



## Representando o conhecimento da inflamação crônica através da Ontocancro

Jéssica A. Bonini<sup>1</sup>, Rômulo M. Stringhini<sup>1</sup>, Laís Falcade<sup>1</sup>, Giovani R. Librelotto<sup>1</sup>

<sup>1</sup>Pós-graduação em Informática – Universidade Federal de Santa Maria (UFSM)  
Caixa Postal 5082 – 97195-000 – Santa Maria – RS – Brasil

{jbonini, rstringhini, librelotto}@inf.ufsm.br,  
laisfalcade@hotmail.com

**Abstract.** *The Ontocancro project was created in order to integrate data from public databases of genes related to cancer. Throughout the research discovered the concept of chronic inflammation and its relation to diseases like alzheimer's, type 2 diabetes and the cancer itself. Thus, it was decided to update the ontology unifying the study of cancer of the previous versions of Ontocancro and specifications of chronic inflammation, entering new information about the diseases mentioned above. The update of the ontology to version 3.0, presented in this work, followed the steps proposed by Noy and McGuinness and even used the Protégé tool. In addition, updated to the relational database to reflect that knowledge inserted in the new version of the ontology. The results are presented other existing ontologies data compared to the ontology data developed in this article.*

**Resumo.** *O projeto Ontocancro foi criado com objetivo de integrar dados, de bases de dados públicas, de genes referentes ao câncer. Ao longo das pesquisas descobriu-se o conceito da inflamação crônica e sua relação com doenças como alzheimer, diabetes tipo 2 e o próprio câncer. Assim, optou-se por atualizar a ontologia unificando o estudo do câncer das versões anteriores da Ontocancro e as especificações da inflamação crônica, inserindo informações novas sobre as doenças citadas anteriormente. A atualização da ontologia para a versão 3.0, apresentada neste trabalho, seguiu os passos propostos por Noy e McGuinness e ainda utilizou a ferramenta Protégé. Além disso, atualizou-se o banco de dados relacional para que esse refletisse os conhecimentos inseridos na nova versão da ontologia. Como resultados são apresentados dados de outras ontologias existentes comparados com os dados da ontologia desenvolvida nesse artigo.*

### 1. Introdução

Mutações e lesões no DNA ocorrem de forma espontânea ou devido a exposição a condições externas. Diversas vias são responsáveis pelo reparo do DNA e o funcionamento incorreto dessas é fator decisivo para o desenvolvimento do câncer. Nesta situação, ocorre um comprometimento da integridade genômica e o favorecimento de mutações em oncogenes e genes supressores tumorais [Barbie and Frank 2009]. A origem do câncer está no desenvolvimento anormal de uma parte das células que começam a crescer desordenadamente, sem cumprir o ciclo natural das células. O projeto da Ontocancro surgiu a partir da necessidade de integrar dados, de diferentes bases de dados públicas, de genes relacionados ao câncer. A ontologia Ontocancro 1.0 visa prover dados centralizados e tornar possível a análise coesa das informações

[Cabral et. al 2011]. Na Ontocancro 2.0, somente vias curadas foram inseridas na base de dados e ocorreu a divisão das vias em subvias e o agrupamento dos genes com maior proximidade [Nascimento et. al 2012]. Ainda na versão 2.0, foi adicionado um cálculo estatístico de expressividade dos genes nas vias [Pereira et. al 2013].

A descoberta do conceito da inflamação crônica (*Inflammaging*) e sua relação com doenças como Alzheimer, diabetes tipo 2 e o próprio câncer motivou a atualização da ontologia para a Ontocancro 3.0. O *Inflammaging* é definido como o processo inflamatório crônico decorrente do envelhecimento. Doenças como o Alzheimer e a Diabetes tipo 2 podem ter sua origem no estado de inflamação crônica [Franceschi et. al 2007]. Existe uma relação entre a liberação de citocinas pró-inflamatórias e o processo de produção anormal de proteínas e a presença de marcadores no sangue de pacientes portadores das doenças citadas, levando a crer que por vezes a origem destas está mesmo ligada ao processo inflamatório advindo do envelhecimento [Giunta et.al 2008]. Além disso ligando a inflamação crônica ao câncer está o fato de que a deterioração natural do sistema imunológico produzido pelo envelhecimento contribuí para a ocorrência de tumores [Agondi et. al 2012].

Considerando a ligação da inflamação crônica com o projeto Ontocancro, o objetivo desse trabalho é unificar em uma ontologia o estudo do câncer das versões anteriores da Ontocancro e as especificações do termo *Inflammaging*. Utilizando as amostras das doenças advindas do processo inflamatório (*samples*), vias metabólicas (*pathways*) e os genes disponíveis na base de dados do projeto e nas bases de dados públicas Hugo, a qual contém todos os genes presentes no ser humano, NCI e Reactome, que trazem as vias metabólicas, e o GEO, que disponibiliza amostras de tecido de diversas doenças.

Este trabalho está estruturado da forma apresentada a seguir: a seção 2 apresenta a revisão bibliográfica; a seção 3 apresenta a construção da ontologia, foco principal deste trabalho; a seção 4 apresenta a atualização do banco de dados relacional; e, por fim, na seção 5 mostra-se a comparação entre ontologias inseridas no contexto do artigo.

## 2. Fundamentação Teórica

Esta seção destina-se à apresentação de conceitos teóricos necessários para a compreensão deste trabalho.

### 2.1. Ontologia Ontocancro

A integração dos dados provenientes de estudos de redes complexas de interações moleculares, responsáveis pelo comportamento das células, exige o desenvolvimento de novas ferramentas. Neste contexto, pesquisadores da área de física, biologia e informática, desenvolveram uma ontologia para auxiliar no estudo da expressão gênica de redes biológicas de genes envolvidos no câncer, a Ontocancro [Cabral et. al 2011].

A Ontocancro 1.0 tinha como objetivo principal estruturar o máximo de dados de genes das vias de manutenção de estabilidade genômica em uma base de dados curada pelos próprios pesquisadores. Esta base de dados é constituída por informações retiradas de diferentes bancos de dados públicos e padronizadas através da ontologia [Cabral et. al 2011].

A necessidade de atualização na Ontocancro 1.0 surgiu a partir de análises que apontaram um grande número de genes presente nas vias o que dificultava a análise de

expressão dos mesmos. Além disso a falta de curagem das vias em seus repositórios de origem resultavam muitas vezes em distorções nos resultados. Para solucionar estes problemas optou-se por separar as vias em subvias, agrupando genes com maior proximidade, e por reestruturar a base de dados inserindo apenas vias curadas. A Ontocancro 2.0 traz vias diretamente ligadas à formação do câncer, retiradas da base de dados NCI, que contém informações curadas sobre vias metabólicas, interações biomoleculares e processos celulares [Nascimento et. al 2012].

Mais tarde, foi acrescentado ao estudo da Ontocancro 2.0, um cálculo de expressividade dos genes nas vias. Esse cálculo permite a integração das informações referentes às doenças estudadas e as vias de manutenção do genoma, traçando um perfil dessas vias e sua atividade nas fases do câncer [Pereira et. al 2013].

## 2.2. Inflamação Crônica

O envelhecimento traz consigo o aumento na produção de citocinas pró-inflamatórias e o enfraquecimento dos sistemas que compõe o corpo humano. Entre eles estão o sistema endócrino e imunológico. Com o envelhecimento das células a resposta do sistema às inflamações diminui e isso pode levar a uma inflamação crônica, o *Inflammaging* [Hunt et. al 2010].

O *Inflammaging* é o estado de inflamação crônica que ocorre com o envelhecimento e a longevidade, e é ocasionado pelo desequilíbrio entre as redes inflamatórias e anti-inflamatórias [Franceschi et. al 2007]. O estado inflamatório é consequência direta da imunossenescência, a qual é considerada um estado desregulado da função imunológica e que permite a exposição a patologias relacionadas com a idade. A capacidade de renovação das células do sistema imunológico diminui a medida que a idade avança. Dessa forma, há uma alteração na população de linfócitos B e das células T, causando também uma alteração em seus comportamentos. Além do corpo estar propenso a adquirir doenças, existe uma redução significativa a resposta vacinal [Agondi et. al 2012].

Doenças como o Alzheimer e a Diabetes tipo 2 podem ter sua origem no *Inflammaging* [Franceschi et. al 2007]. O Alzheimer é uma doença neurodegenerativa que ocorre devido ao dano irreversível das células cerebrais e ocasiona perda de funções cognitivas como memória e linguagem. Uma das principais alterações apresentadas por pacientes com Alzheimer é a produção anormal da proteína beta-amilóide [Lucena 2007]. Existe uma relação entre a liberação de citocinas pró-inflamatórias e o processo de produção da proteína beta-amilóide, levando a crer que em certas situações o Alzheimer está ligado ao *Inflammaging* [Giunta et.al 2008].

A Diabetes tipo 2 ocorre quando o organismo não consegue lidar com a insulina produzida pelo pâncreas. Ou ainda, quando a produção de insulina é insuficiente para controlar o nível de glicose no sangue [Cavalcanti and Guzen 2012]. A resposta imunológica é caracterizada pela presença de marcadores no sangue e um desses é a IL-6, citocina presente também, em maior concentração, no organismo pessoas portadoras da Diabetes tipo 2. Dessa forma, acredita-se que essa patologia também seja decorrente de uma inflamação sistêmica atribuída ao desequilíbrio do sistema imunológico, o que caracteriza o *Inflammaging* [Tonet and Nóbrega 2007].

Ligando o *Inflammaging* ao câncer, foco central do projeto Ontocancro, está o fato de que a imunossenescência causada pelo processo de envelhecimento contribuí

para a ocorrência de tumores. A combinação de eventos internos, como o aumento de células inflamatórias e o desenvolvimento de radicais livres, com eventos externos, como a exposição a diversos patógenos ao longo do tempo, leva a alterações no sistema imune e conseqüentemente ao desenvolvimento do câncer [Agondi et. al 2012].

### 3. Ontocancro 3.0

A ontologia modelada para representar o *Inflammaging* [Franceschi et. al 2007] e as atualizações realizadas sob a Ontocancro 2.0 são apresentadas nesta seção. A Ontocancro 2.0 trazia informações sobre o câncer, já a Ontocancro 3.0 adiciona duas novas doenças ao estudo, o Alzheimer e a Diabetes tipo 2. Estas doenças estão inseridas no contexto do *Inflammaging*, que trata do estado inflamatório crônico que surge com o envelhecimento

Para a construção da nova ontologia proposta utilizou-se a ferramenta *Protégé*, amplamente utilizada na criação e edição de ontologias, pois possibilita o desenvolvimento rápido e fácil de todos os níveis de ontologias [Pereira 2015]. O desenvolvimento da ontologia seguiu cinco passos baseados na metodologia de Noy e McGuinness (2001): definição do domínio e escopo da ontologia; criação e organização hierárquica das classes presentes na ontologia; propriedades de dados e objetos; definição das restrições destas propriedades; e criação das instâncias das classes. A seguir são detalhadas estas etapas.

#### 3.1. Domínio e escopo da ontologia

Visando a atualização dos dados da Ontocancro buscou-se novas possibilidades de estudo do Câncer. Durante estas pesquisas, descobriu-se que o Câncer quando não tem sua origem na hereditariedade pode ser desenvolvido por um processo inflamatório [Barreto et. al 2010]. Encontrou-se também o estudo que trata do *Inflammaging*, que causa doenças como o Alzheimer e Diabetes tipo 2 [Franceschi et. al 2007]. Dessa forma, a ontologia desenvolvida busca mapear o conhecimento genético sobre o estudo do *Inflammaging*, disponibilizando informações sobre as doenças inseridas neste contexto.

As doenças elencadas para o domínio em questão 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. Assim, as informações presentes na ontologia foram retiradas de bases de dados públicas, como o Hugo, que contém todos os genes presentes no corpo humano, o NCI e o Reactome, que trazem as vias metabólicas, e o GEO, que disponibiliza amostras de tecido de diversas doenças.

#### 3.2. Classes e Instâncias

Na versão 1.0 da Ontocancro, as duas classes principais encontradas são a *Entity* e a *UtilityClass*, as quais referem-se ao estudo dos dados genéticos voltados para o Câncer baseado na proposta de Cabral et al. (2011) e ao estudo sobre o *ViaComplex*, aplicação que constrói mapas de redes de expressão gênica [CASTRO et al. 2009], respectivamente. Na Ontocancro 2.0 estão presentes três das principais classes propostas no estudo de Cabral et. al (2011): *Genes*, *Interaction* e *Pathways* e seu foco é o estudo da barreira anticâncer, através da segmentação das vias em subvias [Halazonetis et. al 2008] e do desenvolvimento de um cálculo de atividade e diversidade relativa

responsável por determinar os genes mais ativos ou presentes em cada doença.

A classe *Genes* apresenta os genes mais significativos no processo cancerígeno e esses são vinculados a uma via metabólica representada pela classe *Pathways*. Por último a classe *Interaction* apresenta a combinação via a via, possibilitando a divisão em grupos: vias de apoptose, vias de ciclo celular, vias de estabilidade genômica e vias de reparo do DNA. Essas vias dão sustentabilidade aos mecanismos que auxiliam na barreira contra a propagação do câncer [Pereira et. al 2013]. Na atualização da Ontocancro 2.0 ainda foram inseridas duas classes novas: *Affymetrics* e *Series*. A primeira faz a relação entre os genes e as doenças, já a segunda disponibiliza amostras das doenças retiradas da base de dados GEO.

Para a atualização da ontologia, Ontocancro 3.0, o foco principal do projeto, que era a visualização do câncer em todos os aspectos, foi alterado. Assim, como citado anteriormente, o novo objetivo é o estudo do *Inflammaging*. As doenças inseridas no contexto do *Inflammaging* têm especificações diferentes quanto às vias metabólicas e os genes com maior expressividade. Nesta atualização a classe *Series* deu lugar a classe abstrata *Inflammaging*, da qual deriva uma subclasse para cada doença estudada. *Samples*, anteriormente uma subclasse de *Series*, foi definida como classe e essa é ligada às doenças ou séries, a partir de relacionamentos e instâncias. As classes *Affymetrics* e *Genes* continuam sendo as classes bases da ontologia.

As instâncias representam os objetos no domínio de interesse e cada uma deve pertencer a uma determinada classe. No projeto Ontocancro, as informações necessárias para o estudo são pesquisadas em bases de dados e instanciadas nas suas classes de referência. As vias metabólicas encontradas nas células estão disponíveis em bases de dados como NCI e o Reactome, já os dados sobre as doenças e as amostras de tecidos normais ou com alterações são disponibilizadas no GEO. Por último, dados sobre todos os genes existentes no corpo humano podem ser encontrados na base de dados HUGO.

### 3.3. Propriedades de dados e objetos

As propriedades são responsáveis por estabelecerem uma relação entre dois indivíduos ou entre indivíduo e dados. Um exemplo da segunda relação é o nome de determinado indivíduo [Lóscio 2012]. Na classe *Genes* pode-se citar como principal propriedade o *entrezGene*, que representa o identificador central do gene. Pode-se citar ainda as propriedades *hugoName* e *nciName*, que caracterizam as bases de dados HUGO e NCI, e *approvedSymbol*, *approveSymbol2* e *previousSymbols*, que determinam as versões e atualizações dos símbolos de cada gene. A classe *Genes* determina a classificação de cada gene apresentado nos estudos das doenças, dessa forma é a classe com mais propriedades de dados na ontologia proposta.

A classe *Pathways* apresenta propriedades importantes sobre as vias metabólicas. As principais são *db\_name*, *pathway\_name* e *pathway\_type*, e indicam a base de origem da via, seu nome e a qual grupo pertence, respectivamente. A classe *Affymetrics* faz relação com a classe *Genes* através da propriedade *entrezGene* e com a classe *Samples* através da *affymetrics*. Esta última propriedade é representada em *Samples* por *id\_ref*. No caso das propriedades de objeto pode-se citar a relação entre as classes *Affymetrics*, *Samples* e *Inflammaging*. Esta relação permite que múltiplas instâncias de *affymetrics* estejam em uma amostra (*sample*) e que uma doença (classe *Inflammaging*) contenha múltiplas amostras. É possível ainda que uma mesma propriedade seja usada por mais de uma classe, nesse caso é necessário determinar restrições.

### 3.4. Restrições

No caso da ontologia proposta neste trabalho serão analisadas três tipos de restrições, os mesmos encontrados em *Web Ontology Language* (OWL): restrições de quantificador, restrições de cardinalidade e restrições de valor.

As restrições de quantificador são aquelas onde uma classe contém indivíduos os quais atendem a uma determinada propriedade. Existem dois quantificadores: o universal, o qual determina que todos os indivíduos relacionados devem ser de uma determinada classe; e o existencial, o qual descreve que há pelo menos um indivíduo relacionado. Um exemplo disso, é que todo *gene* deve conter pelo menos um valor de *affymetrics*, no relacionamento entre as classes *Genes* e *Affymetrics*.

Já as restrições de cardinalidade são utilizadas para restringir a quantidade de valores que uma propriedade pode assumir. Na ontologia pode-se citar o exemplo onde cada gene deve conter no mínimo um *affymetrics*, ou seja, a cardinalidade é definida como 1. A última restrição é a de valor, esta descreve indivíduos que possuem pelo menos um relacionamento com outros indivíduos especificados por uma propriedade. O exemplo que pode ser apresentado é que uma amostra pertence somente a uma doença.

### 3.5. Ontologia construída

Na atualização da ontologia foram inseridas ainda 18 pathways referentes ao processo inflamatório: *IL10 Anti-inflammatory Signaling Pathway*, *IL6-mediated signaling events*, *IL8- and CXCR1-mediated signaling events*, *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*. Assim, somando esse valor às 37 vias já encontradas no projeto, obteve-se um total de 55 vias metabólicas após a atualização. Como citado anteriormente, as duas principais bases de dados consultadas para a inserção das vias foram o NCI e o Reactome, dando preferência sempre a primeira pois essa apresentava dados mais concisos a partir da curagem feita pela equipe do projeto. O número de genes também foi modificado, de 896 passou para 1105 genes.

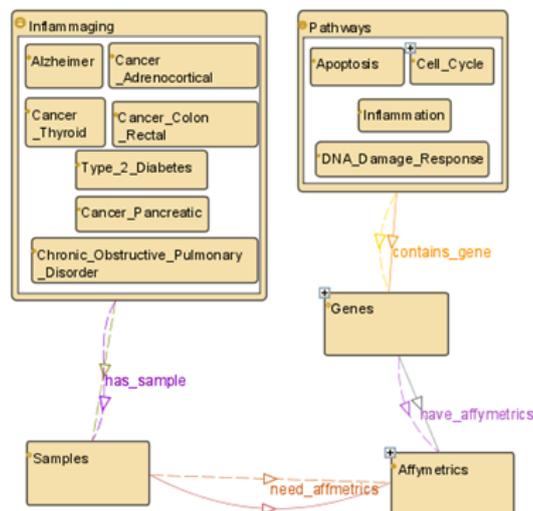


Figura 1. Ontocancro 3.0.

Como apresentado na Figura 1, a classe abstrata *Inflammaging* subdivide-se em uma subclasse para cada doença para que possa ser representada as vias metabólicas e os genes com maior expressividade de cada doença. As informações das amostras (*samples*) das doenças estão inseridas na classe *Samples*. O relacionamento entre *Samples* e *Inflammaging* restringe que cada amostra deve pertencer a uma única doença. A classe *Affymetrics* relaciona-se com *Samples* através do atributo *entrezgene*, e permite saber quais genes fazem parte do grupo de genes presentes nas amostras, levando em consideração que as amostras são organizadas a partir do identificador *affymetrics*. A classe *Genes* contém os dados relativos aos genes pertencentes às vias metabólicas (dados extraídos da base de dados HUGO) e também relaciona-se com a classe *Affymetrics*. A classe *Pathways* guarda os genes em grupos ou famílias metabólicas e assim, relaciona-se com a classe *Genes*.

#### 4. Atualização do banco de dados relacional segundo a ontologia construída

A partir da ontologia criada anteriormente, atualizações foram realizadas no banco de dados relacional do projeto. A ontologia auxiliou na representação do conhecimento dos dados para o desenvolvimento do modelo ER do banco de dados descrevendo, principalmente, conjuntos, objetos básicos, suas propriedades e como estes se relacionam entre si. A Figura 2 mostra o modelo relacional após a atualização.

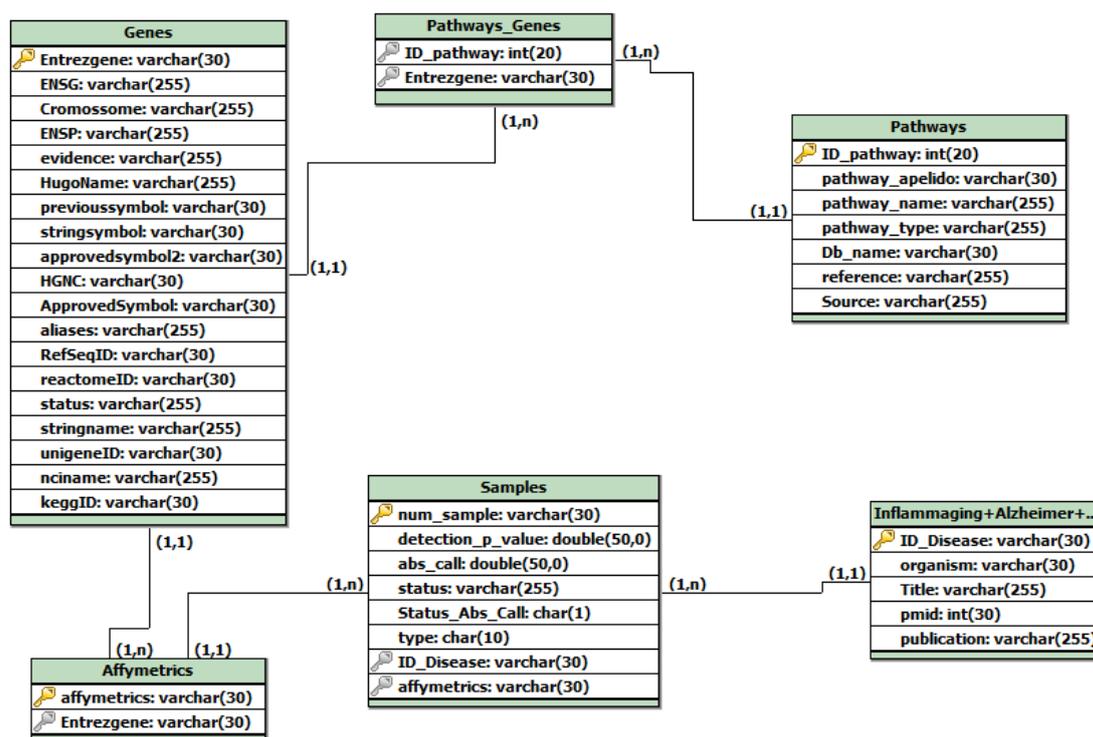


Figura 2. Modelo relacional do banco após atualização das tabelas presentes

##### 4.1. Atualização das Tabelas

As novas classes inseridas na ontologia foram a classe abstrata *Inflammaging* e a classe *Samples*, assim estas deram origem a tabelas de mesmo nome no modelo relacional do banco de dados. A tabela *Inflammaging* contém todas as doenças, levando em conta que os atributos e instâncias são os mesmos para todas. A Figura 3 apresenta a tabela

*Inflammaging*. As informações contidas nesta tabela são referentes ao estudo ao qual pertence cada amostra (*sample*).

**Tabela 1. Tabela *Inflammaging*.**

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

A tabela *Samples* guarda as informações sobre as amostras das doenças e possui ligação com a tabela *Inflammaging*, permitindo a ligação com todas as doenças e restringindo a ligação de cada amostra com apenas uma doença. Na atualização foram adicionadas duas tuplas de ligação à tabela: *id\_disease*, identifica a doença a qual pertence a amostra; *affymetrics*, responsável por unir as amostras com os genes. A Figura 4 apresenta a tabela *Samples*.

**Tabela 2. Tabela *Samples*.**

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

#### 4.2. Atualização das instâncias

A Ontocancro 2.0 continha informações exclusivamente do câncer. As séries encontradas nessa versão eram referentes a quatro tipos de carcinoma: Câncer Adrenocortical (GSE10927), Câncer de Pâncreas (GSE19650), Câncer de Tireoide (GSE27155), Câncer de Cólon Retal (GSE4183), e Transtorno Pulmonar Obstrutivo Crônico – DPOC (GSE8581); além de dados de expressão no tecido pulmonar de pacientes com DPOC moderada, fumantes saudáveis e não fumantes (GSE37768). As amostras continham tecidos normais, de carcinoma e de adenoma. Com a inserção do estudo do *Inflammaging* duas novas doenças foram adicionadas.

Além das outras séries, a Ontocancro 3.0 conta com amostras da Diabetes tipo 2 (GSE26168) e do Alzheimer (GSE29652). Estas doenças são encontradas no site do GEO que disponibiliza informações das patologias e amostras reais de tecidos em vários estágios clínicos (saudável, estado inicial ou estado avançado). Para a atualização foi preciso verificar se as informações necessárias estavam contidas nas amostras disponibilizadas pelo GEO. Essas informações são apresentadas na Tabela 1.

**Tabela 3. Informações necessárias das amostras.**

<i>id_ref</i>	Sonda de estudo
<i>value</i>	Valor da expressão sobre a sonda
<i>status_abs_call</i>	Transcrito presente, ausente ou marginal
<i>p_value</i>	Nível de significância da amostra

No caso dessas informações não estarem presentes nas amostras retiradas do Geo, faz-se necessário a normalização das mesmas. Para isso optou-se pelo uso do programa R, desenvolvido para cálculos estatísticos e gráficos, e de *Scripts* desenvolvidos pelo próprio grupo. O Alzheimer possuía todas as informações necessárias a partir do GEO, ao contrário das amostras da Diabetes tipo 2, que necessitaram da normalização a partir dos *scripts* e do programa R. Para esta normalização foi necessário realizar o download dos arquivos .cel do GEO e realizar diversas inserções no programa a partir dos *scripts* criados até encontrar os dados necessários. Após essa normalização ainda é preciso inserir mais quatro informações, Tabela 2, encontradas no site do GEO, necessárias para completar a tabela *Samples*.

**Tabela 4. Informações complementares tabela *Samples*.**

<i>id_disease</i> ( <i>num_serie</i> )	Doença a qual pertence a amostra
<i>num_sample</i>	Indica de qual amostra da doença são os resultados
<i>status</i>	Estado da amostra (tecido normal ou estágio avançado da doença)
<i>type</i>	Situação da amostra para o cálculo da diversidade relativa

## 5. Comparação entre Ontologias

Realizou-se uma pesquisa sobre ontologias e doenças relacionadas ao *Inflammaging*, analisando as formas de disponibilização e operação das mesmas com o estudo de genes e *pathways*, além da organização ontológica desses estudos. Verificando assim, a disponibilidade de ontologias voltadas para o estudo apresentado neste artigo. 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*. A Tabela 5 denota as propriedades das ontologias encontradas, podendo assim ter uma visão geral do que as mesmas representam, qual seu intuito e o que realmente desejam alcançar.

As seguintes ontologias, além da construída nesse artigo, são analisadas: *Advancing Clinico-genomic Trials (ACGT Master Ontology)*, *Systems Biology Ontology (SBO)*, *TEDDY*, *Agreement Maker* e *Neuropsychological Test Ontology (NTO)*. Entre as ontologias, somente a Ontocancro 3.0, *ACGT Master* e NTO tratam especificamente de dados relacionados às doenças, as demais trazem expectativas diferentes quanto a criação e a atualização de uma ontologia já existente referente ao estudo biológico.

**Tabela 5. Propriedades das Ontologias Pesquisadas.**

Ontologia	Tipo de Estudo	Sintaxe	Download	Inferência/ <i>Reasoning</i>	Classes	Propriedades de Objetos	Propriedades de Dados
-----------	----------------	---------	----------	---------------------------------	---------	----------------------------	--------------------------

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

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. Tem como objetivo uma única ontologia relacionada a câncer na qual se torne possível a integração semântica, distribuição e armazenamento de dados clínicos e genômicos.

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. As ontologias SBO e TEDDY são relacionadas à área de modelagem computacional. Visam o aperfeiçoamento de outras ontologias, possibilitando extrair informações mais significativas e facilitando o mapeamento dos termos utilizados.

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 seis tipos de câncer, 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 do estado de inflamação crônica.

## 6. Conclusão

A Ontocancro é um projeto que integra ramos da saúde e informática visando o desenvolvimento de estudos na área da genética, observando alterações nas vias metabólicas causadas pelo câncer. A descoberta do *Inflammaging*, inflamação crônica com origem no envelhecimento das células e desequilíbrio do sistema imunológico, trouxe consigo a possibilidade de atualização da ontologia para unificar os dados das versões anteriores da Ontocancro e as especificações do termo *Inflammaging*. As pesquisas mostraram que doenças como o Câncer, foco principal do projeto, o Alzheimer e a Diabetes tipo 2 podem ser consequentes do processo inflamatório.

A Ontocancro 3.0 possibilitou a disponibilização de uma maior quantidade de dados, importantes para a compreensão do processo celular do Câncer, Alzheimer e Diabetes tipo 2. Através da comparação entre as 37 vias referentes ao câncer, já presentes no banco antes da atualização, e as 18 vias referentes ao processo inflamatório, inseridas no banco durante a atualização. Além disso, comparou-se a ontologia desenvolvida e outras ontologias inseridas no mesmo contexto, mostrando que a Ontocancro 3.0 é a única que possui o *Inflammaging* como parte do estudo e que possui várias doenças mapeadas, mostrando informações genéticas de uma série de pacientes a partir do estudo do estado de inflamação crônica.

## Referências Bibliográficas

- Agondi, R., Rizzo, L., Kalil, J. and Barros, M. (2012) “Imunossenescência-Immunosenescence”. *Revista Brasileira de Alergia e Immunopatologia*, v. 35, n. 5, p. 169-176. São Paulo. Disponível em <<http://www.sbai.org.br/revistas/vol355.asp>>, Acesso: Maio de 2016.
- Barbie, D. and Frank, D. (2009) “Farmacologia do Câncer: Síntese, Estabilidade e Manutenção do Genoma”. *Princípios de farmacologia: a base fisiopatológica da farmacoterapia*, 2ª edição, artigo n. 38, p. 631-654. Rio de Janeiro.
- Barreto, R., Pereira, G., Costa, L. and Cavalcanti, H. (2010) “O Duplo Papel da Inflamação no Surgimento das Lesões Cancerígenas”. *Revista Brasileira de Ciências da Saúde*, v. 14, n. 4, p. 107-114. João Pessoa, Paraíba.
- Cabral, H.; Librelotto, G., Simão, E., Sinigaglia, M., Castro, M. and Mombach, J. (2011) “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, p. 82-91. Passo Fundo, Rio Grande do Sul. Disponível em <<http://www.upf.br/seer/index.php/rbca/article/viewFile/1323/1074>>, Acesso: Maio de 2016.
- Cavalcanti, J. and Guzen, F. (2012) “Influência das Proteínas Beta-amiloide e TAU na Doença de Alzheimer”. *Facene/Famene*, v. 10, n. 1, p. 57-60.
- Castro, M., Filho, J., Dalmolin, R., Sinigaglia, M., Moreira, J., Mombach, J. and Almeida, R. (2009) “ViaComplex: software for landscape analysis of gene expression networks in genomic context”. *Oxford Journals, Bioinformatics*, v. 35, n. 11, p. 1468-1469. Oxford, Inglaterra. Disponível em

- <<http://bioinformatics.oxfordjournals.org/content/25/11/1468.long>>, Acesso: Maio de 2016.
- Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M., Invidia, L., Celani, L., Scurt, M., Cevenini, E., Castellani, G. and Salvioli, S. (2007) “Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans”. *Mechanisms of Ageing and Development*, v.128, n. 1, p. 92-105.
- Giunta, B., Fernandez, F., Nikolic, W., Obregon, D; Rrapo, E., Town, T. and Tan, J. (2008) “Inflammaging as a prodrome to Alzheimer’s disease”. *Journal Neuroinflammation*. Disponível em <<http://jneuroinflammation.biomedcentral.com/articles/10.1186/1742-2094-5-51>> Acesso: Maio de 2016.
- Halazonetis, T., Gorgoulis, V. and Bartek J. (2008) “An oncogene-induced DNA damage model for cancer development”. *Journal Science*, v. 319, n. 5868, p. 1352–1355. New York, N.Y.
- Hunt, K., Walsh, B., Voegell, D. and Roberts, H. (2010) “Inflammation in aging. Part 2: implications for the health of older people and recommendations for nursing practice”. *Biological Research For Nursing*, v. 11, n. 3, p. 253-260.
- Lucena, J. (2007) “Diabetes Mellitus Tipo 1 e Tipo 2”. Trabalho de Conclusão de Curso, Farmácia, Centro Universitário das Faculdades Metropolitanas Unidas (FMU). São Paulo, Brazil.
- Lóscio, B. (2012) “Dados, Integração de Dados e Dados Interligados”. II Workshop de Introdução a Engenharia de Ontologias e Web Semântica. Brasília, Distrito Federal. Disponível em <[http://dados.gov.br/wp/wp-content/uploads/2012/10/workshop\\_BSB\\_02.pdf](http://dados.gov.br/wp/wp-content/uploads/2012/10/workshop_BSB_02.pdf)>, Acesso: Maio de 2016.
- Nascimento, K., Bastiani, E. and Librelotto, G. (2012) “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*, v. 16, n. 30, p. 177-192. Bagé, Rio Grande do Sul.
- Noy, N. and McGuinness, D. (2001) “Ontology Development 101: A Guide to Creating Your First Ontology”. *International Semantic Web Working Symposium*. Stanford University, Stanford, Califórnia.
- Pereira, J. (2015) “Protégé, Ferramenta de Edição de Ontologias”. Apresentação baseada na apresentação de Drummond, N., Horridge, M. and Knublauch, H., 8th International Protégé Conference. Manaus, Amazônia.
- Pereira, R., Henriques, P. and Librelotto, G. (2013) “Desenvolvimento de uma ferramenta para a análise de vias de estabilidade genômica”. Tese de Mestrado, Departamento de Bioinformática, Universidade do Minho. Braga, Portugal.
- Tonet, A. and Nóbrega, O. (2007) “Associação do polimorfismo - 174 G/C da região promotora do gene IL-6 com fatores de risco cardiovascular em idosas brasileiras”. Tese de Mestrado, Strictu Sensu em Gerontologia, Universidade Católica de Brasília. Disponível em <[http://www.bdttd.ucb.br/tede/tde\\_busca/arquivo.php?codArquivo=668](http://www.bdttd.ucb.br/tede/tde_busca/arquivo.php?codArquivo=668)>, Acesso: Maio de 2016.