

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**INTEGRANDO NOVAS FUNÇÕES AO SOFTWARE
MEDUSA DE ANÁLISE GÊNICA EM VIAS
ONTOCANCRO NO ESTUDO DO CÂNCER**

TRABALHO DE GRADUAÇÃO

JÔNATHAN LUIZ DA SILVEIRA ALVES

**Santa Maria, RS, Brasil
2010**

**INTEGRANDO NOVAS FUNÇÕES AO SOFTWARE
MEDUSA DE ANÁLISE GÊNICA EM VIAS ONTOCANCRO
NO ESTUDO DO CÂNCER**

por

Jônathan Luiz da Silveira Alves

Trabalho de Graduação apresentado ao Curso de Ciência da
Computação da Universidade Federal de Santa Maria (UFSM, RS),
como requisito parcial para a obtenção do grau de
Bacharel em Ciência da Computação

Orientador: Prof. Dr. Giovani Rubert Librelotto

**Trabalho de Graduação N° 291
Santa Maria, RS, Brasil
2010**

**Universidade Federal de Santa Maria
Centro de Tecnologia
Curso de Ciência da Computação**

A Comissão Examinadora, abaixo assinada,
aprova o Trabalho de Graduação

**INTEGRANDO NOVAS FUNÇÕES AO SOFTWARE
MEDUSA DE ANÁLISE GÊNICA EM VIAS ONTOCANCRO
NO ESTUDO DO CÂNCER**

elaborado por
Jônathan Luiz da Silveira Alves

como requisito parcial para obtenção do grau de
Bacharel em Ciência da Computação

COMISSÃO EXAMINADORA:

Giovani Rubert Librelotto, Dr.
(Presidente/Orientador)

Juliana Kaizer Vizzotto, Dr^a. (UFSM)

José Carlos Merino Mombach, Dr. (UFSM)

Santa Maria, 04 de janeiro de 2010

Dedico este trabalho aos meus pais, Sônia da Silveira Alves e Honório Luiz Alves, que sempre confiaram e apoiaram todas as minhas decisões.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela oportunidade de estar no mundo, pelas conquistas e sonhos realizados.

Aos meus pais, Sônia e Honório Alves, por todo amor e confiança, sem eles nada disso seria possível. E também à minha irmã Jenniffer e meu irmão Júnior, que sempre estiveram do meu lado.

À minha namorada Rossana Schmitt, pelas palavras de incentivo nas horas difíceis e pelo amor e companheirismo constante.

Ao professor Giovani Rubert Librelotto, por ter aceitado o desafio de ser meu orientador neste trabalho, assim como sua amizade.

Ao grande amigo e profissional Éder Maiquel Simão, pelo apoio e suporte durante toda esta jornada de pesquisa.

Por fim, a todos meus verdadeiros amigos e aos colegas de curso, especialmente ao Diego Friedrich e Juliano da Costa, que torceram por mim e estiveram presentes nesta caminhada.

"A dor é temporária, pode durar um minuto, uma hora, um dia, ou um ano, mas em algum momento ela passa e outra sensação toma o seu lugar. Mas se eu desistir, a dor fica para sempre."

Lance Armstrong

RESUMO

Trabalho de Graduação
Curso de Ciência da Computação
Universidade Federal de Santa Maria

INTEGRANDO NOVAS FUNÇÕES AO SOFTWARE MEDUSA DE ANÁLISE GÊNICA EM VIAS ONTOCANCRO NO ESTUDO DO CÂNCER

Autor: Jônathan Luiz da Silveira Alves

Orientador: Prof. Dr. Giovani Rubert Librelotto

Local e data da defesa: Santa Maria, 04 de janeiro de 2010

Visando contribuir com as pesquisas no estudo do câncer o projeto Ontocancro, da área de bioinformática, reuniu informações relativas a genes envolvidos no processo carcinogênico, dos bancos de dados de genes NCI-Nature, BioCarta, KEGG, Reactome, Prosite, GO e STRING. Sendo assim a base de dados do projeto é constantemente acessada por pesquisadores da área. Este trabalho tem como objetivo programar uma ferramenta que facilite a busca destas informações no banco de dados do projeto, de modo que demande menos tempo ao pesquisador localizando as informações. Outra ferramenta muito utilizada pelos profissionais desta área é o software de código aberto Medusa, usado para visualizar as redes de interação de proteínas, a partir de dados obtidos pelos cientistas. A ferramenta gera um grafo que mostra as interações entre as proteínas, além de outras informações sobre cada proteína da rede. Por isso este trabalho visa programar novos procedimentos a esta ferramenta, integrando ela à base de dados do Ontocancro, agregando assim novas funções que facilitam ainda mais a busca e identificação das vias envolvidas nas amostras.

Palavras-chave: câncer; bioinformática; vias; Ontocancro; Software Medusa.

ABSTRACT

Undergraduation Final Work
Undergraduation in Computer Science
Federal University of Santa Maria

INTEGRATING NEW FUNCTIONALITIES TO THE SOFTWARE MEDUSA FOR GENIC ANALYSIS IN ONTOCANCRO PATHWAYS FOR THE CANCER STUDY

In order to contribute with researches for the cancer study, the project Ontocancro, from the bioinformatics area, gathered information related to genes involved in the carcinogenic process from the genes database of NCI-Nature, BioCarta, KEGG, Reactome, Prosite, GO and STRING. So the project database is continuously accessed by researchers from the field. This paper has as objective programming a tool that makes this information research easier in the project database, in order that it demands less time from the researcher once trying to find the information. Another useful tool for the professionals from this field is the open code software called Medusa, used to visualize the protein interaction networks from samples obtain by the scientists. The tool processes a graph that shows the protein interactions with each other and also other information regarding each network protein. For that, this paper aims to program new procedures to this tool, integrating it to the Ontocancro database, aggregating new functionalities to make the pathways samples research and identification easier.

Keywords: cancer; bioinformatics; pathways; Oncocancro; Software Medusa.

LISTA DE FIGURAS

Figura 1 – Estrutura do DNA (NLM, 2009).	19
Figura 2 - Site de acesso aos dados do Ontocancro.....	22
Figura 3 – Informações sobre a via de apoptose no site da Ontocancro.	23
Figura 4 - Arquivo texto da via em uma planilha eletrônica.....	23
Figura 5 – Formato do arquivo gerado no STRING.....	24
Figura 6 – Captura da interface do Medusa com o grafo gerado.	25
Figura 7 - Tabela para edição do grafo no Medusa.....	26
Figura 8 - Medusa com a identificação das interações ativada.....	26
Figura 9 – Modo de navegação entre as vias.....	28
Figura 10 – Interface do Navegador Ontocancro.	29
Figura 11 – Arquivo lido pelo parser para gerar a estrutura da árvore.....	30
Figura 12 – Rede de interação identificando as vias (SIMÃO, 2009).....	32
Figura 13 - Fluxo das aplicações utilizadas pelos pesquisadores.....	33
Figura 14 - Redução do fluxo de aplicações com a nova função implementada.....	34
Figura 15 - Resultado obtido com a nova função.....	34
Figura 16 - Menu adicionado à ferramenta.	35
Figura 17 – Arquivo que relaciona via e proteína.....	36
Figura 18 – Grafo com as vias identificadas.	37
Figura 19 - Paleta para seleção de nova cor à via.	38
Figura 20 - Interface do Medusa com as proteínas de apoptose selecionadas.	39
Figura 21 - Grafo com as proteínas de apoptose excluídas.....	40
Figura 22 – Arquivo usado para calcular a atividade relativa.	42
Figura 23 – Graduação de cores para a atividade relativa.....	43
Figura 24 – Atividades relativas expressas no Medusa.	44
Figura 25 - Proteínas de atividade relativa entre 0.50 e 0.55 selecionadas.....	45
Figura 26 - Arquivo das atividades relativas ordenado em uma planilha eletrônica. .	46

LISTA DE ABREVIATURAS E SIGLAS

GLP	Licença Pública Geral
DNA	Ácido desoxirribonucléico
GEO	Gene Expression Omnibus
HTML	HyperText Markup Language
JRE	Java Runtime Environment
NCBI	National Center for Biotechnology Information
NHGRI	National Human Genome Research Institute
NLM	U.S. National Library of Medicine
RNA	Ácido ribonucléico
SGBD	Sistema Gerenciado de Banco de Dados
SO	Sistema Operacional

SUMÁRIO

INTRODUÇÃO	13
1 BIOINFORMÁTICA.....	16
1.1 Genes, DNA, proteínas e vias	18
1.2 Ferramentas para bioinformática	19
2 ONTOCANCRO	21
3 SOFTWARE DE ANÁLISE GÊNICA MEDUSA.....	24
4 IMPLEMENTAÇÃO DO NAVEGADOR ONTOCANCRO E AS NOVAS FUNÇÕES DO SOFTWARE MEDUSA.....	27
4.1 Navegador Ontocancro.....	27
4.1.1 Escolha da Interface.....	27
4.1.2 Desenvolvimento da aplicação.....	28
4.2 Função que identifica as vias do Ontocancro no Medusa	32
4.2.1 Alterações no Medusa para receber a nova função	35
4.2.2 Carregar os dados do Ontocancro	36
4.2.3 Identificar e representar as vias	36
4.2.4 Selecionar e apagar todos os genes de uma via	39
4.3 Identificar a atividade relativa das proteínas no Medusa	41
4.3.1 Alterações no Medusa para receber a nova função	41

4.3.2 Carregar os dados da atividade relativa.....	42
4.3.3 Identificar a atividade relativa das proteínas	42
4.3.4 Selecionar e deletar os genes de acordo com a atividade relativa	45
4.3.5 Gerar arquivo com as atividades relativas	46
4.4 Testes e dificuldades nas implementações.....	47
5 CONSIDERAÇÕES FINAIS.....	49
REFERÊNCIAS.....	51

INTRODUÇÃO

Atualmente a medicina oncológica vem ganhando novas armas na batalha contra o câncer. É crescente o número de pesquisas, estudos e técnicas de combate à doença, tanto no Brasil quanto no mundo. O elevado número de dados gerados e a necessidade de processamento em pesquisas com mapeamento genético fortaleceu o surgimento de uma nova área chamada bioinformática. A bioinformática consiste na aplicação de técnicas computacionais para resolver problemas direcionados a outras áreas de estudo, como a física, a química ou a biologia.

Um desafio para os pesquisadores é desenvolver técnicas de integração de dados que suporte o crescimento contínuo dos bancos de dados biológicos. O grande número de informações necessárias para relacionar às redes genéticas, no combate as doenças, necessita cada vez mais ferramentas para integrar essas informações que estão distribuídas em vários bancos de dados biológicos. Com a finalidade de integrar proteínas envolvidas no processo carcinogênico desenvolveu-se o projeto Ontocancro (LIBRELOTTO, 2009). Baseado em uma ontologia que organiza o conhecimento da área em questão e integra-o com dados relativos à transcriptoma e interatoma dos genes em vias obtidas a partir dos bancos de dados NCI-Nature, BioCarta, KEGG, Reactome, Prosite, GO e STRING. Para visualização das redes de interação de proteínas, os pesquisadores utilizam um software chamado Medusa (HOOPER, 2005). Que é uma ferramenta interativa muito utilizada em uma série de projetos científicos da área. Além de visualizar, o pesquisador pode manipular a rede, incluindo ou excluindo nós e arestas, ou seja, as proteínas e suas interações.

Na utilização do software Medusa, os pesquisadores inserem um grupo de interações entre proteínas, o objetivo é descobrir e identificar a quais vias pertencem cada gene envolvido no grafo e para isto é necessário pesquisar nos dados do Ontocancro. Hoje o banco Ontocancro conta com nove vias, ou seja, conjuntos de proteínas responsáveis por determinadas funções, cinco vias de reparo, duas de morte celular (apoptose), uma de ciclo celular e uma de estabilidade cromossômica,

somando um total de 1236 proteínas, além de outros grupos de genes ligados a morte celular e ao processo carcinogênico.

Outra análise importante no estudo de células cancerígenas é descobrir quais genes são mais ativos e menos ativos no tecido afetado. Estes dados estão disponibilizados em listas com milhares de genes, o que torna a pesquisa um verdadeiro “garimpo”.

Tendo em vista o trabalho exaustivo dos pesquisadores, este trabalho iniciou-se tendo como principais objetivos:

- Desenvolver uma ferramenta, para ser inserida no *site* do projeto Ontocancro, que facilite a busca de informações das vias relacionadas na base de dados do projeto.

- Implementar ao software Medusa uma função que identifica as vias Ontocancro na rede de interação de proteínas.

- Implementar ao software Medusa uma função que calcula e expressa à atividade relativa de cada proteína envolvida no grafo.

O desenvolvimento de uma ferramenta para a busca de informações nas vias relacionadas no *site* do Ontocancro é importante, pois agiliza o processo de pesquisa por parte dos profissionais que constantemente acessam a base de dados do projeto.

O motivo de se implementar novas funções a ferramenta de análise gênica Medusa, e não criar outro aplicativo com este objetivo, deve-se ao fato de que esta ferramenta já vem sendo utilizada, e com sucesso, em vários projetos de pesquisa. Outro fator determinante foi que o software Medusa é distribuído de forma livre e código aberto, podendo ser livremente modificado principalmente para o uso acadêmico.

Sendo assim foram desenvolvidas novas funções a ferramenta Medusa, agregando mais funcionalidades, que diminuem o trabalho dos profissionais que a utilizam para análise gênica, agilizando o processo de busca e manipulação de proteínas e suas interações. Além disso, foi desenvolvido um *applet* para navegação entre as vias relacionadas no projeto Ontocancro, o que facilita a procura de proteínas e suas vias no site do projeto.

Este trabalho encontra-se organizado da seguinte maneira: no capítulo 1 são apresentados conceitos sobre a área de bioinformática, onde este trabalho está inserido. Já no capítulo 2 diz respeito ao projeto Ontocancro, sobre seu banco de

dados e o portal do projeto, para onde foi desenvolvido o *applet* de navegação. O capítulo 3 trata da ferramenta Medusa, suas funcionalidades e objetivos de funcionamento antes de receber as novas funções implementadas neste trabalho.

No capítulo 4 encontram-se informações relativas ao desenvolvimento deste trabalho. Desde o *applet* navegador, com sua proposta de interface e benefícios, até as novas funções implementadas e agregadas ao software Medusa, que vieram a facilitar o processo de pesquisa na área de estudos do processo cancerígeno. Trazendo também os testes e dificuldades encontradas no desenvolvimento destas funcionalidades. Enquanto que no capítulo 5 fala sobre as considerações finais fazendo um apanhado geral sobre o trabalho.

1 BIOINFORMÁTICA

A bioinformática é uma disciplina relativamente nova, sua trajetória inicia-se em 1953 quando os cientistas James Dewey Watson e Francis Crick utilizando uma estrutura de ferro e madeira modelaram a dupla hélice para representar a estrutura da molécula de DNA (SETUBAL, 2003). O trabalho publicado na revista científica Nature em 24 de abril de 1953 foi um dos grandes marcos na história da biologia no século 20. Pois a molécula de DNA, que em português tem o nome de ácido desoxirribonucléico, contém o código de hereditariedade de cada ser. Pelo modelo proposto por Watson e Crick, ela é constituída por duas cadeias paralelas de nucleotídeos unidos em sequência, dispostas no espaço helicoidalmente. Este modelo permitiu entender também como ocorrem às mutações celulares, como ocorre a replicação da molécula e a própria linearidade da codificação da mensagem genética. Posteriormente surgiram métodos de sequenciamento dos polímeros do DNA, permitindo o estudo das formas mais simples que o constitui.

Tais polímeros passaram a constituir os principais objetos de estudo de uma nova ciência a Biologia Molecular, desde então mais de 18 milhões dessas sequências já foram produzidas e estão disponíveis nos bancos de dados publicados (FILHO, 2002).

Já na década de 90, juntamente com a evolução da informática, começaram a surgir sequenciadores automáticos de DNA aumentando consideravelmente a quantidade de sequências genéticas a serem armazenadas, conseqüentemente exigindo cada vez mais recursos computacionais como armazenamento e interpretação dos resultados obtidos. Ou seja, somente quando os computadores estavam suficientemente poderosos para processar milhões e milhões de sequências podem-se obter resultados significativos nas pesquisas com o genoma humano. Setubal (2003) faz uma análise ao observar que se o sequenciamento automático do DNA tivesse chegando com uns 20 anos de antecedência, não haveria computadores com poder de processamento suficiente para dar conta dos dados gerados. Para se ter idéia na década de 70 um computador de grande porte tinha alguns kilobytes de memória não seria capaz de processar sequer o genoma de um vírus, que pode chegar a 20 kilobases, ou 20.000 bases ou ainda 20 KB.

Com os computadores mais velozes e de menor custo, juntamente com as pesquisas nas áreas da biologia molecular gerando milhares de informações para serem catalogadas e organizadas nascia a fusão destas ciências, a bioinformática, uma ciência que envolve diversas linhas de conhecimento, como a engenharia de software, a matemática, física, química, estatística, a ciência da computação e a biologia molecular (FILHO 2002, p. 1).

Os primeiros profissionais que surgiram nessa área eram principalmente formados por médicos, biólogos e físicos, ou seja, basicamente de diferentes áreas da biologia. Porém eram profissionais que tinham certo conhecimento em informática. Alguns projetos eram compostos por profissionais das mais diversas áreas da biologia e da informática. Segundo Filho (2002) existia certa dificuldade de comunicação entre biólogos e cientistas da computação, pois como o biólogo leva em consideração resultados como incertezas e erros que podem ocorrer na prática, o cientista da computação procura uma solução direta para um problema bem definido.

Logo, desta variante de métodos de pesquisa, necessitou-se de um novo perfil de profissional que deveria ter o conhecimento suficiente de ambas as áreas e que poderia encontrar os problemas biológicos reais e quais seriam suas opções viáveis para o desenvolvimento de uma solução na abordagem computacional dos problemas avaliados. Estes profissionais são chamados Bioinformatas. Um exemplo claro do uso da bioinformática são as pesquisas nas áreas biológicas e genômicas que geram terabytes de informação, logo a demanda de um robusto Sistema de Gerenciamento de Bancos de Dados (SGBD) para gerenciar todas essas informações se faz extremamente necessário.

No início, o arquivamento de dados nesta área era realizado por grupos de pesquisa individuais, motivado pelo interesse da ciência naqueles dados, com o aumento da demanda por profissionais e equipamentos, juntamente com uma maior ênfase em computação, os arquivamentos passaram a ser de responsabilidade de projetos de grande escala (LESK, 2008, p.141).

De acordo com a NHGRI (2009) o sucesso do projeto Genoma Humano que inicialmente sequenciou em torno de 25.000 genes e constituindo pouco mais 20% do total do material genético humano. Por isso, a bioinformática se vê ainda mais influente. Mesmo que por limitações tecnológicas atuais aproximadamente 1% do genoma ainda não possa ser sequenciado.

Desta forma, é considerável a quantidade de benefícios que a aplicação da informática nas diversas áreas da biologia pode trazer às pesquisas. Desde uma simples planilha eletrônica para organizar os dados obtidos em pesquisas até um simulador que pode representar os efeitos de uma doença e assim testar possíveis curas.

1.1 Genes, DNA, proteínas e vias

Segundo o projeto Genoma (1990 apud MAGATÃO, 2008, p15) um gene pode ser definido como uma unidade fundamental, física e funcional da hereditariedade. O gene é um segmento de um cromossomo a que corresponde um código distinto, uma informação para produzir uma determinada proteína ou controlar uma característica, por exemplo, a cor dos olhos. Cada gene é formado por uma sequência específica de ácidos nucléicos (onde estão contidas as informações genéticas).

Existem dois tipos de ácidos nucléicos: o DNA ou ácido desoxirribonucléico, ilustrado na Figura 1, e o RNA também conhecido por ácido ribonucléico. Analisando o DNA e o RNA dos diversos pontos de vistas, para um químico o DNA e o RNA são longos polímeros de unidades simples (monômeros) de nucleotídeos, porém o RNA é formado por moléculas muito inferiores que as do DNA. Do ponto de vista biológico o DNA é um composto cujas moléculas contêm as instruções genéticas que coordenam o funcionamento de todos os seres vivos e o RNA é definido como responsável pela síntese de proteínas na célula.

As proteínas são compostos orgânicos responsáveis por determinadas funções na célula. Por exemplo, a queratina é a principal proteína do cabelo humano, dando-lhe rigidez, consistência e elasticidade. E também existem os anticorpos que são proteínas responsáveis pela defesa do organismo.

Geralmente um gene codifica uma proteína, mas existem exceções onde um gene pode codificar mais de uma proteína. Por existir este relacionamento de um para um entre gene e proteína muitas vezes o gene é identificado pela proteína que ele ativa.

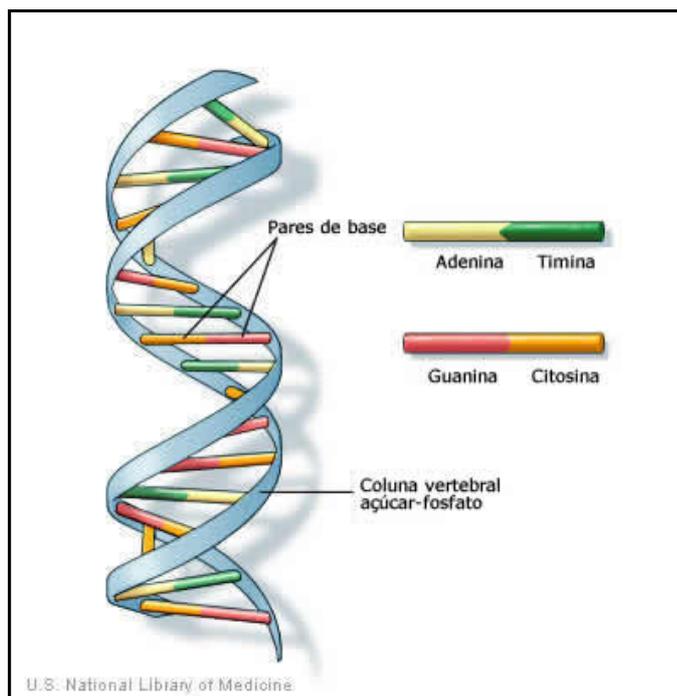


Figura 1 – Estrutura do DNA (NLM, 2009).

Cada via é formada por um grupo de proteínas responsáveis por uma determinada função. Um exemplo é uma via de reparo, que ativa proteínas responsáveis pela função de reparar a célula ou parte do DNA.

1.2 Ferramentas para bioinformática

O grande marco da bioinformática foi o mapeamento do genoma humano, que gera um elevado volume de informações armazenadas nos bancos de dados computacionais e que posteriormente são fontes de estudos para profissionais que trabalham nas áreas envolvidas. Dito isso, fica claro o quanto os SGBDs são ferramentas de grande importância para a bioinformática. Além deles existem outros aplicativos que vieram para facilitar a pesquisa nesta área. Algumas ferramentas como os SGBDs, planilhas eletrônicas, processadores de texto são usadas nas mais diversas áreas e não foram desenvolvidas especialmente para a área de bioinformática, mas que estão sendo muito bem aproveitadas. Com a preocupação dos novos profissionais inseridos nessa área, já estão presentes novas ferramentas específicas para serem usadas pelos pesquisadores. Segue alguns exemplos de ferramentas aplicadas à bioinformática:

-ViaComplex: Aplicativo de código aberto para análise de expressão gênica (CASTRO, 2009). O usuário poderá utilizar essa ferramenta para verificar a distribuição da atividade em um grafo, ou fazer um levantamento estatístico da atividade da rede.

-Bioinformatics Tools: Trata-se de uma coleção de aplicativos para gerar, formatar e analisar sequências de DNA e proteínas.

-PubMed: Desenvolvido pelo National Center for Biotechnology Information (NCBI, 2009) e mantido pela National Library of Medicine, o PubMed é um banco de dados para pesquisa bibliográfica em mais de 19 milhões de artigos médicos.

Além destas podemos citar ainda o Projeto Ontocancro e o Software Medusa que serão descritos nos próximos capítulos deste trabalho.

2 ONTOCANCRO

O projeto Ontocancro consiste numa base de dados que reúne informações de genes e vias envolvidas no processo carcinogênico, onde foram filtrados e catalogados aproximadamente 1428 genes distribuídos nas 130 vias, todos estes dados foram extraídos dos principais bancos de dados públicos de genes: NCI-Nature, BioCarta, KEGG, Reactome, Prosite, GO e STRING.

Devido à falta de consenso na definição dos conjuntos de genes e das vias de cada base de dados, o projeto apresenta um padrão próprio para as vias Ontocancro, que estão relacionadas da seguinte forma:

- Apoptose – 491 genes
- Reparação por excisão de base (BER) – 44 genes
- Ciclo celular(CC) – 286 genes
- Estabilidade cromossômica (CS) – 76 genes
- Apoptose expandida – 955 genes
- Reparo por recombinação homologa (HR) – 34 genes
- Reparo de bases mau pareadas (MMR) – 28 genes
- União terminal não homóloga – 14 genes
- Reparação por excisão de nucleotídeos (NER) – 51 genes

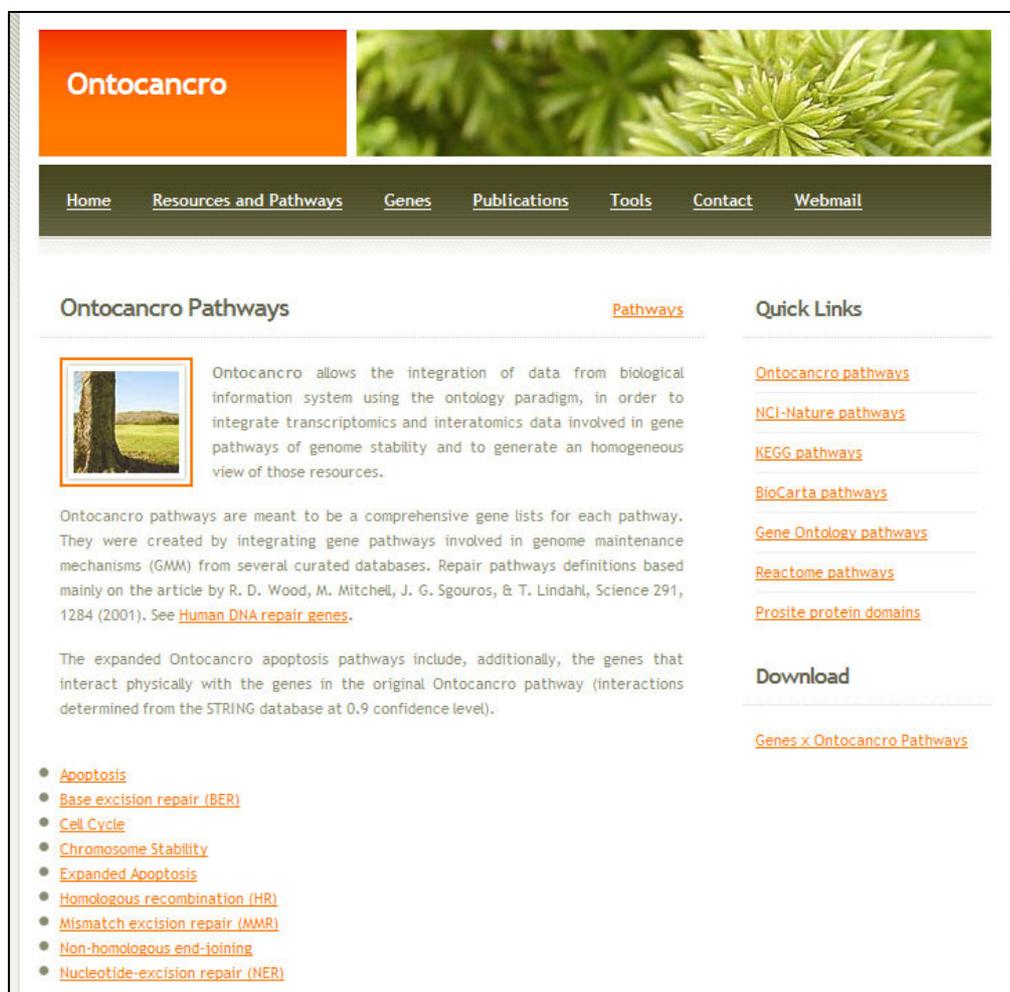
A disfunção destas vias pode provocar inúmeros tipos de anomalias celulares, uma delas, e que relacionamos neste trabalho, é o câncer.

As vias se dividem de acordo com a sua função. Por exemplo, na via de reparação por excisão de bases (BER) encontra-se um conjunto de genes envolvidos diretamente no reparo das bases do DNA, assim como as outras vias de reparação, tais como, reparação por excisão de nucleotídeos (NER), reparo de bases mau pareadas (MMR), reparo por recombinação homologa (HR) entre outras, também auxiliam na manutenção e no reparo do DNA.

Em outros casos aparecem as vias de ciclo celular (CC), responsáveis pelo processo de duplicação do DNA. Outra via importante é a via de estabilidade cromossômica (CS), que coordena todas as proteínas responsáveis pela estabilidade dos cromossomos. Se algumas destas vias não são capazes de corrigir

ou atuar de forma adequada, entre em função a via de apoptose, que tem como função destruir a célula ou então o DNA. É ela que é responsável pelo suicídio celular, também conhecido por morte celular programada.

Os dados relacionados a estas vias são acessados na base de dados do Ontocancro, que é possível a partir do endereço eletrônico <http://www.ontocancro.org>, frequentemente utilizado pelos pesquisadores que estudam o comportamento dos genes relacionados ao câncer. No site do Ontocancro, estes profissionais recebem as informações já selecionadas para o contexto que estão estudando, conforme mostra a Figura 2.



Ontocancro

Home Resources and Pathways Genes Publications Tools Contact Webmail

Ontocancro Pathways [Pathways](#)

 Ontocancro allows the integration of data from biological information system using the ontology paradigm, in order to integrate transcriptomics and interactomics data involved in gene pathways of genome stability and to generate an homogeneous view of those resources.

Ontocancro pathways are meant to be a comprehensive gene lists for each pathway. They were created by integrating gene pathways involved in genome maintenance mechanisms (GMM) from several curated databases. Repair pathways definitions based mainly on the article by R. D. Wood, M. Mitchell, J. G. Sgouros, & T. Lindahl, Science 291, 1284 (2001). See [Human DNA repair genes](#).

The expanded Ontocancro apoptosis pathways include, additionally, the genes that interact physically with the genes in the original Ontocancro pathway (interactions determined from the STRING database at 0.9 confidence level).

Quick Links

- [Ontocancro pathways](#)
- [NCI-Nature pathways](#)
- [KEGG pathways](#)
- [BioCarta pathways](#)
- [Gene Ontology pathways](#)
- [Reactome pathways](#)
- [Prosite protein domains](#)

Download

- [Genes x Ontocancro Pathways](#)

- [Apoptosis](#)
- [Base excision repair \(BER\)](#)
- [Cell Cycle](#)
- [Chromosome Stability](#)
- [Expanded Apoptosis](#)
- [Homologous recombination \(HR\)](#)
- [Mismatch excision repair \(MMR\)](#)
- [Non-homologous end-joining](#)
- [Nucleotide-excision repair \(NER\)](#)

Figura 2 - Site de acesso aos dados do Ontocancro.

O site também possibilita ao usuário baixar os arquivos textos com todas as informações das vias, que pode ser lido e processado em planilhas eletrônicas ou editores de texto. Isto facilita a extração das informações por outros aplicativos, tornando o acesso aos dados mais dinâmico, como mostra a Figura 3. Pode-se

ainda selecionar um gene pertencente a uma determinada via para obter informações detalhadas sobre o mesmo.

Symbol	Name	EntrezGene
BCL6	B-cell CLL/lymphoma 6	604
RIPK1	receptor (TNFRSF)-interacting serine-threonine kinase 1	8737
RIPK2	receptor-interacting serine-threonine kinase 2	8767
ROCK1	Rho-associated, coiled-coil containing protein kinase 1	6093
RPS3	ribosomal protein S3	6188
RPS3A	ribosomal protein S3A	6189
RTKN	rhotekin	6242
RYR2	ryanodine receptor 2 (cardiac)	6262
BID	BH3 interacting domain death agonist	637
SATB1	SATB homeobox 1	6304
SCG2	secretogranin II (chromogranin C)	7857
CCL2	chemokine (C-C motif) ligand 2	6347
SFN	stratifin	2810
BMX	BMX non-receptor tyrosine kinase	660

Figura 3 – Informações sobre a via de apoptose no site da Ontocancro.

A base de dados do Ontocancro reúne diversas informações sobre cada gene catalogado. Estas informações estão descritas também no arquivo texto da via, que pode ser visualizado e mais facilmente descrito quando aberto em uma planilha eletrônica, conforme a Figura 4.

	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Hugo Name	UniGene Name	GeneOntology ID	ENSG	KEGG ID	Affymetrics	Approved Symbol	Previous Symbols	NCI Name	Evidence	Chromosome	Status	Reactome	Aliases
2	caspase 3, a	Caspase 3, apo	GO:0030264	GO	ENSG:hsa+836	202763_at	2	CASP3	CASP3	IGI IMP	4q34	Approved	REACT_57	CPP32, CF
3	caspase 7, a	Caspase 7, apoptosis-related cyst	ENSG:hsa+840	207181_s_a	CASP7			CASP7			10q25	Approved	REACT_57	MCH3, CM
4	chaperonin c	Chaperonin containing TCP1, subu	ENSG0000011548	200877_at	CCT4			CCT4			2	Approved		Cctd
5	zinc finger an	Zinc finger and BTB domain contai	ENSG0000017895	213299_at	2	ZBTB7A	ZBTB7	ZBTB7A			19p13.3	Approved		FBI-1, LRF
6	v-fos FBJ mu	V-fos FBJ murine osteosarcoma vi	ENSG0000017034	209189_at	FOS			FOS			14q24.3	Approved		c-fos, AP-1
7	v-akt murine (V-akt murine th)	GO:0006924	GO	ENSG:hsa+207	207163_s_a	AKT1		AKT1	EXP IMP		14q32.32-c	Approved	REACT_57	RAC, PKB
8	FK506 binding protein 12-rapamycin associated		ENSG0000019879	202288_at	2	FRAP1	FRAP, FRAP2	FRAP, FRAP2			1p36	Approved	REACT_45	MTOR, RA
9	X-linked inhibitor of apoptosis		ENSG:hsa+331	206536_s_a	XIAP		API3, BIRC4	X-linked inhibitor of a			Xq25	Approved	REACT_57	hILP
10	baculoviral IA Effector cell pep	GO:0006916	GO:	ENSG00000008968	202094_at	2	BIRC5	API4		IDA IMP,	17q25	Approved		EPR-1
11	integrin, beta	Integrin, beta 1 (fibronectin recept	ENSG0000015009	1553530_a	ITGB1		FNRB, MSK12, M	ITGB1			10p11.2	Approved	REACT_60	CD29, GPII
12	jun D proto-oncogene		ENSG0000013052	203751_x_a	JUND			JUND			19p13.2	Approved		AP-1
13	phosphoinosi	Phosphoinositide-3-kinase, catalyt	ENSG:hsa+5291	212688_at	2	PIK3CB		PIK3C1			3q21-qter	Approved		
14	D4, zinc and double PHD fingers family 2		ENSG0000013388	202116_at	DPF2		REQ	D4, zinc and double			11q13.1	Approved		ubi-d4

Figura 4 - Arquivo texto da via em uma planilha eletrônica.

3 SOFTWARE DE ANÁLISE GÊNICA MEDUSA

A ferramenta Medusa foi desenvolvida por Sean Hooper e Peer Bork. Segundo Hooper (2005) trata-se de um aplicativo desenvolvido na linguagem Java, livre para uso acadêmico, e distribuído sob a licença pública GPL. Tem como funcionalidade principal a visualização do grafo de interação de proteínas. Os dados bem como o código-fonte desta ferramenta encontram-se disponíveis no endereço eletrônico <http://coot.embl.de/medusa/>.

Para gerar o grafo, o pesquisador necessita buscar no STRING as informações necessárias. Segundo Jensen (2008) o STRING consiste de um banco de dados específico de interação de genes, já está na sua versão 8.0 e abrange cerca de 2,5 milhões de proteínas de 630 organismos. Pode ser acessado pelo endereço eletrônico <http://string-db.org>.

O procedimento para abrir um grafo no Medusa funciona da seguinte forma: o pesquisador insere no banco de dados STRING, uma amostra de proteínas que deseja analisar. O STRING gera um arquivo texto com todos os dados sobre as interações daquele grupo de proteínas. Para visualização e manipulação do grafo descrito neste arquivo é necessário que o pesquisador abra ele na Medusa.

As informações contidas no arquivo gerado na base de dados da STRING são: posição da proteína no plano, cor, interações e outros dados de cada proteína da amostra como representado na Figura 5. Todos esses dados estão catalogados e disponíveis para o pesquisador no banco de dados STRING.

ERCC8	0.14421238	0.559562399	c 196,239,117	406405: DN
TFAP2A	0.743910418	0.353403566	c 117,117,239	"422832: Tr
(AP-2 transcription factor) (Activator protein 2) (AP-2); Sequence- elements to regulate transcription of selected genes. AP-2 factors l spectrum of important biological functions including proper eye, fat including MCAM/MUC18, C/EBP alpha and MYC [...]"				
RACGAP1	0.545520886	0.707860616	c 117,239,151	"410669: Rac
and may play a role in the microtubule-dependent steps in cytokines through mechanisms other than regulating Rac GTPase activity. Also involved in regulating spermatogenesis and in the RACGAP1 pathway in				
<i>Proteína</i>	<i>Plano XY</i>	<i>Cores em RGB</i>	<i>Observações</i>	

Figura 5 – Formato do arquivo gerado no STRING.

Depois de lida as informações necessárias para gerar o grafo, o Medusa permite que o usuário visualize e modifique o grafo aberto, conforme mostra a Figura 6, podendo também salva-lo no arquivo texto para posteriores visualizações.

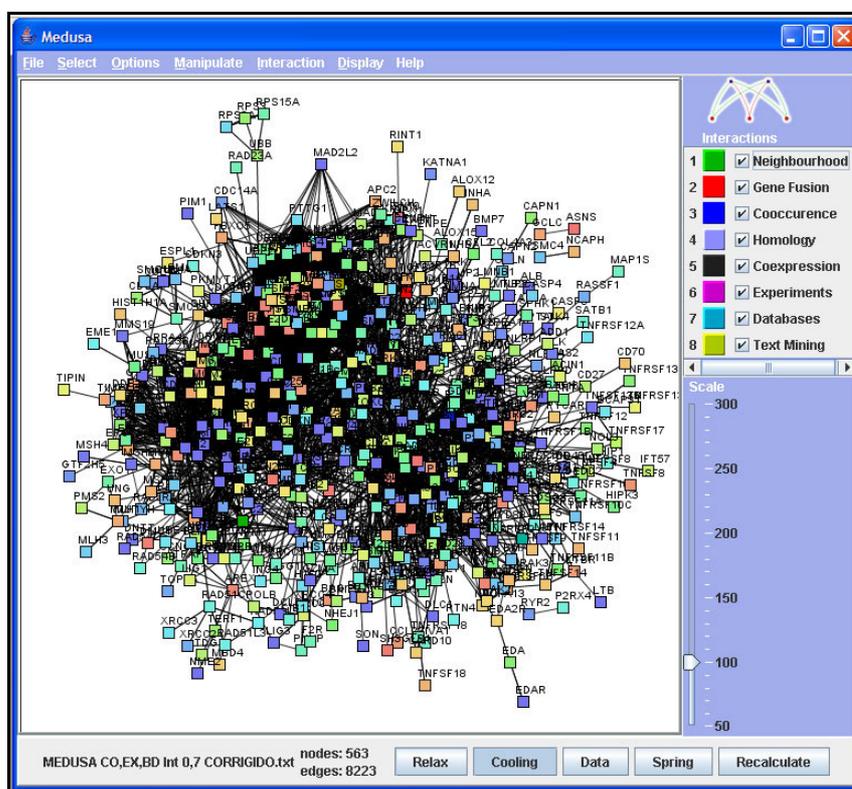
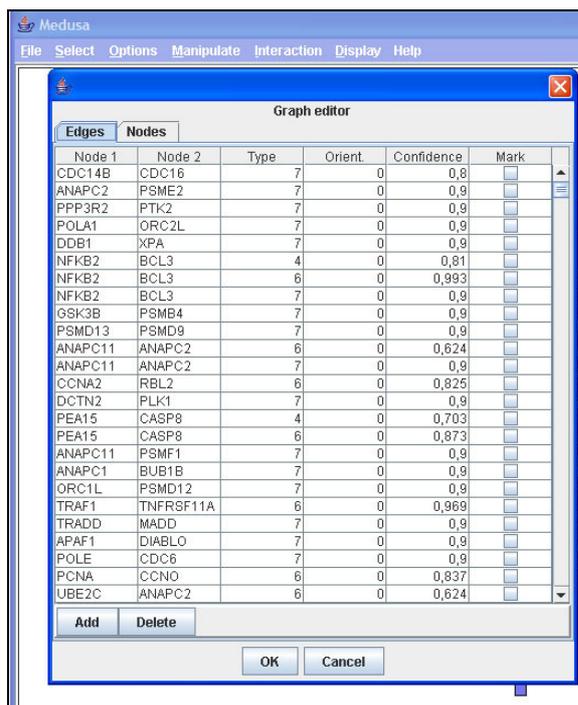


Figura 6 – Captura da interface do Medusa com o grafo gerado.

O Medusa permite manipular o grafo de várias maneiras. No menu *Manipulate* da ferramenta pode-se encontrar opções como:

- *Change node shape*: usado para alterar o formato do nodo para circulo, triangulo, retângulo ou diamante;
- *Delete nodes*: nesta opção é possível apagar os nodos selecionados;
- *Edit Graph*: abre uma tabela, conforme Figura 7, que permite adicionar novos nodos e/ou alterar dados dos existentes.

Outra funcionalidade que o Medusa possibilita é de visualizar as interações das proteínas de acordo com o tipo de interação. Assim como no STRING, o Medusa identifica as interações pela cor do índice no lado direito do grafo.



The screenshot shows the 'Graph editor' window in Medusa. It contains a table with columns for 'Node 1', 'Node 2', 'Type', 'Orient.', 'Confidence', and 'Mark'. Below the table are 'Add' and 'Delete' buttons, and 'OK' and 'Cancel' buttons at the bottom.

Node 1	Node 2	Type	Orient.	Confidence	Mark
CDC14B	CDC16	7	0	0,8	<input type="checkbox"/>
ANAPC2	PSME2	7	0	0,9	<input type="checkbox"/>
PPP3R2	PTK2	7	0	0,9	<input type="checkbox"/>
POLA1	ORC2L	7	0	0,9	<input type="checkbox"/>
DDB1	XPA	7	0	0,9	<input type="checkbox"/>
NFKB2	BCL3	4	0	0,81	<input type="checkbox"/>
NFKB2	BCL3	6	0	0,993	<input type="checkbox"/>
NFKB2	BCL3	7	0	0,9	<input type="checkbox"/>
GSK3B	PSMB4	7	0	0,9	<input type="checkbox"/>
PSMD13	PSMD9	7	0	0,9	<input type="checkbox"/>
ANAPC11	ANAPC2	6	0	0,624	<input type="checkbox"/>
ANAPC11	ANAPC2	7	0	0,9	<input type="checkbox"/>
CCNA2	RBL2	6	0	0,825	<input type="checkbox"/>
DCTN2	PLK1	7	0	0,9	<input type="checkbox"/>
PEA15	CASP8	4	0	0,703	<input type="checkbox"/>
PEA15	CASP8	6	0	0,873	<input type="checkbox"/>
ANAPC11	PSMF1	7	0	0,9	<input type="checkbox"/>
ANAPC1	BUB1B	7	0	0,9	<input type="checkbox"/>
ORC1L	PSMD12	7	0	0,9	<input type="checkbox"/>
TRAF1	TNFRSF11A	6	0	0,969	<input type="checkbox"/>
TRADD	MADD	7	0	0,9	<input type="checkbox"/>
APAF1	DIABLO	7	0	0,9	<input type="checkbox"/>
POLE	CDC6	7	0	0,9	<input type="checkbox"/>
PCNA	CCNO	6	0	0,837	<input type="checkbox"/>
UBE2C	ANAPC2	6	0	0,624	<input type="checkbox"/>

Figura 7 - Tabela para edição do grafo no Medusa.

Para ativar esta função basta ir ao menu *Display*, e depois clicar em *Interactions*. Assim as interações do grafo serão facilmente identificadas pela cor, conforme mostra a Figura 8.

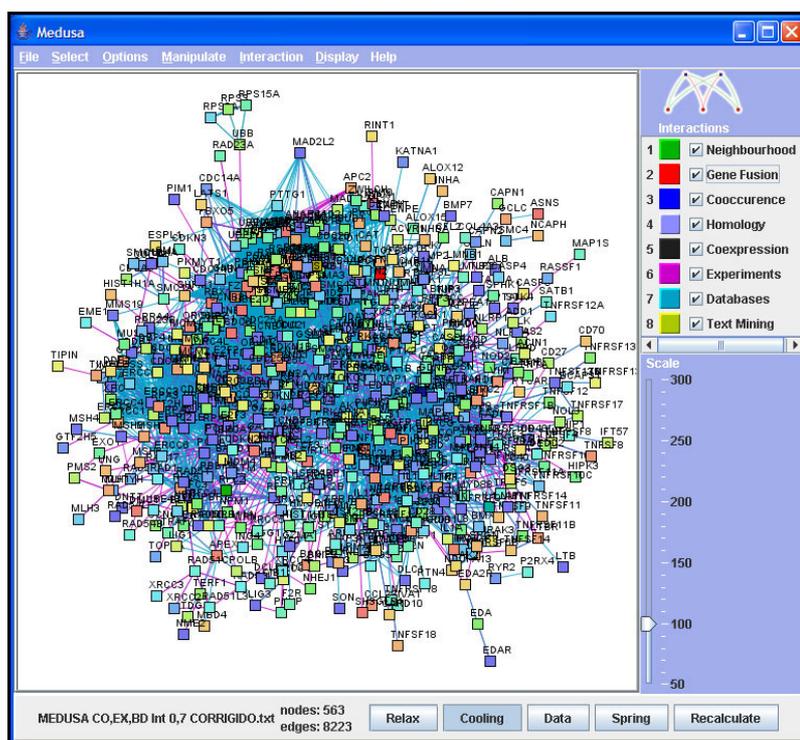


Figura 8 - Medusa com a identificação das interações ativada.

4 IMPLEMENTAÇÃO DO NAVEGADOR ONTOCANCRO E

AS NOVAS FUNÇÕES DO SOFTWARE MEDUSA

Este capítulo descreve as ferramentas desenvolvidas. Apresenta o módulo desenvolvido para ser inserido no site da Ontocancro que facilitará a navegação entre as vias relacionadas na base de dados, também descreve as novas funcionalidades desenvolvidas e agregadas ao software Medusa. Inicialmente, apresentando uma visão geral de cada implementação com suas funcionalidades, principais etapas e decisões do processo de desenvolvimento. Por fim, descreve os testes feitos nas ferramentas e as dificuldades encontradas no processo de desenvolvimento.

4.1 Navegador Ontocancro

A busca por informações no site da Ontocancro é frequente por parte dos profissionais da área oncológica. Geralmente necessita-se buscar muitos dados por vez, e com a navegação por páginas acaba tornando o processo mais demorado. Logo, existe a necessidade de uma ferramenta onde o usuário visualize uma maior quantidade de dados. O Navegador Ontocancro possibilita este trabalho de maneira interativa e eficiente.

4.1.1 Escolha da Interface

Na escolha da interface de interação com o usuário, foi necessário analisar as diversas formas que os dados poderiam ser apresentados. Deveria ser uma interface interativa e intuitiva que o usuário tivesse familiaridade com o visual, tornando a pesquisa ainda mais rápida.

A forma encontrada para representar os dados do Ontocancro foi o formato de “árvore” ou *tree*, semelhante ao utilizado nos gerenciadores de arquivos que estão presentes na maioria dos sistemas operacionais atuais. Sendo assim, uma

interface conhecida pelos usuários que já utilizam alguma ferramenta para gerenciamento de arquivo.

Como mostrado na Figura 9 a interface se torna intuitiva, pois mesmo não mostrando diretamente todos os dados, o usuário tem a noção de que clicando sobre as pastas, existem mais informações para serem mostradas.

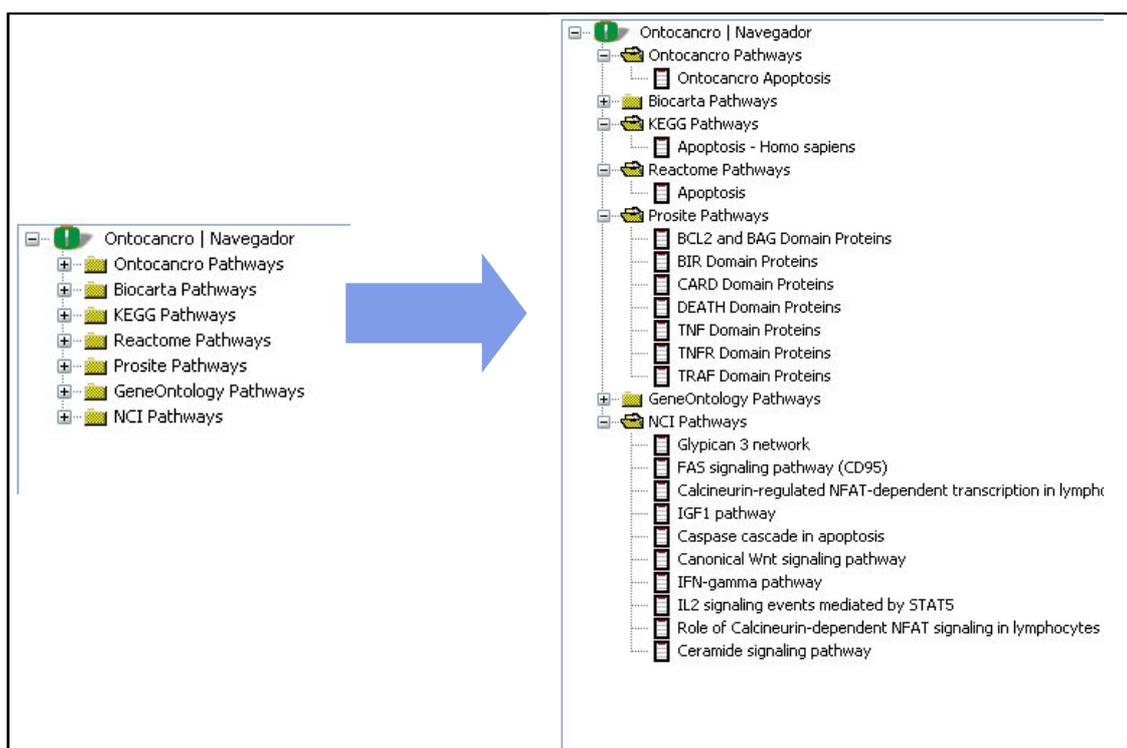


Figura 9 – Modo de navegação entre as vias.

Primeiramente, o usuário tem a lista de todas as bases de dados relacionadas pelo Ontocancro. Clicando sobre a pasta de cada base o usuário abre a lista de vias envolvidas no contexto do projeto Ontocancro. O usuário ainda tem a possibilidade de clicar sobre a via desejada e obter a relação de todos os genes pertencentes a este grupo, juntamente com todas as informações do mesmo, como mostra a Figura 10. Desta maneira, tem-se apenas uma tela com todas as informações necessárias para a pesquisa.

4.1.2 Desenvolvimento da aplicação

Como o objetivo foi criar um módulo de navegação para ser inserido no site

da Ontocancro, a forma escolhida foi um *applet* desenvolvido na linguagem de programação JAVA. Por ser uma linguagem aceita na maioria dos *browsers* de internet, e também por possibilitar a utilização da ferramenta de maneira off-line caso for necessário.

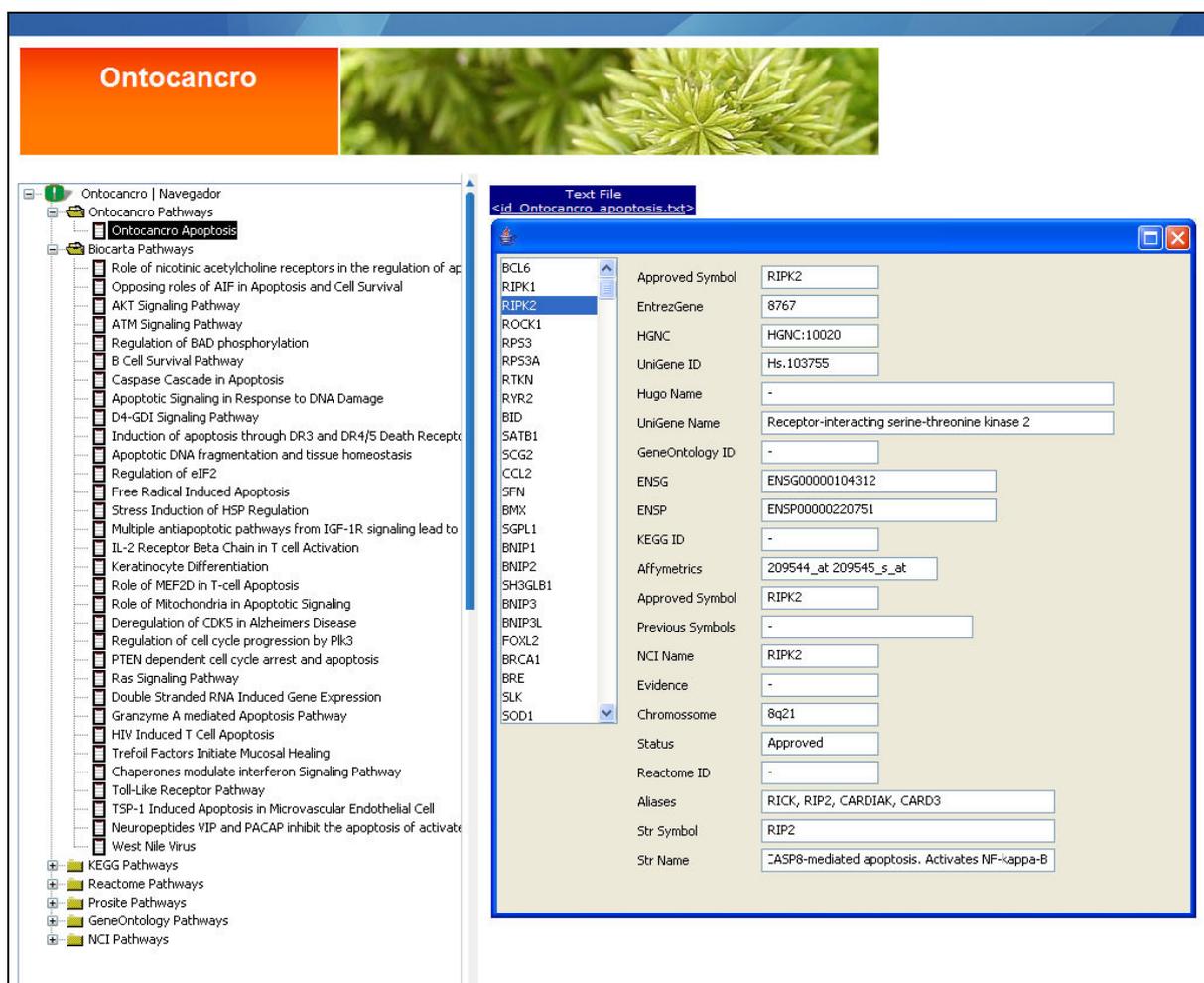


Figura 10 – Interface do Navegador Ontocancro.

No desenvolvimento deste navegador foram criadas as seguintes classes:

- *OntoNodo*: Classe que estrutura cada nodo da árvore. Recebendo como parâmetros todos os dados necessários para cada nodo: título, identificador da via, ícones de aberto e fechado.
- *OntoArvore*: Esta classe que define a árvore de navegação, para isto ela implementa a interface *TreeSelectionListener* do pacote *java.swing.event*.
- *OntoTreeIconeRenderer*: Classe criada para manipular as imagens dos ícones na árvore de navegação.
- *OntoAppletNavegador*: É a classe principal do applet que será visualizado.

Para estruturar a árvore de navegação da ferramenta, com seu nodo raiz e os nodos folhas, foi gerado um arquivo a partir da base de dados do Ontocancro, como mostra a Figura 11. E na classe *OntoArvore* foi utilizado um *parser* para ler este arquivo.

```

0 "Ontocancro | Navegador" ""
1 "Ontocancro Pathways" "" "0" "1"
  2 "Ontocancro Apoptosis" "id_OntocancroPathway_Apoptosis" "2" "2"
1 "Biocarta Pathways" "" "0" "1"
  2 "Role of nicotinic ... of apoptosis" "id_Biocarta_h_achpathway" "2" "2"
  2 "Opposing roles of AIF ... and Cell Survival" "id_Biocarta_h_aifpathway" "2" "2"
  2 "AKT Signaling Pathway" "id_Biocarta_h_aktpathway" "2" "2"
  2 "ATM Signaling Pathway" "id_Biocarta_h_atmpathway" "2" "2"
  2 "Regulation of BAD phosphorylation" "id_Biocarta_h_badpathway" "2" "2"
  2 "B Cell Survival Pathway" "id_Biocarta_h_bcellsurvivalpathway" "2" "2"
  2 "Caspase Cascade in Apoptosis" "id_Biocarta_h_caspasepathway" "2" "2"
  2 "Apoptotic Signaling in ..." "id_Biocarta_h_chemicalpathway" "2" "2"

1 "KEGG Pathways" "" "0" "1"
  2 "Apoptosis - Homo sapiens" "id_Kegg_apoptosis--homo-sapiens--human-" "2" "2"

1 "Reactome Pathways" "" "0" "1"
  2 "Apoptosis" "id_ReactomePathway_apoptosis" "2" "2"

1 "Prositate Pathways" "" "0" "1"
  2 "BCL2 and BAG Domain Proteins" "id_PrositatePathway_bcl2-and-bag-domain-proteins" "2" "2"
  2 "BIR Domain Proteins" "id_PrositatePathway_bir-domain-proteins" "2" "2"
  2 "CARD Domain Proteins" "id_PrositatePathway_card-domain-proteins" "2" "2"
  2 "DEATH Domain Proteins" "id_PrositatePathway_death-domain-proteins" "2" "2"

```

Figura 11 – Arquivo lido pelo parser para gerar a estrutura da árvore.

A hierarquia da árvore é determinada no primeiro *token* lido pelo *parser*, que é representado por um número inteiro. Da seguinte forma: o nodo raiz determinado com o *label* “Ontocancro | Navegador” é identificado pelo número inteiro 0 (zero), suas ramificações são representadas pelo número 1 (um), e assim por diante. Ou seja, de forma geral o nodo pai é N e o filho é N+1, conseqüentemente a árvore torna-se dinâmica, podendo ter várias ramificações, porém todas partirão do nodo raiz.

Depois de lido o identificador da hierarquia a próxima informação obtida é o nome do nodo, que deverá estar representado entre aspas duplas (Por exemplo, “Pathways”), este é o mesmo texto que será descrito no *label* que identificará o nodo na interface do navegador.

O seguinte dado a ser lido do arquivo é o identificador da via, que também deverá estar entre aspas duplas. Este identificador é quem determinará o arquivo contendo a lista de cada gene pertencente à via selecionada e todas as suas informações. Caso seja um nodo da árvore que não representa uma via de genes, a

informação deve aparecer vazia. Os dois dados seguintes após o identificador referem-se a imagens dos ícones que a interface mostra quando um grupo de vias é selecionado. Estas imagens estão contidas em um *arraylist*, que é carregado com os parâmetros descritos no código da página HTML onde está inserido o *applet* do navegador. Como no exemplo:

```
<APPLET CODE="OntoAppletNavegador.class"
        NAME="OntoNavegador"
        HEIGHT=2000 WIDTH=196>
<PARAM NAME="datafile" VALUE="dadosontocancro.txt">
<PARAM NAME="bgcolor" VALUE="FFFFFF">
<PARAM NAME="font" VALUE="Courier">
<PARAM NAME="point" VALUE="11">
<PARAM NAME="images" VALUE="./images">
<PARAM NAME="targetframe" VALUE="_top">
<PARAM NAME="rootimage" VALUE="root.gif">
<PARAM NAME="rootimagefecha" VALUE="root_.gif">
<PARAM NAME="leaveimageslist" VALUE="fclos.jpg,fope.jpg,pag.jpg">
</APPLET></BODY></HTML>
```

O código HTML fornece alguns parâmetros para o funcionamento do *applet* em questão. O parâmetro chamado “leaveimagelist” contém a lista de imagens que serão carregadas para gerar os ícones do navegador. Os parâmetros que identificam o arquivo lido pelo *parser* para gerar a árvore, e o diretório onde estão as imagens dos ícones são o “datafile” e o “images” respectivamente. Com estes parâmetros identificados, o *applet* estará pronto para rodar na página HTML.

Quando selecionada uma determinada via no navegador no *frame* ao lado é descrito todos os genes pertencentes à via em questão. Os dados apresentados neste *frame* estão contidos no arquivo indicado pelo identificador da via, que é o mesmo arquivo texto que o usuário pode baixar do site do Ontocancro. Logo, foi desenvolvido um *parser* para ler este arquivo, que depois de lido suas informações são distribuídas nos campos do *frame*. O primeiro campo é uma lista de todos os genes da via selecionada, onde o pesquisador pode clicar sobre determinado gene e então obter as informações sobre o mesmo.

4.2 Função que identifica as vias do Ontocancro no Medusa

O grafo que a Ferramenta Medusa gera, a partir do arquivo obtido na base de dados do STRING, tem a vantagem de ser dinâmico. Porém, as informações obtidas não são suficientes para identificação das vias que as proteínas pertencem. E estas informações demandam uma busca em outras bases de dados. Muitas vezes este trabalho era feito pelo próprio pesquisador, o que demandava muito tempo de pesquisa, pois identificar uma amostra de proteínas em um grafo que contém em média 800 nodos exigia dias pesquisando. Para então obter o resultado conforme a Figura 12 (SIMÃO, 2009).

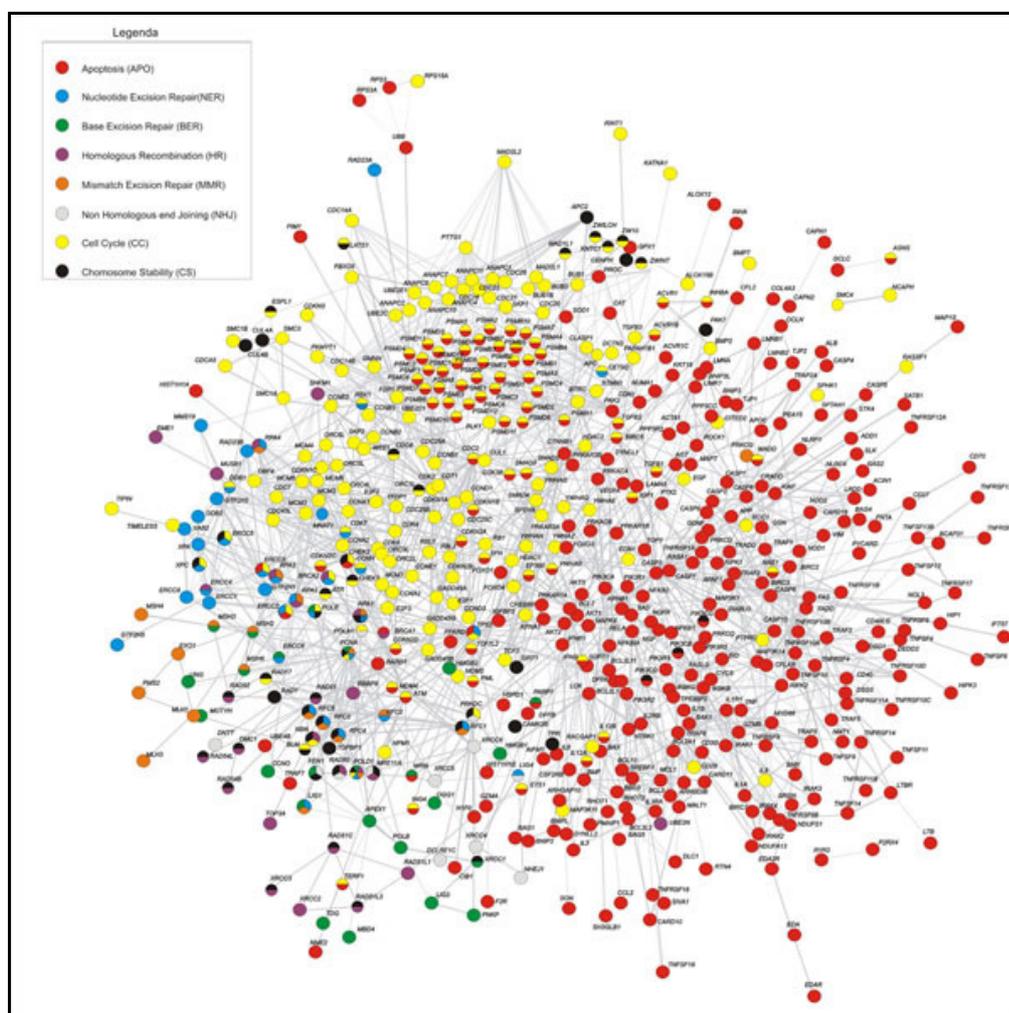


Figura 12 – Rede de interação identificando as vias (SIMÃO, 2009).

Para isto o pesquisador necessitava abrir simultaneamente, no Medusa e em uma planilha eletrônica, o arquivo gerado no STRING contendo um número N de

proteínas. Na planilha eletrônica eram filtrados os dados necessários de cada proteína, e então era necessário pesquisar as N proteínas em todas as vias da base de dados do Ontocancro, uma por uma, pois cada proteína pode pertencer a mais de uma via. Depois que todas as vias envolvidas na amostra fossem catalogadas. O grafo aberto no Medusa era redesenhado e pintado em um aplicativo de desenho, exigindo um processo quase manual. E a partir dos dados filtrados na planilha eletrônica criava-se um índice para identificar as vias no grafo. Como se pode perceber todo este trabalho exigia um grande fluxo de aplicações, conforme ilustra a Figura 13. Para evitar este processo repetitivo e exaustivo ao pesquisador, neste trabalho implementou-se a integração das informações que o Medusa obteve do banco de dados STRING com a base de dados das vias mapeadas pelo Ontocancro.

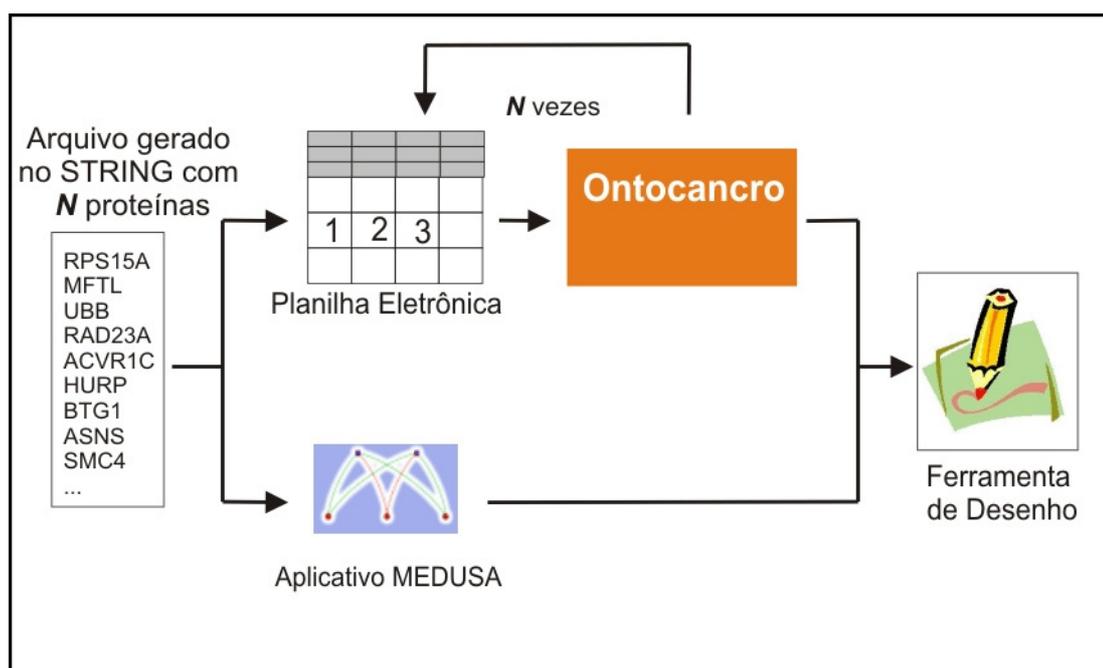


Figura 13 - Fluxo das aplicações utilizadas pelos pesquisadores.

Desta maneira reduziu o número de aplicações necessárias para este processo de pesquisa, ilustrado na Figura 14. Pois com esta nova funcionalidade agregou-se ao Medusa processos que eram feitos nas outras ferramentas. Agora, basta o pesquisador abrir o arquivo gerado no STRING juntamente com um arquivo contendo os dados das vias e suas proteínas. Para então obter o resultado, em alguns segundos, com a identificação das proteínas e as suas respectivas vias. Como na Figura 15 - Resultado obtido com a nova função.

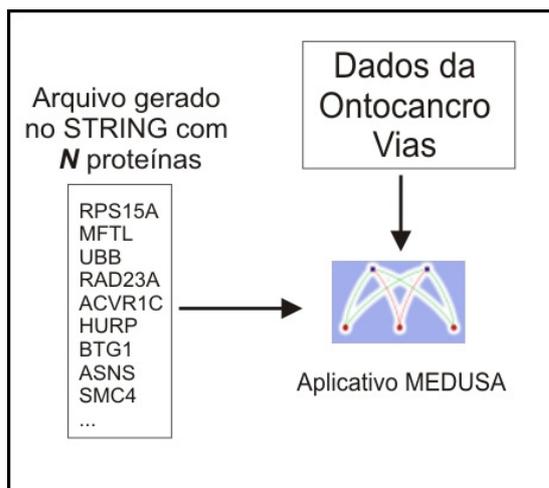


Figura 14 - Redução do fluxo de aplicações com a nova função implementada.

Sendo assim, automatizou-se ainda mais o trabalho de pesquisa, evitando possíveis erros que podem ocorrer em um trabalho manual. Como por exemplo: o grande número de nodos em um grafo, algum deles pode facilmente passar despercebido aos olhos do pesquisador no meio de tantos envolvidos.

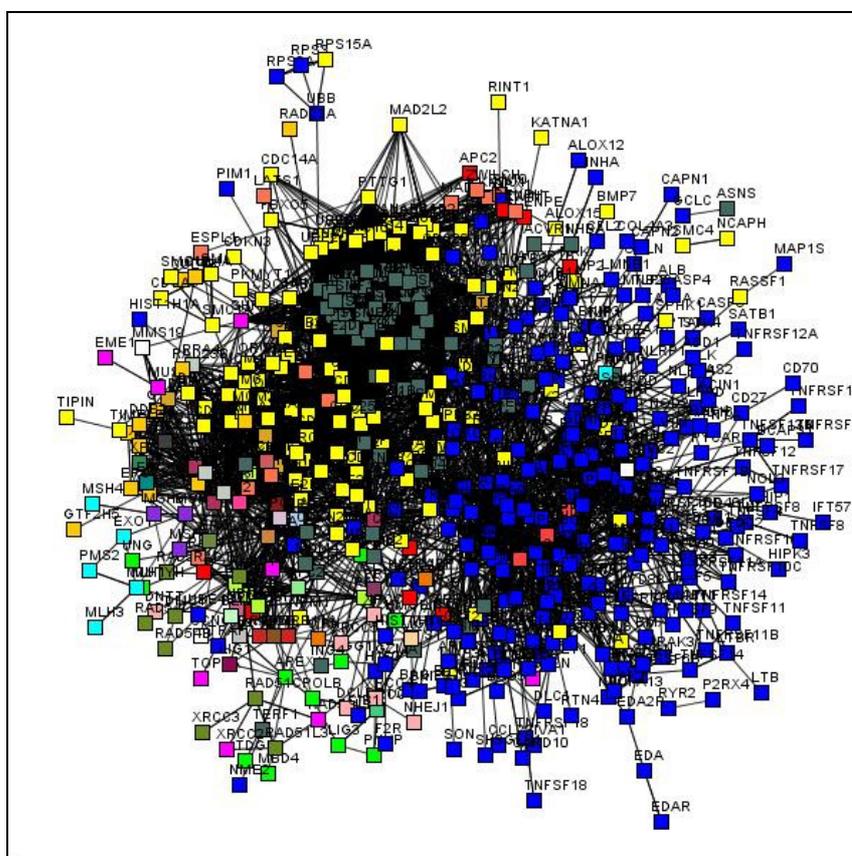


Figura 15 - Resultado obtido com a nova função.

4.2.1 Alterações no Medusa para receber a nova função

Para iniciar a implementação desta funcionalidade na ferramenta, necessitou-se a alteração no código de algumas classes do Medusa. A primeira a ser alterada foi a classe *Node* do pacote de desenvolvimento *medusa.graph*, esta é a classe que determina todos os atributos e métodos dos nós envolvidos no grafo. Nela foi criada a variável *pathway*, que carrega o nome da via a qual o nodo pertence, e os métodos para manipulação da mesma: *setPathway*, que recebe o nome da via, e *getPathway*, que retorna uma variável com o nome da via.

Na classe Principal *MedusaFrame* do pacote *medusa* foi criado o menu de acesso a esta nova função. Da seguinte maneira: No menu *Options* foi criado o sub-menu “Ontocancro”, para acesso as novas funções implementadas neste trabalho. Neste sub-menu o item “Ontocancro Pathways” é o que referencia à função que identifica as vias, conforme mostra a Figura 16.

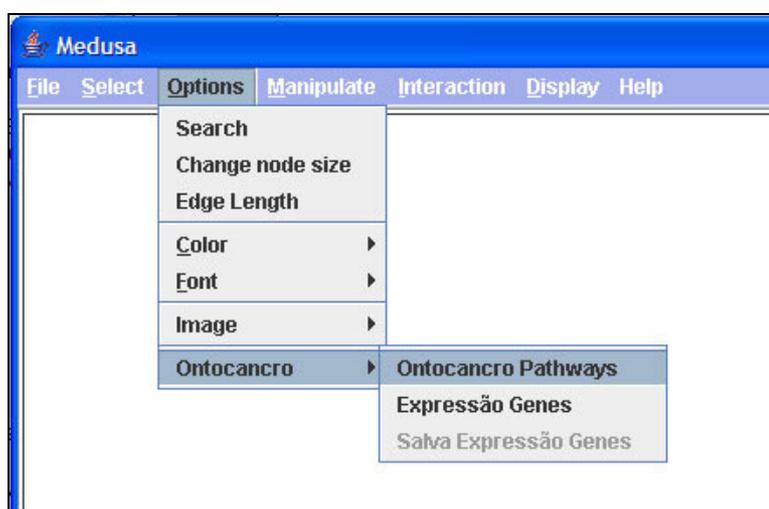


Figura 16 - Menu adicionado à ferramenta.

Outra alteração necessária foi na classe *DisplayPanel* do pacote *medusa.display*, esta classe é responsável por manipular o grafo aberto no Medusa, onde foi implementada a função *repaintOntoPath*. Esta função é onde consta a maior parte da implementação deste trabalho e que está descrita nos próximos itens.

4.2.2 Carregar os dados do Ontocancro

A partir dos dados obtidos no Ontocancro, foi gerado um arquivo com o relacionamento das vias e suas proteínas, no formato texto, por ser de fácil manipulação e integração dos dados. Este arquivo disponibiliza um relacionamento por linha, contendo o nome da via e o nome da proteína.

Para ler este arquivo foi desenvolvido um *parser* na função *repaintOntoPath*, onde o primeiro token lido no arquivo é o nome da via e o próximo o nome da proteína, conforme Figura 17.

APO	AIFM1
APO	AIFM2
NER	POLE2
NER	POLE3
NER	POLE4
BER	WRN
BER	XRCC1
MMR	EXO1
MMR	LIG1
MMR	MLH1
HR	POLD3
HR	POLD4
HR	RAD50
EJ	RAD50
EJ	XRCC4
CS	APC2
CS	ARID4B
CC	YWHAQ
CC	ZWILCH

Figura 17 – Arquivo que relaciona via e proteína.

4.2.3 Identificar e representar as vias

O fato de algumas proteínas estarem envolvidas em mais de uma via, e o modo como a ferramenta Medusa disponibiliza o grafo, determinou o modo de representação destas vias. A base de dados do Ontocancro Vias envolve um total de 9 vias, que foram chamadas de vias principais. Conforme o *parser* lê o arquivo com as vias e suas respectivas proteínas, pode ocorrer a possibilidade de uma proteína que já pertence à determinada via “X” ser encontrada em outra via “Y”, é criado então uma nova sub-via que leva o nome de “X-Y”, esta sub-via que é tratada como uma nova via, para os casos em que outra proteína venha a pertencer às mesmas vias envolvidas nesta sub-via. Para cada via e sub-via identificada a função

determina uma cor diferente, para posteriormente ser usado como um índice.

Para carregar essas informações foi criada a classe *OntocancroPathways* dentro do pacote de desenvolvimento *medusa*. Esta classe é um *arraylist*, que recebe todas as proteínas da via, e possui os atributos *pathNome* e *pathColor* e os métodos de manipulação dos atributos: *setPathName*, *setPathColor*, *getPathName* e *getPathColor* para identificação da via no grafo.

Depois de carregar todas as vias do arquivo, então é percorrido os nodos do grafo no Medusa fazendo uma pesquisa em todas as proteínas e atribuindo a cor do nodo identificado para a cor da via relacionada.

Quando o grafo foi analisado, a função gera um índice ao lado da ferramenta para que o usuário possa identificar as proteínas expressas no grafo pela cor da via a qual pertencem. Para mostrar este índice na tela foi criada a classe *OntocancroPathwaysDialog*, que tem como base um *JFrame* com a tabela de cores das vias relacionadas no grafo, conforme mostra na Figura 18.

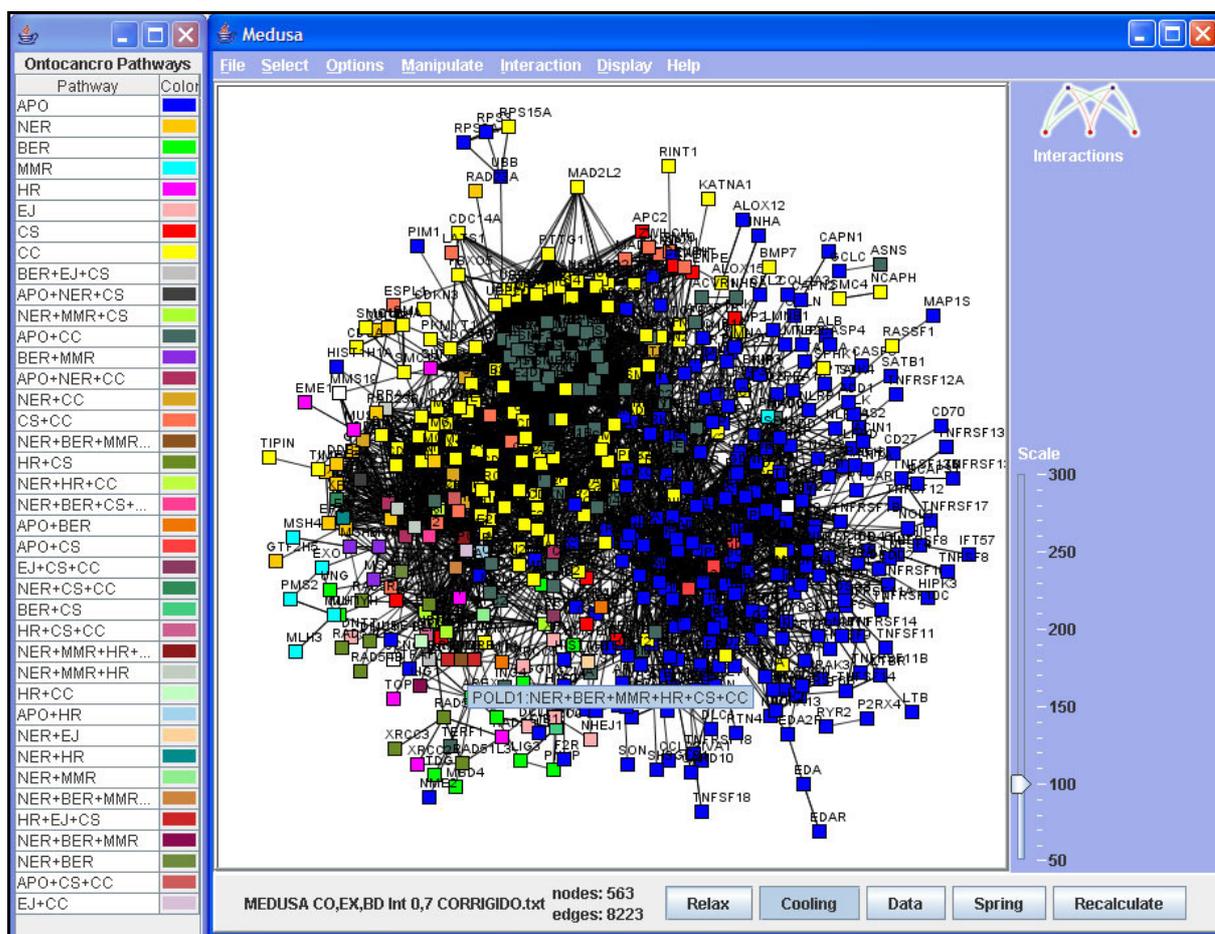


Figura 18 – Grafo com as vias identificadas.

Para que não ocorra confusão na identificação dos índices, assim que gerado o índice das vias, é desabilitado o índice de interações. O qual por padrão já está inserido na interface da ferramenta, podendo ser ativado novamente no menu *Display*, no item *Interactions*.

Muitas vezes o modo como os nodos estão distribuídos no grafo dificultam a identificação dos mesmos, fazendo com que o usuário necessite alterar a escala do grafo no Medusa para identificá-los. Com este objetivo, foi desenvolvida nesta função a possibilidade do pesquisador passar o mouse sobre o nodo para obter informações como o nome da proteína e a via a qual ela pertence.

Para tornar a função ainda mais completa adicionou-se a possibilidade do usuário trocar as cores relacionadas à determinadas vias, caso exista a necessidade.

Por exemplo, pode ocorrer que a cor direcionada para uma via esteja muito parecida com a cor de outra, o que causaria confusão para o usuário identificá-las. Logo, para alterar a cor da via, basta o usuário clicar sobre determinada cor no índice e terá assim uma paleta de cores, um objeto da classe *JColorChooser* que pertence ao pacote *javax.swing*, para selecionar a cor desejada (Figura 19).

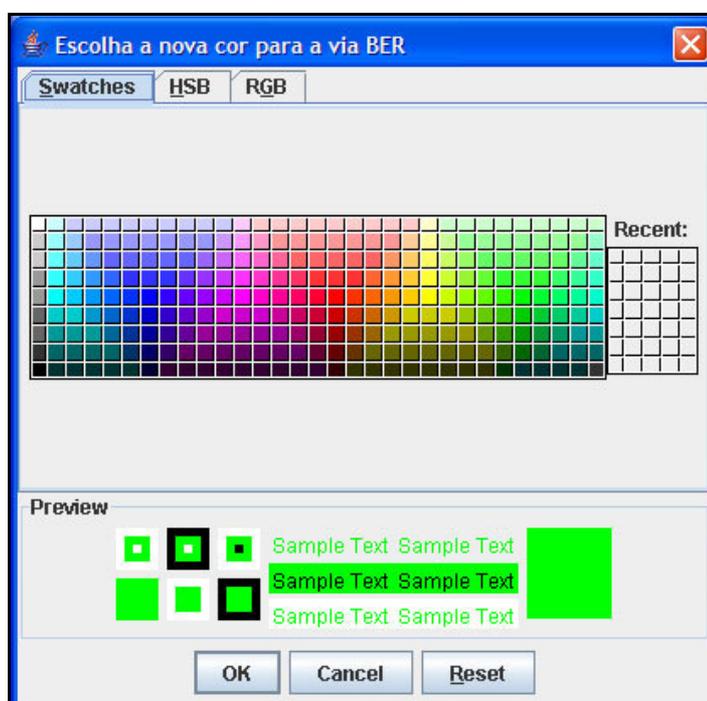


Figura 19 - Paleta para seleção de nova cor à via.

Quando a cor é selecionada, é chamada a função *findChangePathwayColor*

que recebe como parâmetros o nome da via selecionada e a nova cor. Esta função, também implementada neste trabalho, percorre o grafo em busca dos nodos que representam as proteínas da via e, quando encontrados, são alterados para a nova cor.

4.2.4 Selecionar e apagar todos os genes de uma via

Durante a implementação deste trabalho, notou-se a possibilidade de melhorar a função que identifica as vias Ontocancro no grafo do Medusa. Para isto foi desenvolvida uma função que seleciona todas as proteínas de uma determinada via no grafo gerado. Semelhante ao mostrado na Figura 20.

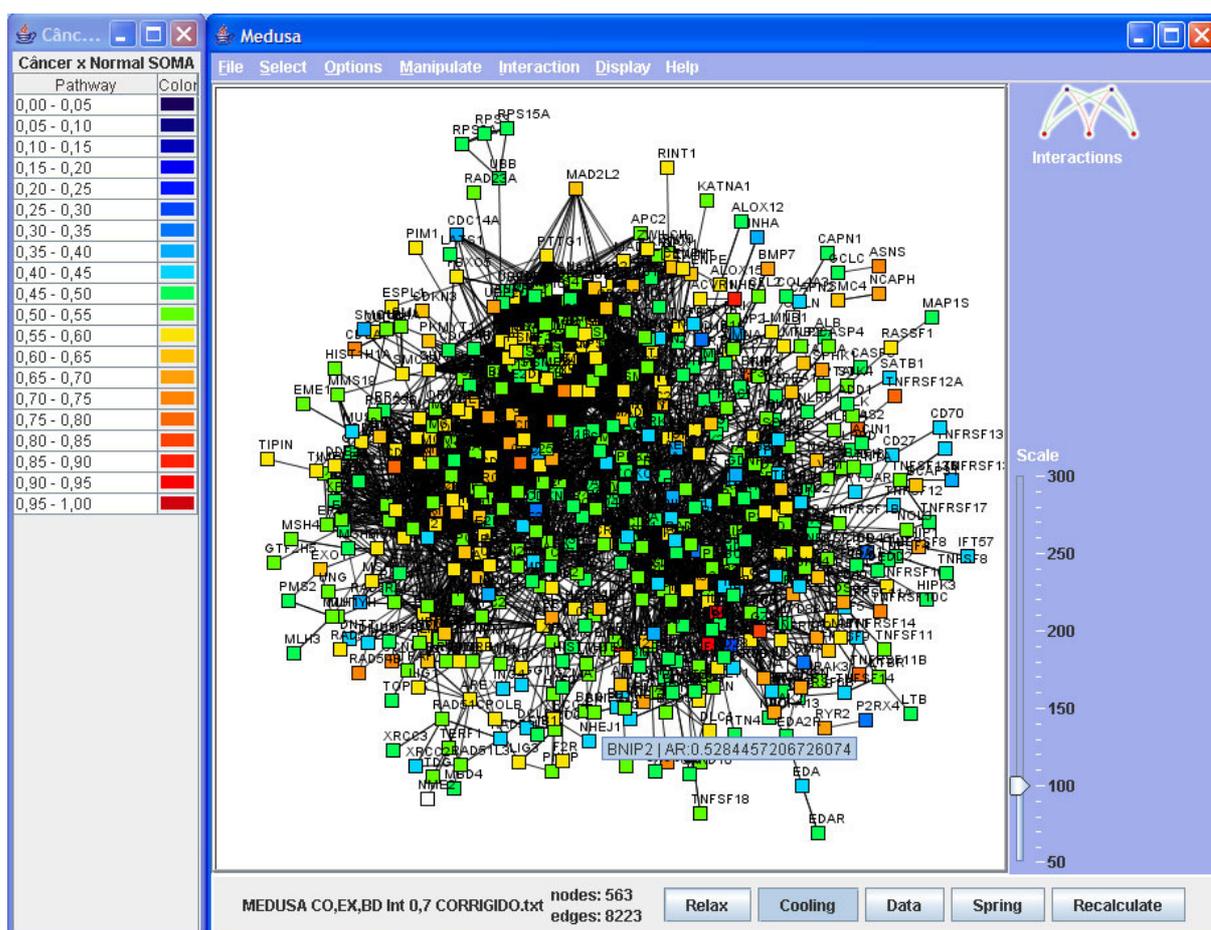


Figura 20 - Interface do Medusa com as proteínas de apoptose selecionadas.

Implementou-se a função *findPathway* que recebe o nome da via selecionada e retorna o valor booleano *true* quando é identificada todas a proteínas da via

selecionada. A função inicia percorrendo todo o grafo e para cada objeto da classe *Node* compara o valor retornado no método *getPathway()* com a via selecionada. Quando é verificado que o nodo pertence à via indicada, é então chamado o método *setFixed(true)* para o nodo em questão, sendo que este método já estava implementado pelos autores da ferramenta na classe *Node*. A função *findPathway* foi implementada na classe *DisplayPanel* do pacote *medusa.display*.

Para o usuário utilizar esta função basta clicar na via desejada, indicada no índice gerado pela função que identifica as via ontocancro no grafo.

Desta maneira é possível selecionar todos os nós da via e então apagá-los. Possibilitando assim, análises com proteínas pertencentes a determinadas vias. Para isto, o usuário seleciona a via desejada e então utiliza a função *Delete Nodes* no menu *Manipulate* do Medusa, conforme mostra a Figura 21.

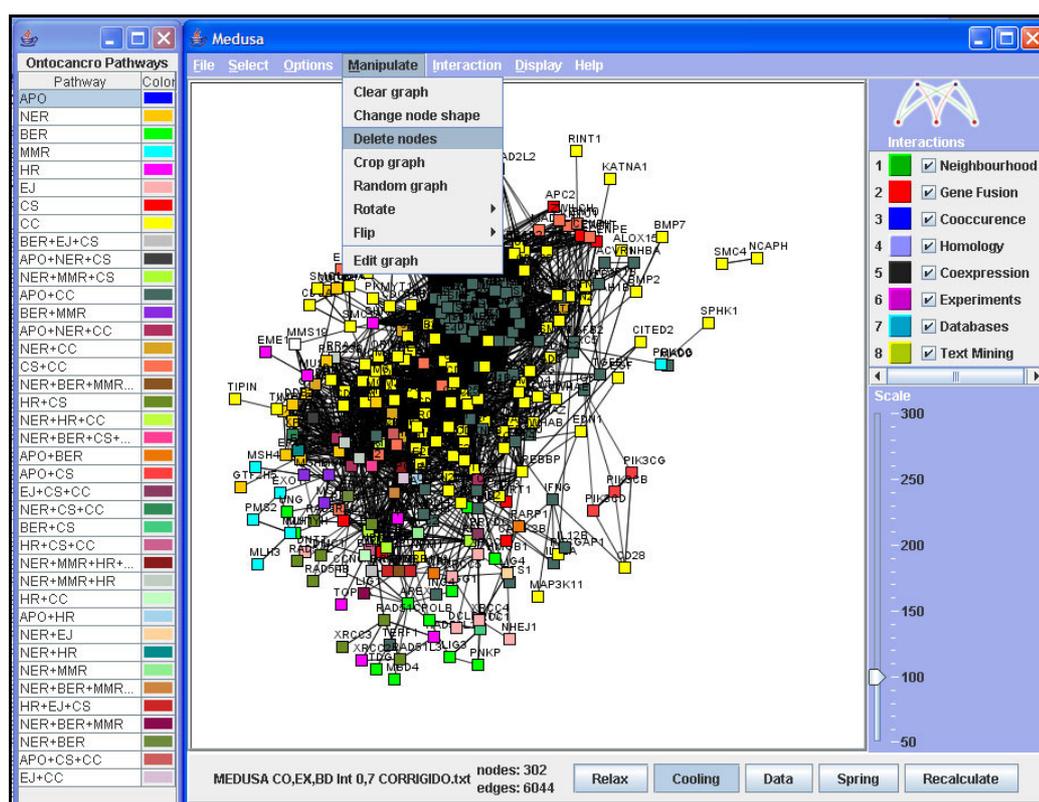


Figura 21 - Grafo com as proteínas de apoptose excluídas.

O Medusa permite que se aplique outro processo aos nodos selecionados, no menu *Manipulate* tem-se ainda a opção de alterar o formato dos nodos selecionados, no item *Change node shape*. Os formatos permitidos são quadrado (padrão), círculo, triângulo e diamante.

4.3 Identificar a atividade relativa das proteínas no Medusa

Outra análise necessária para os pesquisadores é identificar o grau de atividade relativa entre as proteínas envolvidas na amostra carregada no Medusa, identificando se uma proteína está mais ativa que a outra para casos particulares. Por exemplo, caso de amostras de células retiradas de tecidos com inflamação, adenoma ou câncer.

Este tipo de pesquisa também demanda o uso de mais ferramentas como planilhas eletrônicas para selecionar as proteínas desejadas e fazer o cálculo da expressão. Neste trabalho implementou-se a função que identifica a atividade relativa entre as proteínas no software Medusa, apresentando um resultado dentro das disponibilidades que o grafo apresentado na ferramenta oferece.

4.3.1 Alterações no Medusa para receber a nova função

Assim como na função que identifica as vias no grafo, para esta nova função também foi necessário alterar o código de algumas classes implementadas no Medusa.

Na classe *Node* acrescentou-se uma nova variável do tipo *Double* chamada de *atR*, que é carregada com o dado da atividade relativa da proteína que o nodo representa. Também implementou-se os métodos *setATr* e *getATr* para manipulação desta variável.

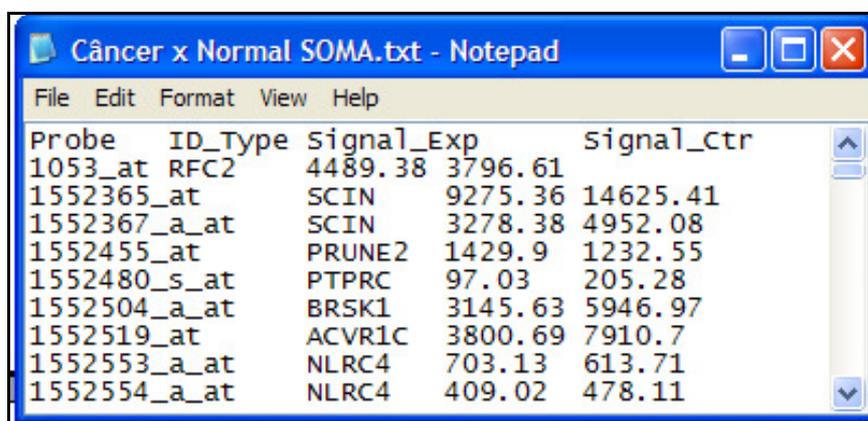
Igualmente na classe principal *MedusaFrame* do pacote *medusa* foi criado o menu de acesso a esta função, ficando dentro no menu Ontocancro junto com o item Ontocancro Pathways e representado pelo item “Expressão Genes”.

Na classe *DisplayPanel* do pacote *medusa.Display*, foi implementada a função *repaintExpr*, sendo esta a qual determina a funcionalidade implementada à ferramenta. E que está descrita nos itens a seguir.

4.3.2 Carregar os dados da atividade relativa

Estão disponíveis para os pesquisadores, no site da GEO os arquivos de expressão contendo resultados de estudos em tecidos, células ou DNA. Estes arquivos contêm dados suficientes para expressar a atividade dos genes envolvidos no grafo. Sendo estes dados, o identificador da prova, o nome da proteína, o sinal de expressão e o sinal de controle, conforme mostra a Figura 22.

Para ler este arquivo foi implementado um *parser* na função *repaintExpr*. Para receber os dados lidos pelo *parser* desenvolveu-se a classe *GeneExpr* que tem como atributos: o nome, o sinal de expressão e sinal de controle.



Probe	ID_Type	Signal_Exp	Signal_Ctr
1053_at	RFC2	4489.38	3796.61
1552365_at	SCIN	9275.36	14625.41
1552367_a_at	SCIN	3278.38	4952.08
1552455_at	PRUNE2	1429.9	1232.55
1552480_s_at	PTPRC	97.03	205.28
1552504_a_at	BRSK1	3145.63	5946.97
1552519_at	ACVR1C	3800.69	7910.7
1552553_a_at	NLRC4	703.13	613.71
1552554_a_at	NLRC4	409.02	478.11

Figura 22 – Arquivo usado para calcular a atividade relativa.

Pode ocorrer de algumas proteínas aparecerem mais de uma vez no arquivo. E para calcular a atividade relativa, todas essas ocorrências devem ser levadas em conta. Para isso os atributos de sinal de expressão e o de sinal de controle da classe são do tipo arraylist de valores *Double*. Podendo assim, ser carregado vários valores nestes campos.

4.3.3 Identificar a atividade relativa das proteínas

Depois de carregado todas as informações contidas no arquivo, é então calculado a atividade relativa para cada proteína. A fórmula da atividade relativa é dada pela média dos sinais de expressão da proteína dividida pela soma da média

dos sinais de expressão com a média dos sinais de controle da proteína, de acordo com a seguinte fórmula:

$$AR = \frac{\text{Média(Expressão)}}{\text{Média(Expressão)+Média(Controlo)}}$$

Para isto foi necessário criar três métodos na classe que retornam um valor do tipo ponto flutuante:

-*getMedExp()*: Retorna a soma dos valores do *arraylist* de sinais de expressão da proteína dividido pelo tamanho do vetor;

-*getMedCtr()*: Semelhante ao *getMedExp()*, porém soma os valores do *arraylist* de sinais de controle da proteína e divide pelo tamanho do vetor;

-*getActR()*: Retorna o valor da atividade relativa da proteína, dividindo o valor de *getMedExp()* pela soma de *getMedExp()* com *getMedCtr()*.

O valor retornado pelo *getActR()* é o grau de atividade relativa da proteína. Para isto foi criada uma tabela de forma gradual de cores para que depois de identificado os valores das atividades, os nodos referentes às proteínas sejam coloridos de acordo com essa tabela. A tabela consiste em uma graduação de 0 a 1, para pontos flutuantes, desta maneira é possível representar infinitos valores. O índice identifica os mais ativos com cores mais fortes como vermelho, e os menos ativos com a cor azul, de acordo com a Figura 23.

Pathway	Color
0,00 - 0,05	Dark Blue
0,05 - 0,10	Blue
0,10 - 0,15	Light Blue
0,15 - 0,20	Very Light Blue
0,20 - 0,25	Cyan
0,25 - 0,30	Light Cyan
0,30 - 0,35	Light Blue
0,35 - 0,40	Blue
0,40 - 0,45	Light Blue
0,45 - 0,50	Light Green
0,50 - 0,55	Green
0,55 - 0,60	Yellow-Green
0,60 - 0,65	Yellow
0,65 - 0,70	Orange
0,70 - 0,75	Light Orange
0,75 - 0,80	Orange
0,80 - 0,85	Red-Orange
0,85 - 0,90	Red
0,90 - 0,95	Dark Red
0,95 - 1,00	Red

Figura 23 – Graduação de cores para a atividade relativa.

Assim como na função que identifica as vias Ontocancro no Medusa, foi desabilitado o índice de Interações. Para que não ocorra confusão na identificação com os dois índices ao mesmo tempo na interface. Podendo o índice de Interações ser ativado novamente no menu *Display* e habilitando item *Interactions*.

A graduação ocorre dentro de uma margem de 0,05. Esta margem determina as cores que a atividade da proteína será representada pelo valor da sua atividade relativa.

Quando todos os valores são carregados, percorrem-se os nodos do grafo para identificar as proteínas envolvidas. Para cada proteína encontrada é relacionada sua atividade e então é colorida de acordo com a tabela de cores referente ao valor da atividade relativa. A Figura 24 mostra o resultado obtido no Medusa com esta função. Possibilita ainda que o pesquisador passe o mouse sobre o nodo desejado e obtenha na tela o nome da proteína e o valor de sua atividade relativa.

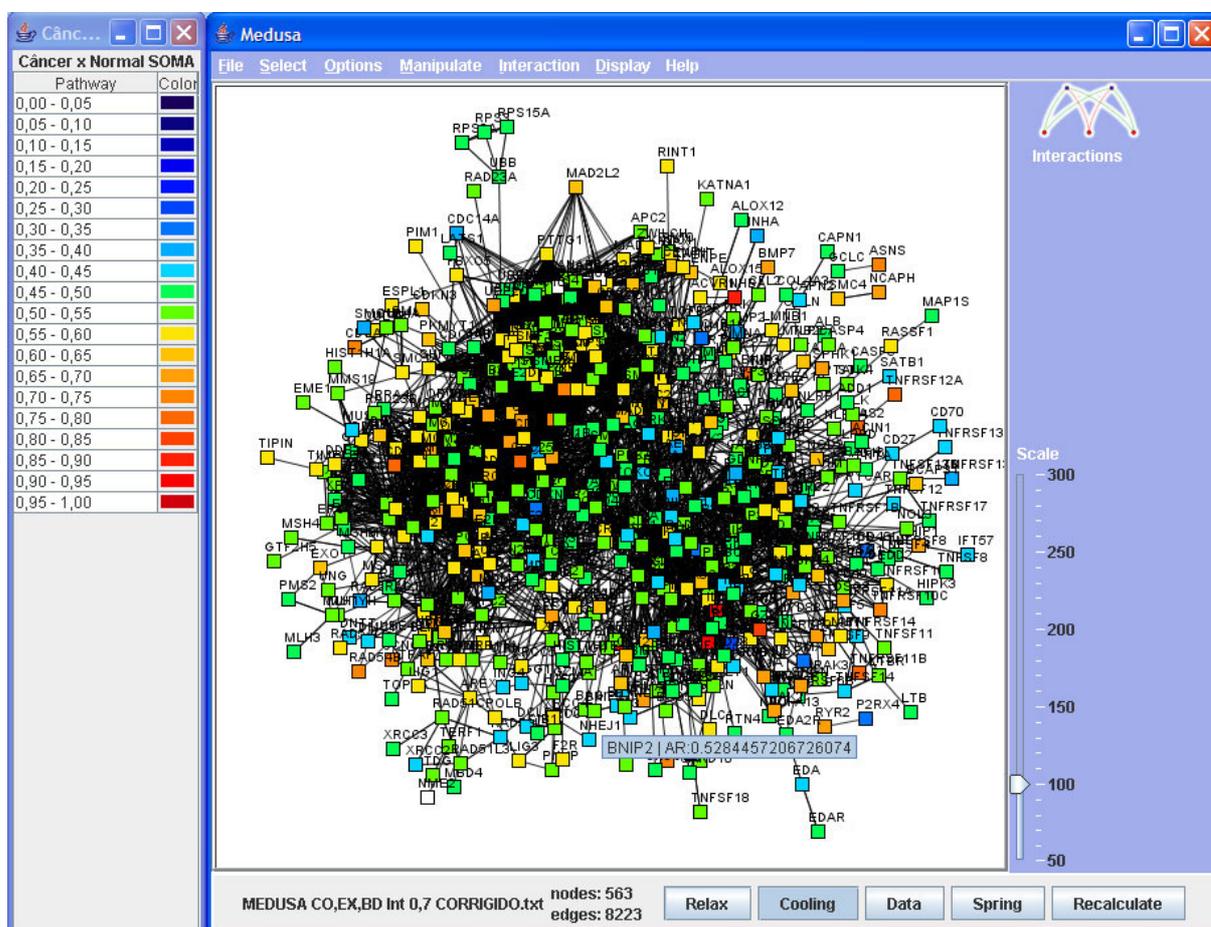


Figura 24 – Atividades relativas expressas no Medusa.

4.3.4 Selecionar e deletar os genes de acordo com a atividade relativa

Com a implementação da função que identifica o grau de atividade relativa das proteínas envolvidas no grafo, observou-se a necessidade de selecionar as proteínas inseridas em uma determinada margem do índice de atividades.

Sendo assim, implementou-se a função *findExpr* na classe *DisplayPanel* do pacote *medusa.display*. A função é chamada quando o usuário clica sobre um item na tabela do índice de atividade relativa. Inicia-se a procura em todos os nodos do grafo, comparando a atividade relativa obtida de cada nodo pelo método *getAtR*, implementado neste trabalho. Quando a *atividade relativa* do nodo em questão pertence a margem selecionada é chamada a função *setFixed* do nodo, para selecioná-lo. Quando terminada a busca, todos os nodos que foram selecionados estão prontos para serem manipulados (Figura 25).

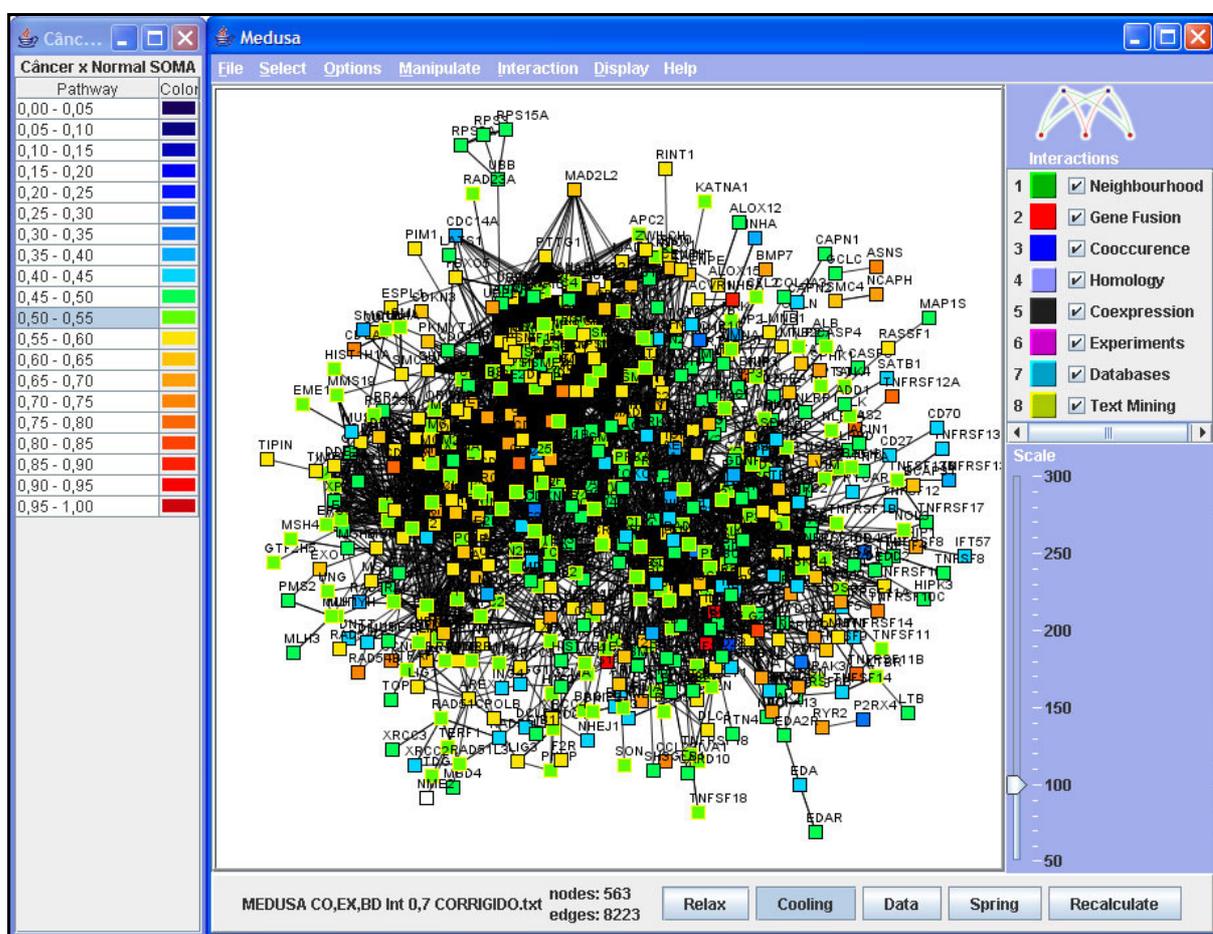


Figura 25 - Proteínas de atividade relativa entre 0.50 e 0.55 selecionadas.

Logo, é possível utilizar a função *Delete Nodes* do menu *Manipulate* para apagar os nodos selecionados. Tornando-se uma funcionalidade muito útil nos casos em que se deseja fazer uma análise do grafo retirando as proteínas menos expressas, por exemplo.

4.3.5 Gerar arquivo com as atividades relativas

Tendo em vista a necessidade de analisar as atividades relativas obtidas na amostra carregada no Medusa, criou-se uma função que gera um arquivo texto com a relação de todas as proteínas envolvidas no grafo com suas respectivas atividades relativas.

Esse arquivo só é possível ser gerado depois de carregado o arquivo contendo os dados com os sinais de expressão e controle. Para isto, o item “Salvar Expressão Genes” presente no menu Ontocancro só é habilitado depois que os dados de atividade relativa já foram gerados.

Depois de gerado, o arquivo pode ser manipulado em editores de texto ou planilhas eletrônicas. Possibilitando assim ordenar ou filtrar o arquivo da maneira que desejar. Por exemplo (Figura 26), ordenar a lista de proteínas pela atividade relativa, tendo no topo da lista a proteína mais ativa na amostra. O que é muito importante para a análise de uma grande amostra de proteínas.

	A	B	C	D
1	IL8	0.9791642427444458		
2	IL1B	0.9260755181312561		
3	IL6	0.90819251537323		
4	BCL2A1	0.907027542591095		
5	INHBA	0.8886247873306274		
6	GZMB	0.8003050088882446		
7	CDC45L	0.7826240658760071		
8	TNFRSF11	0.7781051397323608		
9	GAS2	0.7751649618148804		
10	BNIP3	0.7662078142166138		
11	CDC25B	0.757648229598999		
12	TNFRSF12	0.7556145191192627		
13	ASNS	0.7333166003227234		
14	RAD54B	0.7236078977584839		
15	PLK1	0.7203256487846375		
16	CDC2	0.7168808579444885		

Figura 26 - Arquivo das atividades relativas ordenado em uma planilha eletrônica.

4.4 Testes e dificuldades nas implementações

Todos os testes realizados nos itens implementados neste trabalho foram feitos também pelos pesquisadores que utilizarão estas ferramentas. O navegador Ontocancro foi testado nos principais *browsers* do mercado, como: Mozilla Firefox 3.5.6, Internet Explorer 7.0 e Opera 10.10, todos com o JRE instalados. Em todos os casos os objetivos foram alcançados com o navegador.

Para os testes das novas funções desenvolvidas no *software* Medusa foi utilizado um arquivo gerado no STRING com 563 proteínas e 8223 interações. E os equipamentos utilizados para realizar os testes possuíam as seguintes configurações de *hardware*:

- Processador Intel Centrino Duo 1.66 GHz, 1 GB de Memória RAM, 120 GB de espaço em disco;
- Processador Intel Centrino 1.6 GHz, 512 MB de Memória RAM, 60 GB de espaço em disco;
- Processador Intel Atom 1.3 GHz, 2GB de Memória RAM, 250 GB de espaço em disco.

Em todos os testes a ferramenta obteve um desempenho favorável.

Algumas dificuldades surgiram durante a implementação deste trabalho. Como foi mencionado, no desenvolvimento do *applet* navegador para o site do Ontocancro, uma das questões era a *interface* de interação com o usuário. A maneira como as informações estariam disponíveis para o pesquisador era um requisito importante para o sucesso da ferramenta.

Para isto, a ferramenta sofreu algumas modificações desde o início do desenvolvimento. Primeiramente foi desenvolvida como uma aplicação *off-line* para uso local. Neste momento era possível apenas abrir e visualizar os dados dos genes nas vias disponíveis no site do projeto Ontocancro, através dos arquivos texto.

Assim, verificou-se que o formato de árvore seria uma ótima escolha. Por ser uma interface semelhante à disponível na maioria dos gerenciadores de arquivo, que estão presentes nos principais SO atuais. Desta maneira o usuário tem certa familiaridade com a utilização da ferramenta, sem a necessidade de pré-apresentações.

No desenvolvimento da função que identifica as vias Ontocancro na ferramenta Medusa, realizou-se um estudo do código-fonte do *software* e suas classes para determinar onde seriam feitas as modificações na ferramenta. Os arquivos do código-fonte do Medusa estão bem organizados em pacotes de desenvolvimento, porém envolvem muitos arquivos. E isto demandou certo tempo para descobrir quais as classes deveriam ser alteradas.

Outra questão no desenvolvimento desta funcionalidade foi o fato de que uma proteína poderia pertencer para uma ou mais vias. Criando assim certa complexidade no procedimento implementado. A solução encontrada foi criar sub-vias com o nome das vias principais que as formam, e assim relacioná-las com os nodos das proteínas pertencentes a elas.

No momento de identificar as vias surgiu uma dúvida sobre a quantidade de cores que poderiam ser usadas, pois seria extenso o número de sub-vias. Para solucionar este problema, foi determinado cores para as 9 (nove) vias principais, e as sub-vias são cores diferentes entre elas. Também criou-se a possibilidade do usuário trocar as cores determinadas pela função. Tornando assim, a função mais completa.

Já na implementação da função que identifica a atividade relativa das proteínas envolvidas na amostra, era necessário fazer os cálculos da atividade relativa de determinada proteína com base em todas as provas que a mesma existisse no arquivo. Para isso foi necessário relacionar primeiro todas as provas que aquela determinada proteína aparecesse, para então calcular as médias de expressão e controle.

Também houve dúvidas quanto ao modo como representar as atividades relativas no grafo. O modo como o grafo está disposto na ferramenta não possibilitaria uma análise precisa da atividade relativa. Para isto foi utilizado uma tabela com graduação de cores, sendo o azul escuro representando os nodos menos ativos e o vermelho para os mais ativos. E quando o usuário seleciona um grau de atividade relativa no índice a função seleciona todos os nodos inseridos neste grau. Facilitando a identificação.

Por final, uma funcionalidade que não estava prevista no trabalho era gerar um arquivo com a lista das proteínas envolvidas na amostra e suas respectivas atividades relativas. Mas durante o desenvolvimento da função, viu-se a importância para os pesquisadores que a ferramenta oferecesse também este procedimento.

5 CONSIDERAÇÕES FINAIS

O câncer é uma das doenças mais preocupantes hoje em dia, com isso o investimento na busca de uma cura demanda muitas pesquisas nesta área. E a bioinformática agrega ferramentas que, cada dia mais, tornam-se imprescindíveis na contribuição destas pesquisas que se beneficiam com suas aplicações.

Uma destas ferramentas é o projeto Ontocancro, que se tornou uma importante fonte de pesquisa para os profissionais que estudam o processo carcinogênico. Por esta importância, este trabalho oferece novas facilidades na utilização da base de dados do Ontocancro. Agora, com o navegador Ontocancro que aqui foi implementado, o processo de estudo dos genes e vias relacionadas com o câncer se tornou mais ágil. Uma vez que os pesquisadores podem visualizar todas as informações das vias na mesma tela do computador, sem a necessidade de abrir outras páginas da *web*. Procedimento este, que se tornava repetitivo ao pesquisador.

Outra importante aplicação na área de bioinformática é o software Medusa, que foi desenvolvido para visualizar e manipular os grafos gerados a partir da base de dados do STRING. Apesar de ser uma ferramenta muito utilizada nesta área, muitas vezes os usuários precisavam trabalhar com outros aplicativos para concluir procedimentos de análise no grafo que o próprio Medusa gerava. Sendo esta uma ferramenta de código aberto, foi um fator que contribuiu para que se implementasse na própria aplicação novas funções que os pesquisadores necessitavam para que a ferramenta se tornasse mais completa.

Funções estas que agilizaram o processo de pesquisa e análise das amostras de proteínas. A primeira função implementada no Medusa, foi a que identifica as vias Ontocancro no grafo. Esta função vem a diminuir o número de aplicativos que eram necessários para o pesquisador obter o resultado final para a análise da amostra. Pois necessitava da utilização de planilhas eletrônicas e ferramentas de desenho para filtrar os dados e então redesenhar o grafo com as vias identificadas. Logo, a nova função implementada neste trabalho automatiza o procedimento que levaria dias ao pesquisador.

A função realmente cumpre com seu objetivo e faz ainda mais, pois fornece suporte para que o pesquisador possa manipular as vias no grafo, alterando a cor, o formato e inclusive deletando os nodos que representam as proteínas das vias.

Também foi possível agregar ao Medusa uma função que pode expressar no grafo a atividade relativa das proteínas envolvidas na amostra. Este tipo de análise, que também necessitava o uso de outras ferramentas, demandava tempo ao pesquisador tendo que abrir a mesma amostra em vários aplicativos para obter dados da atividade relativa das proteínas. Com a implementação feita neste trabalho da função que calcula a atividade relativa destas proteínas, o pesquisador pode rapidamente carregar o arquivo das provas com sinais de expressão e sinais de controle, e em alguns segundos obter de modo visual o grau de atividade relativa das proteínas envolvidas no grafo.

O desenvolvimento desta função permitiu que se fosse além do que foi proposto neste trabalho. Pois permite visualizar a atividade relativa das proteínas representadas pelos nodos no grafo, e também permite que o usuário manipule e/ou apague os nodos que estão envolvidos em determinados graus de atividade relativa, caso seja necessário. Além disso a função disponibiliza que seja gerado um arquivo contendo a lista de proteínas envolvidas na amostra com suas respectivas atividades relativas, vindo a contribuir ainda mais para os profissionais que estudam esta área. Pois este arquivo pode ser trabalhado em outros aplicativos, como planilhas eletrônicas para ordenar seus dados e/ou filtrá-los.

Nas alterações implantadas no Medusa, tomou-se o cuidado para que não alterasse o desempenho da ferramenta. Nos testes feitos no aplicativo, após as alterações, é possível verificar que obteve-se resultado satisfatório inclusive em processadores como o Intel Atom, desenvolvido para rodar aplicações que não necessitem um alto poder de processamento.

Tendo em vista os procedimentos que os profissionais, que estudam o comportamento do câncer, necessitavam fazer para obter o resultado final nas análises, conclui-se que o trabalho desenvolvido aqui está contribuindo muito com estes pesquisadores.

REFERÊNCIAS

CASTRO, M. A.; FILHO, J. L.; DALMOLIN R. J.; SINIGAGLIA M.; MOREIRA J. C.; MOMBACH J. C., de ALMEIRA R. M. **ViaComplex: software for landscape analysis of gene expression networks in genomic context**, Bioinformatics 2009. Disponível em <<http://lief.if.ufrgs.br/pub/biosoftwares/viacomplex>>. Acesso em: 11 de outubro de 2009.

FILHO, F. C.; PROSDOCIMI, F.; CERQUEIRA, G. C.; BINNECK, E.; SILVA, A. F.; REIS, A. N.; JUNQUEIRA, A. C. M., SANTOS, A. C. F.; JÚNIOR, A. N.; WUST, C. I.; KESSEDJIAN, J. L.; PETRETSKI, J. H., CAMARGO, L. P., FERREIRA, R. G. M.; LIMA, R. P.; PEREIRA, R. M.; JARDIM, S.; SAMPAIO, V. S., FOLGUERAS-FLATSCHART, A. V.. **Bioinformática: Manual do Usuário**, 2002. Disponível em: <<http://www.biotecnologia.com.br/revista/bio29/bioinf.pdf>>. Acesso em: 11 de outubro de 2009.

HOOPER, S. D.; BORK P. **Medusa: a simple tool for interaction graph analysis**. Bioinformatics 2005. Disponível em <<http://coot.embl.de/medusa>>. Acesso em: 11 de outubro de 2009.

JENSEN L. J.; KUHN M.; STARK M.; CHAFFRON S.; CREEVEY C.; MULLER J.; DOERKS T.; JULIEN P.; ROTH A.; SIMONOVIC M.; BORK P.; von MERING C. **STRING 8--a global view on proteins and their functional interactions in 630 organisms**, 2008. Disponível em <<http://string.embl.de/>>. Acesso em: 11 de outubro de 2009.

LESK, A. M. **Introdução à Bioinformática**, 2. ed. – Porto Alegre, Artmed, 2008.

LIBRELOTTO, G. R. ; MOMBACH, J. C. M. ; SINIGAGLIA, M. ; SIMÃO, E. ; CABRAL, H. B. ; CASTRO, M. A. A. **An Ontology to Integrate Transcriptomics**

and Interatomics Data involved in Gene Pathways of Genome Stability. BSB 2009, LNBI 5676, pp. 164-167, 2009.

MAGATÃO, Maria G. S.; JÚNIOR, Euclides F. S. **Educação para a Ciência: Uma proposta de intervenção diferenciada no ensino de biologia.** 2008. Disponível em <<http://www.diaadiaeducacao.pr.gov.br/portals/pde/arquivos/350-4.pdf>>. Acesso em 20 nov. 2009.

NCBI. **PubMed.gov**, rev. out. 2009. Disponível em <<http://www.ncbi.nlm.nih.gov/pubmed>>. Acesso em 20 nov. 2009.

NHGRI. **The Human Genome Project Completion: Frequently Asked Questions**, 2009. Disponível em <<http://www.genome.gov/11006943>>. Acesso em 20 de nov. 2009.

NLM. **The DNA Structure**, 2009. Disponível em <<http://ghr.nlm.nih.gov/handbook/illustrations/dnastructure>>. Acesso em 22 de nov. 2009.

SETUBAL, J. C.; Revista eletrônica ComCiência.br . **A origem e o sentido da bioinformática**, 2003. Disponível em <<http://www.comciencia.br/reportagens/bioinformatica/bio10.shtml>>. Acesso em 08 de out. 2009.

SIMÃO, E. M. ; SINIGAGLIA, M. ; MOMBACH, J. C. M.; CASTRO, M. A. A. ; ALMEIDA, R. M. C.. **Estudando alterações funcionais de vias de estabilidade genômica em pré-câncer usando o software ViaComplex.** In: II Escola Brasileira de Bioinformática, 2009, Porto Alegre RS. BSB 2009.