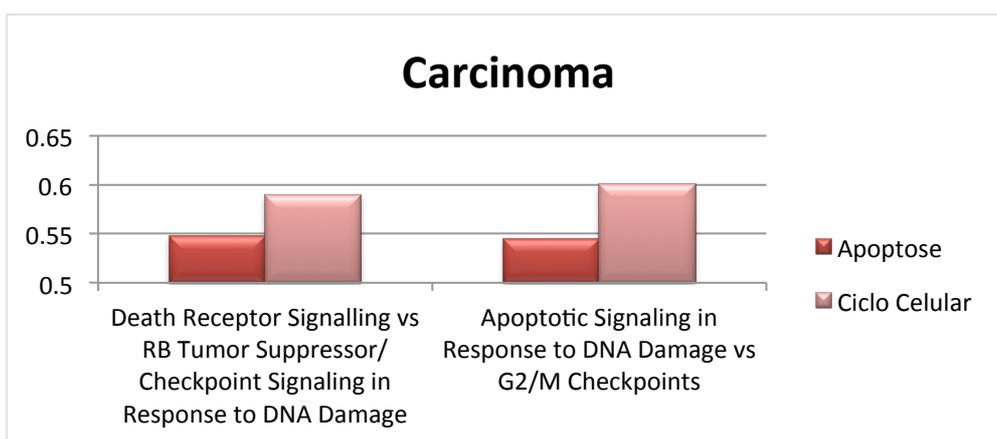


Figura 5 - Gráfico resultante da análise de tecidos cancerosos (carcinomas).

Realizou-se assim, uma comparação de abordagens entre a Ontocancro e a ferramenta *ViaComplex* para inferir o nível de diferenciação entre os métodos utilizados, uma vez que a abordagem desenvolvida pela Ontocancro leva em consideração o identificador EntrezGene (código universal que caracteriza o gene em diversas bases de dados de bioinformática), enquanto a ferramenta *ViaComplex* utiliza como identificador o HugoName, que pode sofrer alterações a medida que são feitos ajustes no nome do gene de acordo com suas características.

A tabela 1 descreve na primeira coluna as vias metabólicas associadas aos mecanismos de manutenção do genoma (GMM): NER (*Nucleotide Excision Repair*), BER (*Base Excision Repair*), MMR (*Mis-match Repair*), HR (*Homologous Recombination*), EJ (*Non-homologous End Joining*) e CS (*Chromosome Stability*); na segunda e na terceira coluna os valores de atividade relativa encontrados em cada abordagem e na última coluna a diferença entre esses valores.

É possível notar que a diferença entre as vias relacionadas com os GMM dá-se somente na terceira casa decimal, ou seja, a diferença é pouco relevante o que permite concluir que a abordagem utilizada pela Ontocancro é válida.

Tabela 1 - Diferença entre abordagens: Ontocancro e ViaComplex [Pereira 2011].

| Vias | Adenoma Ontocancro | Adenoma ViaComplex | Diferença |
|------|--------------------|--------------------|--------------|
| NER | 0.503323378 | 0.5086 | -0.005276622 |
| BER | 0.501151644 | 0.5054 | -0.004248356 |
| MMR | 0.502682944 | 0.5033 | -0.000617056 |
| HR | 0.500873387 | 0.5024 | -0.001526613 |
| EJ | 0.5 | 0.5033 | -0.0033 |
| CS | 0.501154162 | 0.5033 | -0.002145838 |

5. Conclusão

A Ontocancro 2.0 permite a integração de dados relacionados às vias dos Mecanismos de Manutenção do Genoma (GMM), auxiliando na investigação do funcionamento das redes metabólicas de genes envolvidos com o câncer, pois aliado a tais informações disponibiliza um algoritmo para análise estatística dos valores de atividade e diversidade relativa, que são métodos quantitativos de expressão de genes usados na biologia sistêmica. Essa análise quantitativa permite comprovar o modelo qualitativo proposto por [Halazonetis et al. 2008] que descreve uma barreira de evolução tumoral, em tecidos pré-cancerosos, responsável pela ativação das vias metabólicas de Apoptose e de Resposta aos Danos, mas quando rompida dá origem a um tecido maligno.

A partir da aplicação da abordagem utilizada pela Ontocancro 2.0 e dos seus métodos estatísticos é possível traçar um perfil das vias dos mecanismos de manutenção do genoma (GMM) verificando as diferenças encontradas a partir da análise da expressão de genes. Este artigo propõe demonstrar como a Ontocancro pode comprovar quantitativamente os resultados obtidos pelos estudos de Halazonetis.

Como trabalhos futuros, estão sendo avaliados outros métodos de análise de dados obtidos com a técnica de sequenciamento RNASeq, da empresa Illumina, para comparar ao método utilizado até o momento, que se baseia nos dados disponibilizados pela empresa Affymetrix.

Referencias

- Castro, M. A. A., Filho, J. L. R., Dalmolin, R. J. S., et al. (2009). ViaComplex: software for landscape analysis of gene expression networks in genomic context. *Bioinformatics (Oxford, England)*, v. 25, n. 11, p. 1468–9.
- Castro, M. A. A., Mombach, J. C. M., De Almeida, R. M. C. and Moreira, J. C. F. (2007). Impaired expression of NER gene network in sporadic solid tumors. *Nucleic acids research*, v. 35, n. 6, p. 1859–67.
- Galamb, O., Spisák, S., Sipos, F., et al. (2010). Reversal of gene expression changes in the colorectal normal-adenoma pathway by NS398 selective COX2 inhibitor. *British journal of cancer*, v. 102, n. 4, p. 765–73.

- GEO (2013). Gene Expression Omnibus. Disponível em <http://www.ncbi.nlm.nih.gov/geo/>
- Giordano, T. J., Au, A. Y. M., Kuick, R., et al. (2006). Delineation, functional validation, and bioinformatic evaluation of gene expression in thyroid follicular carcinomas with the PAX8-PPARG translocation. *Clinical cancer research*, v. 12, n. 7 Pt 1, p. 1983–93.
- Giordano, T. J., Kuick, R., Else, T., et al. (2009). Molecular classification and prognostication of adrenocortical tumors by transcriptome profiling. *Clinical cancer research*, v. 15, n. 2, p. 668–76.
- Gohlmann, H. and Talloen, W. (2009). *Gene Expression Studies Using Affymetrix Microarrays*. New York: Chapman and Hall/CRC. p. 359
- Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. [N. Guarino & R. Poli, Eds.]In *Formal ontology in conceptual analysis and knowledge representation*. . Kluwer Academic.
- Halazonetis, T. D., Gorgoulis, V. G. and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science (New York, N.Y.)*, v. 319, n. 5868, p. 1352–55.
- Hiraoka, N., Yamazaki-Itoh, R., Ino, Y., et al. (2011). CXCL17 and ICAM2 are associated with a potential anti-tumor immune response in early intraepithelial stages of human pancreatic carcinogenesis. *Gastroenterology*, v. 140, n. 1, p. 310–21.
- Hood, L.; Heath, J. R.; Phelps, M. E.; Lin, B. *Systems Biology and New Technologies Enable Predictive and Preventative Medicine*, Science, v.306, p. 640-43, 2004
- Librelotto, G. R., Mombach, J. C., Sinigaglia, M., et al. (2009). *An Ontology to Integrate Transcriptomics and Interatomics Data Involved in Gene Pathways of Genome Stability*. Springer Berlin Heidelberg. v. 5676p. 164–167
- Loeb, L. A. (2001). A mutator phenotype in cancer. *Cancer research*, v. 61, n. 8, p. 3230–9.
- Pereira, R. T. (2011). Desenvolvimento de uma Ferramenta para a Análise de Vias de Estabilidade Genômica.
- Pereira, R. T. (2012). Utilizando a Ontocancro para traçar o perfil das vias de manutenção da estabilidade genômica. WIM 2012. Disponível em http://www.imago.ufpr.br/csbc2012/anais_csbc/eventos/wim/index.html
- Simão, É. M., Cabral, H. B., Castro, M. A. A., et al. (2010). Modeling the Human Genome Maintenance network. *Physica A: Statistical Mechanics and its Applications*, v. 389, n. 19, p. 4188–4194.
- Simão, É. M. (2012). Dinâmica da Transição Pré-Câncer para Câncer: Estudo da Expressão de Vias de Manutenção do Genoma. Universidade Federal de Santa Maria.
- Watson, J. D. (1990). The human genome project: past, present, and future. *Science (New York, N.Y.)*, v. 248, n. 4951, p. 44–9.