

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**

**A ONTOLOGIA ONTOCANCRO 3.0: REPRESENTANDO O
CONHECIMENTO DOS PROCESSOS INFLAMATÓRIOS**

DISSERTAÇÃO DE MESTRADO

Laís Falcade

Santa Maria, RS, Brasil

2015

A ONTOLOGIA ONTOCANCRO 3.0: REPRESENTANDO O CONHECIMENTO DOS PROCESSOS INFLAMATÓRIOS

Laís Falcade

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em
Informática da Universidade Federal de Santa Maria (UFSM, RS), como requisito
parcial para obtenção do grau de

Mestre em Informática.

Orientador: Dr. Giovani Rubert Librelotto

Santa Maria, RS, Brasil

2015

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**

A Comissão Examinadora, abaixo assinada, aprova a Dissertação de Mestrado.

**A ONTOLOGIA ONTOCANCRO 3.0: REPRESENTANDO O
CONHECIMENTO DOS PROCESSOS INFLAMATÓRIOS**

elaborada por

Laís Falcade

como requisito parcial para obtenção do grau de Mestre em Informática

Comissão Examinadora

Dr. Giovani Rubert Librelotto

(Presidente/Orientador)

Eder Maiquel Simão

(Pós-Doutor/ Centro Universitário Franciscano)

Luís Alvaro de Lima Silva

(Doutor/Universidade Federal de Santa Maria)

Santa Maria, RS, agosto de 2015.

“Se enxerguei mais longe, foi porque estava sobre os ombros de gigantes.”

Isaac Newton

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.”

Marthin Luther King.

AGRADECIMENTOS

Ao Prof. Dr. Giovani Rubert Librelotto, por ter me aceitado como sua aluna e orientanda, pelos seus ensinamentos e pelo suporte em todos os momentos difíceis que passei durante o período do mestrado, principalmente, por acreditar em mim.

Aos meus colegas do grupo de pesquisa da Ontocancro, em especial a Karlise Soares Nascimento por todas as vezes que precisei da sua opinião e dos seus esclarecimentos, durante e após a realização do seu mestrado no projeto. Também o Prof Dr. José Carlos Merino Mombach, o qual também agradeço os ensinamentos e orientações, pois foram de fundamental importância para a elaboração desta dissertação.

Ao Josmar Nuernberg, secretário do PPGI, por todas as vezes que precisei da sua ajuda.

Aos demais professores do PPGI pelos ensinamentos compartilhados.

À toda minha família, por estarem sempre ao meu lado, me incentivando e consolando, sempre com muito amor, principalmente a minha irmã que nunca desistiu de mim, sofreu junto comigo em todos os altos e baixos, derrotas e vitórias.

Aos amigos, que se mantiveram ao meu lado, mesmo quando eu não podia sair para tomar um chimarrão.

Ao meu amor, Grégori Braian Confortin, que mesmo eu não lhe dando toda a atenção que ele merecia, ainda achou que podia ficar comigo para sempre, te amo!

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Informática
Universidade Federal de Santa Maria

A ONTOLOGIA ONTOCANCRO 3.0: REPRESENTANDO O CONHECIMENTO DOS PROCESSOS INFLAMATÓRIOS

Autora: Laís Falcade

Orientador: Dr. Giovani Rubert Librelotto

Data e Local da Defesa: Santa Maria, 28 de agosto de 2015

No processo natural da vida, as pessoas nascem, crescem, envelhecem e morrem. A degeneração das células que incondicionalmente ocorre nesse processo é a incapacidade celular de se renovar, e dessa incapacidade resultam diversas doenças. Um tema pesquisado nesta área é o *Inflammaging*, que trata do estado inflamatório crônico que surge com o envelhecimento. Esta dissertação desenvolveu uma ontologia voltada ao estudo do *Inflammaging* unificada à *Ontocancro*. Propõe-se um mapeamento do conhecimento existente no âmbito genético do estado inflamatório e uma análise comparativa entre as vias do processo natural de desenvolvimento da célula com as vias de inflamação inseridas. Para verificar visualmente estas ligações, utilizou-se o *String*, um processador de ligação genética que disponibiliza grafos de relação entre genes, mostrando sua intensidade de relação. Com a comparação entre duas vias distintas, detectou-se que os genes de intersecção entre elas possuem ligações de grande intensidade num grau de 90% de confiabilidade dado pelo processador *String*.

Palavras-chave: Ontologia. Ontocancro. *Inflammaging*.

ABSTRACT

Dissertação de Mestrado
Programa de Pós-Graduação em Informática
Universidade Federal de Santa Maria

THE ONTOLOGY ONTOCANCRO 3.0: REPRESENTING THE KNOWLEDGE OF INFLAMMAGING

Autora: Laís Falcade

Orientador: Dr. Giovani Rubert Librelotto

Data e Local da Defesa: Santa Maria, 28 de agosto de 2015

In the natural process of life, people are born, grow, age and die. The degeneration of the cells that occurs in this process is unconditionally the inability to renew cell, and this failure is that live various diseases. A topic in this area is Inflammaging, which deals with the chronic inflammatory state that comes with aging. This paper developed of an ontology oriented to the study of Inflammaging unified to Ontocancro, It is proposed a mapping of existing knowledge in the genetic framework of inflammation and a comparative analysis between the tracks of the natural process of cell development with inflammation pathways inserted. To visually verify these connections used the String, a genetic link processor that provides graphs relationship between genes, showing the intensity ratio. With comparison between two pathways; detected genes of intersection between them have intensive connections to a degree of 90% reliability.

Keywords: Ontology. Ontocancro. Inflammaging.

LISTA DE FIGURAS

Figura 1 – Ontologia 1.0.....	23
Figura 2 – Ontocancro 2.0.....	25
Figura 3 – Classes da Ontocancro 3.0.....	30
Figura 4 – Propriedades de Objetos da Ontocancro 3.0.....	33
Figura 5 – Restrição de propriedades definida na classe Pathways.....	34
Figura 6 – Consulta de genes presentes em uma sample.....	36
Figura 7 – Classe Alzheimer em OWL.....	37
Figura 8 – Restrição da classe Genes com relação a classe <i>Affymetrics</i> em OWL.....	38
Figura 9 – Inserção da via Rb Tumor supressor Check. P. Sign. In Response to Damage em OWL.....	38
Figura 10 – Modelo relacional da atualização do banco de dados do banco de dados.....	41
Figura 11 – Pesquisa no GEO do câncer nas glândulas adrenais.....	45
Figura 12 – Pesquisa de vias no NCI-Pathways.....	47
Figura 13 – Pesquisa de vias no Reactome.....	48
Figura 14 – Script SQL de pesquisa da comparação no banco de dados.....	60
Figura 15 – Paridade da via Replicative Senescence.....	62
Figura 16 – Comparação das demais vias de inflamação.....	63
Figura 17 – Definição da String para a via <i>Replicative Senescence</i>	65
Figura 18 - Definição da String para a via Processing of DNA DSB ends Recruitment of Repair and Sig. Proteins.....	66
Figura 19 – União das vias (inflamação com reparo de DNA).....	67

Figura 20 – Definição da String para a via Hr Repair of Replication-Independent DSB.....	68
Figura 21 – União das vias Replicative Senescence e Hr Repair of Replication-Independent DSB.....	69

LISTA DE TABELAS

Tabela 1 – Tabela Genes do Banco de Dados Relacional.....	42
Tabela 2 – Tabela <i>Pathways</i> do Banco de Dados Relacional.....	43
Tabela 3 – Tabela <i>Samples</i> do Banco de Dados Relacional.....	43
Tabela 4 – Tabela <i>Inflammaging</i> do Banco de Dados Relacional.....	44
Tabela 5 – Propriedades das Ontologias.....	56
Tabela 6 – resultados obtidos através do script SQL.....	61

SUMÁRIO

1	Introdução	13
2	Bioinformática e <i>Inflammaging</i>	16
2.1	Tipos de doenças.....	17
2.2	Bancos de dados Biológicos	19
2.3	Sumário do Capítulo	21
3	ontocancro 1.0 e 2.0: representando o conhecimento referente ao câncer	23
3.1	Ontocancro 1.0	23
3.2	Ontocancro 2.0	25
3.3	Sumário do Capítulo	27
4	Ontocancro 3.0.....	28
4.1	Domínio e escopo da ontologia.....	28
4.2	Classes	29
4.3	Propriedade de dados.....	32
4.4	Propriedades de Objeto	33
4.5	Restrições	35
4.6	Instâncias.....	36
4.7	Consultas	37
4.8	Ontologia Ontocancro 3.0 em OWL	38
4.9	Sumário do Capítulo	40
5	Mapeamento da ontologia Ontocancro para banco de dados relacional.....	41
5.1	Definição da estrutura do banco de dados da Ontocancro	41
5.2	Atualização das instâncias para o banco de dados de acordo com o <i>Inflammaging</i>	45

5.3 Sumário do Capítulo	50
6 Comparação de ontologias relacionadas ao câncer.....	51
6.1 <i>Advancing Clinico-genomic Trials</i> (ACGT Master Ontology).....	51
6.2 <i>Systems Biology Ontology</i> (SBO)	52
6.3 TEDDY	54
6.4 Agreement Maker	54
6.5 <i>Neuropsychological Test Ontology</i> (NTO).....	56
6.6 Comparação entre as ontologias	56
6.7 Sumário do capítulo	59
7 Análise dos dados inseridos na Ontocancro	60
7.1 Relação das novas vias da ontologia com as existentes na Ontocancro 2.0	60
7.2 Comparação das vias da Ontocancro 3.0	61
7.3 Sumário do capítulo	64
8 Validação da ontologia proposta com o String	66
8.1 Replicative Senescence X Processing of DNA DSB ends Recruitment of Repair and Sig. Proteins	66
8.2 Replicative Senescence X Hr Repair of Replication-Independent DSB	69
8.3 Sumário do capítulo	72
9 Considerações finais	73

1 INTRODUÇÃO

Atualmente, é possível encontrar várias pesquisas sobre o envelhecimento, a preocupação com a longevidade e, acima de tudo, o bem-estar da população. Um tema que vem sendo pesquisado nos últimos anos nesta área é o *Inflammaging*, que trata do estado inflamatório crônico que surge com o envelhecimento (FRANCESCHI et. al, 2007). Dentre as doenças que podem ser catalogadas nessas especificações, é possível citar o Alzheimer, a Diabetes do Tipo 2, o Câncer (quando não hereditário), o Parkinson, a Artrite, dentre outras. O projeto Ontocancro, um ontologia que estuda o Câncer há seis anos, tem como foco principal as vias metabólicas alteradas no processo evolutivo da doença (NASCIMENTO et al., 2012) e que atualmente dispunha de um estudo paralelo sobre o Alzheimer (MOMBACH et al. 2015).

Ontologia pode ser considerada como um conjunto de axiomas lógicos projetados para explicar o significado de um vocabulário pretendido, podendo especificar uma conceituação única de forma muito indireta, uma vez que só podem aproximar-se de um conjunto de modelos destinados, e este conjunto pode ter apenas uma fraca caracterização de uma conceituação (GUARINO, 1998).

Este estudo visa a criação de uma ontologia para unificar os dados do projeto Ontocancro com as especificações do termo *Inflammaging* proposto por Francheschi et. al (2007) a fim de permitir um maior entendimento do mesmo. Este trabalho dá continuidade ao estudo tanto do Câncer, como do trabalho sobre Alzheimer desenvolvido pelo grupo de pesquisa da Ontocancro, que dispunha de informatas, físicos e biólogos que desenvolvem um trabalho interdisciplinar na busca pelo conhecimento.

A nova ontologia determinada Ontocancro 3.0 abrange o estudo de *Inflammaging*, o câncer mapeado na Ontocancro em suas versões anteriores, outras duas doenças causadas pelo processo inflamatório, o Alzheimer e a Diabetes do Tipo 2, levando em consideração as amostras das doenças advindas do processo

inflamatório (*samples*), as vias metabólicas (*pathways*) e os genes que estão disponíveis na Ontocancro e em outros bancos de dados, como o HGNC ou Hugo (o qual contém todos os genes da raça humana, o *National Cancer Institute* – NCI-*Pathways* (que traz as vias metabólicas) e o *Gene Expression Omnibus* – GEO (que apresenta diversas doenças e suas amostras de tecido).

O método de pesquisa desenvolvido nesse trabalho pode ser delineado como exploratório utilizando estudo de caso (SILVEIRA e CÓRDOVA, 2009). A pesquisa exploratória tem como foco principal “proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses”. Segundo Gil (2007), a grande maioria dessas pesquisas envolve: “(a) levantamento bibliográfico; (b) entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado; e (c) análise de exemplos que estimulem a compreensão”.

Quanto ao estudo de caso, Gil (2007) afirma que esta modalidade de pesquisa é amplamente usada nas ciências biomédicas e sociais. Por sua vez, Alves-Mazzotti (1998) cita que os exemplos mais comuns para esse tipo de estudo são os que focalizam apenas uma unidade, como um indivíduo, um pequeno grupo, uma instituição, e no nosso caso um termo da biologia, o *Inflammaging*.

Este trabalho está dividido da seguinte forma: no Capítulo 2 realiza-se um estudo sobre a bioinformática e o *Inflammaging*, sendo visualizados os tipos de doenças abrangidas neste trabalho e os bancos de dados que trazem as informações sobre o assunto. No Capítulo 3 expõe-se as versões anteriores da Ontocancro. No Capítulo 4 apresenta-se a reformulação da ontologia Ontocancro a partir do estudo sobre o *Inflammaging*. No Capítulo 5 visualiza-se o mapeamento da ontologia desenvolvida para o banco de dados disponível, envolvendo a estrutura e inserção das informações a partir da relação das classes da ontologia com as tabelas do banco de dados. No Capítulo 6 disponibiliza-se uma pesquisa comparativa entre ontologias voltadas para a bioinformática. No Capítulo 7 desenvolveu-se uma análise dos dados dispostos na Ontocancro a partir da atualização da ontologia, a fim de verificar quais genes se fazem presentes tanto nas vias de inflamação inseridas quanto nas vias metabólicas do processo natural das células, seguido do Capítulo 8

que realiza a validação dessas informações a partir do *String* que gera um grafo de ligação entre genes de duas vias distintas, mostrando a compatibilidade entre as mesmas. No Capítulo 9 deixam-se as considerações finais sobre o trabalho e, por fim, as referências utilizadas para este trabalho.

2 BIOINFORMÁTICA E *INFLAMMAGING*

A bioinformática surge como ciência na segunda metade da década de 1990, surgindo da necessidade de analisar de maneira mais eficiente, as sequências de DNA que alimentavam os bancos de dados (VOGT, 2003) (SANTOS, 2004).

Ela pode ser definida como uma modalidade que abrange todos os aspectos de aquisição, processamento, armazenamento, distribuição, análise e interpretação da informação biológica. Através da combinação de procedimentos e técnicas da matemática, estatística e ciência da computação são elaboradas várias ferramentas que auxiliam a compreender o significado biológico representado nos dados genômicos (SANTOS e OTEGA, 2006).

Com o grande número de informações geradas pelo Projeto Genoma e o sequenciamento das bases do DNA, o armazenamento e gerenciamento deste grande número de informações se tornou necessário, sendo criadas as Bases de Dados ou repositórios. Esses repositórios servem para “proporcionar consultas aos dados baseada nos metadados, com a capacidade de filtrar, juntar e concatenar através de conjuntos de dados heterogêneos” (PEZZINI, 2009). A partir desses repositórios genéricos, criaram-se vários bancos de dados que disponibilizam uma seleção de dados mais específica.

Atualmente o assunto que vem se destacando na área de bioinformática é o *Inflammaging*, que segundo Franceschi et al. (2007), define o estado de inflamação crônica que ocorre com o envelhecimento: “esse processo inflamatório é caracterizado por ser de pequena intensidade, sistêmico, crônico” (ROCHA et al., 2010), sistêmico porque é contínuo e regular; e crônico porque vem de longa data; acontece repetida e frequentemente. No entanto, Hunt et al. (2010) refletem que “apesar de nocivas, as mudanças que contribuem para o *Inflammaging* são parte do processo normal do envelhecimento”.

No processo natural da vida, as pessoas nascem, crescem, envelhecem e morrem. A degeneração das células que incondicionalmente ocorre nesse processo,

é na verdade, a incapacidade celular de se renovar, e dessa incapacidade resultam diversas doenças. Sabendo que todos os seres humanos envelhecem e, por consequência disto, estão propícios a processos inflamatórios mais frequentes do que na juventude, a busca da longevidade é um fascínio para muitos.

O *Inflammaging* nada mais é que um processo inflamatório, no qual todas as pessoas sofrem em consequência do envelhecimento. Essas consequências são doenças que se instauram a partir de um estado de imunodeficiência (baixa imunidade), causado pela própria senescência celular. Essa imunosenescência é causada pela alteração populacional das células T, ou linfócitos T, que são especializadas na imunidade celular e fazem parte dos glóbulos brancos. Essas células contêm um receptor específico chamado TRC, ou seja Receptor de Células T, que permite o reconhecimento de uma grande gama de antígeno para doenças. “A redução na diversidade das células T causa diminuição em respostas imunes perante antígenos novos, com os quais o indivíduo ainda não entrou em contato” (HUNT et. al., 2010).

Nesse sentido, pode-se verificar que com o envelhecimento, a produção de diversas células responsáveis por partes imprescindíveis ao estado de equilíbrio do organismo diminui, sendo resultado de muitos fatores. São células de imunidade que não se reproduzem mais como deveriam, grupos de genes que alteram sua atividade e que deveriam sofrer a apoptose (morte celular programada) e que não acontece, o ciclo celular e o reparo de DNA ficam falhos. Todos esses fatores influenciam no desenvolvimento de doenças, e a maioria dessas doenças vem do estado inflamatório causado pelo envelhecimento (ROCHA et al., 2010).

2.1 Tipos de doenças

As doenças analisadas neste trabalho incluem o câncer estudado pela Ontocancro, e que quando não é hereditário faz parte do processo inflamatório de tecidos corporais, o Alzheimer e a Diabetes do tipo 2 que também foram citadas por Franceschi et. al. (2007) como doenças advindas do *Inflammaging*.

Tendo em vista o Câncer e seu desenvolvimento, por motivos variados, uma célula em particular sofre algum tipo de mutação, que leva a um processo de multiplicação desordenada da mesma, causando um acúmulo em certa região do corpo. Esse acúmulo pode ser chamado de doença neoplásica ou Câncer (INCA, 2011).

Segundo Fulop et al. (2015) Moro e Horn (2013), a imunosenescência que ocorre pelo *Inflammaging*, pode contribuir para o aumento da incidência de tumores, sendo ele responsável pelo aumento de células inflamatórias e o desenvolvimento de radicais livres que levam ao câncer.

Na Doença de Alzheimer (DA), o primeiro aspecto a ser percebido é a falta de memória, seguido de dificuldades verbais e motoras, todos esses sintomas vem acompanhados de “distúrbios comportamentais, agressividade, hiperatividade, irritabilidade e depressão” (SERENIKI e VITAL, 2008) (ATHIÉ, 2010). O Alzheimer é uma doença neurodegenerativa associada à idade que atinge o cérebro de maneira progressiva e irreversível, levando a incapacitação do doente.

Segundo Giunta et al. (2008), o *Inflammaging* serve como um sinal anunciador ou um fator agravante para a doença do Alzheimer, sendo que tratamentos imunológicos utilizados nessa doença podem ser aplicados a outras doenças neurodegenerativas como a doença de Parkinson e Esclerose Múltipla. A característica principal que difere as pessoas que adquirem essas doenças com o passar dos anos das que não adquirem, são os mecanismos “*anti-inflammaging*”, que contrariam o efeito do tempo, sendo chamados pelo autor de “bom envelhecimento”.

A partir desse contexto, pode-se afirmar que as pessoas, ao passo que envelhecem, estão suscetíveis a doenças crônicas como o Alzheimer. Ela pode surgir a partir da imunocenesência agravada pelo *Inflammaging*, sendo que indivíduos com mecanismos anti-envelhecimento podem diminuir as chances de adquirir esse tipo de patologia.

A diabetes do tipo 2 (*Diabetes mellitus type 2*), também conhecida como diabetes tardia, é uma doença metabólica que tem como caracterização elevar o nível de glicose no sangue com o objetivo de dar resistência a insulina e conseqüentemente a sua insuficiência. Os sintomas clássicos dessa doença são muita sede, incontinência urinária e fome constante (WISSE, 2014).

“O controle da doença consiste basicamente em uma dieta regrada, uso de medicamentos orais ou injetáveis e na sua grande maioria, não exige o uso de insulina” (FERREIRA e CAMPOS, 2014). Segundo próprio autor, a maior dificuldade encontrada no tratamento dessa doença é manter a restrição alimentar.

Segundo Tonet (2007), a diabetes do tipo 2 possui um agente causador decorrente da “inflamação sistêmica e subclínica” também entendida como *Inflammaging*. Além disso, o desequilíbrio do sistema imunológico inato pode provocar o declínio da suscetibilidade à insulina, colaborando para o desenvolvimento da doença.

2.2 Bancos de dados Biológicos

O estudo biológico manipula grande quantidade de dados todos os dias, para isso, a tecnologia veio para agregar possibilidades maiores e mais ágeis quanto ao manuseio, a busca e a segurança ao guardar esse emaranhado de informações. Nesse contexto da bioinformática, pode-se encontrar vários repositórios de dados, bancos de dados e ontologias que auxiliam no trabalho de biólogos.

Os bancos de dados biológicos, em sua maioria, são públicos com informações técnicas, pois trazem em seu conteúdo informações biológicas curadas, ou seja, analisadas por um especialista da área, de forma manual ou não, sobre ramos diversos da biologia.

Em se tratando da biologia molecular, os bancos de dados avançaram mais a partir do término do projeto Genoma Humano em 2003, que sequenciou os três bilhões de bases que compõem o organismo humano.

Dentre os bancos de dados primários ou repositórios de dados, que possuem sequências de nucleotídeos, e que dão origem a outros bancos estão o NCBI – (*National Center for Biotechnology Information*) (WHEELER et al., 2006), banco de dados americano de sequências de DNA e proteínas, o EBI – (*European Bioinformatics Institute*) (LABARGA et al., 2007), banco de dados europeu de sequências de DNA, o DDBJ – (*DNA Data Bank of Japan*) (MIYAZAKI et al., 2003), banco de dados japonês de sequências de DNA, e o PDB – (*Protein Data Bank*) (BERNSTEIN et al., 1977) que armazena estruturas tridimensionais resolvidas de proteínas.

Esses repositórios oferecem em seus sites ligação para outros bancos gerados a partir de seu conteúdo, facilitando o acesso a dados específicos curados e armazenados em outras bases de dados.

O NCBI traz em sua base, estudos específicos como o GenBank, e o UniGene que disponibilizam dados curados manualmente sobre as funções de genes, e ligação com informações curadas em outros bancos, como a UniProt, que é um “consórcio que visa fornecer anotação relevante e curada de proteínas, baseada em dados de proteômica (principalmente), genômica e transcriptômica” (DANTAS, 2013), e o Reactome, o qual demonstra uma representação detalhada de processos biológicos e suas interações.

O *NCI-Nature* (SCHAEFER et al., 2009), um banco de dados com uma grande coleção de vias e suas interações, possui em sua base, informações curadas sobre interações biomoleculares e índices de processos celulares montados em vias de sinalização das doenças de Câncer. Advém de um projeto de colaboração entre o *National Cancer Institute* (NCI) e *Nature Publishing Group* (NPG). O *NCI-Nature Pathway* é um banco de interação financiado pelo governo americano, sendo de utilização pública e sem restrições

O banco de dados *Kegg Pathways* (KANEHISA e GOTO, 2000), traz em sua proposta, uma via metabólica e o processo ocorrido desde as células normais até o estágio avançado de uma doença, tudo em forma de mapa, contendo os principais

genes e as vias metabólicas normais ao processo, como vias de manutenção genômica, ciclo celular e apoptose.

O GEO – *Gene Expression Omnibus* (EDGAR et al., 2002) está disponível na *Internet* e armazena seus dados com base em três fontes principais: as plataformas (*platforms*), as amostras das doenças (*samples*) e as séries ou doenças. Ele disponibiliza aos pesquisadores uma interface *web* bastante amigável e de fácil exploração.

Dentre os bancos de dados encontrados na internet, há os que se utilizam de ontologias para seu desenvolvimento e tratamento de dados. Esses bancos buscam um padrão para a utilização dos dados curados por eles e por outros bancos, facilitando assim o acesso a informações.

Entre esses bancos, encontra-se o GO (*Gene Ontology Project*) (ASHBURNER et al., 2000) que fornece um vocabulário controlado para descrever genes e produtos gênicos de um organismo e o HUGO ou HGNC (*Gene Nomenclature Committee*) (POVEY et al., 2001) que criou uma ontologia da nomenclatura dos genes. No HUGO localizam-se todos os dados sobre os genes que são necessários para compreendê-los e que são analisados juntamente com sua família em uma via metabólica.

2.3 Sumário do Capítulo

Neste primeiro capítulo realizou-se uma pesquisa teórica sobre o *Inflammaging*, visando apresentar de forma classificada e padronizada as informações sobre o mesmo. Para este estudo, buscou-se definir os bancos de dados que dispunham de tais informações. Dentre os bancos de dados biológicos encontrados, priorizou-se o *National Cancer Institute* – NCI e o *Reactome*, que disponibilizam dados de vias metabólicas envolvidas no processo de desenvolvimento das doenças e dos estágios normais de cada célula, o HUGO, que disponibiliza todos os genes encontrados nos seres humanos, e o GEO que traz a doença com suas especificações, como as *samples* ou amostras de pacientes reais.

No próximo capítulo poderá ser visualizado o estudo das ontologias voltadas para a bioinformática, como é o caso da Ontocancro foco deste trabalho. Este capítulo apresentará também uma comparação entre essas ontologias.

3 ONTOCANCRO 1.0 E 2.0: REPRESENTANDO O CONHECIMENTO REFERENTE AO CÂNCER

Tendo em vista que os genes que fazem parte do processo inflamatório ainda não tenham sido unificados através de uma ontologia, a qual possibilita a padronização de informações (GRUBER, 1993), é de fundamental importância uma representação formal desses dados para facilitar o seu estudo por pesquisadores. Para que este estudo possa ser visualizado de forma mais direta, é necessária uma classificação dessas informações, tanto biológicas quanto genéticas, possibilitando que cientistas da área obtenham resultados com maior rendimento.

Este capítulo apresenta uma explanação sobre o processo de desenvolvimento da ontologia Ontocancro, como começou e como estava em suas versões prévias.

3.1 Ontocancro 1.0

A Ontocancro é um trabalho que vem sendo desenvolvido por pesquisadores da área da biologia, física e medicina em conjunto com cientistas da computação. Na etapa da Ontocancro 1.0 (CABRAL et al., 2011), realizou-se um estudo para centralizar os dados a fim de permitir uma análise consistente de informações extraídas de bancos de dados públicos referentes ao estudo de genes humanos potenciais ao câncer, buscando a padronização dos dados encontrados através de uma ontologia. Na Figura 1 pode-se visualizar a ontologia desenvolvida por Cabral (et al., 2011).

Ao final desse processo, chegou-se a um repositório de dados unificados que permitia a integração de redes de interação molecular do câncer com dados de expressões de genes envolvidos (PEREIRA et. al 2012), totalizando 1434 genes distribuídos em 131 vias de manutenção do genoma envolvidas no ciclo celular, resposta ao dano do DNA, apoptose e senescência.

3.2 Ontocancro 2.0

Na Ontocancro 2.0 (NASCIMENTO et al., 2012) realizaram-se investigações do processo de ativação da barreira anticâncer proposto pelo pesquisador Halazonetis et al. (2008), utilizando métodos quantitativos de expressão de genes e vias produzidos por técnicas usadas na biologia sistêmica (SIMÃO, 2012).

Nesta segunda etapa acrescentou-se a Ontocancro um cálculo estatístico que verificava as redes de manutenção do genoma, identificando as vias metabólicas que se encontravam mais ativas na fase inicial do câncer, possibilitando resultados que comprovassem a expressão de genes no impedimento do processo de evolução tumoral proposto por Halazonetis et al. (2008).

Outro fator importante para a atualização da Ontocancro surgiu da necessidade de integração de suas vias metabólicas com doenças relacionadas com o câncer, de forma a apresentar os genes relativamente expressos em processos cancerígenos. Tal reestruturação pode ser vista na Figura 2 e foi dividida em três momentos. O primeiro momento iniciou com a inclusão dos dados oriundos da técnica de microarranjos da empresa *Affymetrix*. O segundo momento ocorreu com a inclusão do algoritmo para cálculo da atividade e diversidade relativa desenvolvido por Pereira et al. (2012) e o terceiro momento ocorreu com a inclusão das informações da análise das expressões gênicas obtidas pela técnica de sequenciamento RNAseq da empresa *Illumina*.

são encontrados através do atributo *entrezgene*, que nada mais é que uma identificação única de cada gene, podendo ser encontrado a partir de um buscador chamado Entrez, mantido pelo NCBI. As *pathways* ou vias metabólicas dispostas na Ontocancro podem ser encontradas no banco de dados Reactome e no NCI.

3.3 Sumário do Capítulo

Neste capítulo apresentou-se as versões anteriores da Ontocancro. Com o estudo dessas versões e a análise sobre o termo *Inflammaging*, foi proposta a criação de uma nova ontologia de forma que biólogos que estudam os casos possam visualizar e ter uma padronização desses dados, facilitando o seu manuseio e extração de informações.

Nos capítulos que seguem, poderá ser visualizada a nova ontologia proposta por este estudo, bem como a sua adequação ao site já disponível para utilização de biólogos e físicos que trabalham no projeto e demais interessados.

4 ONTOCANCRO 3.0

Este capítulo apresenta a ontologia modelada para representar o estudo sobre o *Inflammaging* proposto por Franceschi et al. (2007), bem como uma forma de atualização da Ontocancro 2.0 que trazia informações sobre o câncer. Para o desenvolvimento da ontologia optou-se pela ferramenta Protégé (HORRIDGE, 2011), uma das mais utilizadas para criação e edição de ontologias.

O Protégé “é considerado uma das melhores ferramentas de construção de ontologias, possibilitando a interação entre outros sistemas que utilizam base de conhecimento para resolver problemas de um domínio” (LICHTENSTEIN e SIGULEM, 2008), é rápido e de fácil desenvolvimento, além de ter um amplo material de apoio e ser uma ferramenta *open source*, desenvolvido na Universidade de Stanford nos Estados Unidos, pelo departamento de informática médica.

A construção da ontologia se deu através dos seguintes passos, baseados na metodologia proposta por Noy e McGuinness (2001): (a) definição do domínio e escopo da ontologia; (b) criação e organização hierárquica das classes presentes na ontologia; (c) propriedades de dados, as quais têm como objetivo caracterizar suas entidades, bem como as propriedades de objetos, cujo papel é estabelecer relações entre as classes; (d) estabelecer as restrições destas propriedades; (e) criar as instâncias das classes; e (f) demonstrar algumas consultas possíveis. Nos subcapítulos a seguir serão apresentadas todas estas etapas.

4.1 Domínio e escopo da ontologia

A partir da necessidade de atualização dos dados da Ontocancro, realizou-se um estudo nos bancos de dados biológicos que disponibilizam as informações já curadas (analisados por especialistas em biologia), a procura de novas possibilidades de apresentação do estudo do câncer. Nesta pesquisa, encontrou-se a versão de Franceschi et al. (2007) sobre o *Inflammaging*, o qual trata da inflamação crônica que advém do envelhecimento causando doenças como o Alzheimer e

Parkinson. Outro estudo encontrado foi realizado por Barreto et al. (2011) onde afirmou-se que o câncer, ponto chave do projeto Ontocancro, quando não advém da hereditariedade pode ser desenvolvido por um processo inflamatório, que favorece o crescimento desordenado das células causando a patologia.

No contexto desse trabalho, a ontologia tem como proposta mapear o conhecimento existente no âmbito genético sobre o estudo de *Inflammaging*, disponibilizando conhecimentos sobre doenças caracterizadas neste processo que foram definidas a partir da revisão bibliográfica já descrita, e também utilizando informações obtidas através de reuniões com profissionais da área biomédica. As doenças especificadas para este domínio foram o Alzheimer, a Diabetes tipo 2, o COPD ou Transtorno Pulmonar Obstrutivo Crônico, o Câncer de Tireoide, de Pâncreas, de Cólon Retal e de Glândulas Adrenocorticais.

As informações dispostas na ontologia foram retiradas de bases de dados já curadas como é o caso do NCI e o Reactome que trazem as *pathways* ou vias metabólicas encontradas nas células, o GEO, que disponibiliza os dados sobre as doenças e suas amostras de tecido normal ou com alterações e o HUGO que disponibiliza os dados sobre todos os genes encontrados no corpo humano.

4.2 Classes

Na Ontocancro 1.0, o foco principal era reunir os dados sobre o câncer em uma ontologia, para que o estudo do mesmo pudesse ser realizado a partir de uma mesma base de dados. Na ontologia desenvolvida para a Ontocancro 1.0 existem duas classes principais: a *Entily* e a *UtilityClass*. A segunda refere-se a um estudo paralelo sobre o *ViaComplex* (CASTRO et al., 2009), uma aplicação *open source* que constrói mapas de redes de expressão gênica, já a primeira classe principal refere-se ao estudo dos dados genéticos voltados para o câncer da proposta de Cabral et al. (2011). No trabalho de Cabral, as classes principais são *genes*, *interaction*, *pathways* e *physicalentily*, dentre essas, as três primeiras migraram para a Ontocancro 2.0. Nesta, o foco principal era expor a barreira anticâncer proposta por Halazonetis et al.

(2008) sendo desenvolvido um cálculo de atividade e diversidade relativa que determinava quais os genes mais ativos ou presentes em cada doença.

A classe *Gene* expõe os genes mais significativos do processo cancerígeno, sendo eles vinculados a uma família ou via metabólica representada a partir da classe *Pathways*. A classe *Interaction* traz a combinação de via a via, possibilitando a divisão em grupos, esta separação apresenta quatro famílias, as vias de apoptose, as vias de ciclo celular, as vias de estabilidade genômica e as vias de reparo de DNA.

A classe *PhysicalEntity* trazia as informações de termos biológicos da genética como o DNA e o RNA. Essa classe não foi agregada na atualização do projeto por não trazer informações específicas para a proposta da Ontocancro 2.0, que era de incluir um estudo sobre os transcriptomas.

Na nova ontologia foram inseridas duas novas classes, a classe *Affymetrics* que relaciona os genes às doenças a partir das *Samples* – subclasse da classe *Série* –, ela contém o identificador da sonda obtida a partir da técnica de microarranjos exemplificado nos trabalhos de Pereira et al. (2012) e Nascimento et al. (2012) e a classe *Séries* que disponibiliza as amostras do estudo das doenças retiradas do banco de dados GEO. Esta atualização manteve o estudo sobre o *ViaComplex* descrito acima, sem fazer nenhuma alteração.

Para a atualização da ontologia nesta dissertação, foi modificado o objetivo do projeto que era o visualizar o câncer em todos os aspectos, para a inserção do estudo sobre a inflamação crônica causada pelo envelhecimento celular, o *Inflammaging*, que inclui o câncer, mas não somente ele. Muitas outras doenças fazem parte dessa família, as quais já foram citadas anteriormente, e cada uma delas tem especificações diferentes quanto às vias metabólicas e genes com maior expressividade nas alterações celulares. Pode-se visualizar as principais classes da nova ontologia na Figura 3.

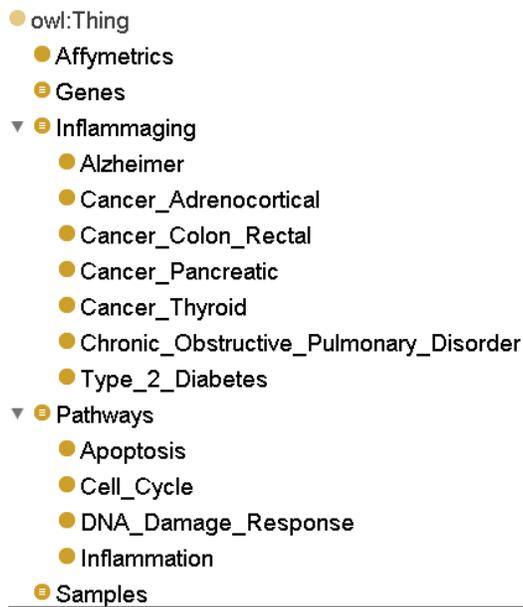


Figura 3 – Classes da Ontocancro 3.0

Nesta nova ontologia, comparada a versão anterior da Ontocancro, pode-se visualizar que não há uma classe *Série*, mas sim uma classe abstrata denominada *Inflammaging*. Desta classe abstrata dividiram-se as subclasses para cada tipo de doença estudada na Ontocancro, sendo possível instanciar mais de um estudo para cada doença. Este processo ocasionou a separação da classe *Sample* de sua classe pai (*Série*), sendo elas ligadas a partir de uma relação/ restrição dentro da ontologia.

A escolha dos estudos das doenças a serem abrangidas pelo projeto estão de acordo com pesquisas bibliográficas e reuniões com o grupo da biologia que faz parte da Ontocancro.

Há também a modificação na classe *Sample*, que era uma subclasse de *Série*. Nessa atualização, a subclasse *Sample* foi definida como classe, sendo ligada às doenças, ou séries, a partir de relacionamentos e instâncias, facilitando a união da mesma com a classe *Affymetrics*.

As classes básicas do projeto continuam sendo *Pathways* e *Genes*, por tratarem do estudo principal do grupo. Este estudo aborda a alteração que um ou mais genes causam a uma via metabólica, como a alteração dessa via ou *pathway*

pode afetar uma *sample* ou amostra de tecido humano, podendo assim chegar a uma futura prevenção da doença a partir de seu estágio inicial de alterações.

4.3 Propriedade de dados

As propriedades de dados de uma classe são as características utilizadas para descrevê-las, diferenciando seus indivíduos. A classe *Gene*, por exemplo, possui as propriedades de *entrezGene*, *hugoName* e *nciName*. Logo, suas subclasses também possuem essas propriedades.

O *entrezGene*, da classe *Genes*, vem do Entrez (*Global Query Cross-Database Search System*) que é um metabuscador mantido pelo *National Center for Biotechnology Information* (NCBI). Entrez inclui registros de genomas que foram totalmente sequenciados. Representa o resultado curado e automatizado da integração de dados do projeto Sequência de Referência do NCBI – *RefSeq* (MAGLOTT, 2007). Na Ontocancro ele representa o identificador principal do gene. O *hugoName* e o *nciName* são os nomes estabelecidos pelas bases de dados da HUGO e do NCI respectivamente.

No caso da ontologia desenvolvida neste trabalho, a classe *Genes* é que apresenta o maior número de propriedade de dados, pois determina a classificação de cada gene disposto nos estudos das doenças. Dentre as propriedades encontram-se o *approvedSymbol*, o *approveSymbol2*, e o *previousSymbols* todos advindos do banco de dados do HUGO, que apresenta as informações gerais de todos os genes. Essas propriedades determinam as várias versões ou atualizações dos símbolos de cada gene, tendo em vista que esses símbolos podem ser diferentes em bases de dados diversas, apesar de tratarem do mesmo gene.

Também são especificados os identificadores desses genes para cada banco de dados estudado a partir da versão 1.0 da Ontocancro, uma vez que o objetivo principal dessa versão era construir uma ontologia que unificasse em uma só base todas as informações relevantes no estudo do câncer. As propriedades de Identificador são: o *hgnc*, o *keggID*, o *reactomeID*, o *refseqID*, e o *unigenelD*

advindos dos bancos de dados HUGO, KEGG, *Reactome*, *RefSeq* e *Unigene*, respectivamente, nos quais também são identificadores.

Na classe *Pathways* podemos encontrar as características de cada via metabólica, como de que base de dados foi retirado (*db_name*), qual seu nome, seu apelido ou abreviatura de nome (*pathway_name* e *pathway_apelido* respectivamente), e qual a família ou grupo pertence (*pathway_type*).

A classe *Inflammaging* não possui propriedade de objetos por ser determinada como abstrata. Dentro desta classe foram inseridas classes filha para cada tipo de doença estudada, como a classe *Alzheimer* e *Cancer_Thyroid*. As classes filha possuem suas propriedades de objetos que identificam qual o estudo (*title*), uma publicação ou artigo sobre o estudo (*pmid*), um identificador *PubMed* que compreende mais de vinte e quatro milhões de citações de literatura biomédica de revistas de ciências da vida e livros on-line.

Na classe *Sample* pode-se citar o *p_value*, um identificador de atividade do gene. Ele mostra o quanto um gene está ativo ou não. A propriedade *id_ref* desta mesma classe representa a propriedade *affymetrics* da classe *Affymetrics*. Partindo disso, a classe *Affymetrics* apresenta duas propriedades de dados, a *affymetrics* e o *entrezgene*, anteriormente explicitados; sendo assim, esta classe faz a relação das classes *Samples* com *Genes*.

4.4 Propriedades de Objeto

As propriedades de objetos de uma ontologia têm o papel de relacionar indivíduos entre as classes. Para exemplificar o uso de relações na ontologia criada, uma doença tem amostras e estas amostras são definidas pelas sondas de pesquisa *affymetrics*. Para sabermos quais *affymetrics* fazem parte de uma doença, relacionamos a classe *Samples* com a classe *Affymetric* e a classe *Samples* com a classe *Inflammaging*. como por exemplo, uma *Samples* contém (*contains_affymetrics*) múltiplas instâncias de *affymetrics* e uma doença da classe

Inflammaging (contains_samples) contém múltiplas instâncias de *Samples*. A Figura 4 apresenta as propriedades de objetos ou relacionamentos da ontologia.

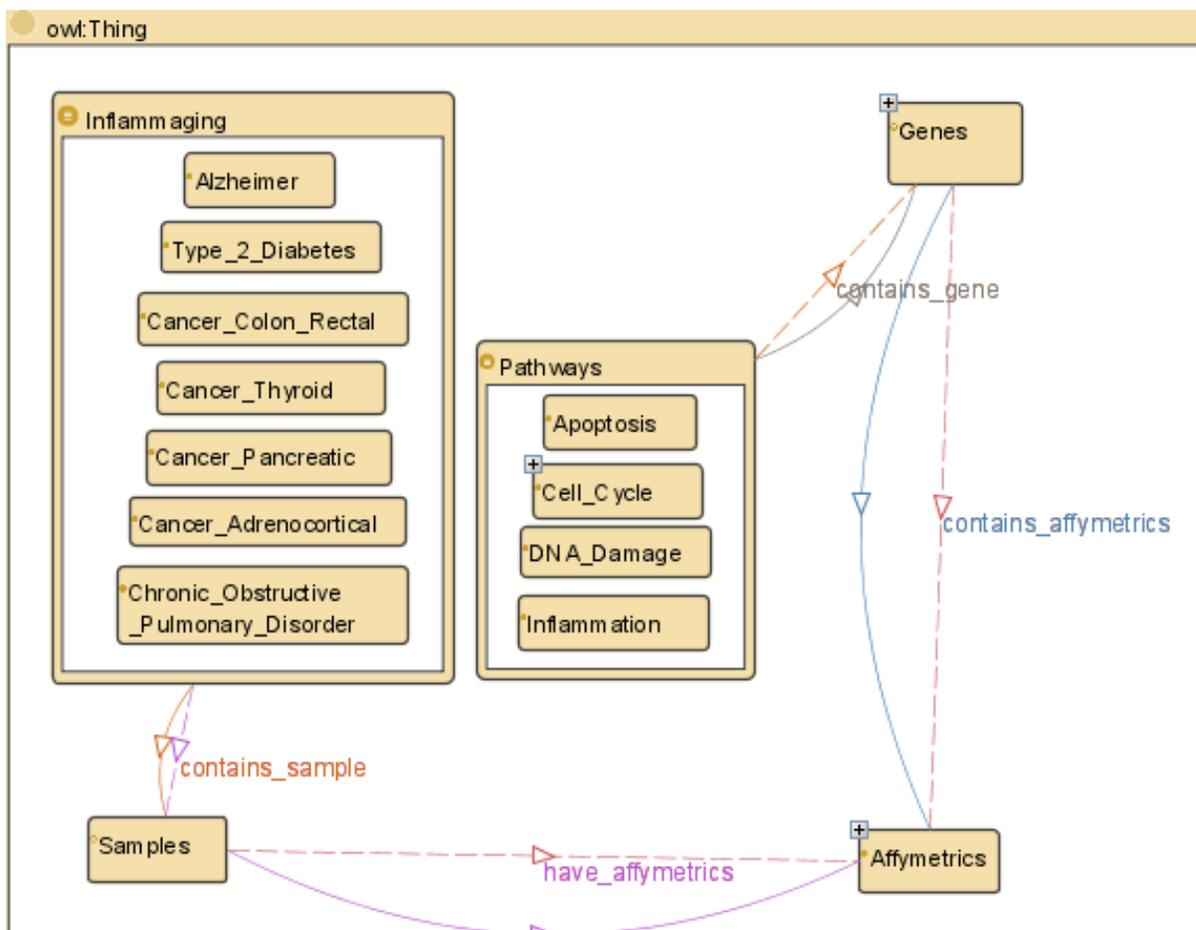


Figura 4 – Propriedades de Objetos da Ontocancro 3.0

Como pode ser visualizado na Figura 4, cada doença possui suas determinadas amostras de tecido, e cada amostra contém sondas (*Affymetrics*). As outras classes também possuem relacionamentos que determinam suas ligações.

Pela expressividade da linguagem, algumas propriedades podem ser utilizadas por mais de uma classe. Podemos dizer que uma *sample* contém várias instâncias de *affymetrics*, bem como também as *pathways* contém várias instâncias de genes. Para tanto, é necessário determinar restrições nas classes da ontologia, afim de evitar incoerências nesse tipo de caso.

4.5 Restrições

As restrições têm a função de limitar os indivíduos que pertencem a determinadas classes. Logo, é possível criar classes abstratas que irão englobar diversos indivíduos que satisfazem determinada restrição. Em Web Ontology Language (OWL) existem três tipos de restrições de propriedades: restrições de quantificador, restrições de cardinalidade e restrições de valor (HORRIDGE, 2011). O primeiro tipo é usado quando os indivíduos de uma classe devem possuir pelo menos um valor (*someValuesFrom*), ou somente valores de determinado conjunto (*onlyValuesFrom*).

Um exemplo pode ser visto no relacionamento entre a classe *Genes* e *Affymetrics*, no qual todo *gene* deve conter pelo menos um valor de *affymetrics* através da propriedade *contains_affymetrics*. Esta restrição pode ser visualizada na Figura 5.

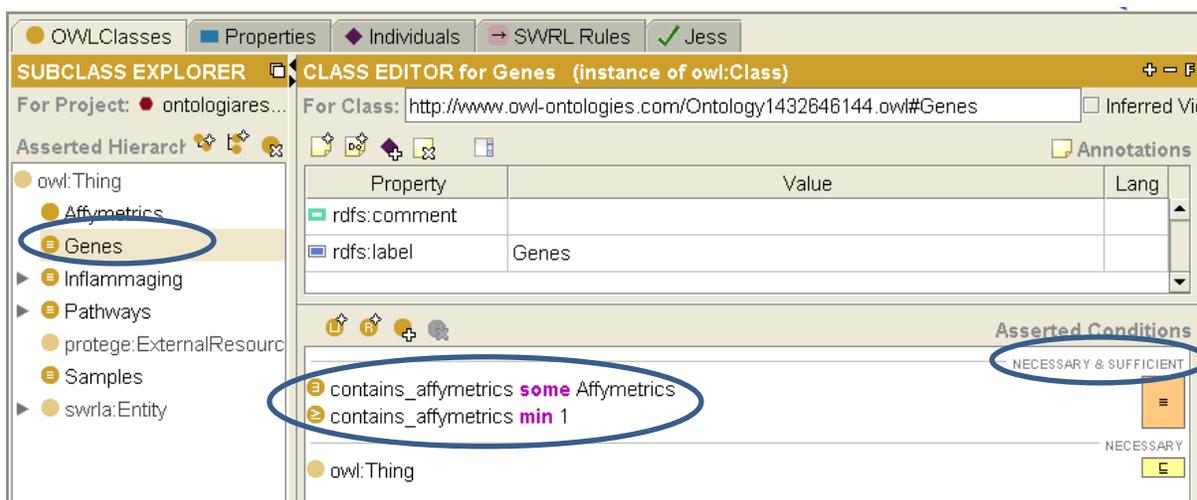


Figura 5 – Restrição de propriedades definida na classe *Pathways*.

Por sua vez, as restrições de cardinalidade são utilizadas para restringir a quantidade de valores que uma propriedade pode assumir. Na restrição acima, cada gene deve conter no mínimo um *affymetrics*, ou seja, a cardinalidade definida é

exatamente 1 *cardinality*. Ainda podem existir restrições de cardinalidade mínima *minCardinality* e máxima *maxCardinality*.

As restrições de valor descrevem indivíduos que estão relacionados a outros indivíduos especificados por uma propriedade. Trata-se de uma situação diferente da restrição de quantificador, onde os indivíduos descritos pela restrição estão relacionados a qualquer indivíduo de uma classe, através de propriedade específica. Na ontologia definida, pode ser determinada uma restrição de valor para as *samples*, onde uma amostra pertence a somente uma doença.

4.6 Instâncias

O último passo para o desenvolvimento da ontologia versa a criação das instâncias. As instâncias em uma ontologia representam os objetos do domínio em questão. Cada instância dentro da ontologia deve pertencer a uma determinada classe, e os indivíduos do mesmo tipo devem ser colocados em uma mesma classe. Por exemplo existem dois genes de *Pathways* diferentes, no entanto os dois serão alocados na classe *Genes*.

Segundo o estudo dos pesquisadores do projeto, as informações visadas para a busca de conhecimentos são pesquisadas nas bases de dados biológicas citadas na seção 2.2 desta dissertação e instanciadas nas suas classes de referência.

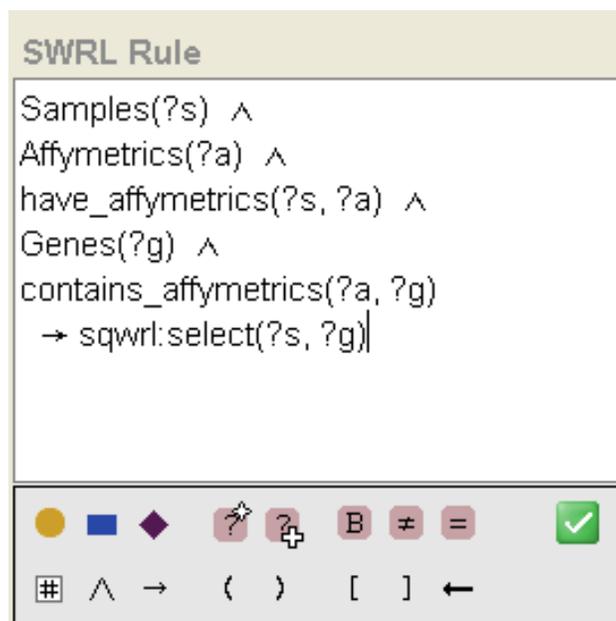
No caso das doenças, o banco de dados utilizado é o GEO, onde podem ser encontradas as séries ou doenças apresentadas na classe *Inflammaging*. É também no GEO que são encontradas todas as amostras ou *samples* dispostas na classe *Samples*. Para as vias, existem duas bases de dados principais, o Reactome e o NCI, delas são retiradas as propriedades de dados de cada via, como o id, o nome, o apelido e o tipo. Nesses bancos são encontrados também as instâncias das propriedades de objetos, que para as *pathways* são os genes relacionados. Como cada via tem acoplada a si os seus genes correspondentes, maiores informações sobre eles, para a disposição na tabela *Genes*, é realizada uma pesquisa desses

genes na base de dados HUGO, assim podem ser encontradas mais detalhadamente as especificações como *previousSymbol* e *approvedSymbol*.

4.7 Consultas

Em ontologias, as consultas são utilizadas para obter dados relevantes ao domínio que está sendo representado. No caso da Ontocancro, determinados tipos de consultas podem ser realizadas, como a verificação de uma relação de genes presentes em vias, bem como em quais doenças esse gene está envolvido.

Para saber quais vias contém um determinado gene, pode-se escolher a classe *Pathways*, determinar a propriedade contém gene (*contains_gene*) e escolher qual gene desejamos encontrar a partir de seu ID. Nas consultas também é possível verificar informações que não estão disponíveis na mesma classe, como é o caso da consulta realizada no Protégé, apresentada na Figura 6.



```
SWRL Rule
Samples(?s) ^
Affymetrics(?a) ^
have_affymetrics(?s, ?a) ^
Genes(?g) ^
contains_affymetrics(?a, ?g)
→ sqwrl:select(?s, ?g)
```

Figura 6 – Consulta de genes presentes em uma *sample*

Esta consulta busca através do *affymetrics* (?a) presentes nas *samples* (?s) os seus respectivos genes (?g). Apresenta como resultado os genes presentes nas

samples (sqwrl:select(?s, ?g)). As consultas podem ser guardadas para serem utilizadas após as atualizações da ontologia e inserções de novas informações.

A consulta é criada na aba *SQWRL Rules* do Protégé e após salva pode ser executada a partir do *plugin do Jess* instalado dentro do programa. O *Jess Tab* é um mecanismo de regras rápido e leve escrito em Java e é capaz de "raciocinar" usando regras declarativas fornecidas pelo usuário, regras como a descrita na Figura 6.

4.8 Ontologia Ontocancro 3.0 em OWL

Como visto anteriormente, para desenvolver a ontologia foi utilizado a ferramenta Protégé, que possibilita organizar a ontologia de maneira mais amigável, contendo uma interface. No entanto, todo arquivo gerado pelo Protégé pode ser exportado para um arquivo XML/OWL, que define todas as alterações propostas pela interface, no entanto em forma de código.

```
</owl:Class>
  <owl:Class rdf:ID="Alzheimer">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Alzheimer</rdfs:label>
    <rdfs:subClassOf rdf:resource="#Inflammaging"/>
  </owl:Class>
```

Figura 7 – Classe Alzheimer em OWL

No código expresso na Figura 7, visualiza-se a definição da classe “Alzheimer” que é uma subclasse de *Inflammaging*. No código da Figura 8, visualiza-se a restrição apresentada na Figura 5. Nos trechos destacados em negrito, apresenta-se que um gene deve conter valores de *affymetrics*, com uma cardinalidade mínima de um indivíduo.

```

<owl:Restriction>
  <owl:onProperty>
    <owl:ObjectProperty rdf:ID="contains_affymetrics"/>
  </owl:onProperty>
  <owl:someValuesFrom rdf:resource="#Affymetrics"/>
</owl:Restriction>
[...]
```

```

  <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
  >1</owl:minCardinality>
</owl:Restriction>
```

Figura 8 – Restrição da classe Genes com relação a classe *Affymetrics* em OWL

No código da Figura 9 poderá ser visualizada a inserção da *pathway* *Rb_Tumor_suppressor_Check._P._Sign._in_Response_to_Damage* do tipo ciclo celular. Esta inserção contém o gene *_1026* disposto no banco de dados sobre esta via e o *affymetrics* vinculado a este gene.

```

<Cell_Cycle rdf:ID="Rb_Tumor_suppressor_Check._P._Sign._in_Response_to_Damage">
  <references rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >http://www.biocarta.com/pathfiles/h_rbPathway.asp</references>
  <contains_gene>
    <Genes rdf:ID="_1026">
      <ensp rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
[...]
```

```

    <contains_affymetrics>
      <Affymetrics rdf:ID="_202284_s_at">
        <entrezgene
        </Genes>
    </contains_gene>
```

Figura 9 – Inserção da via Rb Tumor supressor Check. P. Sign. In Response to Damage em OWL

A ontologia desenvolvida em OWL possui 1823 linhas de código, incluindo todas as especificações de classes, subclasses, propriedades de classe e de objetos, restrições, consultas e algumas instâncias. Para um arquivo total, com todas as instancias do projeto, o arquivo OWL passa de 10 mil linhas de código, sendo a

parte mais extensa os arquivos das *samples* das doenças, que chegam a 60 mil *tuplas* por tabela.

4.9 Sumário do Capítulo

Este capítulo apresentou detalhadamente a ontologia criada para atualização do projeto Ontocancro com o intuito de inserir o estudo sobre *Inflammaging* no foco do grupo de pesquisa. Além das classes criadas, foram discutidas as propriedades tanto de dados quanto de objetos e a criação de instâncias.

A partir dessa nova proposta foram atualizados os dados do projeto e inseridas as informações do *Inflammaging*. Como essa inserção e atualização foi realizada é o foco o próximo capítulo deste trabalho.

5 MAPEAMENTO DA ONTOLOGIA ONTOCANCRO PARA BANCO DE DADOS RELACIONAL

A partir da ontologia desenvolvida no capítulo anterior, definiu-se o banco de dados relacional contendo os dados obtidos pelo projeto. Este banco é desenvolvido em *phpMyAdmin* (DELISLE, 2009), um Sistema Gerenciador de Banco de Dados – SGBD desenvolvido em PHP para a administração do MySQL na *web* facilitando a sua ligação com o site, que desde a primeira versão da Ontocancro estava disposto na internet para o acesso aos dados pelo grupo de pesquisa e demais interessados no estudo.

Este capítulo expõe o banco de dados em concordância com a ontologia apresentada no capítulo anterior, sendo apresentados o modelo relacional e a distribuição das tabelas no banco de dados.

5.1 Definição da estrutura do banco de dados da Ontocancro

Tendo em vista que a pesquisa já dispunha de um banco de dados, o mesmo foi atualizado para ficar de acordo com a ontologia desenvolvida para o estudo do *Inflammaging*. Na Figura 10 verifica-se o modelo relacional da atualização do banco de dados.

Na especificação relacional do banco de dados, foi criada uma tabela para as doenças relacionadas ao *Inflammaging*. Na ontologia a classe *Inflammaging* é abstrata, sendo ela pai de todas as doenças relacionadas, foi descrito uma especialização contendo uma tabela por hierarquia, ou seja, todas as tabelas referentes às especializações de uma entidade genérica são fundidas em uma única tabela (HEUSER, 1998). Portanto, no banco de dados, as doenças da classe *Inflammaging* ficarão expressas em uma única tabela (*Inflammaging*) do banco de dados, tendo em vista que os atributos ou instâncias são os mesmos para cada uma.

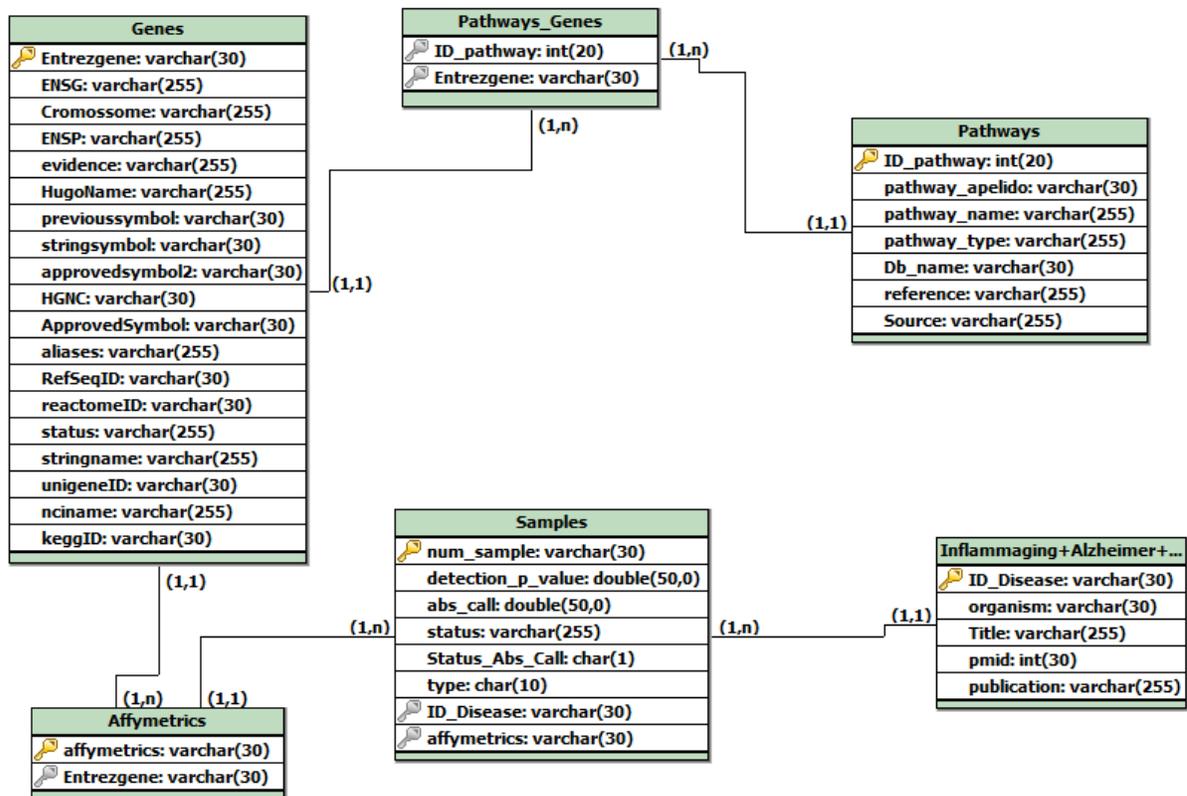


Figura 10 – Modelo relacional da atualização do banco de dados do banco de dados

A classe *Samples* da ontologia tornou-se uma tabela do banco de dados que guardará as informações sobre as amostras das doenças, para tanto, sua tabela possui ligação com a tabela da classe abstrata *Inflammaging*, logo, está ligada a todas as doenças, sendo que toda a doença incluída no banco de dados terá as suas amostras correspondentes, e uma amostra deverá se relacionar com apenas uma única doença segundo as restrições apresentadas na Seção 4.5.

A tabela *Affymetrics* representa a classe de mesmo nome da ontologia, esta tabela possibilitará saber quais genes (atributo *entrezgene*) faz parte do grupo de genes encontrado nas amostras (ou tabela e classe *Samples*), uma vez que os dados das amostras são organizados a partir do identificador *affymetrics*. A classe *Genes* estará disposta na tabela *Genes*, o qual faz ligação com a tabela *Affymetrics*,

como detalhado anteriormente, e também com a tabela *Pathways*, que guardará os genes em grupos ou famílias de vias metabólicas.

As propriedades dos objetos da ontologia podem ser verificados a partir da tabela *Pathways_Genes* que determina a propriedade *contains_gene* presente na classe *Pathways* e também pelas chaves estrangeiras dentro das tabelas *Affymetrics* e *Samples*. Abaixo serão apresentadas as tabelas principais do banco de dados, sendo especificado seus atributos e o objetivo de cada informação ali depositado.

A tabela Genes contém os dados extraídos do banco de dados HUGO relativos aos genes pertencentes às vias metabólicas selecionadas para esta pesquisa. Ela é a representação da classe Genes da ontologia. Sua estrutura pode ser visualizada na Tabela 1.

Tabela 1 – Tabela Genes do Banco de Dados Relacional

Tabela Genes	Nesta tabela são encontradas as informações selecionadas referentes aos genes presentes nas <i>pathways</i> e nas <i>samples</i> da Ontocancro.	
Atributos	Tipo	Descrição
EntrezGene	Varchar(30)	Identificador do EntrezGene e do gene na Ontocancro
ENSG	Varchar(255)	Identificador do Gene na base Ensembl
Chromosome	Varchar(255)	Localização do gene no cromossomo.
ENSP	Varchar(255)	Identificador da Proteína na base Ensembl
Evidence	Varchar(255)	Códigos de evidência da base GeneOntology
HugoName	Varchar(255)	Nome do gene aprovado pelo HGNC
PreviousSymbols	Varchar(30)	Símbolos previamente aprovados para o gene
StringSymbol	Varchar(30)	Identificador no banco de dados String
ApprovedSymbol2	Varchar(30)	ApprovedSymbol conferido
HGNC	Varchar(30)	Identificador no banco de dados HGNC
ApprovedSymbol	Varchar(30)	Símbolo aprovado pelo HGNC.
Aliases	Varchar(255)	Outros nomes do gene (sinônimo)
RefSeq_IDs	Varchar(30)	Referência de Sequencia da base NCBI
ReactomeID	Varchar(30)	Identificador do Reactome
Status	Varchar(255)	Situação da pesquisa e definição de nome do gene
StringName	Varchar(255)	Nome no banco de dados String
UniGeneID	Varchar(30)	Identificador no banco de dados UniGene
NCIName	Varchar(255)	Nome do gene no NCI
KeggID	Varchar(30)	Identificador no banco de dados Kegg

A tabela *Pathways* advinda a partir da classe *Parhways* da ontologia, contém os dados das vias metabólicas selecionadas para esta pesquisa. Sua estrutura pode ser visualizada na Tabela 2.

Tabela 2 – Tabela *Pathways* do Banco de Dados Relacional.

Tabela <i>Pathways</i>		
Corresponde às vias presentes na Ontocancro.		
Atributos	Tipo	Descrição
Id_pathway	Int(20)	Identificador das vias na base Ontocancro
Pathway_apelido	Varchar(30)	Nome abreviado da via na base Ontocancro
Pathway_name	Varchar(255)	Nome da via metabólica
Pathway_type	Varchar(255)	Tipo da via: <i>Cell Cycle, DNA Damage Response</i> ou <i>Apoptosis, inflammation.</i>
Db_name	Varchar(30)	Fonte de onde a via foi obtida.
References	Varchar(255)	2º Link de referência da fonte.
Source	Varchar(255)	Link de referência da fonte.

A tabela *Samples* visualizada na classe *Samples* da ontologia agora separada da tabela *Serie*, como era apresentado na versão 2.0 deste projeto, acrescentou duas novas *tuplas* de ligação, a *id_disease* que é a identificação da doença a qual pertence a amostra e *affymetrics*, que une as *samples* com os genes. Esta estrutura pode ser visualizada na tabela 3.

Tabela 3 – Tabela *Samples* do Banco de Dados Relacional

Tabela <i>Samples</i>		
Composta pelos dados obtidos nas amostras das doenças.		
Atributos	Tipo	Descrição
Num_sample	Varchar(30)	Identificação da amostra. Ex. GSM490138
Detection_P_Value	Double(50,0)	Valor de significância do gene na amostra
Abs_call	double(50,0)	Estado de expressão
Status	Varchar(255)	Descrição da amostra.
Status_Abs_Call	char(1)	Estado da amostra: Presente, Ausente Marginal.
Type	Char(10)	Tipo do tecido: controle (saudável) ou doente
ID_disease	Varchar(30)	Série que pertence a amostra. Ex. GSE19650
Affymetrics	Varchar(30)	Identificador Affymetrix

A tabela *Inflammaging* da classe *Inflammaging* da ontologia, traz as informações do estudo a que os valores de cada amostra (*samples*) pertencem, sua estrutura está representada na Tabela 4.

Tabela 4 – Tabela *Inflammaging* do Banco de Dados Relacional

Tabela <i>Inflammaging</i>	Composta pelas informações sobre os estudos das doenças selecionadas	
Atributos	Tipo	Descrição
ID_disease	Varchar(30)	Identificador da Serie.
Organism	Varchar(30)	Organismo que se refere o estudo
Title	Varchar(255)	Nome do estudo.
Pmid	Int(30)	Identificação do artigo no banco de dados <i>PubMed</i> .
Publication	Varchar(255)	Status e data da publicação

Além dessas tabelas, podemos verificar a tabela *Affymetrics* da classe *Affymetrics* da ontologia, que identifica a sonda utilizada no estudo das amostras, ou seja, o gene da amostra, e a tabela *Pathways_Genes*, que faz a ligação dos genes inclusos em cada via metabólica.

5.2 Atualização das instâncias para o banco de dados de acordo com o *Inflammaging*

Com a descoberta do estudo sobre *Inflammaging*, e que o câncer, foco principal da Ontocancro, faz parte do processo inflamatório crônico advindo do envelhecimento natural, optou-se por unificar as duas pesquisas a fim de visualizar as compatibilidades entre os dados do próprio câncer e doenças de outros estudos que abrangem o *Inflammaging*, como é o caso do Alzheimer e da Diabetes do tipo 2.

As doenças importadas para o banco de dados foram a *Type 2 Diabetes mellitus: mRNA and miRNA profiling (GSE26168)* e a *Microarray analysis of the astrocyte transcriptome in the ageing brain: relationship to Alzheimer's pathology and ApoE genotype (GSE29652)*. Ambas selecionadas pelo grupo de pesquisa da Ontocancro, tendo em vista que podem ser encontrados outros estudos sobre estas doenças. Elas encontram-se no site do GEO, que disponibiliza além de especificações relativas as patologias, também traz amostras reais de tecidos que estão em vários estágios clínicos, como tecido saudável, em estado inicial e em estado avançado da doença. Na Figura 11 pode-se visualizar uma pesquisa simples

no GEO, que mostra as especificações e possibilidades dos dados encontrados sobre as doenças.

The screenshot shows the NCBI GEO website interface. At the top left is the NCBI logo, and at the top right is the GEO logo (Gene Expression Omnibus). Below the logos are navigation links: HOME, SEARCH, SITE MAP, GEO Publications, FAQ, MIAME, and Email GEO. The main header shows the breadcrumb path: NCBI > GEO > Accession Display [?] and a user status: Not logged in | Login [?]. Below the header is a search bar with the following fields: Scope: Self, Format: HTML, Amount: Quick, and GEO accession: GSE10927, with a GO button. The main content area displays the details for Series GSE10927, including a link to Query DataSets for GSE10927. The details are as follows:

Status	Public on Jan 22, 2009
Title	Human adrenocortical carcinomas (33), adenomas (22), and normal adrenal cortex (10), on Affymetrix HG_U133_plus_2 arrays
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	Human samples of 33 adrenocortical carcinomas, 22 adrenocortical adenomas, and 10 normal adrenal cortex samples, each from a different patient, had mRNA assays performed using Affymetrix HG_U133_plus_2 arrays, with 54675 probe-sets. We note that the same array data is in GEO series GSE33371, where we assayed the cancer samples for Beta-catenin staining or mutation, and make new comparisons based on those assays. Keywords: disease state analysis

Figura 11 – Pesquisa no GEO do câncer nas glândulas adrenais

As séries das doenças dispostas na Ontocancro 2.0 eram especificamente de câncer, sendo elas Câncer Adrenocortical (GSE10927), Câncer de pâncreas (GSE19650), Câncer de Tireoide (GSE27155), Câncer de Cólon Retal (GSE4183), Transtorno Pulmonar Obstrutivo Crônico – DPOC (GSE8581) e Dados de expressão no tecido pulmonar de pacientes com DPOC moderada, fumantes saudáveis e não fumantes (GSE37768), sendo que as amostras continham tecidos de carcinoma, adenoma e normal.

Para que a atualização do banco pudesse ser realizada, era preciso baixar os dados das amostras de uma série ou doença do site da GEO e verificar se era possível visualizar o *id_ref* (identifica a sonda do estudo), *value* (guarda o valor da expressão sobre esta sonda) *abs_call* (indica se o transcrito está presente, ausente ou marginal) e o *p_value* (indica o nível de significância da amostra). Se todas estas informações não estiverem dispostas nas amostras, é necessário uma normalização.

Esta normalização se dá a partir do programa R, um ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos, desenvolvido pelo departamento de estatística da universidade de Auckland na Nova Zelândia (IHAKA, GENTLEMANE, 1996), e de *Scripts* desenvolvidos pelo grupo de pesquisa para encontrar tais informações. Depois de normalizado, o arquivo gerado terá as quatro informações citadas acima. Para a tabela do banco de dados é necessário completar com mais quatro informações que podem ser encontradas no site GEO que traz a doença.

As informações retiradas do site para completar a tabela das *Samples* no projeto são o *num_serie* (refere-se a doença a qual pertence a amostra), o *num_sample* (informa de qual amostra da doença são aqueles resultados), o *status* (denota o estado da amostra como um tecido normal ou com a doença já avançada) e o *type* (reafirma a situação do *status* de forma mais direta para que o cálculo da diversidade relativa proposta no trabalho de Nascimento (et. al 2012) seja efetuado).

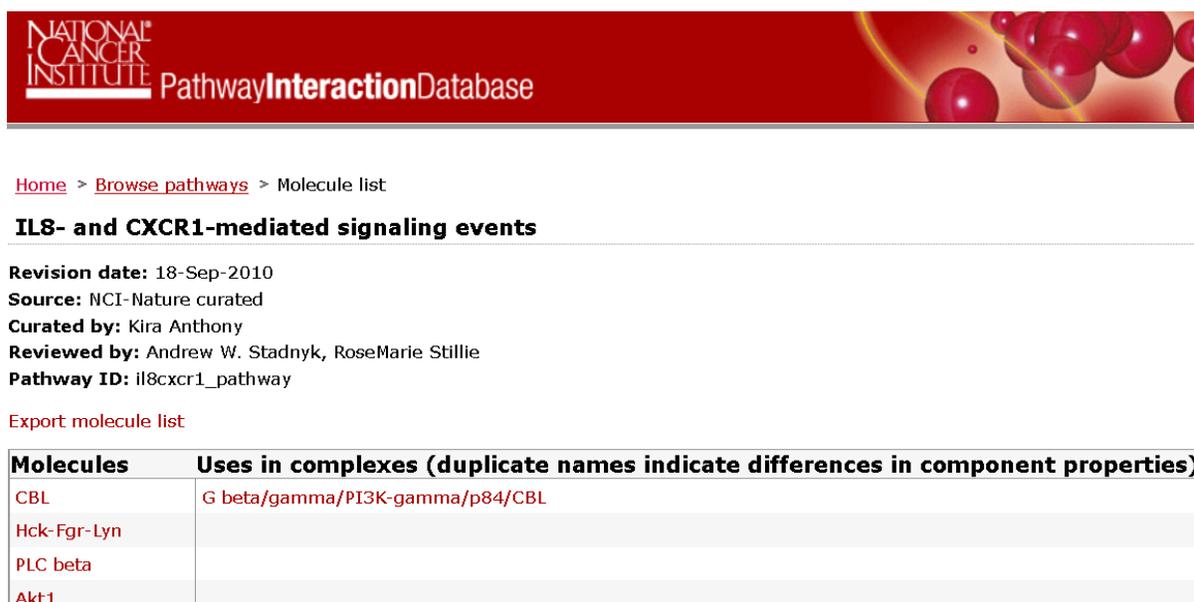
Na tabela *Pathways* também ocorreram inserções, das 37 vias metabólicas encontradas no projeto, passou-se para 55, sendo assim, foram inseridas 18 *pathways* que se referem ao processo inflamatório. Dentre as vias do processo inflamatório podemos citar *IL10 Anti-inflammatory Signaling Pathway*, *IL6-mediated signaling events* e *IL8- and CXCR1-mediated signaling events*, citocinas citadas no processo inflamatório no estudo do Alzheimer (BRITO et al., 2011).

Outras vias inseridas no banco de dados foram: *Canonical NF-kappaB*, *IL8- and CXCR2-mediated signaling events*, *mtor signaling Nemo*, *NF-kappaB pathway*, *nf-kb signaling*, *p38 MAPK signaling*, *p53*, *Regulation_of_p38-alpha_and_p38-beta*, *Replicative Senescence*, *RIG-I*, *SASP*, *Signaling_mediated_by_p38-alpha_and_p38-beta*, *TGF*, *TNF*, todas do tipo inflamação.

Para a inserção dos dados das vias, foram pesquisados dois sites principais, o NCI e o Reactome que são bases de dados públicas. A preferência era dada para o NCI pois apresentava os dados mais concisos a partir da curagem feita pela equipe

de trabalho do projeto. Nas Figuras 12 e 13 pode-se visualizar uma pesquisa aleatória das vias realizada no site do NCI e do Reactome respectivamente.

Na Figura 12, pode-se visualizar a pesquisa no site do NCI *pathways*, a data da revisão, quem fez a curagem dos dados, o nome estipulado para a via e as moléculas ou genes encontrados nessa via metabólica. As moléculas podem ser exportadas a partir do link logo acima da tabela e o arquivo gerado é *.mol* aberto em planilha do Excel.



The screenshot shows the NCI Pathway Interaction Database interface. At the top, there is a navigation bar with the NCI logo and the text 'PathwayInteractionDatabase'. Below this, there is a breadcrumb trail: 'Home > Browse pathways > Molecule list'. The main heading is 'ILS- and CXCR1-mediated signaling events'. Below the heading, there is a list of metadata: 'Revision date: 18-Sep-2010', 'Source: NCI-Nature curated', 'Curated by: Kira Anthony', 'Reviewed by: Andrew W. Stadnyk, RoseMarie Stillie', and 'Pathway ID: il8cxcr1_pathway'. There is a link 'Export molecule list'. Below this is a table with two columns: 'Molecules' and 'Uses in complexes (duplicate names indicate differences in component properties)'. The table contains four rows of data.

Molecules	Uses in complexes (duplicate names indicate differences in component properties)
CBL	G beta/gamma/PI3K-gamma/p84/CBL
Hck-Fgr-Lyn	
PLC beta	
Akt1	

Figura 12 – Pesquisa de vias no NCI-Pathways

A Figura 13 mostra a pesquisa de vias através do banco de dados Reactome, pode ser visualizada a família de vias metabólicas a qual aquela *pathway* pertence, uma explanação geral da via e do processo de curagem, além de um gráfico comentado da via.

The image shows the Reactome pathway browser interface. At the top, it indicates 'Pathways for: Homo sapiens'. On the left, there is an 'Event Hierarchy' with various biological processes listed, including 'DNA Repair' and 'Double-Strand Break Repair'. The main area displays a complex network diagram of the 'Double-Strand Break Repair' pathway. Below the diagram, there are tabs for 'Overview', 'Molecules (25)', 'Structures', 'Expression', 'Analysis', 'Processes', and 'Downloads'. The 'Molecules (25)' tab is active, showing a list of 'Chemical Compounds (2)' and 'Proteins (23)'. The protein list includes UniProt entries: O60934 NBN (5x) and P12956 XRCC6 (7x).

Figura 13 – Pesquisa de vias no Reactome

Para conseguir os dados dos genes encontrados na via a partir da pesquisa no Reactome, é possível realizar o *download* a partir da aba *molecules*, no entanto, é preciso verificar quais os dados encontram-se nessa aba. Somente os dados de *Proteins* são necessários, pois estes tratam dos genes que compõe a via, os *Chemical Compounds* não se referem a genes. Na hora de fazer o download selecionar somente as proteínas. Com a inserção das 18 *pathways*, houve um acréscimo no número de genes, que passou de 896 para 1105 registros.

5.3 Sumário do Capítulo

Este capítulo descreveu como é realizada a atualização e obtenção dos dados dispostos na Ontocancro, como está estruturado o banco de dados que recebe as informações da ontologia e de onde vem tais informações. No próximo capítulo poderá ser visualizada uma comparação de ontologias que apresentam dados sobre câncer, as ontologias destacadas e expostas neste capítulo foram a Ontocancro (NASCIMENTO et al., 2012), a ACGT (BROCHHAUSEN et al., 2011) e a NTO (CUSTÓDIO, 2012).

6 TRABALHOS RELACIONADOS A ONTOLOGIAS E O ESTUDO DAS DOENÇAS DE *INFLAMMAGING*

Para verificar a disponibilidade de ontologias voltadas para o estudo deste trabalho, realizou-se uma pesquisa sobre ontologias e doenças relacionadas ao *Inflammaging*, verificando as formas de disponibilização e operação das mesmas com o estudo de genes e *pathways*, além da organização ontológica desses estudos.

Na literatura encontram-se algumas ontologias relacionadas ao estudo biológico, cada qual com suas particularidades e desenvolvidas para um devido fim. Nesse capítulo serão apresentadas as ontologias encontradas, tendo em vista sua estrutura, objetivos e aplicações.

6.1 *Advancing Clinico-genomic Trials (ACGT Master Ontology)*

Nas pesquisas em respeito ao câncer, são utilizadas ferramentas de alto rendimento que resultam em uma “explosão” de informações e conhecimento sobre os diversos tipos de câncer e seu tratamento. Isto é devido à complexidade deste grupo de doenças que afetam uma porção significativa da população. A falta de estrutura de informação aberta e compartilhada muitas vezes impede que instituições de pesquisa possam visualizar e integrar dados de diferentes níveis de investigação, por exemplo, para integrar dados de estudos voltados para os elementos moleculares de câncer com aqueles focados no que ocorre no nível de órgãos.

A visão do projeto ACGT (*Advancing Clinico-genomic Trials on Cancer – Open Grid Services for Improving Medical Knowledge Discovery*) (BROCHHAUSEN et al., 2011) é contribuir para a resolução destes problemas através de uma ontologia orientada. Seu objetivo final é o desenvolvimento de uma rede de serviços de infraestrutura semântica para facilitar o acesso de forma transparente e segura a bancos de dados heterogêneos distribuídos em vários níveis, com o intuito de

fornecer uma gama de semânticas reutilizáveis, valendo-se de ferramentas abertas para análise dos mesmos, integrando dados clínicos genômicos de vários níveis.

A ACGT é um projeto financiado pela União Europeia, teve início em 2006, foi implementada em OWL, tem sido ampliada para atender as necessidades tanto técnicas quanto clínicas. Representa de forma abrangente o domínio de pesquisa e tratamento de câncer, em especial o câncer de mama, tumor de *Wilms* (GARVIN et al., 1993) e tumor *rhabdoid* (BIEGEL et al., 1999). Atualmente possui 1667 classes, 288 propriedades de objetos, 15 propriedades de dados e 61 indivíduos.

A ACGT consiste na adoção de uma BFO (Ontologia Básica Formal), a taxonomia BFO faz uso de uma base de nível superior que distingue dois tipos de entidades: *continuants* e *ocurrents*. As entidades *continuants* podem ser subdividas em dependente, independente ou região espacial, enquanto as *ocurrents* podem ser subdividas em região temporal, região espaço temporal ou entidade processual. As entidades *continuants* perduram através do tempo, mantendo sua identidade, enquanto as *ocurrents* são entidades que podem acontecer, se desdobram ou desenvolvem com o tempo (COCOS, 2008).

A característica que define *continuants* dependentes é que são conceitos inerentes de outro conceito (ou seja, *continuants* independentes), por exemplo, o tamanho da superfície de um fígado, o sexo de um ser humano, a doença de um organismo, etc. As *continuants* independentes são conceitos definidos pelas dependentes, por exemplo, ser humano, fígado, laboratório, etc. Entidades processuais são entidades que podem ocorrer e sempre depende de uma entidade *continuant*, por exemplo, a vida de um organismo, o processo de meiose, os processos de biópsia, exame clínico, desenvolvimento do tumor, processo de divisão celular, etc (COCOS, 2008).

6.2 Systems Biology Ontology (SBO)

A SBO (COURTOT e JUTY, 2011) (*Systems Biology Ontology* - Ontologia de Sistemas Biológicos) descreve as entidades usadas na modelagem computacional,

fornecendo um conjunto de conceitos inter-relacionados que podem ser utilizados para especificar, por exemplo, o tipo de componente que está sendo representado em um modelo, ou a função destes componentes em descrições de sistemas biológicos. As entidades com termos SBO permite um entendimento explícito e inequívoco do significado destas entidades. Além disso, utilizando os termos SBO em diferentes formatos de representação facilita o mapeamento entre elementos de modelos codificados naqueles formatos.

SBO é atualmente composto por sete ramos de vocabulário: sistemas de descrição de parâmetros, funções participantes, estrutura de modelagem, expressão matemática (cujos termos constituintes referem-se aos três ramos anteriores), a representação de entidades ocorrentes, a representação física da entidade e representação de metadados. Os conceitos são relacionados através da inclusão de relacionamentos ontológicos (subclasses), bem como a construção de expressões matemáticas representadas em *Mathematical Markup Language* (MathML) versão 2 (COURTOT e JUTY, 2011). Se um termo SBO carrega a expressão matemática, então cada símbolo usado dentro desta expressão pode ser definido por outro termo SBO. Este procedimento enriquece a informação obtida quando usado como termos, e presta-se a um maior processamento computacional.

É uma ontologia aberta. Está disponível em diferentes formatos como OBO, OWL e SBO-XML, podendo ser acessado através de *web services*. A documentação relacionada e recursos associados estão disponíveis gratuitamente, é desenvolvida como uma ontologia padrão, respeitando o conjunto de princípios comum de desenvolvimento, como descrito pela OBO.

Termos SBO podem ser utilizados para introduzir uma camada de informações semânticas para a descrição de um modelo padrão, ou para anotação dos resultados de experiências bioquímicas, a fim de facilitar a sua reutilização. No entanto, as representações formais de modelos nem sempre fornecem informações sobre os procedimentos a seguir para analisar e trabalhar com o modelo, podendo gerar resultados diferentes dependendo do procedimento e do algoritmo de simulação.

6.3 TEDDY

TEDDY (COURTOT e JUTY, 2011) é uma ontologia projetada para o alinhamento e comparação de sequências biológicas. É composto por quatro ramos: a classificação dos comportamentos concretos temporais observados em uma simulação (trajetórias), as diversificações e as relações entre os comportamentos, as características de comportamentos específicos e os motivos funcionais geradores de determinados tipos de comportamentos.

TEDDY tem como objetivo fornecer uma ontologia para comportamentos dinâmicos, fenômenos dinâmicos observáveis e elementos de biomodelos e sistemas biológicos. Os termos TEDDY podem ser suficientes para qualificar, com níveis variáveis de detalhe, as características críticas de resultados numéricos obtidos a partir de simulações, bem como aqueles a partir de medições experimentais. Está codificada em OWL.

6.4 Agreement Maker

O AgreementMaker (CRUZ et al., 2009) é um sistema composto de uma grande variedade de métodos de correspondência entre ontologias, abordando diferentes níveis de granularidade dos componentes que estão sendo correspondidos (conceitual *versus* estrutural), quantidade de intervenção manual que eles necessitam (manual *versus* automático), seu uso (sozinho ou composto) e o tipo de componentes para considerar (esquema somente ou esquema e instâncias). Consiste em “definir mapeamentos entre esquema ou elementos da ontologia que são semanticamente relacionados” (CRUZ et al., 2009), considerando alguns domínios e aplicações, como geoespacial, ambiental e biomédica. As informações conceituais destas aplicações são armazenadas em forma de ontologia.

O mapeamento normalmente é feito entre dois esquemas ou duas ontologias de cada vez sendo uma chamada de fonte e outra de alvo. Sua arquitetura permite a composição serial e paralela, onde, respectivamente, a saída de um ou mais

métodos pode ser usada como entrada de outro, ou vários métodos podem ser usados em alguma entrada e então combinados. O conjunto de mapeamentos pode ser o resultado de uma sequência de passos chamados camadas. O processo de correspondência pode ser dividido em dois módulos principais: computação de similaridade e seleção de mapeamento.

Na computação de similaridade (*Similarity Computation*) cada conceito da ontologia fonte é comparado com todos os conceitos da ontologia alvo, então produz duas matrizes de similaridade (uma para classes e outra para propriedades), que contém o valor de cada par de conceitos. Na seleção de mapeamento (*Mappings Selection*), cada matriz gerada pela computação de similaridade é examinada para selecionar somente o melhor mapeamento de acordo com uma dada entrada e a cardinalidade das correspondências.

Para permitir extensibilidade, é adotado o padrão de modelo orientado a objeto a partir da definição do esqueleto do processo de correspondência de uma combinação genérica que difere somente poucas operações na extensão do combinador concreto. Esta abstração minimiza o esforço de desenvolvimento para desacoplar a estrutura de um simples método para a arquitetura do sistema completo, assim permite o reuso tendo uma possível composição de módulos de correspondência.

Todas as operações de correspondência e seus resultados são gerenciados pela interface, na qual permite os usuários executarem e gerenciarem os métodos e seus resultados. Uma vez que os mapeamentos são definidos, eles serão capazes de interagir com os resultados obtidos, por exemplo, para reduzir o número de linhas que descrevem os mapeamentos. Cada linha mostra o valor de *performance* para os métodos associados, assim permitindo a comparação destas com outras linhas. Além disso, os usuários podem modificar os parâmetros de um método em tempo de execução, mudando diretamente seus valores na tabela ou pela seleção de correspondência previamente calculada com a entrada dos métodos que serão aplicados depois.

6.5 Neuropsychological Test Ontology (NTO)

O NTO (*Neuropsychological Test Ontology*) (CUSTÓDIO, 2012) é uma ontologia desenvolvida para o estudo da Doença de Alzheimer (DA), identificando os marcadores de diagnósticos e prognósticos, perfil dos doentes e progressão da doença. Os dados incluem a “história clínica, exames neurológicos, testes neurofisiológicos, avaliação laboratorial e imagens cerebrais” e tem como objetivo principal prever se um doente com problemas cognitivos irá sofrer de DA no futuro.

Os dados de conhecimento que serviram para popular a ontologia estavam dispostos em arquivos Excel relativos a “1642 avaliações médicas efetuadas ao longo de 22 anos por médicos do Instituto de Medicina Molecular (IMM) a 92 dos seus pacientes” (CUSTÓDIO, 2012). Dentre as informações disponibilizadas, havia uma atribuição de um diagnóstico para cada paciente, além da aplicação de testes neuropsicológicos.

A NTO reutiliza ontologias existentes, como a *Human Dementia Ontology* e a *Friend of a Friend*, que permitem especificar conceitos e propriedades da Doença de Alzheimer e de pessoa. Para a visualização dos dados da ontologia foi desenvolvido um framework em Java, implementado sobre uma arquitetura Cliente-Servidor.

6.6 Comparação entre as ontologias

Conforme a pesquisa realizada, observa-se que hoje em dia existem poucos estudos envolvendo ontologias na área de genética, principalmente relacionada as doenças do *Inflammaging*. Entre as ontologias citadas anteriormente, somente a Ontocancro, a ACGT e a NTO tratam especificamente de dados relacionados às doenças; as demais propostas apresentadas trazem expectativas diferentes quanto a criação e a atualização de uma ontologia já existente referente ao estudo biológico. Na tabela 5 poderá ser visualizada a especificação de cada ontologia estudada.

Tabela 5 – Propriedades das Ontologias.

Ontologia	Ontocancro 3.0	ACGT Master	NTO	TEDDY	SOB	Agreement Maker
Tipo de Estudo	Câncer adrenocortical pancreático tireoide e cólon retal Transtorno Pulmonar Obstrutivo Crônico Diabetes tipo 2 Alzheimer	Câncer de Mama, Tumor de <i>Wilms</i> , Tumor de <i>Rhabdoid</i>	Doença de Alzheimer	Realiza comparações de sequencias biológicas com comportamentos dinâmicos	Descreve as entidades usadas na modelagem computacional	Define mapeamentos entre esquema ou elementos das ontologias que são semanticamente relacionados
Sintaxe	OWL	OWL	OWL	OWL	OBO OWL SBO-XML	OWL
Download	Sim	Sim	Não	Sim	Sim	Sim
Inferência/ Reasoning	Não	Não	Sim	Não	Não	Não
Classes	16	1667	17	-	-	-
Propriedades de Objetos	7	288	27	-	-	-
Propriedades de Dados	39	15	10	-	-	-

A Tabela 5 denota as propriedades de cada ontologia comparada, podendo assim ter uma visão geral do que as mesmas representam, qual seu intuito e o que realmente desejam alcançar.

A Ontocancro 3.0 armazena informações de vias genéticas humanas, que possuem várias subvias do ciclo celular, mecanismos de reparo do DNA, apoptose e estabilidade cromossômica, tendo como principal objetivo visualizar a interação dos genes com cada uma das situações citadas, e a partir desse estudo, detectar com antecedência fatores de risco para o desenvolvimento das doenças.

A ACGT possui informações de três tipos de câncer. Essas informações vão desde a parte gênica até o desenvolvimento e tratamento da doença em si, passando pela biologia molecular, processos clínicos médicos até a parte administrativa da doença. Pode ser considerada uma ontologia abrangente por conter todos os estágios da doença, tem como objetivo alcançar a existência de apenas uma ontologia relacionada a câncer na qual se torne possível a integração

semântica, distribuição e armazenamento de dados clínicos e genômicos, rastreando a doença desde a sua detecção até seu tratamento.

A NTO disponibiliza dados sobre a doença de Alzheimer, de estudos do IMM, trazendo dados de pacientes reais, e possibilitando a inserção e novos dados em tempo real, inseridos por médicos com acesso prioritário. Tem o intuito de diagnosticar antecipadamente possibilidades de desenvolvimento da doença por pacientes com doença cognitiva.

As ontologias SBO e TEDDY são relacionadas à área de modelagem computacional, onde a SBO descreve as entidades usadas na modelagem computacional e TEDDY é projetada para a comparação e alinhamento de sequências biológicas. As atividades desenvolvidas por essas duas ontologias, em linhas gerais, visam o aperfeiçoamento de outras ontologias, possibilitando extrair informações mais significativas, facilitando o mapeamento dos termos que estão sendo utilizados.

Pode-se ver pela Tabela 5, que apenas o projeto NTO não possibilita o *download* dos estudos realizados, todos estão no formato OWL. Destes estudos, somente o NTO possui propriedades de inferência e raciocínio, possibilitando a troca de informações entre o banco de dados e os pesquisadores.

De todas as ontologias encontradas, as que mais se aproximam da Ontocancro são a ACGT e a NTO, ricas em classes e relações que descrevem as informações clínicas relacionadas ao estudo das doenças definidas pelo *Inflammaging*, contudo, nenhuma delas tem o *Inflammaging* como parte do seu estudo. A ACGT e a NTO são compostas apenas pela representação do domínio das doenças, no entanto a NTO permite a inserção de novos dados e não apenas a consulta dos mesmos. No caso da Ontocancro, são 6 (seis) tipos de câncer, mais uma doença neurodegenerativa (Alzheimer) e a diabetes do tipo 2, que estão mapeados, mostrando informações genéticas de uma série de pacientes a partir do estudo de *Inflammaging*.

Para o trabalho proposto com *Inflammaging*, os dados clínicos são dispensáveis, pois os estudos sobre o mesmo referem-se a alteração molecular de proteínas, visando a prevenção ou detecção do estágio inicial da doença, e não o diagnóstico e tratamento do paciente.

6.7 Sumário do capítulo

Este capítulo apresentou uma comparação entre ontologias relacionadas ao estudo de doenças como tipos de câncer, diabetes e Alzheimer. Para melhor visualizar as diferenças e semelhanças, foi exposta a comparação entre os trabalhos relacionados, que possibilitou um melhor estudo sobre cada caso.

No próximo capítulo será apresentada uma análise de comparação entre as vias metabólicas. A comparação realizada deu-se através dos genes inseridos nas *pathways* de inflamação com as demais *pathways* dispostas anteriormente no banco que fazem parte do sistema regular do metabolismo celular como a apoptose e o reparo do DNA.

7 ANÁLISE DOS DADOS INSERIDOS NA ONTOCANCRO

Após a inserção das informações sobre as doenças e das vias de inflamação, realizou-se uma comparação das vias metabólicas já dispostas no banco de dados referentes aos processos cancerosos, com as vias de inflamação inseridas a partir do estudo do *Inflammaging* e das doenças referidas nesta dissertação. Neste estudo desenvolveu-se uma comparação entre as vias de apoptose, ciclo celular e reparo de DNA com as novas vias de inflamação inseridas. Este capítulo apresentará os resultados obtidos através dessa comparação, e também como foi realizado este processo.

7.1 Relação das novas vias da ontologia com as existentes na Ontocancro 2.0

Com o intuito de verificar a compatibilidade de cada via metabólica encontrada no banco de dados com o estudo sobre o processo inflamatório, foi realizada uma comparação via a via, visando os genes que as compõem. Esta comparação teve como foco principal verificar a interação entre as vias metabólicas presentes nas doenças e a inclusão das vias do processo inflamatório proposta por este trabalho.

A comparação foi realizada através do banco de dados da Ontocancro utilizando-se um *script* SQL que gerava os genes iguais entre duas vias distintas. Para otimizar o processo, definiu-se um laço de repetição *while*, desta forma, o script gerou todos os resultados de forma automática, sendo comparadas todas as vias. O *script* SQL utilizado para gerar os resultados da comparação das vias metabólicas expressas no banco de dados pode ser visualizado na Figura 14.

```

DECLARE i int DEFAULT 1;
DECLARE j int DEFAULT 1;
WHILE i < 60 DO
BEGIN
    SET i = i+1;
    WHILE j < 60 DO
    BEGIN
        SET j = j+1;
        SELECT * FROM pathways_genes pg
        WHERE pg.entrezgene
        IN ( select entrezgene from pathways_genes
        WHERE id_pathway = j)
        AND pg.id_pathway = i
    END WHILE;
END WHILE;

```

Figura 14 – Script SQL de pesquisa da comparação no banco de dados

Desta análise buscou-se somente as comparações das vias de inflamação com as demais vias, tendo em vista que estão dispostos no banco de dados 18 vias de inflamação que foram comparadas com as 37 vias advindas do estudo do câncer.

7.2 Comparação das vias da Ontocancro 3.0

Os dados obtidos pela análise estão dispostos em valores de porcentagem, tendo como base sempre a via com menor número de genes, ou seja, se uma via tem 50 genes e outra via tem 30 genes e o resultado da pesquisa for 30 genes, temos 100% da via com menos genes na via com maior número de genes. A pesquisa gerou esse resultado via a via.

O Objetivo desta análise foi verificar quais os genes que atuam tanto nas vias metabólicas do processo natural das células quanto no processo inflamatório, verificando qual a representatividade dos genes responsáveis pela inflamação dentro

do ciclo celular, da apoptose e do reparo do DNA. Na Tabela 6 disponibiliza-se os resultados obtidos através da análise.

Com esta tabela podemos verificar que a via com maior número de genes iguais na comparação foi a *Replicative Senescence*, uma via do processo inflamatório definida através de um estudo dentro do grupo de pesquisa da Ontocancro realizado por Pereira et al. (2012).

As vias *Mitotic Spindle Checkpoint* da família *Cell Cycle* e *Granzyme a Mediated Apoptosis Pathway* da família *Apoptosis*, não foram compatíveis com nenhuma das vias de *inflammation* inseridas no estudo da Ontocancro.

Tabela 6 – Resultados obtidos através do script SQL.

	Canonical NF-kappaB	IL10 Anti-inflammatory Signaling Pathway	IL6-mediated signaling events	IL8 and CXCR1-mediated signaling events	IL8 and CXCR2-mediated signaling events	mtor signaling	Nemo	NF-kappaB pathway	nf-kb signaling	p38 MAPK signaling	p53	Regulation_of_p38-alpha_and_p38-beta	Replicative Senescence	RIG-I	SASP	Signaling_mediated_by_p38-alpha_and_p38-beta	TGF	TNF
Base Excision Repair	0%	0%	0%	0%	0%	0%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Double-Strand Break	5%	0%	0%	0%	0%	0%	20%	5%	0%	0%	18%	0%	50%	18%	18%	0%	0%	18%
Fanconi Anemia Pathway	4%	0%	0%	0%	0%	0%	5%	4%	0%	0%	4%	0%	13%	5%	8%	0%	0%	8%
Homologous	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	12%	0%	24%	14%	9%	0%	0%	11%
Hr Repair of	6%	0%	0%	0%	0%	0%	25%	6%	0%	0%	25%	0%	69%	25%	25%	0%	0%	25%
Mismatch Repair	0%	0%	0%	4%	4%	0%	0%	0%	0%	0%	0%	0%	14%	0%	0%	0%	0%	0%
Non-Homologous end	0%	0%	0%	0%	0%	0%	14%	0%	0%	0%	14%	0%	14%	14%	14%	0%	0%	14%
Nucleotide Excision	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	0%	10%	5%	2%	3%	0%	0%
Processing of DNA DSB	8%	0%	0%	0%	0%	0%	33%	8%	0%	0%	33%	0%	92%	33%	33%	0%	0%	33%
ATM signaling	17%	0%	6%	0%	0%	0%	11%	17%	11%	0%	22%	0%	44%	17%	44%	11%	11%	33%
ATR signaling	4%	0%	0%	0%	6%	13%	5%	0%	0%	12%	8%	0%	28%	5%	10%	0%	0%	11%
Cell Cycle Checkpoints	4%	0%	0%	0%	0%	3%	5%	3%	0%	0%	16%	0%	30%	14%	11%	3%	5%	14%
Cell Cycle, Mitotic	4%	0%	2%	0%	6%	7%	0%	0%	0%	12%	12%	0%	34%	9%	5%	0%	11%	0%
Cyclins and Cell Cycle	0%	7%	0%	0%	0%	7%	0%	0%	0%	0%	13%	0%	40%	7%	13%	0%	7%	0%
G1/S checkpoint	5%	0%	0%	0%	0%	10%	5%	5%	0%	0%	24%	0%	48%	19%	33%	5%	11%	10%
G1/S DNA Damage	4%	0%	0%	0%	0%	3%	5%	3%	0%	0%	16%	0%	18%	14%	10%	3%	5%	11%
G2/M checkpoint	5%	0%	0%	0%	0%	14%	5%	5%	0%	12%	18%	0%	36%	14%	32%	5%	5%	18%
Mitotic G2-G2/M Phases	4%	0%	0%	0%	3%	4%	0%	0%	0%	12%	0%	0%	5%	0%	0%	0%	0%	0%
Mitotic M0M/G1 Phases	4%	0%	0%	0%	6%	3%	0%	0%	0%	0%	0%	0%	16%	0%	0%	0%	0%	0%
Mitotic Spindle	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Rb Tumor	8%	0%	0%	0%	0%	15%	8%	8%	0%	8%	0%	0%	62%	23%	31%	8	8%	15%
Regulation of DNA	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	4%	0%	9%	5%	2%	0%	5%	0%
Regulation of Mitotic Cell	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%
S Phase	0%	0%	2%	0%	0%	3%	0%	0%	0%	6%	8%	0%	21%	9%	3%	0%	11%	0%
Apoptosis - Homo	43%	13%	9%	7%	6%	3%	45%	28%	71%	0%	32%	4%	5%	23%	31%	3%	32%	54%
Apoptotic Execution	4%	0%	2%	0%	0%	0%	5%	0%	0%	0%	4%	4%	2%	9%	8%	0%	0%	7%
Apoptotic signaling in	8%	0%	17%	8%	8%	8%	8%	8%	0%	0%	33%	0%	25%	17%	42%	8%	0%	17%
Caspase Cascade in	13%	6%	0%	0%	0%	2%	25%	3%	24%	0%	4%	4%	2%	0%	17%	0%	0%	29%
Death Receptor	15%	8%	0%	0%	0%	0%	8%	0%	38%	0%	0%	8%	0%	0%	31%	0%	0%	31%
Extrinsic Pathway for	15%	8%	0%	0%	0%	0%	8%	0%	38%	0%	0%	8%	0%	0%	31%	0%	0%	31%
Granzyme a Mediated	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Induction of apoptosis	0%	0%	0%	0%	0%	0%	10%	5%	5%	0%	0%	5%	0%	0%	10%	0%	0%	10%
Intrinsic Pathway for	0%	0%	11%	5%	5%	11%	5%	0%	0%	6%	11%	0%	11%	0%	26%	0%	0%	16%
Regulation of Apoptosis	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	0%	0%	0%	0%	0%	0%
TNF Receptor Signaling	22%	13%	3%	0%	0%	0%	15%	17%	38%	0%	0%	15%	3%	0%	18%	0%	5%	21%
tnfr1 Signaling Pathway	8%	8%	8%	0%	0%	0%	8%	0%	31%	0%	0%	15%	0%	0%	31%	0%	0%	31%
tnfr2 Signaling Pathway	10%	0%	0%	0%	0%	0%	10%	30%	20%	0%	0%	10%	0%	0%	10%	0%	0%	10%

Não foram realizadas comparações entre vias da mesma família, pois pelo fato de representarem um mesmo conjunto de genes, sua compatibilidade seria alta, mas não necessariamente representativa para o objetivo deste estudo.

Para melhor visualizar essas compatibilidades, podemos verificar na Figura 15 a vias com maior índice de paridade das vias da Ontocancro, com a via *Replicative Senescence*, que obteve um resultado superior a noventa por cento.

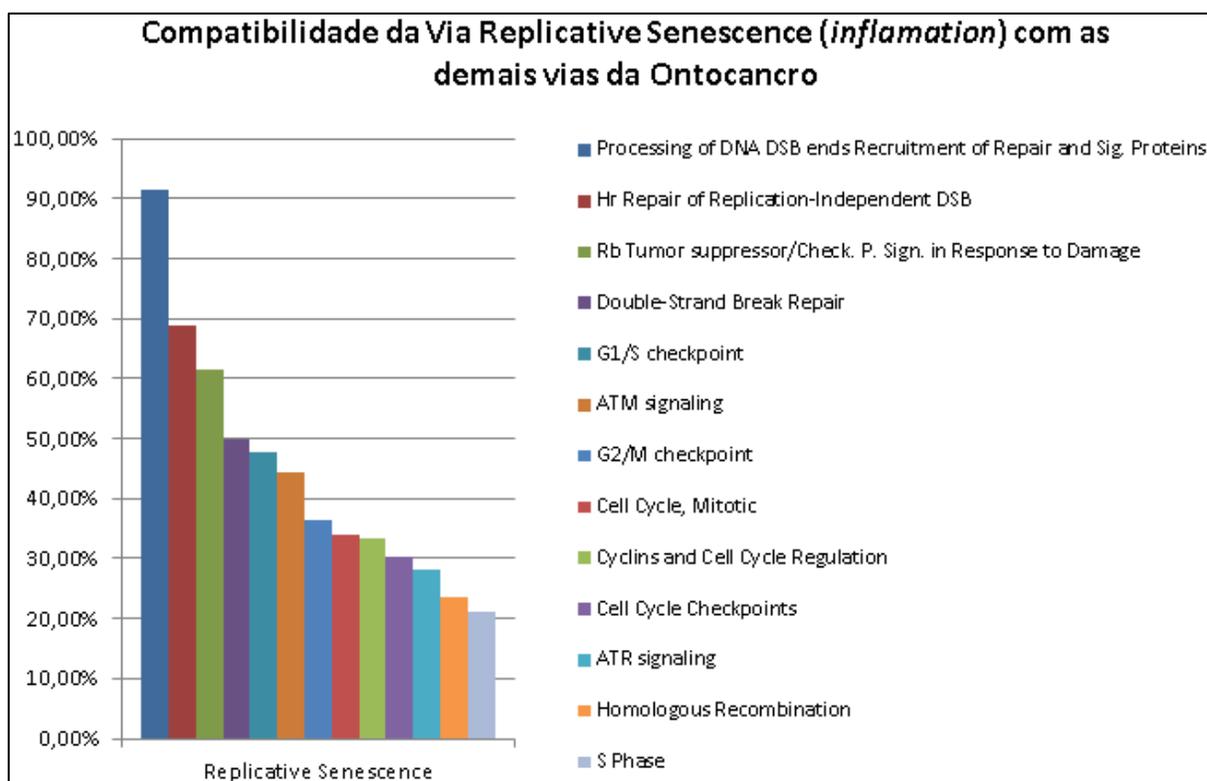


Figura 15 – Paridade da via *Replicative Senescence*

Como pode ser visualizado na Figura 15, das treze vias com maior similaridade com a *Replicative Senescence*, somente quatro estão na linha ou acima dos cinquenta por cento de semelhança, e somente uma via passa dos setenta por cento de paridade. Na Figura 16, pode-se visualizar os demais resultados da análise, que obtiveram menor índice de semelhança na comparação.

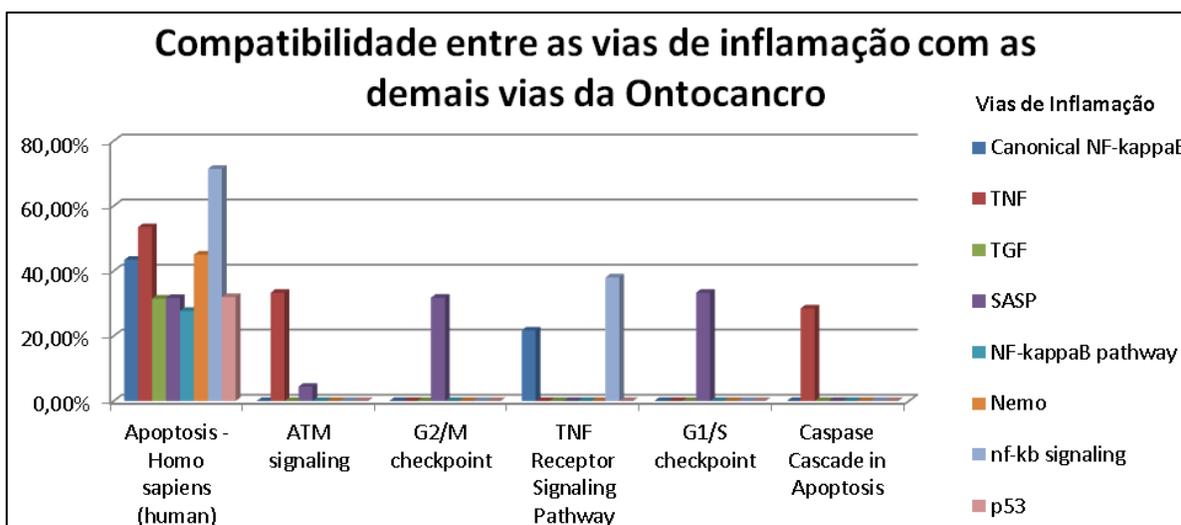


Figura 16 – Comparação das demais vias de inflamação

Na Figura 16 pode-se ver que a via que obteve maior compatibilidade com as demais vias de inflamação foi a *Apoptosis – Homo sapiens (human)*. Esta via foi determinada pelo grupo de pesquisa da Ontocancro a partir de estudos sobre vias da apoptose, sendo que sua maior similaridade pode ser vista com a *Pathway Nf-kb Signaling*, passando dos setenta por cento, no entanto, pelo fato da via *Apoptosis – Homo sapiens (human)* ter um grande número de genes, 163 no total, e a *Pathway Nf-kb Signaling* tendo 12 genes, sua representatividade não seria relevante.

7.3 Sumário do capítulo

Neste capítulo apresentou-se uma análise dos dados realizada após a inserção dos dados no banco de dados da Ontocancro, buscando verificar quais vias dispostas no banco eram mais compatíveis com o estudo proposto pelo *Inflammaging* e suas *pathways* de inflamação. Este estudo possibilitou verificar que a *pathway Replicative Senescence* do tipo inflamação obteve os valores mais elevados de compatibilidade com as demais vias do estudo.

No próximo capítulo será apresentada uma analogia das vias com maior compatibilidade a partir do sistema *String*, que disponibiliza grafos de relação entre

genes, mostrando sua intensidade de relação dentro de uma *pathway* ou na união de mais *pathways* dentro desta pesquisa.

8 VALIDAÇÃO DA ONTOLOGIA PROPOSTA COM O STRING

Para a validação da compatibilidade da análise realizada, optou-se por utilizar o site da *String* (SCHAEFER et al., 2015) que é um banco de dados de interações proteicas, incluindo interações com associações diretas (físicas) e indiretas (funcionais). Ela integra quantitativamente e qualitativamente os dados de interação a partir destas fontes para um grande número de organismos, e as transferências de informação entre esses organismos.

Segundo as Figuras 15 e 16 e a tabela de resultados já apresentadas, foram selecionadas as vias com maior número de genes iguais:

- *Replicative Senescence* (Inflamação) e *Processing of DNA DSB ends Recruitment of Repair and Sig. Proteins* (reparo de DNA),
- *Replicative Senescence* e *Hr Repair of Replication-Independent DSB* (reparo de DNA) para uma análise a partir do processador *String*.

A similaridade entre as vias é de 92% e 69% respectivamente.

8.1 Replicative Senescence X Processing of DNA DSB ends Recruitment of Repair and Sig. Proteins

A via *Replicative Senescence* contém 56 genes e na Figura 17 é apresentada a sua interação, no qual cada gene é representado por um círculo, e as interações gene a gene são apresentadas a partir das linhas de ligação.

Nas pesquisas do *String* é possível definir um grau de confiabilidade que vai de 20% a 90%, sendo que o padrão estipulado é de 40%. Para as análises deste trabalho, optou-se por um grau de confiança de 90%, deixando o resultado mais específico. Sendo assim, pode ser visualizado que os genes HIST1H2A e CSNK2A2 não fazem nenhuma ligação de 90% de confiabilidade com nenhum outro gene da *pathway*. O primeiro gene faz ligação com demais genes da via a partir de um grau

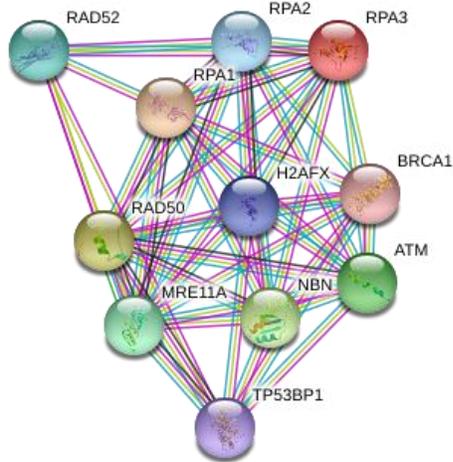


Figura 18 - Definição da String para a via Processing of DNA DSB ends Recruitment of Repair and Sig. Proteins

Nesta via, ao ser aplicado o grau de confiabilidade estipulado para as análises, nenhum gene ficou sem ligação. Como foi definido anteriormente, o resultado da análise partia do pressuposto que a via com menor número de genes estivesse “dentro” da via com maior número de genes, e nesta perspectiva, a segunda via mostrada tem mais de noventa por cento de compatibilidade, em números, somente um gene da via de reparo de DNA não está presente na via de inflamação, este gene é o *RAD52*.

Fazendo-se a união dessas duas vias no *String*, pode-se visualizar que mesmo num grau de 90% de confiabilidade, o gene não presente na união das duas vias, faz ligação com os demais genes do grupo, mas não somente com os genes da sua própria via, ele se une ao gene RPA4 que não faz parte do seu grupo. Este grafo pode ser visualizado na Figura 19.

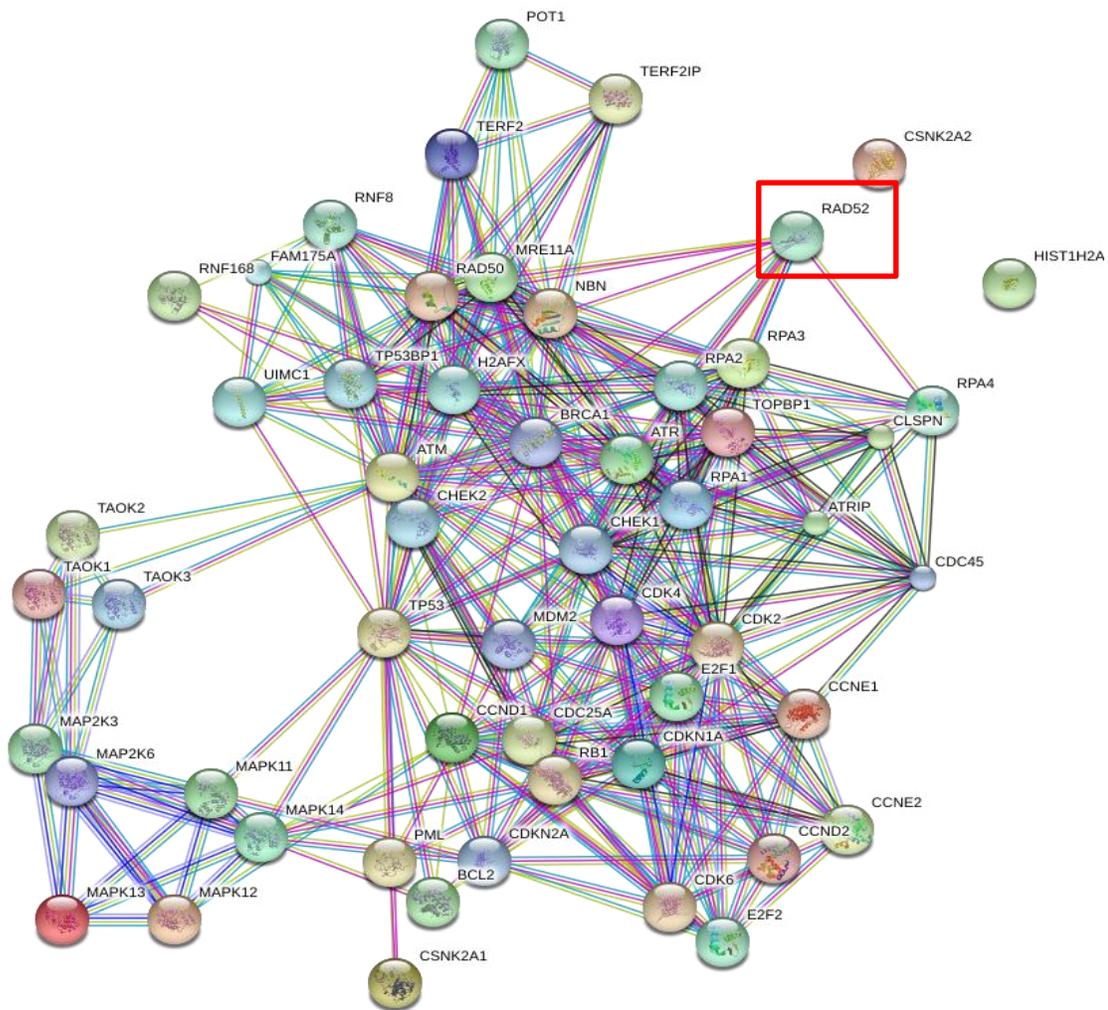


Figura 19 – União das vias (inflamação com reparo de DNA)

Nesta imagem pode-se visualizar também, que os genes do reparo de DNA então localizados mais ao centro do grafo, sendo assim, apresentando maior número de ligações e com maior intensidade. Os genes que não faziam ligação com nenhum outro na Figura 17, também não apresentaram ligação com a união das duas vias.

8.2 Replicative Senescence X Hr Repair of Replication-Independent DSB

Para melhor representar as ligações entre os genes das duas vias, optou-se por uma comparação com menor similaridade com a mesma via de inflamação

(*Replicative Senescence*), que é o caso da via *Hr Repair of Replication-Independent DSB* apresentado na Figura 20.

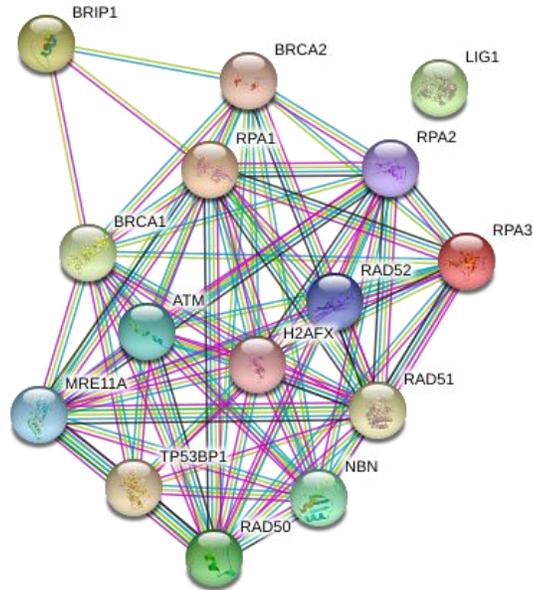


Figura 20 – Definição da String para a via *Hr Repair of Replication-Independent DSB*

Nesta via foi utilizado os mesmos 90% de confiabilidade, apresentando o gene *LIG1* sem nenhuma ligação, no entanto com um grau de confiabilidade de setenta por cento ele apresenta ligações com outros genes da via. Dentre os genes da via, cinco não pertencem a via *Replicative Senescence*, são eles: *BRIP1*, *BRCA2*, *LIG1*, *RAD51* e *RAD52*.

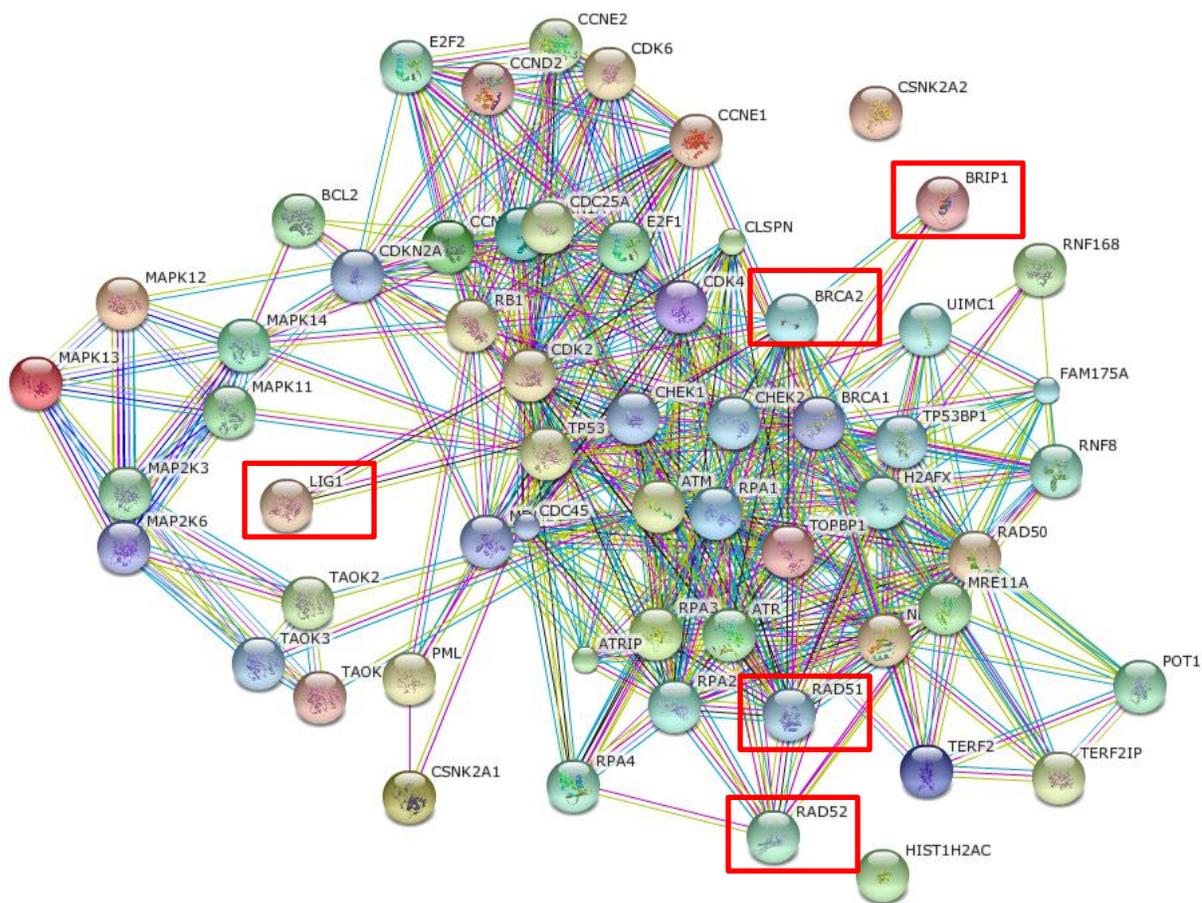


Figura 21 – União das vias Replicative Senescence e Hr Repair of Replication-Independent DSB

Com a união das duas vias, visto na Figura 21, pode-se notar maiores diferenças na compatibilidade dos genes que não estão presentes nas duas vias. Dos cinco genes citados anteriormente, o *BRIP1* continua fazendo as mesmas ligações da sua via original; os genes *BRCA2*, *RAD51* e *RAD52* fazem ligações com genes de fora da sua via original; e o gene *LIG1*, troca totalmente suas ligações, pois quando está somente na sua via, faz ligações de até setenta por cento de confiabilidade com os genes *MRE11A* e *TP53BP1*, no entanto quando é incorporado a via de inflamação, num grau de noventa por cento, não atinge os mesmos genes, mas sim dois outros da via de inflamação (*CDK2* e *CHECK1*).

Com as imagens do *String* pode-se verificar que as vias com maior similaridade também apresentam maior grau de ligação entre seus genes, mesmo eles não estando representados nas duas vias.

8.3 Sumário do capítulo

Este capítulo apresentou uma comparação entre as vias com maior compatibilidade a partir do *String*, mostrando as relações e intensidade de ligação de cada gene em sua via e em vias diferentes. Foram estudadas duas vias (*Processing of DNA DSB ends Recruitment of Repair and Sig. Proteins* e *Hr Repair of Replication-Independent DSB*) que apresentavam compatibilidade superior a sessenta por cento com a via *Replicative Senescence* do processo inflamatório.

No próximo capítulo serão apresentadas as considerações finais deste trabalho, bem como as sugestões de trabalhos futuros para o estudo genético de doenças da Ontocancro.

9 CONSIDERAÇÕES FINAIS

O estudo sobre bioinformática vem crescendo a partir da necessidade de encontrar soluções para doenças que assolam a população como o câncer e a diabetes. A Ontocancro é um desses projetos que desenvolve um estudo na área genética com o intuito de rastrear as ocorrências de alterações em vias metabólicas causadas por doenças. Sendo aplicado há seis anos, trabalhava primeiramente com o câncer, que quando não hereditário, advém de um processo inflamatório.

Nesse sentido, esta dissertação buscou atualizar essa informação no projeto Ontocancro inserindo o estudo sobre *Inflammaging*, definido como o estado inflamatório crônico adquirido com o envelhecimento. Essa atualização permitiu ampliar o banco de dados, a partir de uma nova ontologia que disponibilizou os dados de forma mais abrangente, expandindo as características sobre câncer, presente no projeto e incluindo as propriedades da diabetes do tipo 2 e do Alzheimer, também consequentes do processo inflamatório.

Para que fosse possível desenvolver essa atualização, realizou-se um estudo das doenças que seriam inseridas no projeto e quais as principais vias metabólicas do processo inflamatório. Após esse processo, gerou-se uma nova ontologia através do sistema Protégé. Anteriormente as doenças eram tratadas como instâncias da classe *Series*, porém com a nova disposição passaram a ser classes independentes dentro da classe *Inflammaging*, que passou a ser a classe pai que gerencia todas as doenças.

A classe *Samples* possuía as amostras das doenças e era dependente da classe *Séries*, passando a ser uma classe pai na nova ontologia, sendo ligada a classe *Inflammaging* a partir de restrições. Por fim, a classe *Pathways* tornou-se pai dos diferentes tipos de vias metabólicas, como a *Apoptosis*, *Cell_Cicle* e *Inflammation*.

Com a inserção dos novos dados, a Ontologia passou a abranger oito doenças, sendo elas seis de câncer: Pâncreas, Tireóide, Cólon Retal, Glândulas Adrenocorticais, Transtorno Pulmonar Obstrutivo Crônico - COPD, COPD em

pessoas fumantes e não fumantes, e duas inseridas a partir do *Inflammaging*: Alzheimer e Diabetes do tipo 2, todas inclusas no processo inflamatório crônico. Além disso, foram acrescentadas dezoito vias de inflamação e os genes aumentados de oitenta e seis para mil cento e cinco.

Este estudo contribuiu com uma ontologia que explicitasse de forma mais clara as doenças dispostas na Ontocancro, abrangendo a pesquisa para inserir o estudo sobre *Inflammaging* e incluindo duas novas doenças, o Diabetes do tipo 2 e o Alzheimer. Além disso permitiu uma visualização das similaridades entre as vias metabólicas do processo natural de desenvolvimento do organismo (ciclo celular e apoptose) e as vias metabólicas que influenciam na inflamação crônica. Com essa nova proposta de ontologia e a inserção de novas doenças, houve um acréscimo nos dados inseridos no banco, tanto no número de doenças, vias metabólicas e genes.

Para trabalhos futuros, pretende-se estudar o *plugin* desenvolvido pela GitHub Enterprise, o Ontop (BAGOSI et. al 2014), que serve para vincular a ontologia ao banco de dados de maneira que as informações dispostas na ontologia sejam automaticamente passadas para o banco de dados e vice versa. Esse *plugin* garante a inclusão dos dados no banco utilizando o Protégé nas versões mais recentes do programa, atualizando de forma automática o banco a partir da atualização da ontologia, como também impede a duplicação de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES-MAZZOTTI, A. J.; GEWANDSZNAJDER, F. **O método nas ciências naturais e sociais: pesquisa quantitativa e qualitativa**. São Paulo: Pioneira, 1998.
- ASHBURNER, Michael; BALL, Catherine A.; BLAKE, Judith A.; BOTSTEIN, David; BUTLER, Heather; CHERRY, J. Michael; DAVIS, Allan P.; DOLINSKI, Kara; DWIGHT, Selina S.; EPPIG, Janan T.; HARRIS, Midori A.; HILL, David P.; ISSELTARVER, Laurie; KASARSKIS, Andrew; LEWIS, Suzanna; MATESE, John C. Matese; RICHARDSON, Joel E.; RINGWALD, Martin; RUBIN, Gerald M.; SHERLOCK, Gavin. **Gene Ontology: tool for the unification of biology**. Nat Genet 2000. Doi:10.1038/75556. Disponível em <http://bioinfo2.ugr.es/PDFsClase/GeneOntology2000.pdf> Acesso em 03 de junho de 2015.
- ATHIÉ, Maria Carolina Pedro. **Análise da expressão gênica e proteômica em pacientes com doença de Alzheimer: busca de marcadores periféricos**. Instituto de Biociência da Universidade de São Paulo. Dissertação de mestrado. São Paulo: 2010.
- BAGOSI, Timea; CALVANESE, Diego; HARDI, Josef; KOMLA-EBRI, Sarah; LANTI, Davide; REZK, Martin; RODRÍGUEZ-MURO, Mariano; SLUSNYS, Mindaugas; XIAO, Guohui. **The Ontop Framework for Ontology Based Data Access**. The Semantic Web and Web Science - 8th Chinese Conference, CSWS 2014, ISBN: 978-3-662-45495-4 > DOI: 10.1007/978-3-662-45495-4_6. Disponível em <http://www.ghxiao.org/publications/2014-csws-ontop.pdf> Acesso em 15 de junho de 2015.
- BARRETO, Rosimar de Castro; PEREIRA, Giuseppe A. Scarano; COSTA, Lino João da; CAVALCANTI, Hellen Rose Barreto B. **O Duplo Papel da Inflamação no Surgimento das Lesões Cancerígenas**. Revista Brasileira de Ciências da Saúde, Volume 14 Número 4 Páginas 107-114. 2011. ISSN 1415-2177.

BIEGEL, Jaclyn A.; ZHOU, Jun-Ying; RORKE, Lucy B.; STENSTROM, Cindy; WAINWRIGHT, Luanne M.; FOGELGREN, Benjamin. **Germ-Line and Acquired Mutations of INI1 in Atypical Teratoid and Rhabdoid Tumors.** CANCER RESEARCH 59, 74–79, January 1, 1999. Disponível em <<http://cancerres.aacrjournals.org/content/59/1/74.full.pdf+html>> Acesso em 10 de dezembro de 2013.

BRITO, Ciro José; VOLP, Ana Carolina Pinheiro; NÓBREGA, Otávio de Toledo; JÚNIOR, Fernando Lopes e Silva; MENDES, Edmar Lacerda; ROAS, Aendria Fernanda Castro Martins; BARROS, Jonatas de França; CÓRDOVA, Cláudio. **Exercício físico como fator de prevenção aos processos inflamatórios decorrentes do envelhecimento.** Artigo de revisão. Motriz, Rio Claro, v.17 n.3, p.544-555, jul./set. 2011. Disponível em <http://www.scielo.br/pdf/motriz/v17n3/17.pdf> Acesso em 22 de abril de 2015.

BROCHHAUSEN, Mathias et al. **The ACGT Master Ontology and its applications – Towards an ontology-driven cancer research and management system.** Journal of Biomedical Informatics 8-25. 2011.

CABRAL, Heleno C. B.; LIBRELOTTO, Giovani Rubert; SIMÃO, Éder M.; SINIGAGLIA, Marialva; CASTRO, Mauro A. A.; MOMBACH, José C. M. **O Processamento de uma Ontologia sobre a Integração de Dados de Vias de Interação Molecular Envolvidas em Câncer.** Revista Brasileira de Computação Aplicada. v. 3, n. 1. Passo Fundo: 2011.

CASTRO, Mauro A. A.; FILHO, José L. Rybarczyk; DALMOLIN, Rodrigo J. S.; SINIGAGLIA, Marialva; MOREIRA, José C. F.; MOMBACH, José C. M.; ALMEIDA, Rita M. C. de. **ViaComplex: software for landscape analysis of gene expression networks in genomic context.** Bioinformatics Volume 25, Ed. 11. Pp. 1468-1469. 2009. Disponível em <http://bioinformatics.oxfordjournals.org/content/25/11/1468.long> Acesso em 12 de maio de 2015.

INCA, Instituto Nacional do Câncer. **ABC do câncer: abordagens básicas para o controle do câncer**. Ministério da Saúde. – Rio de Janeiro: Inca, 2011. Disponível em <http://goo.gl/n387Mb> Acesso em 06 de julho de 2015.

COCOS, C. **Design Principles of the ACGT Master Ontology: Examples and Discussion**. Information Society Technologies, Dezembro 2008. Disponível em <http://www.uni-saarland.de/fileadmin/user_upload/Institute/ifomis/documentation/ACGT_D7.7_IFO_MIS_Final.pdf> Acesso em 16 de julho de 2014.

COURTOT, Mélanie; JUTY, Nick; et al. **Controlled vocabularies and semantics in systems biology**. Molecular Systems Biology. 2011.

CRUZ, Isabel F.; ANTONELLI, Flavio Palandri; STROE, Cosmin. **AgreementMaker: Ecient Matching for Large Real-World Schemas and Ontologies**. VLDB '09, August 24-28, 2009, Lyon, France 2009 VLDB Endowment, ACM. Disponível em <<http://disi.unitn.it/~p2p/RelatedWork/Matching/vldb2009-revised.pdf>> Acesso em 27 de outubro de 2013.

CUSTÓDIO, Marco Paulo Machado. **Sistema de informação e extração de conhecimento para análise e gestão de informação médica**. Mestrado em Engenharia Informática e de Computadores. Instituto Superior Técnico Universidade de Lisboa, 2012. Disponível em https://fenix.tecnico.ulisboa.pt/downloadFile/395144681086/Dissertacao_50136.pdf Acesso em 22 de junho de 2015.

DANTAS, Adriana. **Bioinformática: Banco de Dados**. UERGS, Bento Gonçalves, 2013. Disponível em <http://www.slideshare.net/AdrianaDantas2/bioinformtica-arquivo-de-dados> Acesso 01 de julho de 2013.

DELISLE, Marc. **Mastering Phpmyadmin 3.1 for Effective Mysql Management**. Packt Publishing Ltd, 2009.

EDGAR, Ron; DOMRACHEV, Michael; LASH, Alex E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repositior. Nucleic Acids Research, 2002, Vol. 30, issue 1. Disponível em

<http://nar.oxfordjournals.org/content/30/1/207.short> Acesso em 03 de junho de 2015.

FERREIRA, Valceir Aparecido; CAMPOS, Simone Marques Bolonheis de. **Avanços farmacológicos no tratamento do Diabetes tipo 2**. Brazilian Journal of Surgery and Clinical Research – BJSCR. Vol. 8, n. 3, pp. 72 – 78. 2014. Disponível em http://www.mastereditora.com.br/periodico/20141101_221529.pdf Acesso em 10 de fevereiro de 2015.

FRANCESCHI C; CAPRI, M; MONTI, D; GIUNTA, S; OLIVIERI, F; SEVINI, F; PANOURGIA, MP; INVIDIA, L; CELANI, L; SCURT, M; CEVENINI, E; CASTELLANI, GC; SALVIOLI, S. **Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans**. Mech Ageing Dev. 2007; 128:92-105. Disponível em <http://www.ncbi.nlm.nih.gov/pubmed/17116321> Acesso em 27 de setembro de 2014.

FULOP, Tamas; PAWELEC, Graham; DUPUIS, Gilles; KOTB, Rami; FRIGUET, Bertrand; LARBI, Anis. **Immunosenescence, Oxidative Stress, and Cancers**. Cancer Immunology 2015, pp 377-393.

GARVIN, A. Julian; RE, Gian G.; TARNOWSKI, Betty; HAZEN-MARTIN, Debra J.; SENS, Donald A. **The G401 Cell Line, Utilized for Studies of Chromosomal Changes in Wilms' Tumor, Is Derived from a Rhabdoid Tumor of the Kidney**. American journal of Pathology, Vol. 142, No. 2, 1993. Disponível em <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1886739/pdf/amjpathol00074-0031.pdf> Acesso em 10 de dezembro de 2013.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

GIUNTA, B; FERNANDEZ, F; NIKOLIC, WV; OBREGON, D; RRAPO, E; TOWN, T; TAN, J. **Inflammaging as a prodrome to Alzheimer's disease**. J Neuroinflammation. 2008. Disponível em <http://www.ncbi.nlm.nih.gov/pubmed/19014446> Acesso em 27 de setembro de 2014.

- GRUBER, T. R. **Toward principles for the design of ontologies used for knowledge sharing.** In: Formal ontology in conceptual analysis and knowledge representation (Guarino N and Poli R, eds.) Kluwer Academic, Dordrecht, Netherlands, 1993.
- GUARINO, Nicola. **Formal Ontology and Information Systems.** Book Formal Ontology in Information Systems: Proceedings of the 1st International Conference, June 1998, Pages 3 - 15, ISBN:9051993994
- HALAZONETIS, T. D.; GORGOULIS, V. G.; BARTEK, J. **An oncogene-induced DNA damage model for cancer development.** *Science (New York, N.Y.)*, v. 319, n. 5868, p. 1352–55, 7 mar. 2008. Disponível em: <<http://www.sciencemag.org/content/319/5868/1352.abstract>>. Acesso em 7 de abril de 2014.
- HEUSER, Carlos A. **Projeto de Banco de dados.** Editora Sagra, 4ª ed. UFRGS: 1998.
- HORRIDGE, Matthew. **A Practical Guide To Building OWL Ontologies Using Prétége 4 and CO-ODE Tools - Edition 1.3.** 2011. Disponível em https://mariaiulianadascalu.files.wordpress.com/2014/02/owl-cs-manchester-ac-uk-eowltutorialp4_v1_3.pdf Acesso em 12 de maio de 2015.
- HUNT, KJ; WALSH, BM; VOEGELI, D; ROBERTS, HC. **Inflammation in aging. Part 2: implications for the health of older people and recommendations for nursing practice.** *Biol Res Nurs.* 2010; 11: 253-60. Disponível em <http://www.ncbi.nlm.nih.gov/pubmed/19934108> Acesso em 27 de setembro de 2014.
- KANEHISA, Minoru; GOTO, Susumu. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 2000, Vol. 28, issue 1. Disponível em <http://nar.oxfordjournals.org/content/28/1/27.full> Acesso em 03 de junho de 2015.
- IHAKAA, Ross; GENTLEMANA, Robert. R: **A Language for Data Analysis and Graphics.** *Journal of Computational and Graphical Statistics*, Volume 5, Issue 3, 1996. Disponível em <http://goo.gl/yHRqwi> Acesso em 18 de setembro de 2015.

LICHTENSTEIN, Flávio; SIGULEM, D. Criando uma Ontologia em Saúde com a ferramenta Protégé no padrão OWL. Congresso Brasileiro de Informática em Saúde, Campos do Jordão, São Paulo SP: 2008. Disponível em <http://www.sbis.org.br/cbis11/arquivos/992.pdf/>. Acesso em 25 de setembro de 2014.

MAGLOTT, Donna; OSTELL, Jim; PRUITT, Kim D.; TATUSOVA, **Tatiana**. **Entrez Gene: gene-centered information at NCBI**. *Oxford Journals, Nucleic Acids Research*, Volume 35, Edição 1. Pp. D26-D31. 2007. Disponível em http://nar.oxfordjournals.org/content/35/suppl_1/D26.full.pdf+html Acesso em 13 de maio de 2015.

MOMBACH JCM, VENDRUSCULO B, BUGS CA. A Model for p38MAPK-Induced Astrocyte Senescence. PLoS ONE 10(5): e0125217. doi: 10.1371/journal.pone.0125217. 2015. Disponível em <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0125217> Acesso em 17 de setembro de 2015.

MORO, Juliano; HORN, Roberta Cattaneo. **Inflamação e implicações no Câncer**. XVIII Seminário Interinstitucional de Ensino, Pesquisa e Extensão. UNICRUZ, 2013. Disponível em <http://goo.gl/B2enTI> Acesso em 27 de setembro de 2014.

NASCIMENTO, Karlise Soares; BASTIANI, Ederson; LIBRELOTTO, Giovani R. **Ontocancro 2.0: um estudo de caso para a aplicação da ontologia em vias metabólicas ligadas ao processo carcinogênico**. REVISTA DO CCEI – Centro de Ciências da Economia e Informática Vol. 16 Nº30 ISSN 1415-2061. Bagé: 2012. Disponível em http://revista.urcamp.tche.br/index.php/Revista_CCEI/issue/view/Issue/32/pdf_28 Acesso em 12 de maio de 2015.

NOY, Natalya F.; MCGUINNESS, Deborah L. **Ontology Development 101: A Guide to Creating Your First Ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001. Disponível em

http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

Acesso em 12 de maio de 2015.

PEREIRA, Rafael Teodósio; AZEVEDO, Renato P. de; SIMÃO, Eder M.; MOMBACH, Jose C.M.; LIBRELOTTO, Giovani R. **Utilizando a Ontocancro para traçar o perfil das vias de manutenção da estabilidade genômica.** – Congresso da Sociedade Brasileira de Computação. XII Workshop de Informática Médica – WIM 2012. Disponível em http://www.imago.ufpr.br/csbc2012/anais_csbc/eventos/wim/artigos/WIM2012%20-%20Utilizando%20a%20Ontocancro%20para%20tracar%20o%20perfil%20das%20vias%20de%20manutencao%20da%20estabilidade%20genomica.pdf Acesso em 12 de maio de 2015.

PEZZINI, Flávia Fonseca; et al. **Armazenamento e Disponibilização de dados e metadados PPBio/CENBAM.** 2009. Disponível em <http://ppbio.inpa.gov.br/sites/default/files/Armazenamento%20e%20Disponibiliza%C3%A7%C3%A3o%20de%20dados%20e%20metadados.pdf> Acesso em 30 de junho de 2013.

POVEY, Sue; LOVERING, Ruth; BRUFORD, Elspeth; WRIGHT, Mathew; LUSH, Michael; WAIN, Hester. **The HUGO Gene Nomenclature Committee (HGNC).** Human Genetics 2001, Volume 109, Issue 6. Disponível em <http://link.springer.com/article/10.1007%2Fs00439-001-0615-0?LI=true> Acesso em 03 de junho de 2015.

ROCHA, Natália Pessoa; REIS, Helton José; TEIXEIRA, Antônio Lúcio. **Inflammaging: um novo paradigma para as ciências gerontológicas.** Geriatria & Gerontologia, pg 224-228, 2010. Disponível em <http://goo.gl/9BPAhc> Acesso em 27 de setembro de 2014.

SANTOS, Fabrício R. **Sobre Bioinformática, genoma e ciência.** Revista Ciência Hoje vol. 35 nº 209 do Instituto de Ciências Biológicas, UFMG, 2004, p 54 – 57. Disponível em

http://www2.bioqmed.ufrj.br/prosdocimi/divulgacao/Prosdocimi_CHoje.pdf Acesso em 18 de junho de 2013, às 21hs e 06 min.

SANTOS, Fabrício R.; ORTEGA, José Miguel. **Bioinformática aplicada à Genômica**. Manuscrito para capítulo do Biowork IV. Universidade Federal de Minas Gerais. 2006. Disponível em <ftp://www.ufv.br/DBG/material%20curso%20bioinfo/Leitura%20Complementar/artigos/Bioinform%E1tica%20aplicada%20%E0%20Gen%F4mica.pdf> Acesso e 01 de julho de 2013.

SERENIKI, Adriana; VITAL, Maria Aparecida Barbatto Frazão. **A doença de Alzheimer: aspectos fisiopatológicos e farmacológicos**. Revista Psiquiátrica RS, 2008. Disponível em <http://www.scielo.br/pdf/rprs/v30n1s0/v30n1a02s0.pdf> Acesso em 05 de fevereiro de 2015.

SCHAEFER, Carl F.; ANTHONY, Kira; KRUPA, Shiva; BUCHOFF, Jeffrey; DAY, Matthew; HANNAY, Timo; BUETOW, Kenneth H. PID: the Pathway Interaction Database. Nucleic Acids Research, 2009, Vol. 37, issue 1. Doi:10.1093/nar/gkn653. Disponível em http://nar.oxfordjournals.org/content/37/suppl_1/D674.full Acesso em 03 de junho de 2015.

SILVEIRA, Denise Tolfo; CÓRDOVA, Fernanda Peixoto. **A pesquisa Científica. In.: Métodos de Pesquisa. Série Educação a Distância. 1ª ed. Editora UFRGS: 2009.** Disponível em <http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf> Acesso em 14 de abril de 2015.

SIMÃO, É. M. **Dinâmica da Transição Pré-Câncer para Câncer: Estudo da Expressão de Vias de Manutenção do Genoma**. Universidade Federal de Santa Maria, 2012. Disponível em <http://goo.gl/vlbT5k> Acesso em 17 de setembro de 2015.

SZKLARCZYK D.; FRANCESCHINI A.; WYDER S.; FORSLUND K.; HELLER D.; HUERTA-CEPAS J.; SIMONOVIC M.; ROTH A.; SANTOS A.; TSAFOU K. P.; KUHN M.; BORK P.; JENSEN L. J.; VON MERING C. **STRING v10: protein-**

protein interaction networks, integrated over the tree of life. Nucleic Acids 2015. PMID:25352553. Disponível em <http://www.ncbi.nlm.nih.gov/pubmed/25352553> Acesso em 10 de junho de 2015.

TONET, Audrey Cecília. **Associação do polimorfismo -174 G/C da região promotora do gene IL-6 com fatores de risco cardiovascular em idosas brasileiras.** Programa de Pós-Graduação Strictu Sensu em Gerontologia. Brasília 2007. Disponível em http://www.btdt.ucb.br/tede/tde_busca/arquivo.php?codArquivo=668 Acesso em 10 de fevereiro de 2015.

VOGT, Carlos. **Será a bioinformática uma área já constituída?** Com Ciência, Revista Eletrônica de Jornalismo Científico. Unicamp, 2003. Disponível em <http://www.comciencia.br/reportagens/bioinformatica/bio04.shtml> Acesso em 18 de junho de 2013, às 20hs e 46 min.

WISSE, Brent. **Diabetes Tipo 2**, MedLinePlus, 2014. Disponível em <http://goo.gl/bqLUZ> acesso em 05 de Fevereiro de 2015.

WHEELER, David L.; BARRETT, Tanya; BENSON, Dennis A.; BRYANT, Stephen H.; CANESE, Kathi; CHETVERNIN, Vyacheslav; CHURCH, Deanna M.; DICUCCIO, Michael; EDGAR, Ron; FEDERHEN, Scott; GEER, Lewis Y.; KAPUSTIN, Yuri; KHOVAYKO, Oleg; LANDSMAN, David; LIPMAN, David J.; MADDEN, Thomas L.; MAGLOTT, Donna R.; OSTELL, James; MILLER, Vadim; PRUITT, Kim D.; SCHULER, Gregory D.; SEQUEIRA, Edwin; SHERRY, Steven T.; SIROTKIN, Karl; SOUVOROV, Alexandre; STARHENKO, Grigory; TATUSOV, Roman L.; TATUSOVA, Tatiana A.; WAGNER, Lukas; YASCHENKO, Eugene. **Database resources of the National Center for Biotechnology Information.** Oxford Journals Science & Mathematics Nucleic Acids Research Volume 35, 2006. Disponível em <http://goo.gl/ZtBH4T> Acesso em 06 de julho de 2015.

LABARGA, Alberto; VALENTIN, Franck; ANDERSON, Mikael; LOPEZ, Rodrigo. **Web Services at the European Bioinformatics Institute.** Oxford Journals Science &

Mathematics Nucleic Acids Research Volume 35, 2007. Disponível em <http://goo.gl/cGp53y> Acesso em 06 de julho e 2015.

MIYAZAKI, S.; SUGAWARA, H.; GOJOBORI, T.; TATENO, Y. **DNA Data Bank of Japan (DDBJ) in XML**. Oxford Journals Science & Mathematics Nucleic Acids Research Volume 31, 2003. Disponível em <http://goo.gl/OB0yVk> Acesso em 06 de julho de 2015.

BERNSTEIN, Frances C.; KOETZLE, Thomas F.; WILLIAMS, Graheme J.B.; MEYER, Edgar F. Jr.; BRICE, Michael D.; RODGERS, John R.; KENNARD, Olga; SHIMANOUCI, Takehiko; TASUMI, Mitsuo. **The protein data bank: A computer-based archival file for macromolecular structures**. Journal of Molecular Biology, Volume 112, Issue 3, 25 May 1977, Pages 535-542. Disponível em <http://goo.gl/RoNzQD> Acesso em 06 de julho de 2015.