

Utilizando a Ontocancro para traçar o perfil das vias de manutenção da estabilidade genômica

Rafael T. Pereira, Renato P. de Azevedo, Eder M. Simão,
José C.M. Mombach e Giovani R. Librelotto

¹UFSM - Universidade Federal de Santa Maria
Grupo de Sistemas Complexos - CCNE
Programa de Pós-Graduação em Informática - PPGI
Av. Roraima, 1000, 97105-900, Brazil

{rafatp, rparenato}@gmail.com, librelotto@inf.ufsm.br

Abstract. *The Ontocancro project provides a repository of cancer related information where researchers can find interactome and transcriptome data related to the metabolic pathways of the genome maintaining mechanisms from several distinct databases. In this paper not only we describe Ontocancro but we also present an approach to characterise the profile of the metabolic pathways based on the Shannon's Entropy. Such parameter provides a quantitative measure of the activity and diversity from the metabolic pathways.*

Resumo. *O projeto Ontocancro surgiu da necessidade de reunir informações relacionadas ao câncer encontradas em diferentes bases de dados e integrá-las em um único repositório, fornecendo aos pesquisadores dados de interactoma e transcriptoma relacionados as vias metabólicas dos mecanismos de manutenção do genoma (GMM). Além do cruzamento destas informações, este artigo aborda uma metodologia para analisar o perfil destas vias metabólicas utilizando a entropia de Shannon, onde é possível obter um parâmetro quantitativo utilizado para caracterizar a diversidade e a atividade relativa de vias metabólicas.*

1. Introdução

Atualmente, um dos principais temas abordados na Biologia é a compreensão da estrutura e do comportamento de redes complexas de interações moleculares, as quais controlam o comportamento das células [Barabási and Oltvai 2004]. O tamanho e a complexidade dos dados biológicos coletados durante os últimos anos incluem informações que requerem uma abordagem integradora [Golemis 2002], o que força os cientistas a procurarem soluções cada vez mais inovadoras para tratar destes dados. No entanto, esta tarefa se torna difícil à medida que o pesquisador necessita buscar informações em diferentes bases de dados ou então trabalhar com diferentes tipos de dados ao mesmo tempo, o que caracteriza uma das dificuldades encontradas na área da bioinformática atualmente.

O objetivo deste artigo é demonstrar a composição de uma base de dados utilizando uma ontologia que possa permitir a integração de informações relacionadas a transcriptomas e interactomas dos genes que compõem as vias dos mecanismos de manutenção genômica (GMM).

Disfunções nas GMM podem originar câncer [Castro et al. 2009a], disfunções dos mecanismos que regulam o ciclo celular, que realizam o reparo do DNA ou, que são responsáveis pela morte celular programada (apoptose), resultam numa instabilidade genômica e conseqüentemente em desenvolvimento do câncer. Para ser possível o estudo deste tipo de acontecimento é necessário o desenvolvimento de novos métodos, como uso de ontologias e ferramentas de bioinformática que viabilizem o estudo em larga escala de dados moleculares.

As ontologias são cruciais para a manutenção da coerência dos dados em uma coleção de conceitos complexos e seus relacionamentos, permitindo uma especificação explícita da contextualização [Gruber 1993]. A ontologia Ontocancro tem por objectivo auxiliar a investigação do funcionamento das redes metabólicas de genes envolvidos com o câncer [Librelotto et al. 2006].

É proposto então neste artigo a atualização da ontologia, criando a Ontocancro 2.0, de forma a fornecer ao utilizador uma análise estatística do perfil das vias metabólicas relacionadas diretamente com o câncer através da utilização de informações de transcriptomas, o que não estava sendo contemplado na sua versão inicial.

Através dessa nova ontologia, a extração e seleção das informações relativas aos transcriptomas, serão provenientes da base de dados biológica do *Gene Expression Omnibus* (GEO) onde é possível obter dados sobre experimentos realizados através da técnica de microarranjos em amostras de tecidos. O Geo é uma base de dados biológica que possui dados de amostras de vários organismos que apresentam diversas anomalias tais como amostras de câncer e adenoma presentes em seres humanos [Barret and Edgar 2006].

Esta técnica permite suportar o desenvolvimento de uma aplicação *web* que, através da Ontocancro, acede as listas de transcriptomas e interactomas lá presentes e, a partir disto, realiza uma análise estatística, podendo assim verificar o perfil dos GMM em diversos transcriptomas relacionados com o desenvolvimento do câncer. Portanto, a partir da integração dos dados do GEO com a Ontocancro, permite-se descobrir quais vias estão mais ativas quando o câncer está na sua fase inicial.

O presente artigo está dividido da seguinte maneira: Na Seção 2 pode-se encontrar a contextualização e a descrição do projeto Ontocancro, a Seção 2.1 traz o objetivo e explanação sobre a área biológica do projeto, bem como o que motivou a atualização da ontologia e também a atual estrutura da base de dados. A Seção 3 aborda os dados que serão analisados, para que seja possível então, na 3.1 e na Subseção 3.2, através da utilização de métodos matemáticos, traçar um perfil da vias envolvidas nos mecanismos de manutenção do genoma. A Seção 4 traz os resultados alcançados utilizando este tipo de abordagem e por fim a conclusão na Seção 5.

2. A Ontocancro

A Ontocancro é um banco de dados que integra informações de anotação transcriptômica e interatômica de genes envolvidos em vias de manutenção e estabilidade do genoma humano. Ele possui informações obtidas dos bancos NCI-Nature, Gene Ontology (GO), String, Reactome, KEGG, BioCarta, HUGO e UniProt. Esta ferramenta é única atualmente e cobre de maneira mais completa a rede GMM [Sinigaglia et al. 2010].

Em sua versão original, a Ontocancro 1.0 possui em torno de 120 vias genéticas

e informações detalhadas de anotação, parceiros de interação, sondas de microarrays e outras de 1428 genes envolvidos em GMM. A partir de estudos realizados sobre os dados contidos na Ontocancro 1.0, surgiu a necessidade de uma atualização em sua base de dados e a inclusão de novas informações.

Nas vias contidas na Ontocancro 1.0, existem um número muito grande de genes, o que dificulta a análise da expressão dos mesmos, resultando em distorções dos dados analisados, isto porque as vias nem sempre sofriam um processo de curagem nas bases de dados que as mantinham.

Para resolver o problema da grande quantidade de genes presentes em uma mesma via, a estratégia encontrada foi a separação de vias em sub-vias de forma a agregar genes que possuam uma maior proximidade dentro de uma mesma via. Decidiu-se por incluir também na base de dados somente vias que passaram por uma curagem, aumentando assim a fidelidade da análise do perfil das mesmas [Éder M. Simão et al. 2010].

Na composição da Ontocancro 2.0 os dados provenientes das vias metabólicas, quase que em seu total, foram retirados da base de dados NCI-Nature (*National Cancer Institute*) pois neste repositório é curado e nele encontram-se informações sobre vias, interações biomoleculares e processos celulares montados de maneira mais confiável e sustenta uma abordagem para a análise das vias que contribuem para a identificação de alvos moleculares que poderão ser usados em terapia contra o câncer [Schaefer 2006].

2.1. A Ontocancro 2.0 e sua ontologia atualizada

Segundo Halazonetis (2008), os tecidos pré-cancerosos, de um modo geral, apresentam uma diminuição da expressão em vias relacionadas ao índice de proliferação celular (ciclo celular), seguido por um aumento da atividade da via de apoptose e reparo de danos no DNA. Com o aumento da atividade da via de apoptose e reparo do DNA, é definida a barreira de progressão ao câncer. Uma vez que esta barreira é quebrada, há o predomínio do fenótipo mutante, que resulta no acúmulo de mutações em células geneticamente alteradas. Com tal embasamento teórico, o estudo está focado somente nas vias que compreendem os GMM, sendo que todas as vias contidas na base de dados da Ontocancro 2.0 estão ligadas diretamente à formação do câncer.

Sabendo disso, optou-se por incluir na Ontocancro 2.0 apenas as vias que estão envolvidas incisivamente neste processo. A apoptose, o ciclo celular e o reparo por dano ao DNA constituem as sub-vias que dão sustentabilidade aos GMM e são estes os mecanismos que auxiliam na barreira contra propagação do câncer, ou seja, estas sub-vias serão alvos diretamente da análise do perfil de cânceres nas vias que compõem a Ontocancro 2.0.

Ontologias são conhecidas pelo seu poder de representação de um universo de domínio. Grandes benefícios são alcançados, sobre tudo, pelo reuso do conhecimento de domínio.

Uma vez que o domínio da ontologia a ser descrita já era conhecido (Ontocancro 1.0), partiu-se do princípio de utilizá-la para a criação de definição dos novos conceitos e relacionamentos que permitissem a representação de informações relacionadas ao câncer. Para o desenvolvimento da Ontologia fez-se o uso da ferramenta *Protégé*, com licença *open source* disponibilizada pela Universidade de Stanford (*em:*

<http://protege.stanford.edu>).

A ontologia utilizada para montar a Ontocancro 1.0 era composta basicamente por genes e vias (*pathways*), definindo interações entre estas duas entidades. No entanto, de forma a adaptar a estrutura da ontologia de acordo com as modificações realizadas se faz necessária a criação de três novas entidades: sub-vias (*sub-pathway*), amostras (*samples*) e séries (*series*). Cria-se assim uma relação entre os transcriptomas (doenças relacionadas ao câncer) e interactomas (vias). A Figura 1 apresenta o grafo da ontologia atual.

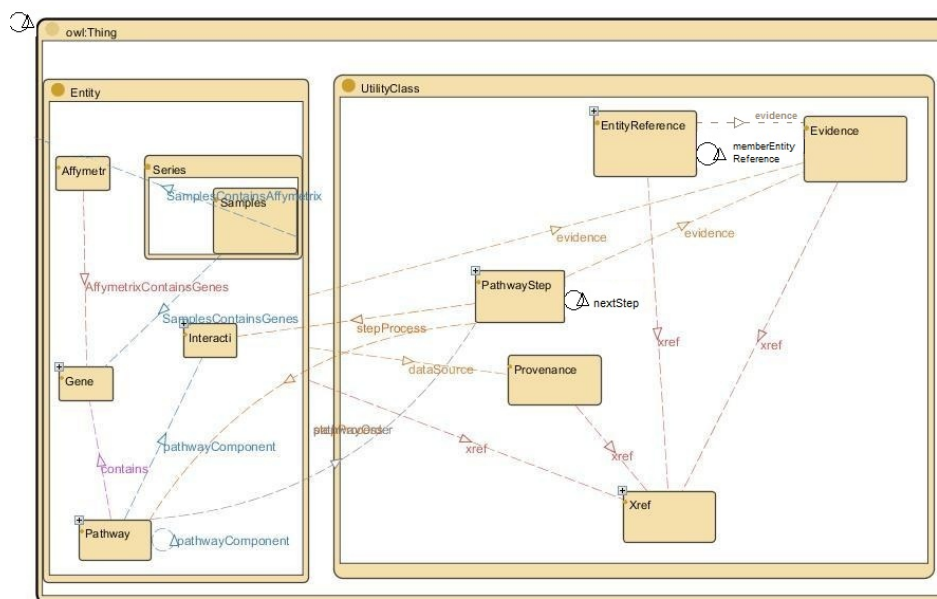


Figura 1. Grafo gerado da ontologia aplicada para a Ontocancro 2.0

Após a ontologia estar devidamente descrita, o próximo passo é realizar a importação dos dados, utilizando a base de dados do GEO. Este repositório tem servido como uma fonte bastante confiável e diversificada de informações sobre experimentos em microarranjos, e dados baseados na genômica funcional, disponibilizando aos pesquisadores maneiras de importação destas informações e ferramentas para o auxílio de análises dos dados [Barret and Edgar 2006].

A partir da modelagem da ontologia, a base de dados da Ontocancro teve que ser remodelada para que os dados das sub-vias, series e amostras pudessem ser inseridos. Nesta etapa utilizou-se como ferramenta de modelação da base de dados o MySQL Server 5.5 (disponível em: <http://dev.mysql.com>)

Para a inserção destes experimentos na base de dados da Ontocancro 2.0, é de suma importância que as amostras contivessem experimentos realizados com tecidos humanos que apresentem alguma indicação de câncer e amostras deste mesmo tecido sem a presença dele. Considera-se que tecidos que apresentem alguma alteração (que indique a possibilidade de haver câncer) são chamados de “experimento” e tecidos normais de “controle”.

3. Análise da expressão dos genes em relação às vias da Ontocancro 2.0

A partir do momento que a base de dados da Ontocancro possui todas as informações pertinentes às doenças a serem estudadas e as vias de manutenção do genoma, o próximo passo é realizar o estudo estatístico que integra estes dois tipos de informação de forma que seja possível traçar o perfil destas vias, relacionado-as com doenças envolvidas no câncer.

Inicialmente, para se calcular o valor da atividade, é necessário fazer a escolha da via que será analisada e também da doença que será relacionada a esta via. As opções de doenças contidas na base até o momento são:

- GSE4183, contendo amostras de tecidos com câncer e normais retiradas do cólon,
- GSE10927, relacionada com o tumor adrenocortical,
- GSE27155, refere-se ao câncer da tireoide e
- GSE19650, contendo amostras de tecidos de células epiteliais durante o processo de formação do câncer pancreático.

Estas doenças foram escolhidas devido ao fato de possuírem amostras de tecidos com câncer e de amostras com tecidos normais.

A análise será realizada através da entropia de Shannon, que permite obter um parâmetro quantitativo usado para caracterizar a diversidade e a atividade relativa de vias metabólicas [Castro et al. 2007]. Para isso, é essencial realizar um processo de filtragem nas amostras a serem avaliadas, pois nem todas apresentam um nível de relevância considerável em seu sinal de expressão dentro do microarranjo, nesta fase também foram selecionadas amostras pertencentes a plataforma Affymetrix que possuem grandes quantidades de sondas ligadas aos genes envolvidos em GMM. Além disso, deve-se levar em consideração também que poderá haver mais de uma sonda para um determinado gene e estes podem se repetir dentro de uma mesma amostra. A seguir serão descritos as etapas para se realizar esta análise.

3.1. Cálculo da atividade

É de suma importância levar em consideração que as sondas podem possuir genes repetidos em algumas amostras. Para contornar este problema, é calculada a média aritmética do valor de expressão de cada gene que se encontra repetido na amostra. Com isso cada gene presente na amostra possui um único valor de expressão.

A partir dos valores de expressão de todos os genes que compõem a amostra realiza-se a média aritmética. O resultado desse cálculo representa o valor da atividade dos genes dentro da amostra. Após o cálculo das médias dos valores de expressão dos genes, é realizado o cálculo da atividade para cada amostra.

Para amostras com tecidos identificados como câncer, o cálculo da atividade é obtido desta forma:

$$N_a^e = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

Onde N_a^e , refere-se ao cálculo da atividade em tecidos com câncer, n é o número total de genes da amostra, e X_i representa os diferentes genes analisados apresentando indícios

de tecidos cancerosos. Para amostras com tecidos identificados como normal.

$$N_a^y = \frac{\sum_{i=1}^n Y_i}{n} \quad (2)$$

Onde N_a^y significa o cálculo da atividade em tecidos normais, n é o número total de genes da amostra, e Y_i representa os diferentes genes identificados como sendo de tecidos normais.

O resultado desta etapa é a média aritmética de cada tipo de amostra separando-as em amostras normais e com câncer, essa média é chamada de cálculo da atividade e será usada para o passo seguinte.

3.2. Cálculo da atividade relativa e da diversidade

Para realizar a análise estatística das interações funcionais entre as amostras e as vias, utilizou-se a entropia de Shannon [Shannon 1948]. Através dessa entropia é possível obter um parâmetro quantitativo usado para caracterizar a diversidade e a atividade relativa de vias [Sinigaglia et al. 2010].

A entropia de Shannon é definida pela seguinte fórmula:

$$H_a = -\frac{1}{\ln(M_a)} \sum_l^{M_a} p(i, a) \ln p(i, a) \quad (3)$$

Onde M_a é o número de genes na via e $p(i, a)$ é a contribuição da probabilidade de cada gene dentro da amostra, dada por:

$$p(i, a) = \frac{s(i, a)}{N_a} \quad (4)$$

Com $s(i, a)$ sendo a atividade do gene e N_a a soma da atividade na via. O termo $\ln(M_a)$ é um fator de normalização que garante $0 \leq H_a \leq 1$. Assim pode-se comparar vias compostas por diversas quantidades de genes.

Levando em consideração como referência o sinal de controle da amostra, defini-se então a diversidade relativa:

$$h_a = \frac{H_a^e}{H_a^e + H_a^y} \quad (5)$$

Onde H_a^e refere-se as diversidades de amostras com alteração (câncer) e H_a^y as amostras consideradas como controle. É válido observar se $0 \leq h_a \leq 1$, e $h_a < 0.5$ implica que $H_a^e < H_a^y$; ou seja, a distribuição do sinal dos genes na via é menor para a amostra alterada do que para o controle, enquanto que $h_a > 0.5$ representa o caso inverso.

Uma abordagem pode ser utilizada para realizar o cálculo da atividade relativa da expressão da via n_a :

$$n_a = \frac{N_a^e}{N_a^e + N_a^y} \quad (6)$$

Com N_a^e referindo-se a soma das atividades das amostras identificadas com alteração e N_a^y a soma das amostras identificadas como controle. Então, se $0 \leq n_a \leq 1$, e $n_a <$

0.5 implica-se que $N_a^e < N_a^y$; sendo assim, a via com amostra alterada é menor que a atividade do controle, enquanto que $n_a > 0.5$ representa o oposto.

O cálculo da diversidade envolve um outro fator do qual o *SQL* não consegue tratar. O fato da fórmula da diversidade ser composta por vários tipos de operadores. Para contornar este problema desenvolveu-se uma função determinística em *SQL*. Funções determinísticas sempre retornam o mesmo tipo de valor quando são chamadas por um conjunto específico de valores de entrada e dado o mesmo estado da base de dados. Existem várias propriedades de funções que podem ser determinadas pelo usuário para indexar os resultados da função [Microsoft 2012].

Esta função se encarrega de realizar o somatório de todas as médias do valor de expressão das amostras de cada gene contido nas vias. O resultado retornado por essa função diz respeito ao elemento N_a referenciado na fórmula (4) conhecido pelo valor da atividade da via. Assim tendo o número total de genes na via, o valor de expressão de cada gene e atividade relativa, aplicou-se a fórmula (3) e obteve-se como resultado a diversidade da via para cada tipo de amostra. Para o cálculo da diversidade relativa fez-se o uso da fórmula (5) obtendo os valores referentes a mesma.

Existe uma aplicação de código aberto chamada *ViaComplex* [Castro et al. 2009b] onde é possível a construção de mapas metabólicos a partir de expressões gênicas, além disso ela fornece também uma ferramenta generalizada para avaliar estas redes baseada na entropia de Shannon.

4. Resultados

Como caso de estudo, optou-se por trabalhar com amostras de tecidos da tireoide retiradas de diferentes pacientes, correspondentes a experimentos contendo 66.849, ensaios contendo: amostras normais, de adenomas e carcinomas. Estas amostras contêm dados de mRNA extraídos através da utilização do microarranjo *Affymetrix HG_U133A*, formado por 22.283 conjunto de sondas [Giordano et al. 2005]. O transcriptoma em causa foi extraído a partir da base de dados do GEO e importado para o banco de dados da Ontocancro onde será realizada a análise.

O conjunto de vias extraídas da base de dados da Ontocancro, para este estudo de caso, é formado por duas vias de apoptose (*Death Signalling; Apoptotic Signaling in Response to DNA Damage*) e duas vias de ciclo celular (*rb Tumor Suppressor/Checkp. Sig. in Response to DNA Damage; G2/M Checkpoints*).

Halazonetis e seus colaboradores [Halazonetis 2008] propuseram um modelo baseado em estudos de tecidos pré-cancerosos onde nestes tecidos há um aumento de atividade nas vias de reparo em danos de DNA e morte celular programada (apoptose), vias que estabelecem uma barreira contra a progressão do câncer.

O caso de estudo descrito a seguir, serve para caracterizar alterações nestas vias, comparando a análise de expressão em transcriptomas de tecidos pré-cancerosos e cancerosos. O estudo realizado com a ferramenta Ontocancro pode ser visto na Figura 2, onde é possível comprovar os estudos realizados por Halazonetis onde sugere que em tecidos pré-cancerosos há um aumento na atividade das vias de apoptose e uma diminuição da atividade das vias de ciclo celular, mostrando que a abordagem utilizada pela Ontocancro 2.0 é condizente.

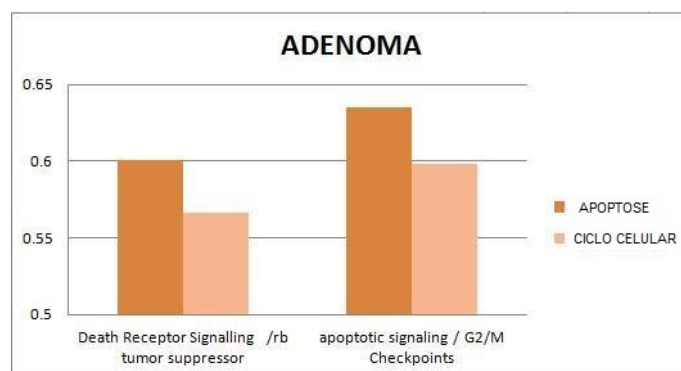


Figura 2. Gráfico resultante da análise de tecidos pré-cancerosos

Halazonetis também propõe que em amostras de tecidos que contenha o câncer, há um aumento da atividade em vias de ciclo celular e uma diminuição da atividade em vias de apoptose. A metodologia aplicada ao caso de estudo reforça tal teoria descrita anteriormente, o que pode ser visto na Figura 3.

Este aumento de atividade em vias de ciclo celular dá-se a partir da não ocorrência de apoptose em células alteradas (com câncer). O organismo então começa a proliferar estas células modificadas aumentando gradativamente o grau do câncer [Castro et al. 2007]

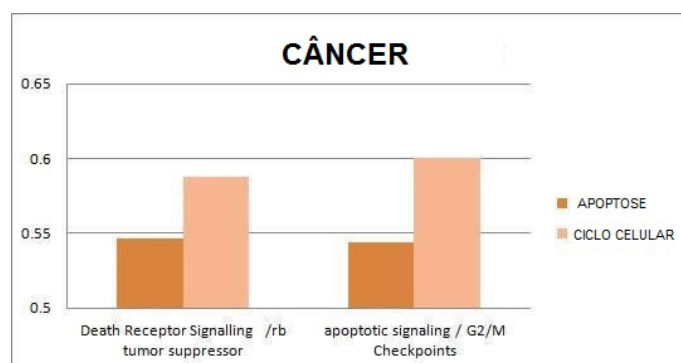


Figura 3. Gráfico resultante da análise de tecidos cancerosos

5. Conclusão

Em seu princípio a Ontocancro era um repositório contendo informações sobre genes e vias metabólicas, que fornecia informações relativas aos genes que são mais ativos na fase inicial do câncer e a que vias metabólicas eles pertenciam. Porém, não havia grande interação com o utilizador; apenas era possível consultar a lista de genes e vias metabólicas que se encontravam nesta base, limitando bastante o uso desta ferramenta.

Desta forma, este artigo proporciona ao projeto uma maior interação com o utilizador, devido a inserção de informações provenientes de experimentos relacionados ao câncer e a análise do perfil das vias ligadas diretamente a esta doença. Para isso, realizou-se então uma atualização na Ontocancro, inserindo informação que eram de fundamental importância e realizando o relacionamento entre estas.

Os cálculos retirados da bibliografia foram, portanto, estudados e implementados. Percebeu-se que através da entropia de Shannon obtém-se um parâmetro quantitativo que permite caracterizar a diversidade e a atividade relativa das vias.

Admite-se também que é possível, através da integração de dados referentes a transcriptoma e interactoma, traçar um perfil das vias metabólicas envolvidas nos mecanismos de manutenção do genoma.

A partir da análise do perfil da vias de estabilidade genômica é possível perceber que as mesmas apresentam desempenhos diferenciados em transcriptomas de adenomas e câncer. Nos estudos com os microarranjos de adenoma em tireoide e câncer de tireoide foi identificado através dos cálculos de atividade relativa e diversidade evidências do que Halazonetis propôs para a barreira de progressão ao câncer [Halazonetis 2008].

Conclui-se também que as vias de estabilidade genômica apresentam diferentes comportamentos em doenças que envolvem alterações genéticas. O estudo do transcriptoma apresentando amostras com tecidos pré-cancerosos e amostras de tecidos com câncer revela alterações nas vias de apoptose e ciclo celular. Para identificar-se transcriptomas de doenças como o câncer, adenoma e doença relacionadas a instabilidade do genoma, pode-se considerar diferenças nas vias GMM. Em câncer as vias de apoptose e ciclo celular deverão aparecer com o seu valor aumentado, em adenomas é possível observar também um aumento do valor em vias de ciclo celular.

Como trabalho futuro a partir deste artigo, pretende-se também integrar a essa nova base de dados, informações referentes à estudos com o novo protocolo de sequenciamento conhecido como RNA-Seq, o qual faz uso do mRNA para gerar milhares fragmentos de sequência em uma única operação. Esses fragmentos, podem ser usados para medir o nível de expressão dos genes e identificar novas variantes de ligações dos genes [Trapnell et al. 2009].

References

- Barabási, A. and Oltvai, Z. (2004). Network biology: understanding the cells functional organization. *Nature Reviews Genetics*, pages 101–113.
- Barret, T. and Edgar, R. (2006). Mining microarray data at ncbi's gene expression omnibus (geo). *National Institutes of Health*, 338:175–190.
- Castro, M. A. A., Filho, J. L. R., Dalmolin, R. J. S., Sinigaglia, M., Moreira, J. C. F., Mombach, J. C. M., and de Almeida, R. M. C. (2009a). Viacomplex: software for landscape analysis of gene expression networks in genomic context. *Bioinformatics*, 25:1468–1469.
- Castro, M. A. A., Filho, J. L. R., Dalmolin, R. J. S., Sinigaglia, M., Moreira, J. C. F., Mombach, J. C. M., and de Almeida, R. M. C. (2009b). Viacomplex: software for landscape analysis of gene expression networks in genomic context. *Bioinformatics*, 25:1468–1469.
- 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*, 35:1859–1867.

- Éder M. Simão, Cabral, H. B., Castro, M. A. A., Sinigaglia, M., Mombach, J. C. M., and Librelotto, G. R. (2010). Modeling the human genome maintenance network. *Physica A*, 389:4188–4194.
- Giordano, T. J., Kuick, R., Thomas, D. G., Misek, D. E., Vinco, M., Sanders, D., Zhu, Z., Ciampi, R., Roh, M., Shedden, K., Gauger, P., Doherty, G., Thompson, N. W., Hanash, S., Koenig, R. J., and Nikiforov, Y. E. (2005). Molecular classification of papillary thyroid carcinoma: distinct *BRAF*, *RAS*, and *RET—PTC* mutation-specific gene expression profiles discovered by dna microarray analysis. *Oncogene*, 24:6646–6656.
- Golemis, E. (2002). *Protein-Protein Interactions - A Molecular Cloning Manual*, pages 623–646. Cold Spring Harbor Laboratory Press, New York.
- Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. In *In Formal Ontology in Conceptual Analysis and Knowledge Representation*, Kluwer Academic Publishers, In Press. Substantial revision of paper presented at THE INTERNATIONAL WORKSHOP ON FORMAL ONTOLOGY. Kluwer Academic Publishers.
- Halazonetis, T. D. (2008). An oncogene-induced dna damage model for cancer development. *Science*, 319:1352–1355.
- Librelotto, G. R., Ramalho, J. C., and Henriques, P. R. (2006). Metamorphosis: A Topic Maps Based Environment to Handle Heterogeneous Information Resources. In *Lecture Notes in Computer Science*, volume 3873, pages 14–25. Springer-Verlag GmbH.
- Microsoft (2012). Microsoft developer network - virtual library. <http://msdn.microsoft.com/pt-br/library/ms178091.aspx>.
- Schaefer, C. F. (2006). An introduction to the nci pathway interaction database. *NCI-Nature Pathway Interaction Database*.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656.
- Sinigaglia, M., Filho, J. L. R., Dalmolin, R. J. S., Moreira, J. C. F., de Almeida, R. M. C., Castro, M. A. A., Mombach, J. C. M., and Librelotto, G. R. (2010). Bioinformatics analysis of gene networks involved in genomic stability and cancer. *Cancer Biomarkers*, 13:34–53.
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). Tophat: discovering splice junctions with rna-seq. *Bioinformatics*, 25:1105–1111.