

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE  
DEPARTAMENTO DE BIOQUÍMICA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS  
BIOLÓGICAS: BIOQUÍMICA**

***Análises da Entropia do Fluxo de Informação em  
Redes de Interação Proteica Associadas as Ferramentas de  
Enriquecimento Funcional Revelam Genes de Interesse  
Prognóstico em Glioblastomas***

Autor: Luís Henrique Trentin de Souza

Orientador: Dr. Alfeu Zanotto Filho

**Dissertação apresentada ao  
Programa de Pós Graduação em  
Ciências Biológicas: Bioquímica  
como requisito para obtenção do  
grau de Mestre em Bioquímica**

**Porto Alegre, Abril de 2015**

## CIP - Catalogação na Publicação

TRENTIN DE SOUZA, LUIS HENRIQUE

Análises da Entropia do Fluxo de Informação em  
Redes de Interação Proteica Associadas as Ferramentas  
de Enriquecimento Funcional Revelam Genes de  
Interesse Prognóstico em Glioblastomas / LUIS  
HENRIQUE TRENTIN DE SOUZA. -- 2015.

74 f.

Orientador: ALFEU ZANOTTO FILHO.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Instituto de Ciências Básicas da  
Saúde, Programa de Pós-Graduação em Ciências  
Biológicas: Bioquímica, Porto Alegre, BR-RS, 2015.

1. BIOLOGIA TUMORAL. 2. BIOLOGIA DE SISTEMAS. 3.  
TEORIA DA INFORMAÇÃO. 4. GLIOBLASTOMA MULTIFORME. I.  
ZANOTTO FILHO, ALFEU, orient. II. Título.

A meus pais, Daniel e Gema,  
“construir para além da permanência”

“Se seques por uma estrada que sabes onde vai dar e decides mudar, lembra-te, quando tomares um caminho desconhecido, que aquela estrada te servira como referência, e que muito do que te dará a firmeza que noutra caminho chegarás onde queres, aprendera com os que já traçou...”

*Próprio punho*

## **Agradecimentos**

Em primeiro lugar, gostaria de agradecer aos meus pais, Daniel e Gema, pelo apoio incondicional e sacrifícios feitos para que eu chegasse até aqui. Agradeço também aos professores da Universidade de Caxias do Sul e da Universidade Federal do Rio Grande do Sul por terem feito parte de minha formação, e em especial ao Professor Dr. José Cláudio “Zé” por ter me convidado a fazer parte de seu grande e diverso grupo, onde também agradeço ao Rodrigo “Xaca” e ao Alfeu pela paciência e os “empurrões” nas horas certas, que contribuíram muito para minha formação, ao Ricardo, “Iva” e Mariel pelas conversas produtivas e improdutivas que tínhamos no tempo em que estávamos no laboratório. Fernandinha e Matheus “Cabeça” pelos puxões de orelha nem sempre necessários, Ao Rafael “Uruguaiana” pela prestatividade, Maurílio e Lucas “12” pela parceria nos momentos de descontração, e aos demais membros do grupo onde os risos eram uma obrigação diária, em complemento a seriedade do trabalho, tornando-se mais do que colegas, amigos.

Grato aos professores Diego, Fábio e Rita por aceitarem fazer parte da banca examinadora deste trabalho, trazendo contribuições bastante significativas para o mesmo. Tenho também muito a agradecer ao pessoal da secretaria do PPG-Bioquímica – Giordano, Cléia, Patrícia e Douglas, pela prestatividade e por estarem sempre dispostos a resolver as burocracias inerentes ao processo do Mestrado.

Agradeço também aos colegas da biologia e amigos que estiveram ao meu lado durante minha formação, em especial ao Emerson e “Filipão”, camaradas para todos os momentos, tanto de demoradas divagações a respeito das minúcias da vida as quais de uma forma ou outra, são abordadas neste trabalho, bem como as descontrações e meditações nas escaladas e campos da vida.

## Índice:

<b>Parte 1</b>	
Lista de Siglas e Abreviaturas	2
Resumo	3
Abstract	4
<b>Capítulo 1: Introdução</b>	
1.1 Gliomas	5
1.2 Biologia Tumoral	6
1.3 Biologia de Sistemas	8
1.4 Redes Biológicas	10
1.5 Entropia em Redes Biológicas	12
<b>Capítulo 2: Objetivos</b>	
Objetivo Geral	14
Objetivos Específicos	15
<b>Parte II</b>	
<b>Capítulo 1: Metodologia</b>	
1.1 Obtenção dos dados:	17
1.1.1 Rede de interação proteica	17
1.1.2 Dados de expressão gênica	18
1.1.3 Filtragem dos dados de expressão	19
1.2 Análise de Entropia	19
1.2.1 Rede ponderada	19
1.2.2 Matriz de Probabilidades	20
1.2.3 Entropia de Rede	21
1.3 Análise Estatística	22
1.4 Análise Funcional	22
1.5 Análise de Expressão diferencial	23
1.6 Análise de Sobrevivência	23

Capítulo 2: Resultados	
2.1 Glioblastomas apresentam aumento na entropia de rede:	24
2.2 O aumento de entropia não é homogêneo ao longo dos genes da rede de Glioblastoma.	26
2.3 O aumento da entropia local de rede aponta rotas biológicas relevantes para a biologia de Glioblastomas.	30
2.4 As rotas identificadas na análise funcional apresentam genes diferencialmente expressos	35
2.4.1 Processos do Sistema Imunitário	36
2.4.2 Destino celular	37
2.4.3 Adesão Celular	39
2.4.4 Processamento de DNA	40
2.4.5 Transdução de Sinal	41
2.4.6 Sinalização Endócrina	43
2.4.7 Processos do Sistema Nervoso	44
2.4.8 Patologias Humanas	45
2.5 Genes diferencialmente expressos nas rotas de alta entropia podem prever o prognóstico de pacientes com GBM	47
Parte III	
Capítulo 1: Discussão	51
Capítulo 2: Conclusão	58
Referências Bibliográficas	61
Colaborações em outros trabalhos:	66
Material Suplementar – Tomo Anexo	

## Parte I

## Lista de Siglas e Abreviaturas

C	Índice de correlação de Pearson
GBM	Glioblastoma Multiforme
HPRD	<i>Human Protein Reference Database</i> - Base de Referência de Proteínas Humanas
k	Conectividade (ou número de vizinhos) de um nó da rede
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i> - Enciclopédia de Kyoto de Genes e Genomas
mRNA	RNA Mensageiro
N	“ <i>Neighbor</i> ” – ou vizinho direto de rede
P	Probabilidade (ou matriz de probabilidades)
S	Entropia Local de Rede
TCGA	The Cancer Genome Atlas - Atlas de Genoma do Câncer
w	Wheight - Peso (em uma associação proteína-proteína)
PCA	<i>Principal Component Analysis</i> – Análise de componente principal

## Resumo

Gliomas são os tumores cefálicos mais comuns na vida adulta, geralmente estão associados a um mau prognóstico. A forma mais agressiva, o Glioblastoma Multiforme (GBM) leva metade de seus pacientes a morte em 12 a 14 meses, e os tratamentos pouco ou nada melhoram esta expectativa. O desenvolvimento de terapias é bastante complexo, uma vez que se trata de um tumor com alta diversidade de células, devido a sua instabilidade genômica. Esta instabilidade interfere no equilíbrio “estequiométrico” da transcrição gênica, implicando alterações na proporção entre proteínas atuantes em um mesmo complexo. Analisando proteínas que participam de um mesmo processo sob uma perspectiva de redes (de interação proteica), pode-se analisar a probabilidade de fluxo de informação entre proteínas conectadas, ponderando pela correlação de expressão gênica, assumindo que altos valores de correlação (de expressão – entre duas proteínas) correspondem a uma maior probabilidade de fluxo de informação entre elas. Utilizando-se de uma medida de entropia chamada entropia local de rede, é possível analisar a organização do fluxo de informação entre uma determinada proteína e as demais com quem esta se relaciona. O aumento nestes valores de entropia é proposto enquanto uma assinatura sistêmica de tumores, tendo sido primeiramente demonstrada em tumores gástricos, de pulmão, bexiga, pâncreas, fígado e do colo uterino. Compreender como a informação flui ao longo de redes de interação proteica em glioblastomas pode trazer novas perspectivas de tratamento deste tumor. Portanto, este trabalho visa avaliar: I se há aumento da entropia local de rede em glioblastomas quando comparados ao tecido sadio; II: quais os genes responderiam pelo maior aumento de entropia e qual o papel biológico destes genes (em que rotas ou processos biológicos estariam atuando), III: se estas rotas ou processos biológicos apresentam alterações de expressão gênica, e IV: se algum dos genes alterados em nível de transcrição possui associação com prognóstico. Foi verificado: I: que Glioblastomas apresentam um aumento significativo da entropia local de rede em comparação a condição fisiológica normal. II: que os genes de maior ganho de entropia atuam em (28) processos biológicos relacionados com a biologia tumoral; III que estes processos biológicos têm alterações transcricionais (quando comparados com a situação fisiológica normal) em alguns de seus genes; IV que dentre todos os genes diferencialmente expressos nos (28) processos biológicos indicados pelos genes de entropia alterada, os genes *PAK6* e *MYC* tiveram correlação com sobrevida em ambas as coortes analisadas. No entanto ao estratificar por idade, verificamos que esta relação se mantinha apenas em pacientes com idade inferior a 50 anos, onde verificou-se que *PLCB1*, *MAPK8*, *CDK6* e *MYD88* também tinha correlação com sobrevida, enquanto a expressão de duas formas da *cinase dependente de cálcio/calmodulina* (*CAMK2A* e *CAMK2B*) associou-se a um prognóstico ruim. Um estudo ainda mais aprofundado destes genes, bem como de outras enzimas chave nos processos identificados são promissores no sentido do desenvolvimento de estratégias de combate a glioblastomas.

## Abstract

Gliomas are the most common form of brain tumor in adults, being generally associated with dismal prognosis. Glioblastoma Multiform (GBM) is the most aggressive glioma subtype and – despite the recent advances in therapy regimens and patient care – median survival remains between 12 and 14 months after diagnosis, and the treatment is basically palliative. The development of GBM therapy is a difficult task, given that GBM possesses high levels of genomic instability, which consequently promotes the formation of a variety of highly proliferative, invasive and chemoresistant cell phenotypes. Such a genomic instability in tumors potentially impact the “stoichiometric” balance between functionally related proteins, mostly due to changes in transcriptional activity with these entities. Analyzing protein relations under a *protein interaction network* perspective allow studying the informational flux probability between linked proteins. It means that linked proteins with positive correlation are more prone to information flux than negatively correlated links. The local network entropy is an index which measures the “disorder” of the informational flux between a protein and its neighbors. Increased local network entropy was recently shown as a systemic hallmark of diverse tumors. However, to best of our knowledge, there are no studies investigating the significance of local network entropy changes in brain tumors thus far. We believe that understanding the network informational flux, and the biological processes affect in this context, can bring new insights on the pathobiology and drugable pathways in GBM. In this study, we aimed to investigate; i) whether the local network entropy of GBM differs from normal brain tissues; ii) which genes displayed increases in entropy and what biological pathways or processes they are involved in; iii) if the identified biological processes/pathways carry differentially expressed genes and; iv) whether some of the differentially expressed belonging to highly entropic pathways are correlated with GBM patients survival. In view of such a aims, our results showed that: i) GBMs showed a significant increase in local network entropy values when compared to non-tumor brain tissues; ii) genes with high entropy played a role in 28 biological processes potentially related to GBM physiopathology; iii) Several genes with the identified pathways were found overexpressed or down-regulated in tumor versus normal brain tissues and; iv) amidst them, the expressions of *PAK6*, *PLCB1*, *MAPK8*, *CDK6* e *MYD88* predicted better prognosis, while overexpression of two Calcium/Calmodulin Kinase isoforms (*CAMK2A* e *CAMK2B*) were correlated to poor prognosis; an effect only observed in patients younger (<50 years-old) at the age of diagnosis. In summary, this study shows that local network entropy in combination with pathway enrichment analysis are a useful strategy to improve our knowledge on the biological alterations as well as genes relevant to prognosis in GBM under a systems biology perspective.

## Capítulo 1 - Introdução

### 1.1 Gliomas

Gliomas são os tumores cefálicos mais comuns na vida adulta. Devido à localização intracraniana e a baixa eficiência dos tratamentos, estão associados a um prognóstico desfavorável. O Glioblastoma multiforme (GBM) é o subtipo mais agressivo dentre os gliomas. A terapia convencional aplicada a pacientes de GBM consiste em: excisão cirúrgica, radioterapia e quimioterapia com temozolomida, no entanto estes tratamentos não possuem eficácia curativa (Stewart, 2002). Cerca de 50% dos pacientes de GBM morrem até 12 a 14 meses após o diagnóstico (Ohgaki & Kleihues, 2007; Wen & Kesari, 2008).

A letalidade de um tumor está diretamente relacionada com o comprometimento de uma função vital relacionada ao tecido “hospedeiro” do mesmo. GBM é classificado pela Organização Mundial da Saúde como um tumor de grau IV, correspondendo ao grau mais avançado de tumor glial. O prognóstico ruim do GBM é consequência do comprometimento de funções do sistema nervoso central, onde parte do tecido é substituída, ou tem seu funcionamento afetado pela massa tumoral em crescimento. A caixa craniana restringe o espaço disponível, sendo que o crescimento do tumor dar-se-á a custa do aumento da pressão intracraniana (relacionada com o sintoma de cefaleia), ou da perda de tecido nervoso.

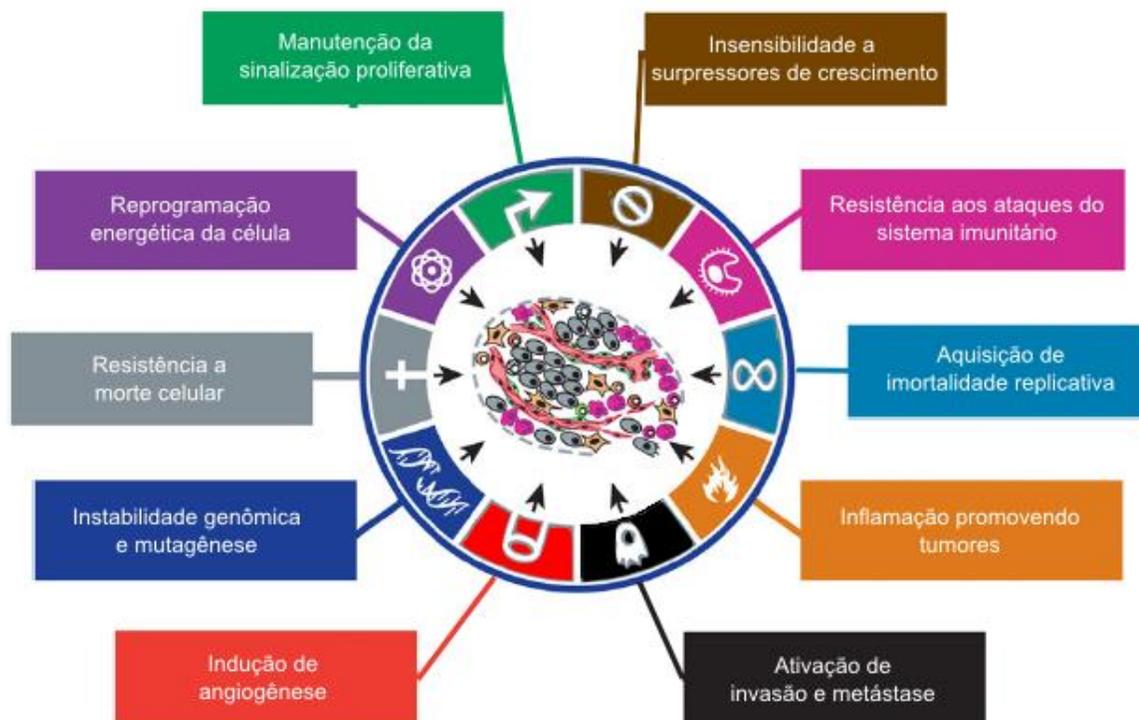
Em nível molecular, o fenótipo aberrante de uma célula tumoral de GBM é consequência de desregulações em uma série de rotas bioquímicas, especialmente

vias de transdução de sinal (Brennan et al, 2009). Em GBM, são descritas uma série de alterações como mutação, amplificação ou superexpressão de genes como *PTEN*, *EGFR*, *PDGFR-a*, *TP53*, e *mTOR* (Soni et al, 2005; Yu et al 2008; Brennan et al 2009). Estas alterações tem impacto em diversas rotas bioquímicas (metabólicas e de transdução de sinal), muitas delas diretamente relacionadas com a instalação e progressão tumoral.

## **1.2 Biologia Tumoral**

A transição do fenótipo de uma célula normal para um fenótipo tumoral é chamada de transformação celular, um processo complexo onde a célula passa a não responder mais a sinais sistêmicos de maneira adequada (Hanahan & Weinberg, 2011). A instalação de um tumor depende da sua capacidade de tolerar ou reverter os mecanismos de defesa do organismo, responsáveis pela eliminação de células aberrantes. Estes mecanismos são comumente chamados “barreiras antitumorais”, e atuam desde o nível celular, através da ativação de vias intrínsecas de morte celular programada, até o nível sistêmico – através do sistema imunitário e mecanismos de reconhecimento de histocompatibilidade (Klein, 2006). As barreiras antitumorais agem enquanto obstáculos seletivos ao desenvolvimento do tumor, em um processo que mimetiza a seleção natural, onde a diversidade de genótipos (e fenótipos) resultantes de processo de mutação provê uma população de células resistentes aos agentes seletivos (barreiras anti-tumorais) (Klein, 2006). Os atributos das células capazes de transpor as barreiras seletivas do organismo e iniciar o estabelecimento do tumor são

denominados “capacidades tumorais” e são descritos na revisão “*Hallmarks of Cancer: The Next Generation*” (Hanahan & Weinberg, 2011)



**Figura 1:** Características fenotípicas adquiridas por uma célula durante os processos de transformação celular e progressão tumoral (adaptado de Hanahan & Weinberg, 2011).

A aquisição de características fenotípicas relacionadas às capacidades tumorais de uma célula resulta de diferentes “estratégias” selecionadas ao longo da evolução do tumor. Estas estratégias podem variar conforme o tecido de origem (que influenciará no tipo celular e no “microambiente tumoral”) e até mesmo em uma mesma massa tumoral, pois resultam de um processo de balanço entre alterações genéticas e epigenéticas de natureza estocástica (mutações de perda ou ganho de função,

deleções, ampliações, silenciamento) e a pressão seletiva dos mecanismos antitumorais (Klein, 2006).

A compreensão do processo de transformação celular é complexa, dado que as alterações acontecem em diversos níveis de organização celular, desde mudanças na informação genética, processamento de sinais (vias de sinalização) ou sinais ambientais (mudanças no pH, alterações no balanço redox ou de quimiocinas). Este encadeamento de sinais dificulta que algum fator isolado (como mutação de um gene ou superexpressão de outro) seja capaz de modelar o desencadeamento do fenótipo tumoral ou sua severidade, o que caracteriza o câncer enquanto um distúrbio complexo (Hornberg et al, 2006), que requer uma interpretação de caráter integrativo de sua informação, desta forma permanecendo por décadas com poucos avanços significativos na sua compreensão e tratamento. A vigência de uma visão científica compartimentalizadora, foi muito importante na construção de conceitos fundamentais da biologia, mas não se mostrou adequada para a compreensão das alterações nos sistemas biológicos que levassem a distúrbios complexos, por carecer de uma perspectiva integrativa (Marashi & Tefagh, 2014).

### **1.3 Biologia de Sistemas**

A biologia de sistemas complexos, ou simplesmente biologia de sistemas é uma ciência que toma elementos da física (de sistemas complexos) e se utiliza de ferramentas e conceitos da teoria de sistemas no estudo de sistemas biológicos. Esta disciplina fundamenta-se em uma visão integrativa onde a compreensão de um sistema

biológico não pode ser obtida através de uma simples função do comportamento de seus componentes, pois da interação entre os componentes de qualquer sistema surgem propriedades emergentes, não observáveis quando analisando seus componentes individualmente (Mast, Ratushny & Aitchison, 2014, Marashi & Tefagh, 2014).

O desenvolvimento desta área do conhecimento foi possibilitado pelo avanço de técnicas de biologia molecular de análises em grande escala (de genômica, transcriptômica e proteômica). Este novo momento das ciências biológicas é intitulado “era das ômicas”, e em menos de duas décadas alcançou uma magnitude de informações antes só conhecida em disciplinas como a astrofísica (Marx, 2013). O advento da biologia computacional e os avanços das capacidades de processamento dos computadores foram fundamentais para expandir a capacidade analítica a tamanha magnitude de dados.

Experimentos em larga escala, via de regra, são subutilizados no que diz respeito ao potencial da informação gerada. Em geral estes experimentos produzem mais dados do que o necessário para responder a pergunta na qual foram concebidos, ou que podem ser utilizados sob um novo questionamento. Devido a isso, é comum a disponibilização destes dados em repositórios públicos, que concentram informações de acordo com sua natureza (como dados de transcriptômica, concentrados em repositórios de expressão gênica, ou proteômica, em bancos de dados de estrutura ou interação proteica). Estando disponíveis, estes dados podem ser analisados sob perspectivas diferentes das empregadas nos estudos que motivaram a geração dos mesmos, como estudos associativos ou de meta-análise.

## 1.4 Redes Biológicas

Sistemas complexos são usualmente representados por diagramas que ilustram as interações (arestas, ou elos) entre os elementos que compõe este sistema (interatores). Estes diagramas de interação, também chamados grafos ou redes de interação, são amplamente empregados no estudo de sistemas biológicos (Barabasi & Oltvai, 2004), ao exemplo das redes tróficas, no caso da ecologia ou redes de interação proteica, na biologia molecular.

O uso de redes biológicas na biologia molecular trouxe novos *insights* sobre como é organizada a informação biológica, e como esta organização pode estar relacionada com a fisiologia da célula (Barabasi & Oltvai, 2004). Redes de interação proteica são uma importante classe de redes biológicas. Estas redes são construídas através de experimentos de varredura em larga escala de interações filtradas por dados de literatura (Prasad, Kandasamy & Pandey, 2009; Stelzl & Wanker, 2009, Brown & Jurisica, 2005). Arestas em redes proteicas representam principalmente interações entre proteínas pertencentes a um mesmo complexo (proteico) ou a um mesmo sistema de modificações pós-traducionais, mediadoras da transdução de sinal.

As análises topológicas de redes de interação proteica têm identificado importantes características relacionadas à sua organização como, por exemplo, a liberdade de escala na distribuição de conectividade (ou grau) de seus nós, onde a maior parte das proteínas se conecta com poucos vizinhos (baixos valores de grau), em contraste a um pequeno número de proteínas altamente conectadas, chamadas “hubs”. Hubs são frequentemente descritos enquanto alterados no contexto tumoral (Hornberg

et al, 2006; Jonsson & Bates, 2006), bem como proteínas com altos valores de centralidade (Taylor et al, 2009, Platzer et al, 2007) sugerindo que medidas topológicas desempenham um papel importante no entendimento da organização da informação biológica necessária para o desempenho da fisiologia normal da célula.

Redes de interação proteica possuem um grande viés atribuído a um maior número de estudos abordando proteínas conhecidamente relacionadas com condições patológicas, interferindo diretamente na estimativa de propriedades topológicas destes nós (Platzer et al, 2007). Isto resulta em um desequilíbrio na disponibilidade de informação ao longo das redes, interferindo, por exemplo, na informação de conectividade entre as proteínas, possivelmente subestimando o número de conexões de proteínas ainda não relacionadas com contextos patológicos. Ainda que, de certa forma, incompletas, as redes proteicas sejam bastante usuais enquanto modelos de estrutura da informação celular, e sua integração com dados de expressão gênica possibilitaram novos *insights* em estudos genômicos de tumores, auxiliando na identificação de módulos funcionais relacionados com a progressão do câncer, bem como marcadores de prognóstico e diagnóstico (Bozdag et al, 2014). A integração de dados de expressão gênica a interatomas pode contornar problemas inerentes a análises puramente de expressão gênica - como, por exemplo, a dificuldade de interpretação de assinaturas de expressão obtidas em análises de expressão diferencial - as quais são cada vez mais entendidas como demasiadamente simplistas, e que vêm sendo gradualmente suplantadas por abordagens baseadas em redes, que procuram identificar sub-redes diretamente implicadas na progressão da doença (Bozdag et al, 2014, Gevaert & Plevritis, 2013). No contexto tumoral, os estudos inferem a presença

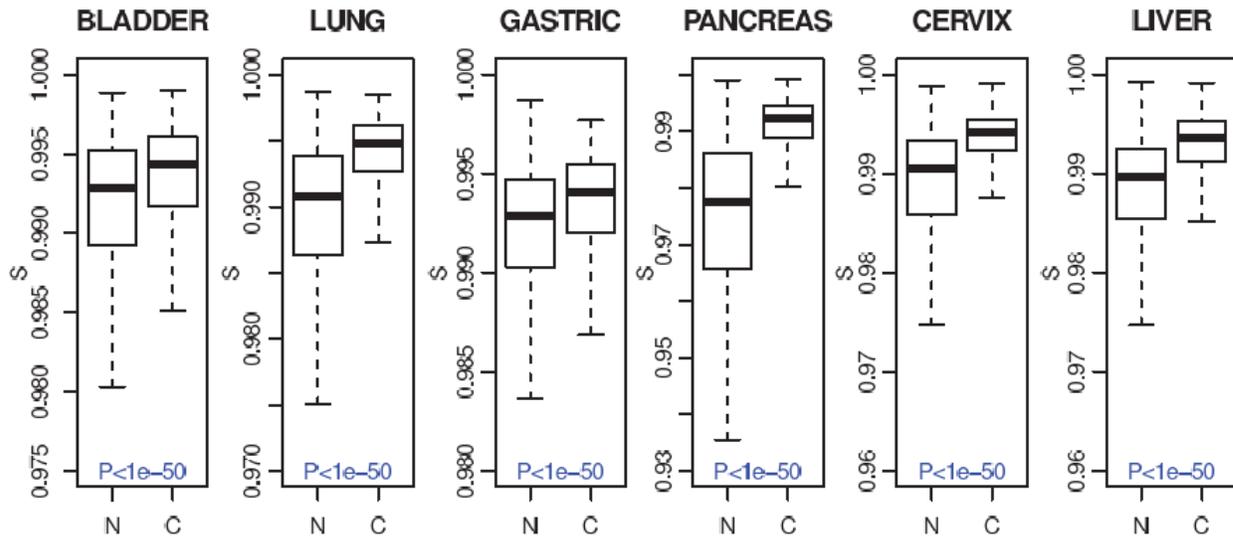
de conexão entre os vértices de uma rede de interação a partir de medidas de correlação de expressão, como uma maneira de representar a reestruturação da rede biológica no contexto tumoral *versus* tecidos saudáveis relacionados.

No entanto, os fenótipos tumorais de pior prognóstico estão geralmente associados a uma maior frequência de alterações genômicas (Carter et al, 2006). O GBM apresenta o pior prognóstico dentre os tumores gliais, apresentando, além das alterações genômicas previamente citadas, altos índices de instabilidade genômica (Milinkovic et al, 2012). Caso estas alterações genômicas tenham efeito sobre a topologia informacional da rede, a exploração de medidas topológicas capazes de analisar redes de interação proteica, integrada a dados de expressão gênica, podem revelar métricas que quantifiquem o impacto destas alterações no fluxo de informação da rede.

## **1.5 Entropia em Redes Biológicas**

A construção de redes de interação proteica com ponderações definidas por uma matriz de probabilidade de fluxo de informação, que leva em consideração a correlação dos níveis de expressão gênica entre proteínas conectadas, serviu de base para a proposição de uma medida de entropia capaz de avaliar o grau de aleatoriedade nos padrões locais de fluxo de informação no entorno de um nó. Esta medida, intitulada Entropia Local de Rede (S) permitiu a comparação de redes geradas a partir de dados expressão gênica originados de diferentes fenótipos.

Tumores mamários metastáticos apresentam valores de  $S$  significativamente maiores que tumores mamários não metastáticos (Teschendorff & Severini, 2010). Em um estudo posterior (West et al, 2010), foi demonstrado um aumento nos valores de  $S$  em tumores gástricos, de pulmão, bexiga, pâncreas, fígado e do colo uterino (figura 2 desta seção), quando comparados com o tecido sadio correspondente.



**Figura 2** – Distribuição dos valores de entropia ( $S$ ) em tumores (nesta ordem) de Bexiga, Pulmão, Estômago, Pâncreas, Colo Uterino e Fígados, onde se verifica um aumento na entropia local de rede destes tumores (C) quando em comparação com o tecido sadio correspondente (N) – representação dos valores de: média (linha), 1º (caixa) e 2º quartil (hastes) – retirada de West & cols. 2012.

Dado que todos os tipos tumorais apontados acima apresentaram valores de  $S$  significativamente mais altos que os respectivos tecidos sadios, West e cols. (West et al, 2010) propuseram que uma maior aleatoriedade no fluxo de informação de rede seria uma assinatura sistêmica dos processos neoplásicos.

## Capítulo 2 - Objetivos

### 2.1 Objetivo Geral

Tendo em vista que os tumores do sistema nervoso, principalmente os Glioblastomas Multiformes (GBM) são enfermidades de consequências bastante severas por ainda carecerem de estratégias terapêuticas eficazes - e que as medidas de entropia de rede foram capazes de avaliar o grau de desordem no fluxo de informação de rede em outros tumores (West et al, 2010) - compreender a biologia de tumores gliais (especialmente o GBM) sob esta perspectiva pode trazer informações relevantes para o entendimento da patofisiologia destes tumores, tão bem quanto, a longo prazo, o desenvolvimento de estratégias terapêuticas eficazes. Uma vez que até o presente momento não se verificou nenhum trabalho científico utilizando a abordagem de entropia de rede em tumores do sistema nervoso. **Este trabalho visa analisar os valores de entropia local de GBMs, comparando-os com tecido sadio e outros tumores de origem nervosa, verificando os quais genes da rede de Glioblastoma apresentam aumento na entropia local quando comparados com o tecido nervoso sadio; e em quais processos biológicos estes genes estariam atuando.**

## 2.2 Objetivos Específicos

- Avaliar a entropia local de rede (S) em coortes de pacientes acometidos por tumores do sistema nervoso, comparando a entropia local da rede de GBM com a de tecido nervoso sadio e com a de outros tumores do sistema nervoso.
- Classificar os genes da rede de GBM de acordo com a diferença nos valores de entropia local da rede em relação ao tecido nervoso sadio, verificando quais genes apresentam maior ganho de entropia.
- Verificar o papel biológico dos genes de maior ganho de entropia em GBM, através da análises de enriquecimento funcional, identificando rotas que possam estar relacionadas com o desenvolvimento do GBM.
- Avaliar o perfil de expressão gênica das rotas potencialmente envolvidas com o fenótipo tumoral do GBM, reveladas pela análise de enriquecimento funcional.
- Analisar, dentre os genes diferencialmente expressos em cada uma das rotas potencialmente envolvidas com o fenótipo tumoral do GBM, a correlação com prognóstico dos pacientes.

## Parte II

## Capítulo 1 – Metodologia

### 1.1 Obtenção dos dados

Para análise da entropia da rede de Glioblastoma Multiforme, foi construída uma rede de interação proteica ponderada por correlações de expressão gênica, conceito similar às redes de expressão diferencial (Bozdag et al, 2010). No entanto, diferente destas, onde a conectividade entre duas proteínas é dada pela significância de uma medida de correlação (informação mútua) entre elas, aqui, a presença de relação entre duas proteínas é dada por uma base de dados de informação de conectividade funcional (Barabasi & Oltvai, 2004). A correlação de expressão gênica apenas atribui um valor de “força de ligação” entre duas proteínas (correspondentes aos genes cuja expressão é analisada) funcionalmente conectadas, em uma rede de interação proteica.

#### 1.1.1 Rede de Interação Proteica

Para obtenção da rede de interação proteica, utilizou-se a base de dados de referência a proteínas humanas (*HPRD – Human Protein Reference Database*). Esta base de dados provê informações de interações (arestas) proteína-proteína (nós, ou vértices), manualmente extraídas em revisões bibliográficas (Prasad et al, 2009).

### 1.1.2 Expressão gênica

A informação de expressão gênica foi obtida de dois repositórios: o *Array Express* (Kolesnikov et al 2015) - mantido pelo Instituto de Bioinformática do Laboratório Europeu de Biologia Molecular<sup>1</sup> (EBI-EMBL); e o *Genome Expression Omnibus* – GEO (Edgar, Domrachev & Lash, 2002), do Centro Nacional de Informação Biotecnológica do Instituto Nacional de Saúde<sup>2</sup> dos Estados Unidos – NCBI-NIH. Buscou-se coortes que: contivessem mais de 30 amostras de Glioblastoma Multiforme, ao menos 8 amostras saudias, e cujas análises tenham sido executadas sob uma mesma plataforma de microarranjo. Três coortes satisfizeram estes critérios (*e-mtab3070* e *e-geod50161* do Array Express, e *gse16011* do GEO), possuindo o número requerido de amostras e sendo desenhadas sob mesma plataforma de quantificação de mRNA (*microarray gene expression chip affymetrix Hgu133plus2*<sup>3</sup>). A tabela 1 traz a descrição resumida destas coortes.

**Tabela 1:** Número de amostras de tecidos saudios e dos diferentes tipos tumorais contidos em cada uma das coortes analisadas. \*\*Meduloblastoma é um tumor do sistema nervoso de origem não glial.

	<b>mtab3073</b>	<b>geod50161</b>	<b>gse16011</b>
<b>Tecido Nervoso</b>	28	13	8
<b>Glioblastoma Multiforme</b>	228	34	159
<b>Astrocitoma</b>	148	0	29
<b>Astrocitoma Pilocítico</b>	0	15	0
<b>Oligodendroglioma</b>	67	0	52
<b>Oligoastrocitoma</b>	0	0	28
<b>Meduloblastoma**</b>	0	22	0

<sup>1</sup> Tradução livre para *European Bioinformatics Institute of the European Molecular Biology Laboratory*;

<sup>2</sup> Tradução livre para *National Center of Biotechnology Information of National Institute of Health*.

<sup>3</sup> *Hgu133plus2* é uma plataforma comercial, desenvolvida pela affymetrix®, que faz a análise seriada de expressão gênica utilizando chips de sondas fluorescentes que hibridizam a sequencias complementares de cDNA sintetizadas a partir do conteúdo de RNAs mensageiros no tecido analisado.

### 1.1.3 Filtragem dos Dados de Expressão

Os dados dos experimentos de expressão gênica de cada coorte foram normalizados com o método RMA, e submetidos a um teste de controle de qualidade, que envolve a análise de coordenada principal (PCA) para verificar os componentes dominantes na variação relacionada com os fenótipos tumorais e não tumorais. A análise de qualidade dos experimentos foi executada no pacote MAQQ (*Micro Array Quality Control*) para ambiente estatístico R v.3.1.2. As amostras que cumpriam os critérios de qualidade foram submetidas às análises posteriores.

## 1.2. Análise de Entropia

### 1.2.1. Rede Ponderada

Utilizando a rede de interação proteica humana obtida no HPRD como base da informação de conectividade, foi ponderado um valor de força ( $w_{ij}$ ) para cada uma das arestas, referenciadas pelo par de proteínas por ela ligadas (I e J), baseado em valores de correlação de Pearson ( $C_{ij}$ ) entre os genes (i e j) correspondentes as proteínas ligadas por esta aresta. Uma vez que os valores de  $C_{ij}$  situam-se entre -1 e 1, realizou-

se uma transformação monotônica (Eq.1):  $C_{ij} \rightarrow w_{ij} \equiv \frac{1}{2}(1 + C_{ij})$  dos valores de C.

Esta transformação permitiu que os valores de  $w_{ij}$  se mantivessem entre 0 e 1, onde valores próximos de 1 representam fortes correlações positivas, 0.5 ausência de correlação e 0 uma correlação fortemente negativa entre as proteínas da aresta.

### 1.2.2 Matriz de Probabilidades:

Com base nas ponderações produzidas pelas correlações de expressão gênica, foi construída uma matriz ( $p$ ) de probabilidades de fluxo de informação:

$$\text{Eq. 2: } p_{ij} = \frac{w_{ij}}{\sum_{j \in N(i)} w_{ij}}$$

Onde  $p_{ij}$  é a probabilidade de fluxo de informação de  $i \rightarrow j$  em função do peso ( $w$ ) atribuído a aresta de  $i$  para  $j$  ( $w_{ij}$ ) ponderado pelos vizinhos ( $N$ ) de  $i$ . Portanto a soma das probabilidades das arestas que saem de  $i$  para  $j$ , onde  $j$  é vizinho de  $i$ , é igual a 1 ( $\sum_{j \in N(i)} p_{ij} = 1$ ) o que caracteriza uma matriz de probabilidades. Esta transformação confere uma maior probabilidade de fluxo de informação para arestas com altas correlações positivas. Assume-se que correlações positivas são necessárias para a transdução do sinal em complexos proteicos ou vias de sinalização, enquanto correlações negativas representam relações inibitórias ou o estado inativo do complexo, portanto, a essas interações são atribuídos os menores valores de fluxo de informação. A correlações fracas ( $C_{ij} \sim 0$ ) são atribuídos valores de probabilidade maiores que em arestas com correlações negativas ( $C_{ij} < 0$ ) pois há uma menor evidência que estes valores anulem a interação (West et al, 2012; Teschendorff & Severini, 2012). Desta forma, a matriz  $p$  descreve como se dá o fluxo de informação de expressão gênica em uma rede de interação proteica, possibilitando a construção de matrizes baseadas no perfil de expressão de mRNA em tecidos nervosos saudáveis –  $p_{ij}(N)$ , e tumorais, e.g.: Glioblastoma multiforme(GBM)  $p_{ij}(T)_{GBM}$ .

### 1.2.3 Entropia de Rede

O conceito de entropia visa avaliar o grau aleatoriedade (também chamado grau de desordem) em um sistema. Uma série de índices se propõe a estimar um valor numérico para este conceito. A medida de entropia utilizada neste trabalho foi o índice de entropia de Shannon (S) e advém da teoria da informação. Desta forma, aplicando o índice de Shannon a matriz de probabilidades  $p$  pode-se estimar o grau de desordem no fluxo de informação local de redes referentes a cada um dos tecidos nervosos analisados neste trabalho. A estimação do valor de entropia local de rede (S) é dada por:

$$S_i = - \frac{1}{\log k_i} \sum_{j \in \mathcal{N}(i)} p_{ij} \log p_{ij}$$

Eq. 3:

Onde  $S_i$  é a entropia de um dado gene  $i$  e  $k_i$  é a conectividade (ou o número de vizinhos) do gene  $i$ . Assim foram computadas as entropias locais da rede de cada um dos tecidos analisados: Tecido Sadio – S(N), Glioblastoma Multiforme – S(T)<sub>GBM</sub>

### 1.3 Análise Estatística

A distribuição de entropia de todas as proteínas nas redes dos tecidos tumorais foi comparada com as dos tecidos normais das respectivas coortes através do teste não-paramétrico “*Willcoxon Rank Sum*”, de acordo com a metodologia utilizada em trabalhos de abordagem similar (West et al, 2012; Teschendorff & Severini, 2012) sobre as diferenças nos valores de  $S$  entre tecidos saudáveis –  $S(N)$  e tumorais –  $S(T)_s$ . Os testes foram executados no ambiente estatístico *R* v.3.1.2, assim como os a geração dos gráficos correspondentes a distribuição de  $S$  das proteínas analisadas.

### 1.4 Análise Funcional

Para análise funcional, foi computada uma curva da diferença na entropia de rede ( $S\Delta$ ) do tecido sadio  $S(N)$  versus Glioblastoma  $S(T)_{GBM}$ :  $S\Delta = S(N) - S(T)_{GBM}$ . Proteínas situadas na cauda (5%) de maior ganho na entropia de rede (menor  $S\Delta$  ou  $\Delta S$ ), bem como seus interatores (vizinhos) diretos (os quais têm influência direta no valor de  $S$  dessas proteínas), foram introduzidas na análise sendo verificadas quanto ao seu papel biológico, com auxílio da ferramenta de anotação funcional - análise de Rota do Banco de Dados de Anotação, Visualização e Descoberta Integrada – DAVID (Huang et al, 2009). Apenas rotas da Enciclopédia de Kyoto de Genes e Genomas – KEGG – foram utilizadas na análise. A ferramenta de anotação funcional estima um valor (índice) de enriquecimento para um determinado conjunto de genes referente a uma rota ou processo biológico. O índice de enriquecimento é dado em função da razão da proporção de genes fornecidos ao número de genes contidos em uma

determina rota bioquímica, pela probabilidade ao acaso desta rota conter os genes pertencentes a esta rota. Esta análise indica quais processos biológicos podem estar sendo afetados pela mudança na entropia.

### **1.5 Análise de expressão diferencial**

Foram selecionadas, dentre as rotas com enriquecimento funcional significativo comum às três coortes avaliadas, aquelas relacionadas ao contexto tumoral. Os genes correspondentes às proteínas atuantes nas rotas selecionadas foram analisados quanto a sua expressão diferencial, com auxílio do pacote *limma* (Modelos Lineares para Análise de *Microarrays*) para o ambiente estatístico R v.3.1.2. Os valores de expressão gênica são representadas em Logaritmo base 2 da mudança de expressão com relação aos valores do tecido normal.

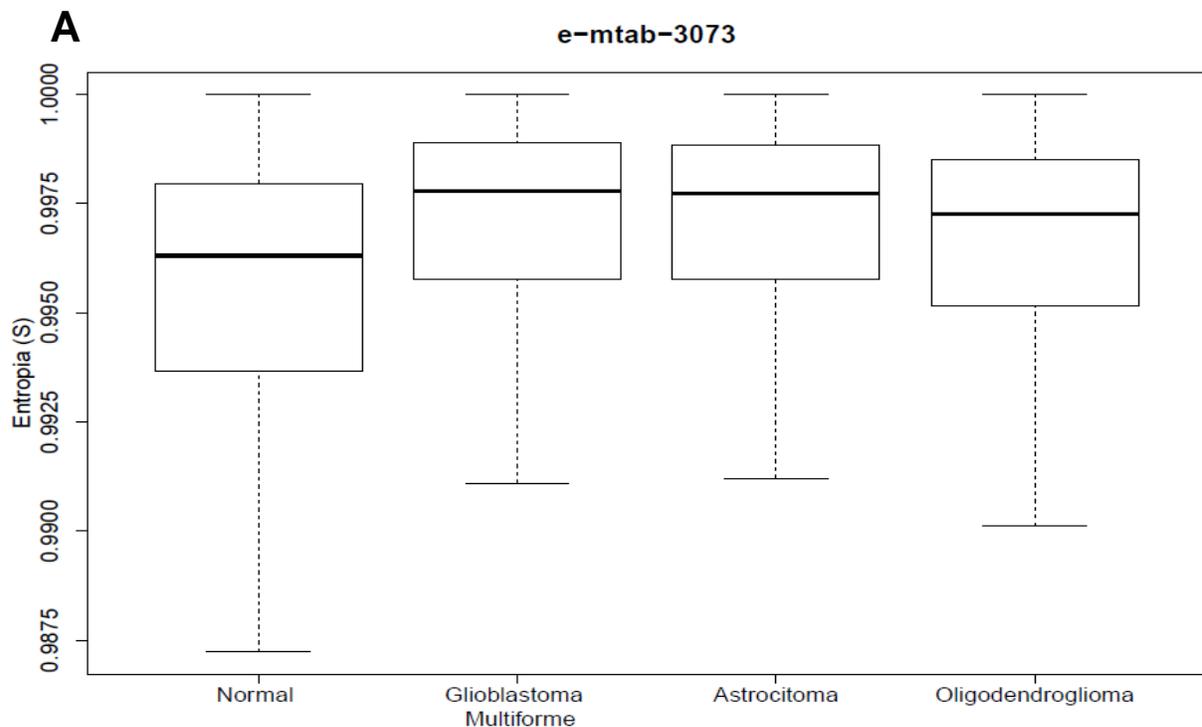
### **1.6 Análise de sobrevivência.**

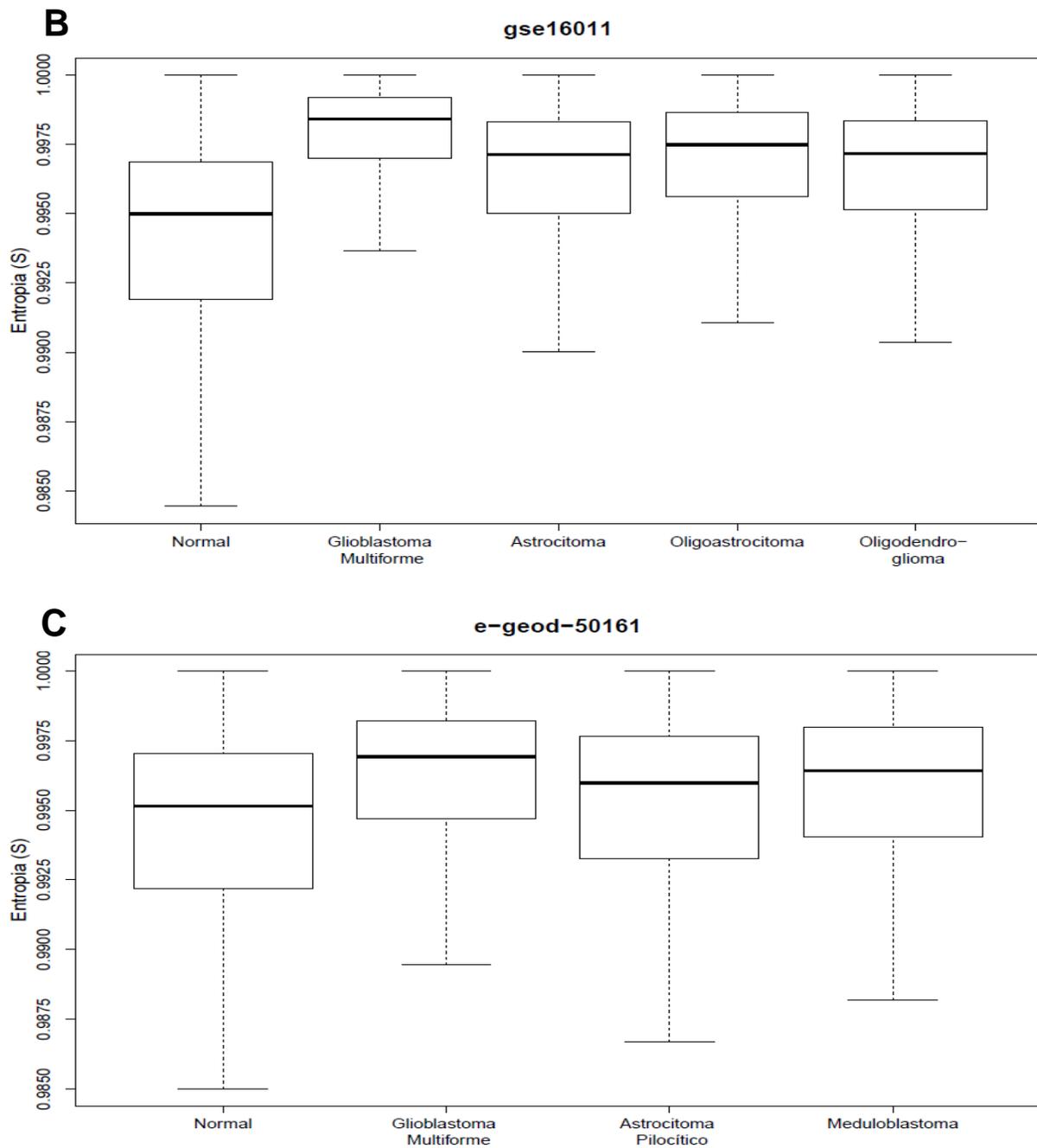
Os genes diferencialmente expressos em cada uma das rotas analisadas foram submetidos à análise de correlação com sobrevida, através da ferramenta *ProgGeneV2* (Goswami & Nakshatri, 2013), com o objetivo de determinar se algum deles possuía capacidade de predição do prognóstico dos GBM. Foram selecionadas coortes de glioblastoma dentre as coortes de tumores cerebrais disponíveis na ferramenta.

## Capítulo 2 – Resultados

### 2.1 Glioblastomas apresentam aumento na entropia de rede:

Inicialmente, observamos que os valores de entropia de rede do glioblastoma multiforme (GBM) encontram-se significativamente aumentados em comparação com o tecido nervoso normal, nas três coortes analisadas; tal efeito foi também observado em gliomas de grau menos agressivo, como os oligodendrogliomas e astrocitomas. A figura 3 mostra a distribuição dos valores de entropia de rede dos tecidos normais, GBMs e outros tumores nervosos contidos em cada uma das coortes analisadas (quadros A,B, e C).





**Figura 3:** Distribuição dos valores de entropia de rede ( $S$ ) de tecidos cerebrais saudáveis, glioblastomas multiformes e outros tipos de tumores associados ao sistema nervoso contidos nas três coortes analisadas (quadro A: *emtab3073*; quadro B: *gse16011*; quadro C: *geod50161*) – representação dos valores de: média (linha), 1º (caixa) e 2º quartil (hastes). As distribuições dos valores de  $S$  dos glioblastomas encontram-se positivamente deslocadas em relação aos valores do tecido sadio e, em menor grau, dos demais tecidos tumorais analisados.

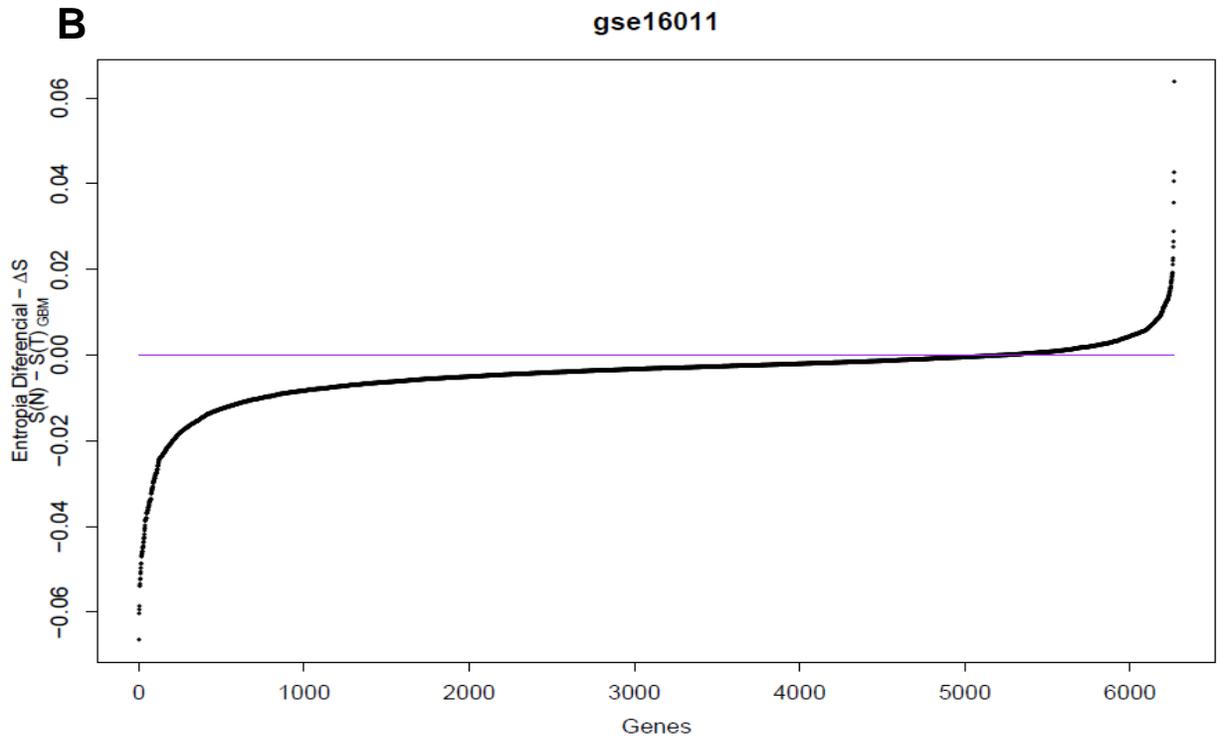
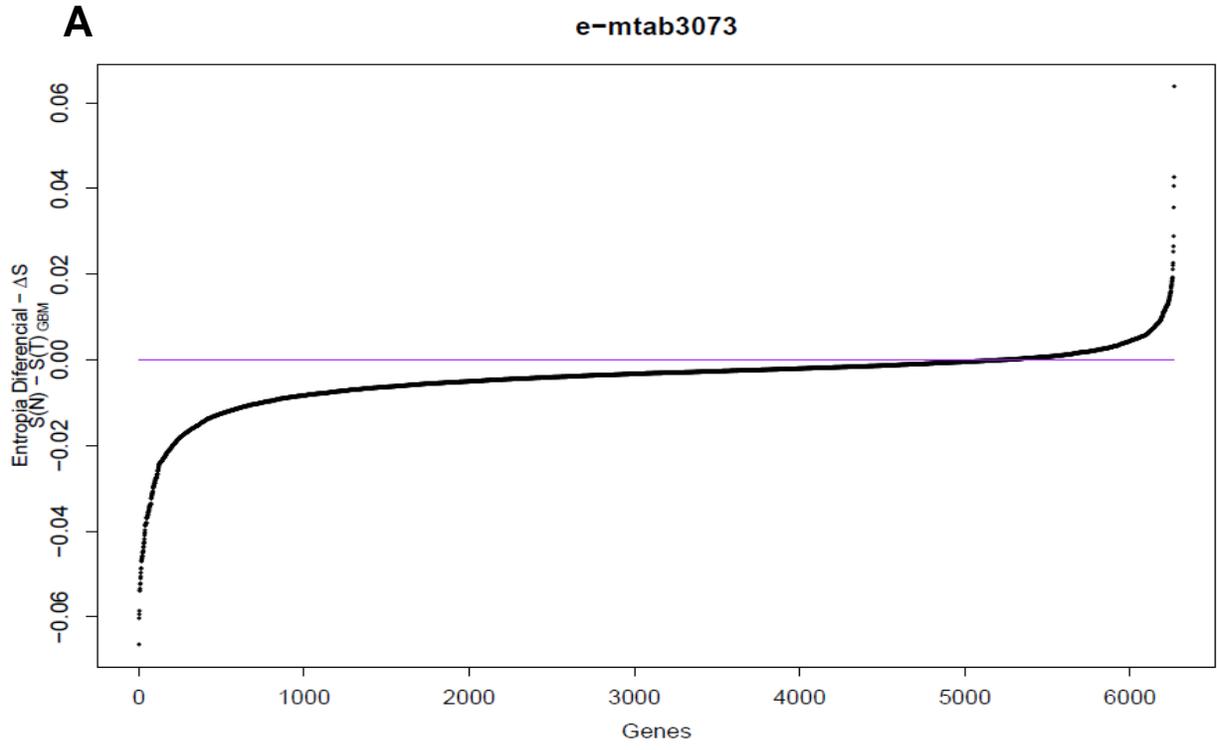
Duas das três coortes analisadas (*emtab3073* e *gse16011*) continham um número satisfatório de amostras de outros subtipos de tumores gliais – oligodendrogliomas e astrocitomas. Em ambas as coortes, o aumento da entropia de rede é significativo na comparação entre GBM e oligodendrogliomas, todavia, somente na coorte '*gse16011*' houve significância na comparação entre as entropias de rede de GBMs e astrocitomas (tabela 2).

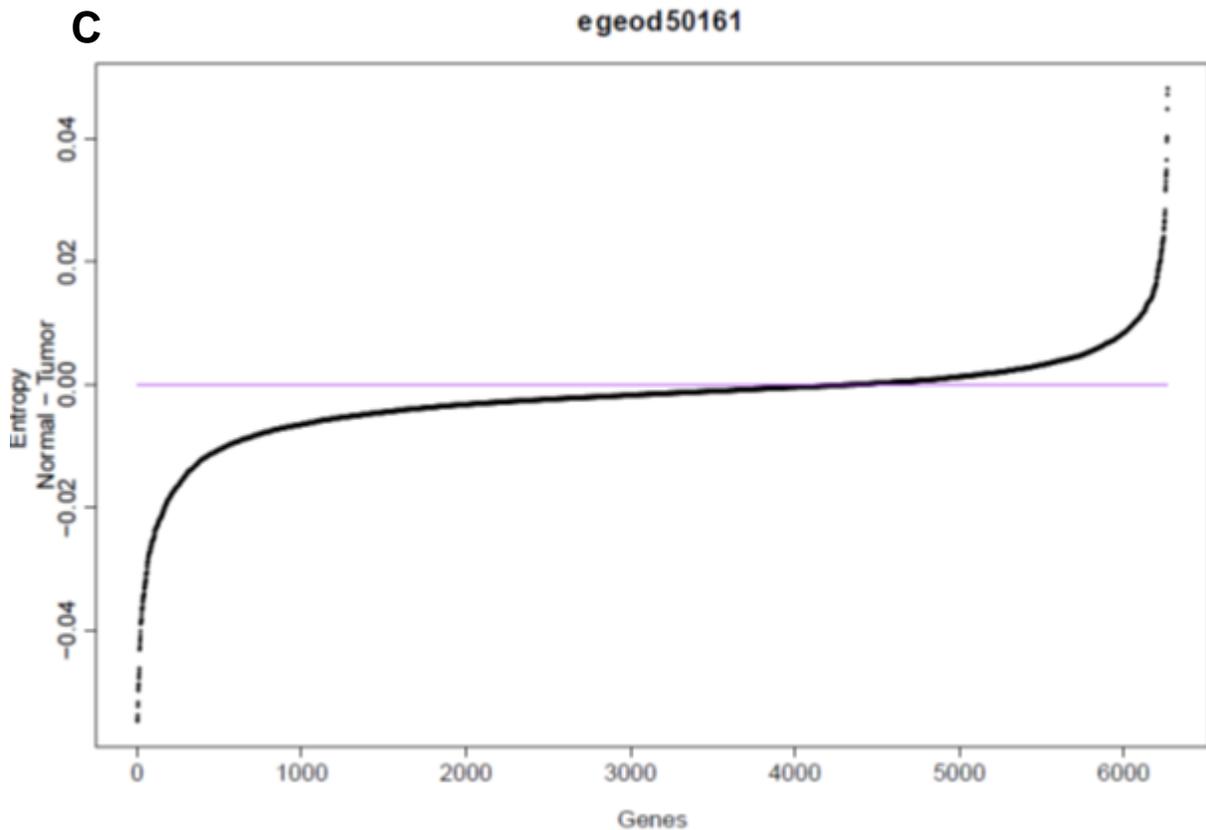
**Tabela 2:** Valores p dos testes *Willcoxon Rank Sum* entre tumores do tipo glioblastoma multiforme (*GBM*) e os demais grupos analisados, o asterisco (\*) denota comparações significativas, enquanto o N.D. denota a não disponibilidade de amostras para aquele tipo de tecido em número satisfatório para a análise.

	<b>mtab3073</b>	<b>gse16011</b>	<b>geod50161</b>
<b><i>GBM vs NORMAL</i></b>	<2.2e-16*	<2.2e-16*	<2.2e-16*
<b><i>GBM vs Astrocitoma</i></b>	0.1048	<2.2e-16*	N.D.
<b><i>GBM vs Oligodendroglioma</i></b>	<2.2e-16*	<2.2e-16*	N.D.

## 2.2 O aumento de entropia não é homogêneo ao longo dos genes da rede de Glioblastoma.

A análise da entropia diferencial entre as redes de GBM e tecido sadio (figura 4) mostra: i) genes que apresentam aumento na entropia (menores valores de  $\Delta S$ ); ii) genes que praticamente não alteram seus valores de entropia ( $\Delta S \cong 0$ ) e; iii) genes que apresentaram uma diminuição na entropia (maiores valores de  $\Delta S$ ).

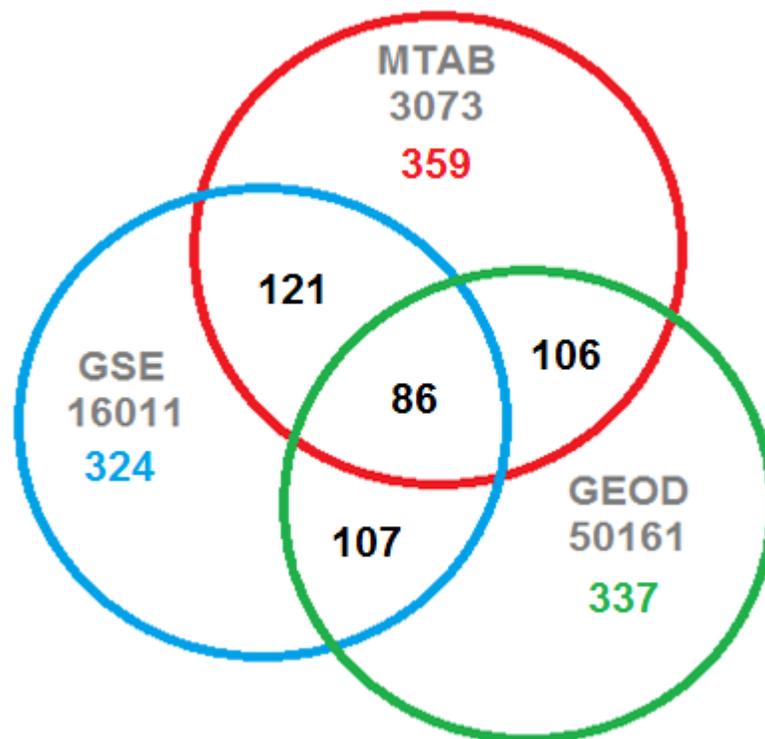




**Figura 4:** Distribuição dos valores de entropia diferencial ( $\Delta S \rightarrow$  dada pela entropia de rede –  $S$  – de glioblastoma, subtraída de  $S$  do tecido normal. Os quadros A,B e C correspondem a distribuição dos valores de  $\Delta S$  em cada uma das coortes analisadas (quadro A: *emtab3073*; quadro B: *gse16011*; quadro C: *geod50161*), sendo que a posição dos genes no eixo X é estabelecida em cada um dos gráficos, ditada pelos valores de  $\Delta S$ .

Em cada uma das coortes, foram selecionados os genes correspondentes aos (5%) menores valores de  $\Delta S$  (tabelas suplementares S1.1, S1.2 e S1.3), para identificação dos processos biológicos possivelmente afetados por um aumento na sua entropia local de rede. Dado que o valor de entropia de rede de um determinado gene é obtido em função do comportamento (expressão) deste gene e de seus vizinhos, as análises posteriores (análises funcionais) foram conduzidas levando em consideração

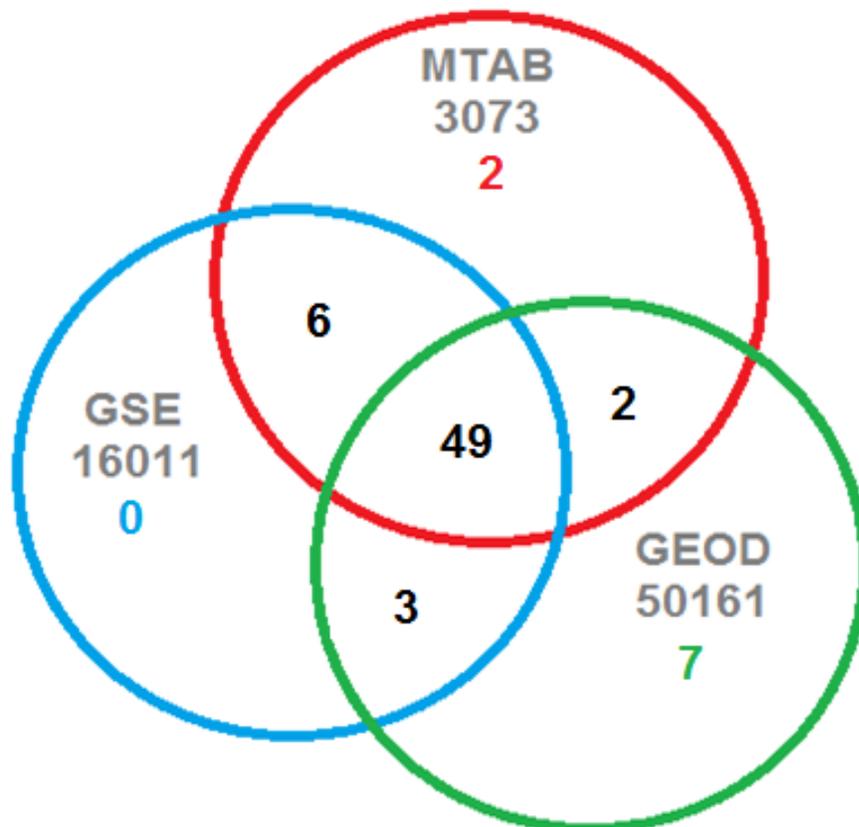
também os vizinhos imediatos (vizinhos de 1ª ordem). As tabelas suplementares S2.1, S2.2 e S2.3 apontam os genes de menor entropia bem como seus vizinhos diretos, representados na figura 5, que mostra, por meio de um diagrama de Venn, o número de genes e sua sobreposição nas listas de cada uma das coortes. Os genes contidos na sobreposição entre as três coortes encontram-se listados na tabela suplementar S2.4.



**Figura 5:** Diagrama de sobreposição (diagrama de Venn) dos genes correspondentes aos maiores aumentos na entropia de rede (5% menores valores de entropia diferencial na comparação de Glioblastoma multiforme – GBM – com o tecido normal) bem como seus vizinhos de primeira ordem (vizinhos diretos), diretamente influentes nos valores de entropia do gene que os conecta – Sobreposição entre as 3 coortes em cerca de 5,9% dos genes relacionados com os maiores aumentos nos valores de entropia local de rede.

### 2.3 O aumento da entropia local de rede aponta rotas biológicas relevantes para a biologia de Glioblastomas.

Considerando apenas as rotas e processos biológicos oriundos da base de dados “Kegg<sup>1</sup>”, a análise funcional apontou em média 60 rotas (ver figura 6) potencialmente afetadas pelo aumento da entropia em cada uma das coortes, havendo uma considerável sobreposição entre as rotas obtidas em cada uma das coortes, conforme detalhado na figura 6.



**Figura 6:** Sobreposição (diagrama de Venn) das rotas identificadas (na comparação de Glioblastoma multiforme – GBM – com o tecido normal) pela análise funcional em cada uma das coortes. Verifica-se uma considerável sobreposição (~71%) entre as rotas identificadas em cada uma das coortes.

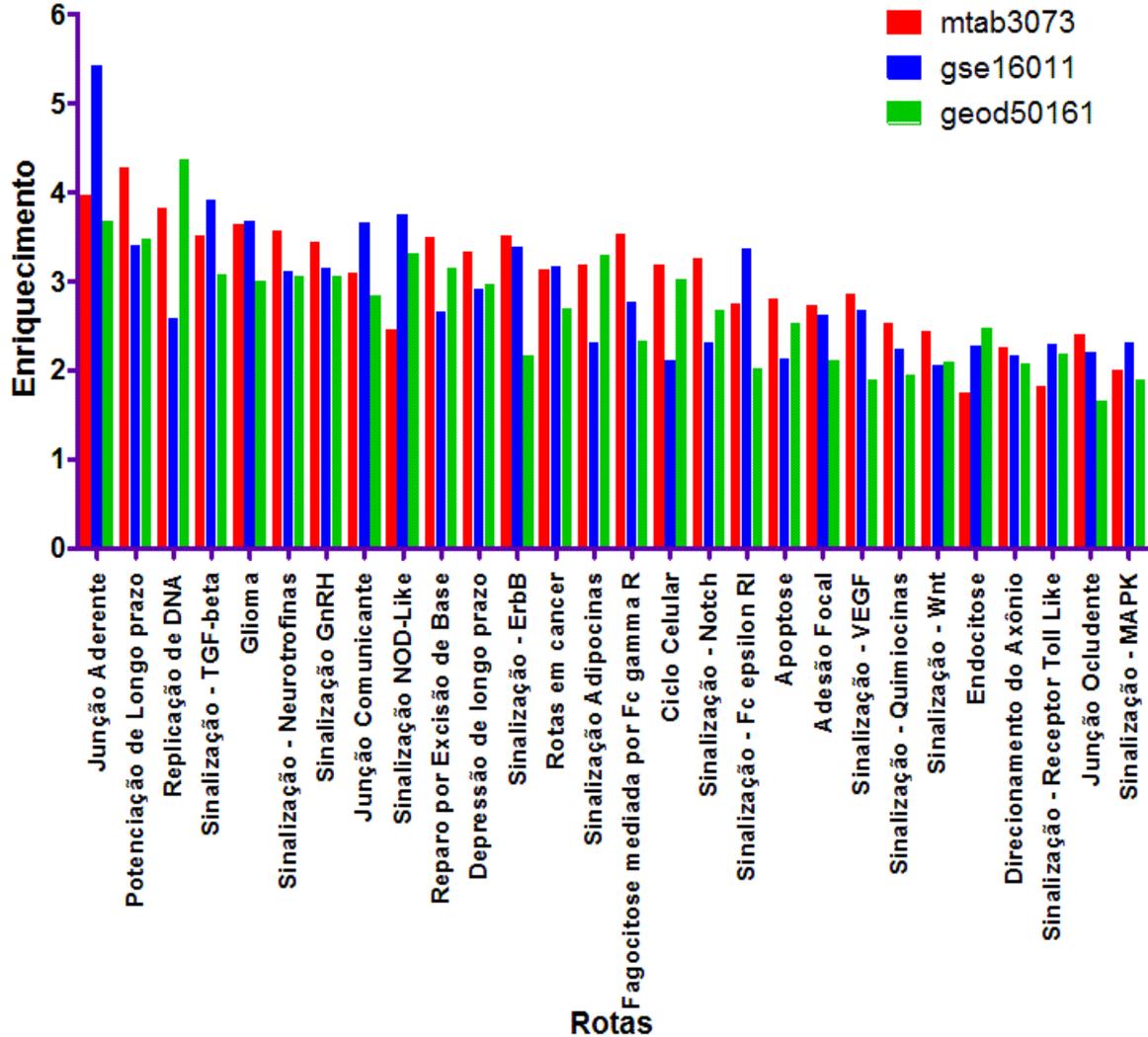
<sup>1</sup> *Kyoto encyclopedia of genes and genomes*: Enciclopédia de Kyoto de genes e genomas (tradução livre); <http://keeg.jp>

As rotas ou processos biológicos identificados para cada coorte bem como a sobreposição representada na figura 6 estão descritos nas tabelas suplementares S3.1 – mtab3073, S3.2 – gse16011, S3.3 – geod50161 e S3.4 – rotas encontradas nas três coortes. Na tabela 3 encontram-se listadas as rotas com alteração de entropia comuns às três coortes e seus respectivos valores de enriquecimento.

**Tabela 3:** Rotas referentes a atuação dos genes com 5% menores valores de entropia e seus vizinhos de rede, comuns às 3 coortes analisadas, excluindo rotas ou grupos de genes com denominações relativas a patologias específicas que não “Cancer” e “Glioma”. Os valores de enriquecimento e significância (valor p) são trazidos para cada uma das coortes. A classificação das rotas nos grandes grupos listados segue o critério do *Kegg* (Os nomes das rotas foram traduzidos, a tabela suplementar S3.5 traz a referência dos nomes originais).

Hierarquia de Processos	Id da Rota (Kegg)	Nome da Rota (traduzido)	mtab3073		gse16011		geod50161		
			Enriq.	Valor p	Enriq.	Valor p	Enriq.	Valor p	
Processamento de Informação Ambiental:	Transdução de Sinal	hsa04370	Sinalização – VEGF	2.85	9.74E-04	2.68	2.77E-03	1.89	1.00E-01
		hsa04310	Sinalização – Wnt	2.43	9.99E-05	2.05	3.56E-03	2.09	2.89E-03
		hsa04012	Sinalização – ErbB	3.51	2.12E-06	3.38	7.50E-06	2.17	2.05E-02
		hsa04010	Sinalização – MAPK	2.00	1.02E-04	2.32	7.64E-07	1.89	6.31E-04
		hsa04330	Sinalização – Notch	3.25	2.88E-03	2.30	7.94E-02	2.68	2.69E-02
		hsa04350	Sinalização - TGF-beta	3.51	2.12E-06	3.91	8.01E-08	3.08	8.80E-05
	Sistema Imunitário	hsa04620	Sinalização - Receptor Toll Like	1.81	6.36E-02	2.30	4.97E-03	2.18	1.07E-02
		hsa04621	Sinalização NOD-Like	2.46	1.82E-02	3.74	2.79E-05	3.30	4.12E-04
		hsa04062	Sinalização - Quimiocinas	2.53	2.90E-06	2.23	1.38E-04	1.94	3.28E-03
		hsa04664	Sinalização - Fc epsilon RI	2.74	1.42E-03	3.37	2.72E-05	2.02	5.65E-02
		hsa04666	Fagocitose mediada por Fc gamma R	3.54	4.91E-07	2.77	3.18E-04	2.32	6.41E-03
	Sistema Nervoso	hsa04730	Depressão de longo prazo	3.32	1.12E-04	2.91	1.32E-03	2.97	1.13E-03
		hsa04720	Potenciação de Longo prazo	4.27	1.85E-07	3.41	8.31E-05	3.47	6.79E-05
		hsa04722	Sinalização - Neurotrofinas	3.57	3.39E-09	3.12	8.04E-07	3.05	2.16E-06
		hsa04360	Direcionamento do Axônio	2.25	1.62E-03	2.16	3.53E-03	2.08	6.96E-03
	Sistema Endócrino	hsa04912	Sinalização GnRH	3.43	8.51E-07	3.15	1.14E-05	3.05	3.25E-05
		hsa04920	Sinalização Adipocinas	3.19	3.14E-04	2.31	2.69E-02	3.29	2.32E-04
	Processos Celulares	Adesão Celular	hsa04510	Adesão Focal	2.74	5.03E-08	2.61	3.92E-07	2.12
hsa04540			Junção Comunicante	3.09	4.72E-05	3.65	5.75E-07	2.83	3.97E-04
hsa04520			Junção Aderente	3.97	2.82E-07	5.42	4.85E-13	3.68	4.16E-06
hsa04530			Junção Ocludente	2.39	3.74E-04	2.19	2.19E-03	1.65	8.04E-02
Destino Celular		hsa04210	Apoptose	2.81	4.28E-04	2.13	2.31E-02	2.53	2.97E-03
		hsa04110	Ciclo Celular	3.18	3.07E-07	2.10	6.12E-03	3.02	2.50E-06
Transporte e Catabolismo		hsa04144	Endocitose	1.74	1.60E-02	2.27	1.05E-04	2.48	9.91E-06
Processamento de Informação Genética	Replicação e Reparo	hsa03410	Reparo por Excisão de Base	3.49	6.51E-03	2.65	7.16E-02	3.15	2.10E-02
		hsa03030	Replicação de DNA	3.82	1.83E-03	2.58	7.89E-02	4.37	3.00E-04
Disfunções Humanas	Cancer	hsa05214	Glioma	3.64	3.88E-05	3.68	3.39E-05	3.00	1.74E-03
		hsa05200	Rotas em cancer	3.12	7.02E-18	3.16	3.61E-18	2.69	5.81E-12

As rotas listadas na tabela 3 estão relacionadas aos seguintes “grandes grupos” de processos biológicos: Processos inflamatórios/imunes; Destino celular; Adesão Celular; Processamento de DNA; Transdução de Sinal; Sinalização Endócrina, bem como grupos de genes diretamente atribuído a fenótipos tumorais. A figura 7 traz um panorama dos valores de enriquecimento das rotas da tabela 3; e, na seção 2.4 desta dissertação, os valores de expressão dos genes de algumas rotas pertencentes a estes grandes grupos de processos biológicos em glioblastomas, serão comparados à situação fisiológica normal, através da análise de expressão gênica diferencial.



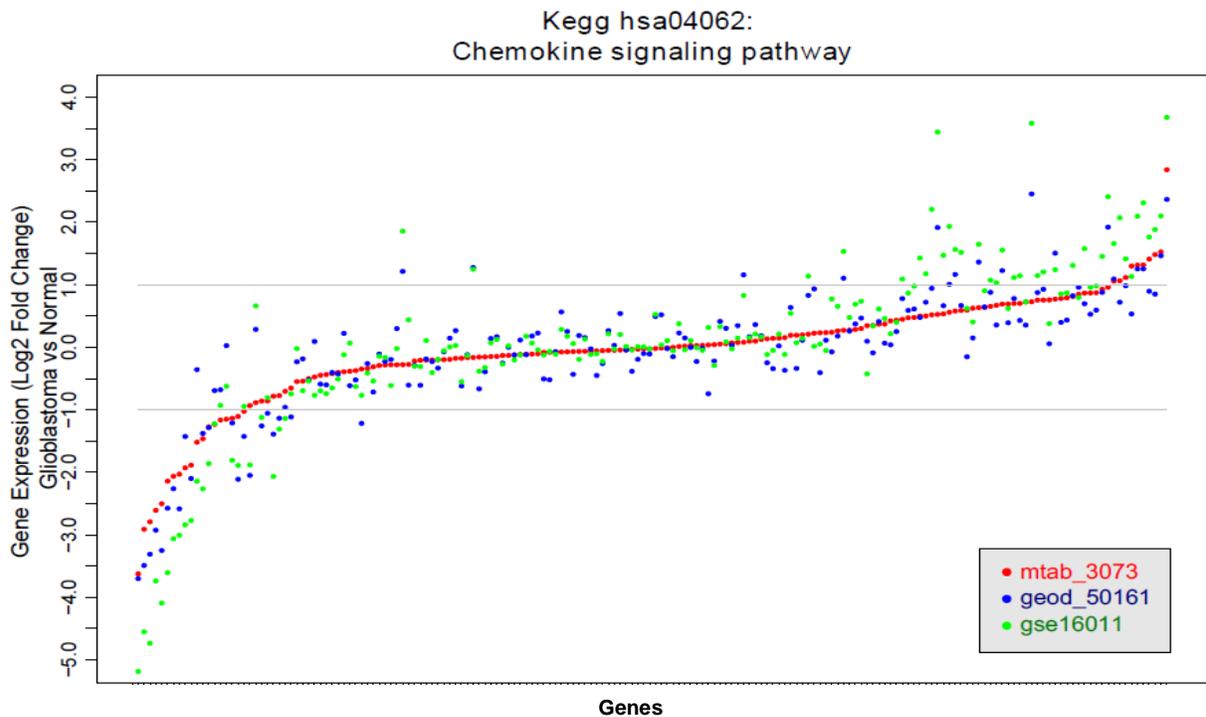
**Figura 7:** Valores de enriquecimento funcional das rotas referentes aos genes selecionados para análise funcional, identificadas em sobreposição nas três coortes analisadas. Apenas rotas comuns às três coortes são mostradas (foram excluídos os processos com denominações de outras patologias não diretamente relacionadas com câncer – e.g. “*vibrio cholerae infection*”). O valor de enriquecimento é dado como uma razão da proporção de genes encontrados sobre a probabilidade de encontra-los ao acaso dentre os genes pertencentes às rotas analisadas.

## **2.4 As rotas identificadas na análise funcional apresentam genes diferencialmente expressos**

As 28 rotas identificadas na análise funcional (apresentadas na figura 7) foram analisadas quanto à expressão diferencial de seus genes. A expressão diferencial é representada graficamente de forma crescente, portanto onde houver genes subexpressos na situação patológica (GBM) em comparação à fisiológica, estes estarão à esquerda do gráfico, enquanto os genes superexpressos estarão à direita e, por conseguinte, ao centro, os genes cuja expressão se encontra mais próxima dos valores do tecido sadio.

Em função do grande número de rotas (28), para cada um dos grandes grupos de processos biológicos (listados abaixo), foram selecionadas rotas para exemplificar o padrão de variação na expressão gênica. Os gráficos correspondentes às demais rotas encontram-se no material suplementar a este trabalho, e encontram-se listados em cada um dos grupos que seguem:

### 2.4.1: Processos do Sistema Imunitário:



**Figura 8:** Panorama da expressão diferencial entre tecidos saudios e glioblastomas para a rota de sinalização por quimiocinas. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).

#### Figuras Suplementares:

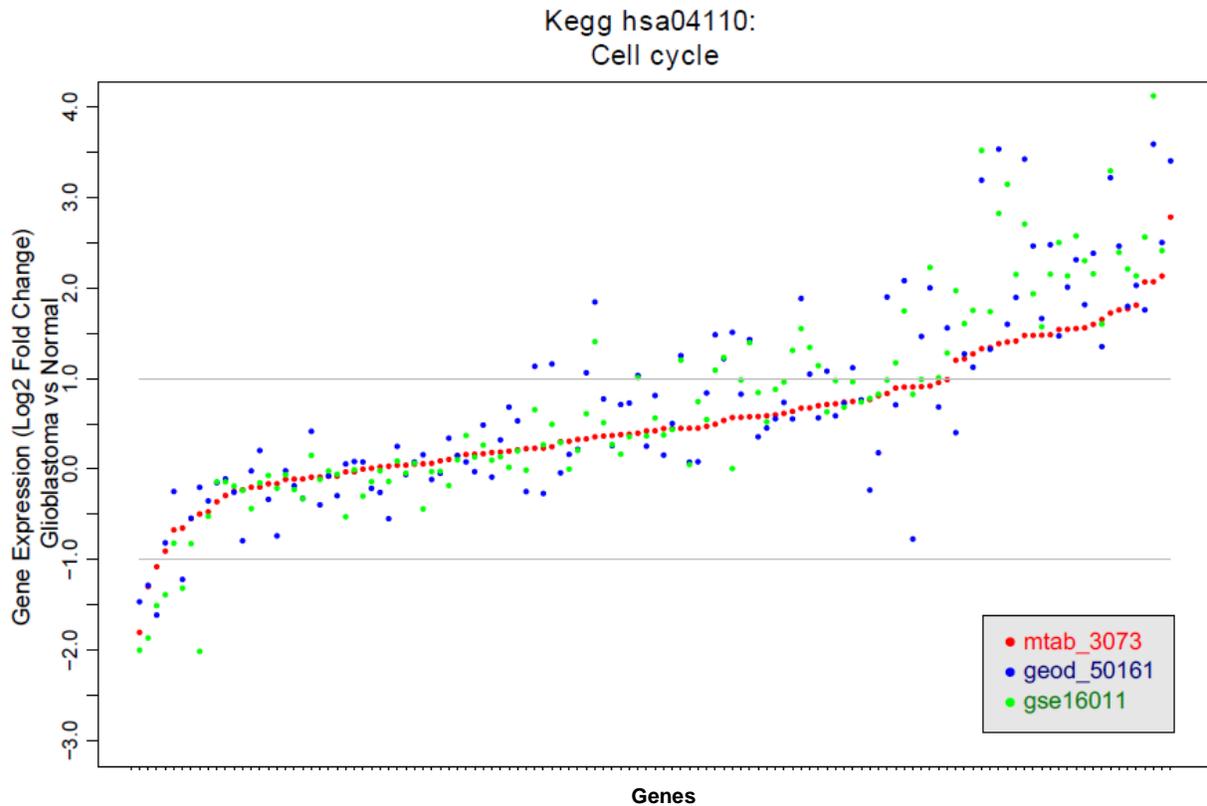
**S4.1.1:** hsa04664: Via de Sinalização Fc epsilon RI;

**S4.1.2:** hsa04666: Fagocitose mediada por Fc gamma R;

**S4.1.3:** hsa04621: Via de Sinalização do receptor NOD-like;

**S4.1.4:** hsa04620: Via de Sinalização do receptor Toll-like.

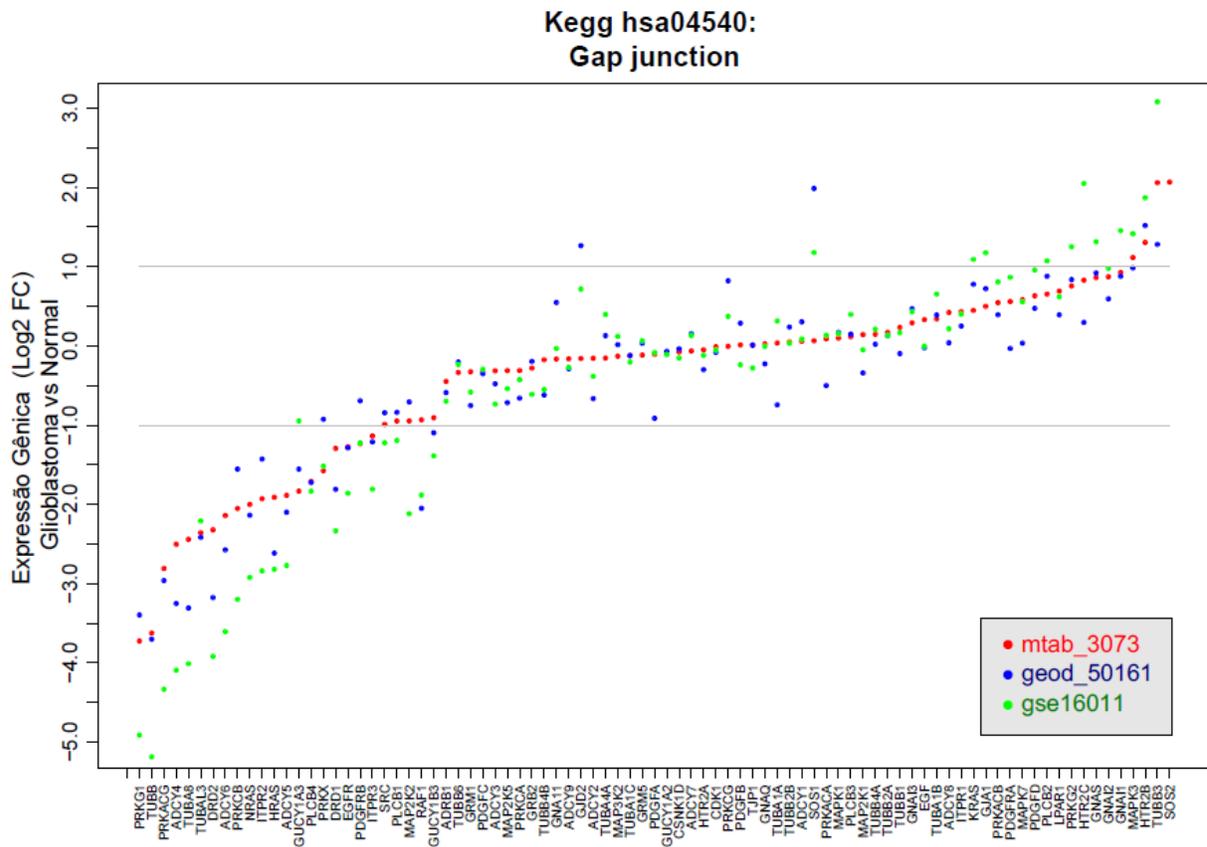
## 2.4.2: Destino celular:



**Figura 9:** Panorama da expressão diferencial entre tecidos sadios e glioblastomas para a rota do ciclo celular. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).



### 2.4.3: Adesão Celular:



**Figura 11:** Panorama da expressão diferencial entre tecidos sadios e glioblastomas no processo de Junção Comunicante. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).

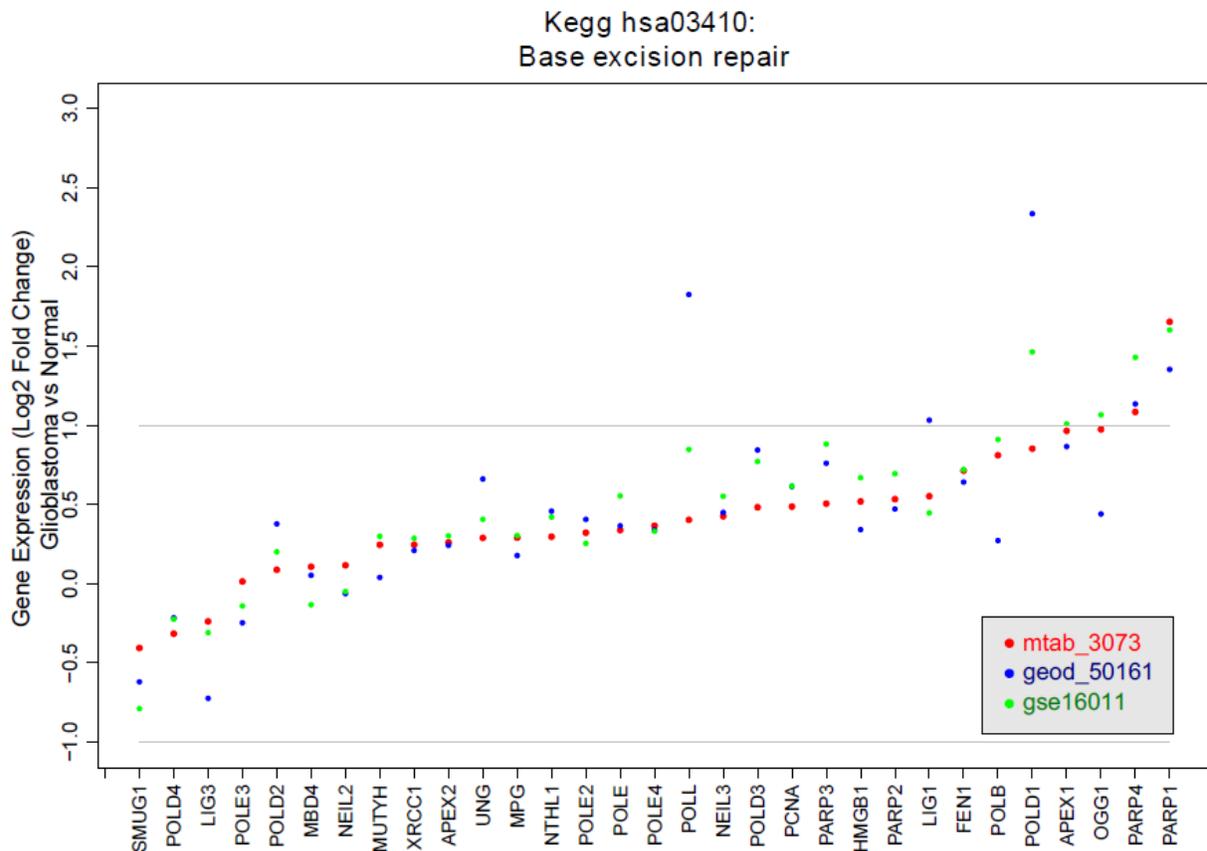
#### Figuras Suplementares:

**S4.3.1:** hsa04520: Junção Aderente;

**S4.3.2:** hsa04510: Adesão Focal;

**S4.3.3:** hsa04530: Junção Ocludente.

#### 2.4.4: Processamento de DNA:

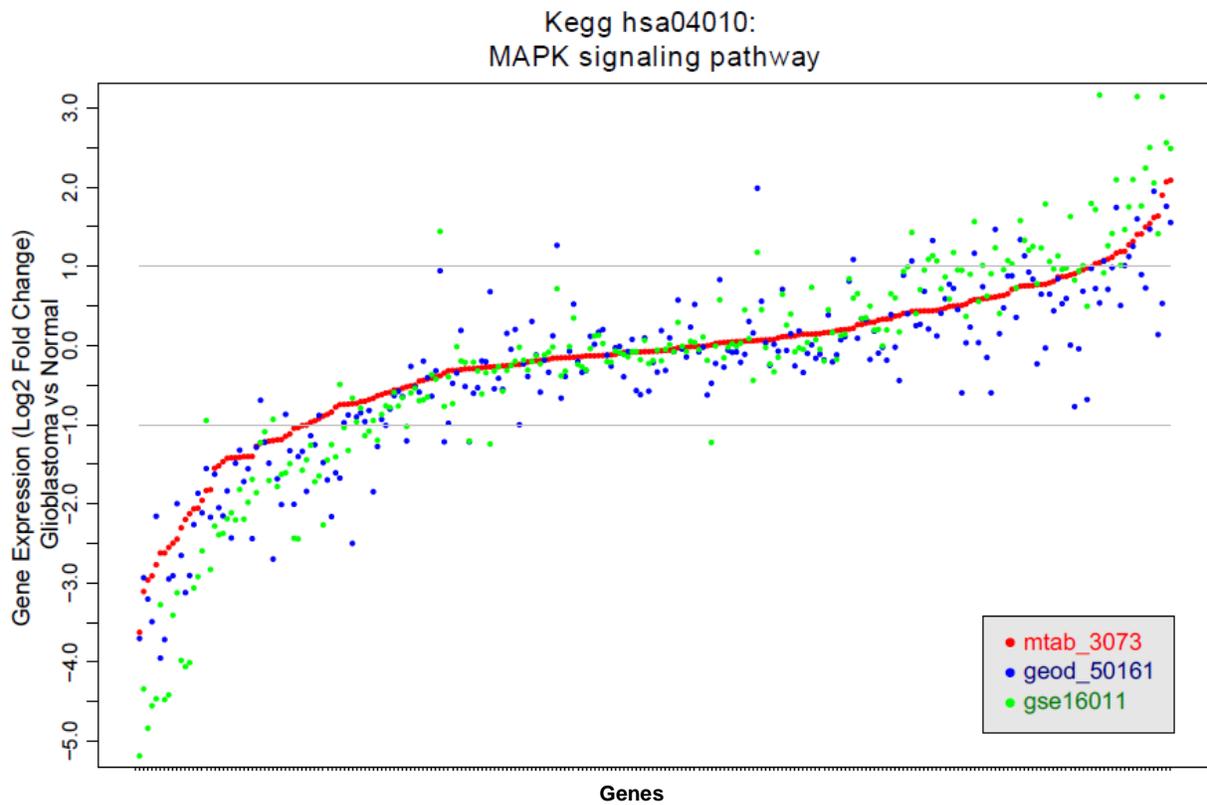


**Figura 12:** Panorama da expressão diferencial entre tecidos saudáveis e glioblastomas para a rota de Reparo por Excisão de Base. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).

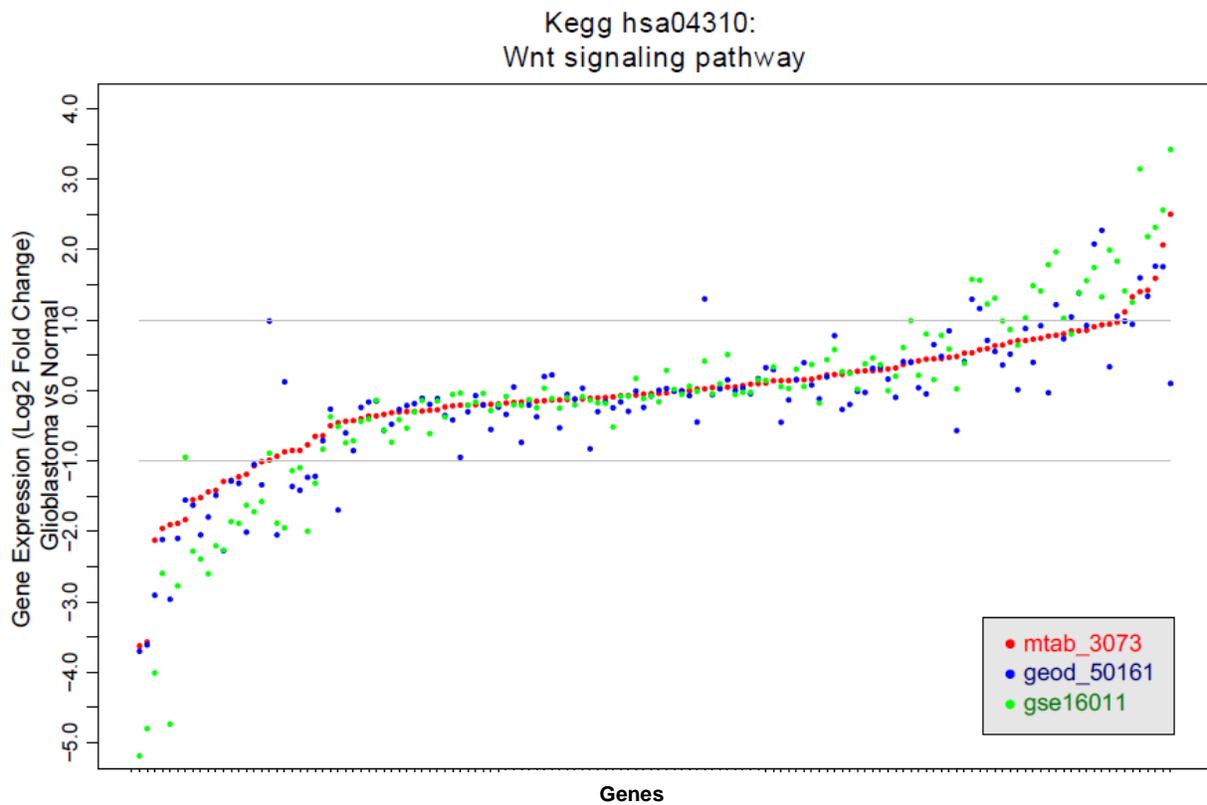
#### Figura suplementar:

##### S4.4.1: hsa03030: Replicação de DNA.

### 2.4.5: Transdução de Sinal:



**Figura 13:** Panorama da expressão diferencial entre tecidos sadios e glioblastomas na via de sinalização de MAPK. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).



**Figura 14:** Panorama da expressão diferencial entre tecidos sadios e glioblastomas na via de sinalização de WNT. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).

#### Figuras suplementares:

**S4.5.1: hsa04012:** Via de sinalização ErbB;

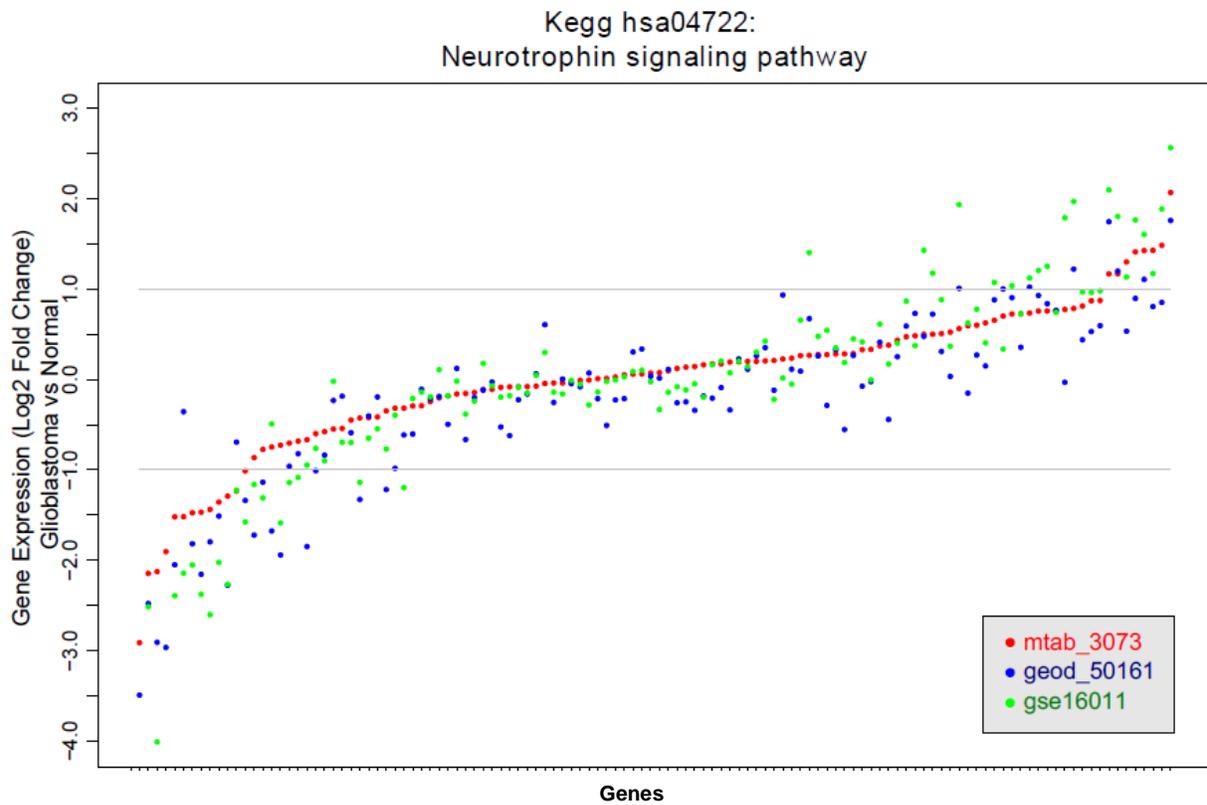
**S4.5.2: hsa04330:** Via de sinalização Notch;

**S4.5.3: hsa04350:** Via de sinalização TGF-beta;

**S4.5.4: hsa04370:** Via de sinalização VEGF.



### 2.4.7: Processos do Sistema Nervoso



**Figura 16:** Panorama da expressão diferencial entre tecidos saudáveis e glioblastomas dos genes da via de sinalização por Neurotrofinas. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).

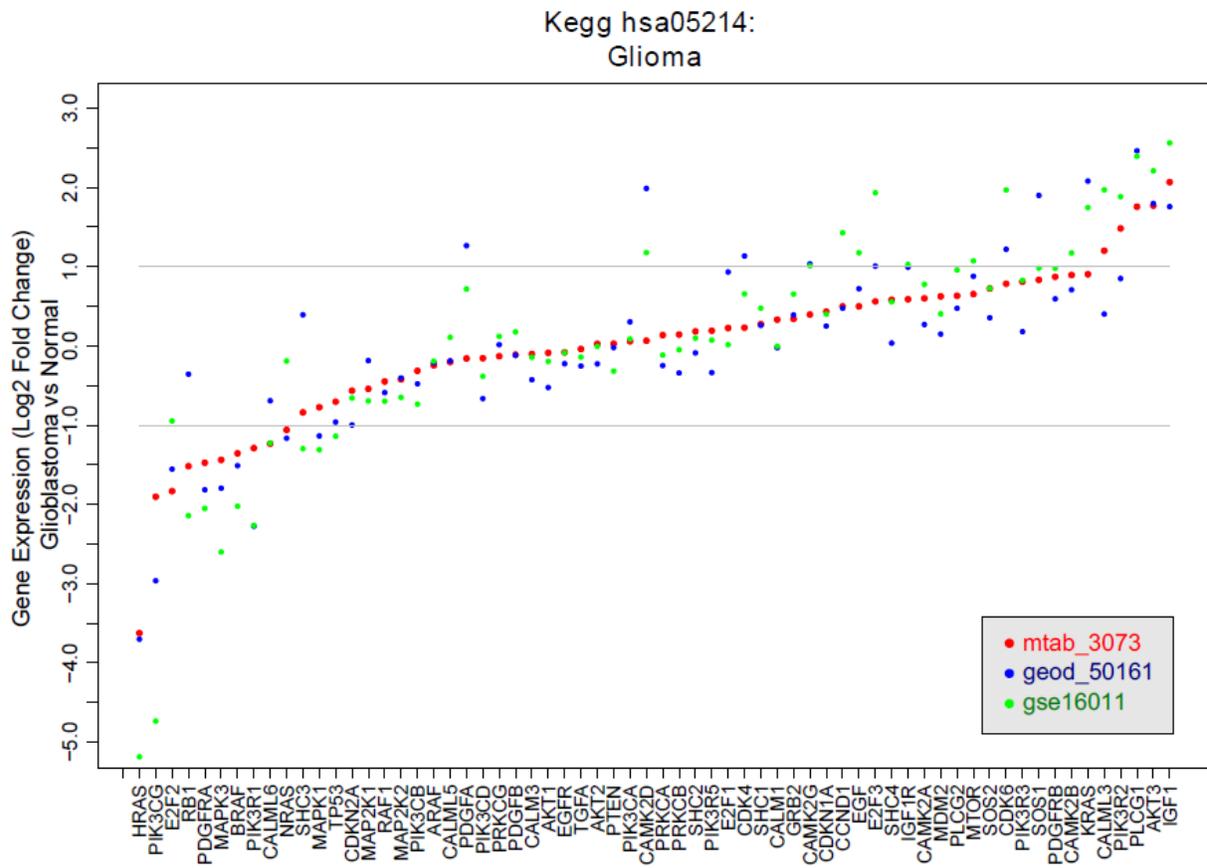
#### Figuras suplementares:

**S4.7.1:** hsa04380: Direcionamento do Axônio;

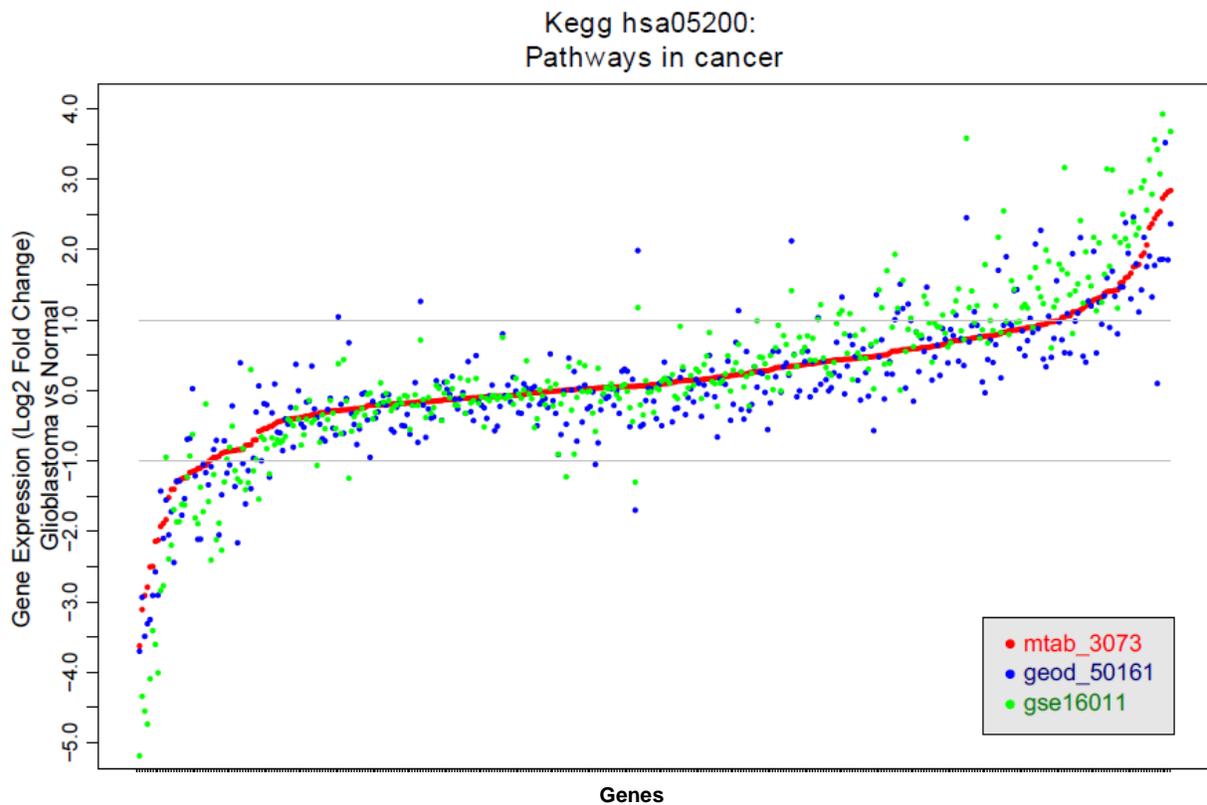
**S4.7.2:** hsa04730: Depressão de Longo Prazo;

**S4.7.3:** hsa04720: Potenciação de Longo Prazo.

## 2.4.8: Patologias Humanas:



**Figura 17:** Panorama da expressão diferencial entre tecidos saudios e glioblastomas dos genes atuantes nas rotas relacionadas com a patologia de gliomas. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).

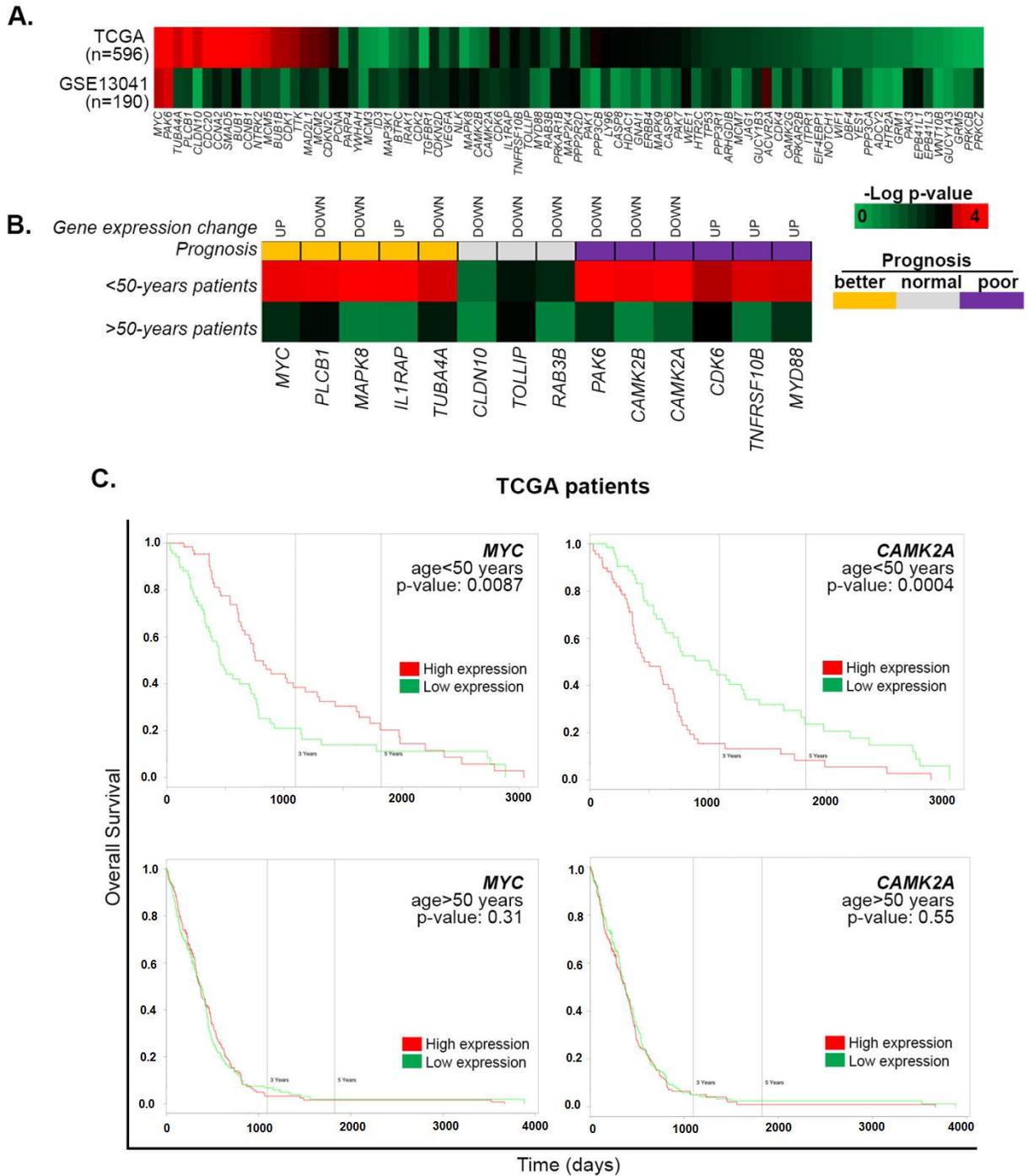


**Figura 18:** Panorama da expressão diferencial entre tecidos sadios e glioblastomas dos genes atuantes nas rotas relacionadas com a patologia tumoral. Cada coorte é representada por uma cor, conforme legenda. No eixo “x”, encontram-se os genes pertencentes à rota, e no eixo “y”, os correspondentes valores de expressão diferencial, dado pelo logaritmo de base 2 da mudança de expressão gênica (entre tecido sadio e tumoral).

Rotas biológicas referentes a outras patologias que não as acima citadas não foram analisadas quanto a sua expressão diferencial, uma vez que reúnem uma série de processos (e, portanto genes/proteínas) comuns aos já analisados neste trabalho bem como peculiaridades em cada doença, onde os genes provavelmente não correspondem aos *hits* da análise funcional.

## **2.5 Genes diferencialmente expressos nas rotas de alta entropia podem prever o prognostico de pacientes com GBM**

Os genes diferencialmente expressos (selecionados pelo critério de expressão diferencial para valor  $p < 0.001$ ) em cada uma das rotas identificadas tiveram seus valores de expressão confrontados com dados de sobrevivência em 2 coortes de pacientes (*GSE13041* – 190 pacientes, *TCGA* – 576 pacientes) utilizando-se a ferramenta *PROGGene* (Goswami & Nakshatri, 2013), a partir da qual identificamos genes cuja expressão encontra-se correlacionada com prognósticos favoráveis ou desfavoráveis (Figura 15A). Os genes com valor prognóstico nesta análise inicial, em pelo menos uma das coortes, foram reanalisados em pacientes estratificados de acordo com critérios clínicos. Nesta etapa, observamos que o oncogene *MYC* e a *PAK6* (*p21 Protein Cdc42/Rac-Activated Kinase 6*) foram os únicos genes cuja expressão foi significativamente correlacionada com prognóstico nas 2 coortes (figuras 15A e B). Verificou-se que as expressões de *MYC* e *PAK6* possuíam correlação positiva com a sobrevida do paciente (quanto maior a expressão, maior é a expectativa de vida do paciente), mas apenas em pacientes com diagnóstico de glioblastoma em idade precoce (Idade < 50 anos) (figura 15B). Outros genes como *PLCB1*, *MAPK8*, *CDK6*, *MYD88* também correlacionaram com a sobrevida nessa população de pacientes (figura 15B). Com alta significância, foi observado que as duas isoformas de calmodulina cinase (*CAMK2A* e *CAMK2B*), que, ainda que se apresentem diminuídas em GBM, quando superexpressas conferem um prognóstico bastante desfavorável em pacientes com idade inferior a 50 anos.



**Figura 19 – quadro A:** Representação em escala de cor de valores de significância (valores  $p$ ) das curvas de Kaplan-Meier (curvas de sobrevivência) de alguns genes nas duas coortes analisadas (GSE13041 – 190 pacientes, TCGA – 576 pacientes) . Em vermelho encontram-se os genes que tiveram associação com prognóstico; **quadro B:** Estratificação das coortes

por idade (com mais ou menos de 50 anos) e reanálise da capacidade de predição de prognóstico, indicando a direção da mudança de expressão (aumento – “*up*” – ou diminuição – “*down*” – da expressão do gene) e o tipo de prognóstico associado (favorável em amarelo e desfavorável em roxo). **Quadro C:** Curvas de sobrevivência dos genes *MYC* e *CAMK2A* na coorte *TCGA*, aplicando-se a estratificação por idade (linha superior, pacientes com menos de 50 anos de idade, e linha inferior, pacientes com mais de 50 anos de idade).

### Parte III

## Capítulo 1 – Discussão

A análise da entropia de rede em tumores do sistema nervoso vem a corroborar a proposição de que o aumento da entropia local de rede é uma assinatura sistêmica de tumores, em conformidade com o que foi demonstrado em outros tipos tumorais (West et al, 2012; Teschendorff & Severini, 2012). Na figura 3 é verificado um deslocamento positivo nas distribuições dos valores de entropia de rede em todos os tumores analisados, em comparação com os tecidos cerebrais saudáveis. Este deslocamento nos valores de entropia é mais pronunciado nos tumores do tipo glioblastoma (GBM), o que acontece também quando comparamos os tumores do tipo glioblastoma a alguns outros tumores de origem glial (tabela 2).

Glioblastomas primários são tumores que apresentam alto grau de instabilidade genômica (Milinkovic et al; 2012). Supõe-se que as diversas aberrações genéticas, consequentes da grande instabilidade genômica característica destes tumores, podem estar por trás das mudanças na expressão gênica que responderiam pelas alterações no modo em que essa informação (de expressão gênica) flui ao longo das redes nestes tumores.

Deleções, ampliações e mutações, que, em tumores, não raramente atingem genes codificantes de proteínas relacionadas com a regulação da transcrição gênica, (ou ainda, com a regulação da degradação do RNA mensageiro – mRNA), desencadeiam alterações nas proporções entre mRNAs de proteínas que interajam entre si em uma rede. A regulação transcricional de um gene é um processo

altamente complexo, envolvendo uma intrincada rede de proteínas, cuja composição é determinada principalmente pelos reguladores transcricionais daquele gene, que estabelecem (e conectam) quais proteínas e vias de sinalização vêm a compor a rede de elementos que regulam a expressão daquele gene. A composição e a proporção entre os elementos de uma rede envolvida na regulação da transcrição de um gene podem variar não somente de gene para gene, mas também em resposta a condições fisiológicas (ou patológicas). Portanto é intuitivo que as alterações transcricionais (e, portanto no fluxo de informação de rede) não acontecem de maneira igual a todos os genes ao longo do genoma de uma célula tumoral.

Dada a natureza estocástica das alterações genômicas que repercutem nas mudanças de padrões transcricionais, a expressão de cada gene é afetada de maneira diferente, o que implica que as alterações no fluxo de informação ao longo da rede também acontecem em diferentes proporções, como pode ser observado na figura 4, onde verifica-se a distribuição dos valores de entropia diferencial ( $\Delta S$ ) na comparação do fluxo de informação de rede de glioblastomas e tecidos saudáveis.

A variação de entropia de rede entre tecidos saudáveis e tumorais acontece de forma não homogênea. São notáveis duas caudas na distribuição, a primeira com menores valores de  $\Delta S$  (maior ganho de entropia), seguida de boa parte dos valores - que se encontram na zona de não variação - e a segunda cauda apresentando valores positivos de  $\Delta S$ , ou seja, alguns genes passam a ter uma maior “organização” (diminuição da entropia) no seu fluxo de informação de rede, contrapondo a tendência geral da distribuição.

A alteração da entropia local de rede ( $S$ ) de um gene é reflexo das mudanças nos seus valores de expressão, bem como de seus vizinhos diretos (ou vizinhos em primeiro grau), uma vez que o valor de  $S$  reflete o fluxo de informação entre este gene e seus vizinhos. Nas tabelas suplementares S1.1, S1.2 e S1.3 estão listados os genes correspondentes aos 5% menores valores de entropia de rede diferencial ( $\Delta S$ ) – ou seja, que apresentam um maior aumento na sua entropia local na rede de Glioblastomas, em comparação ao tecido normal. A figura 5 mostra uma considerável sobreposição (86 genes em comum às 3 coortes) entre os genes correspondentes aos 5% menores  $\Delta S$ , em conjunto com seus vizinhos diretos nas três coortes – referidos como 5%+ $N$  (a listagem dos referidos conjuntos de genes correspondentes a cada coorte encontram-se nas tabelas suplementares S2.1 S2.2 e S2.3). Os genes pertencentes aos conjuntos referenciados (5% +  $N$  em cada coorte) foram submetidos à análise funcional, para verificar quais rotas ou processos biológicos poderiam estar sendo afetados por uma reprogramação no fluxo de informação de rede, sendo apontados pelo aumento na entropia local de rede de seus genes.

Pela análise funcional, foram identificados 59 – *mtab3073*, 58 – *gse16011*, e 61 – *geod50161* – processos biológicos com valores de “enriquecimento” significativos, sendo que a sobreposição entre as rotas ou processos identificados nas três coortes é bastante significativa (49 rotas comuns às três coortes) – figura 6. Esta sobreposição é muito mais expressiva que a sobreposição observada entre os genes propriamente (figura 5). Isto poderia estar relacionado com as diferentes estratégias (alterações em diferentes genes) de se obter um mesmo fenótipo (mediante alterações nos mesmos processos biológicos). Filtrando os processos relacionados com outras patologias não

relacionadas, obtemos uma lista de rotas que tiveram um número considerável de genes envolvidos no aumento de entropia. Estas rotas estão relacionadas com processos de controle de morte e reprodução celular, bem como inflamação, reparo de DNA, transdução de sinal e adesão celular.

A análise funcional dos 5% +  $N$  genes aponta processos biológicos que tiveram seus fluxos de informação de rede “perturbados”. Visando uma melhor compreensão do sentido destas alterações em cada uma das rotas, analisamos o panorama da expressão diferencial dos genes que compõe as mesmas, verificando a presença de genes diferencialmente expressos em cada uma das rotas. A presença destes genes diferencialmente expressos em cada uma das rotas como pode ser observado nas figuras referenciadas na seção 2.4 – resultados – desta dissertação (“parte II → 2.4 *As rotas identificadas na análise funcional apresentam genes diferencialmente expressos*”).

Os processos apontados enquanto alterados pela análise funcional revelaram uma série de rotas possivelmente relacionadas com a fisiopatologia de glioblastomas. Dentre estas rotas encontram-se processos referentes à biologia de tumores (*hsa05200* – câncer) e a gliomas (*hsa05214* – glioma), dentre outros processos, como vias de transdução de sinal (*hsa04370*: Sinalização VEGF; *hsa04310*: Sinalização WNT; *hsa04012*: Sinalização ErbB; *hsa04010*: Sinalização MAPK; *hsa04330*: Sinalização Notch; *hsa04350*: Sinalização - TGF-beta), regulação do ciclo celular, da apoptose, de moléculas de adesão e de vias inflamatórias ou imunes, bem como vias de manutenção da estabilidade genômica ou de resposta a sinalizações endócrinas. Foram identificadas também alterações em nível de entropia em genes pertencentes a

rotas responsáveis por funções específicas do sistema nervoso central como direcionamento do axônio ou potenciações de curto e longo prazo, processos possivelmente afetados negativamente em células de uma massa tumoral. Isso mostra que a análise aqui proposta é útil na identificação de rotas alteradas em processos tumorais.

No trabalho de West e cols. (West et al, 2012) foi verificado um aumento de entropia local de rede em rotas relacionadas com as vias de transdução de sinal. Ao analisar quais rotas possuíam alteração de entropia em seus genes, foi verificado que 6 das 28 rotas que encontram-se alteradas nas três coortes analisadas pertencem a vias de transdução de sinal, corroborando, portanto, com o dado estabelecido pelos referidos autores em outros tipos tumorais (figura 2).

Analisando a expressão gênica destas rotas, foi verificada a presença de genes diferencialmente expressos em cada uma delas, reforçando a ideia da presença de alterações em nível transcricional nestas rotas, além de possíveis desequilíbrios estequiométricos em nível de proteína, conforme apontado pelos valores de entropia de seus genes.

Os genes diferencialmente expressos em cada uma das rotas listadas na tabela 3 foram verificados quanto a sua capacidade de previsão de prognóstico, onde verificou-se, dentre os mais de 100 genes avaliados, apenas o gene *MYC* possuía relação direta com a sobrevida do paciente – em outras palavras, quanto maior a expressão de *MYC*, melhor é o prognóstico (maior a sobrevida) do paciente.

Em um estudo pré clínico (Rajagopalan V et al, 2014) verificou-se que o silenciamento de *MYC* desencadeava o desenvolvimento de um fenótipo ainda mais agressivo por estes tumores, em modelos de ratos Wistar, o que sugere que a atividade deste gene, mesmo estando relacionada com o fenótipo tumoral (Choi JW et al, 2014), é essencial para a contenção da agressividade do glioblastoma. Em um outro trabalho (Cenci T et al, 2012), a expressão de *MYC* foi correlacionada com a expressão do gene *BMI1*, verificando-se que valores mais altos da expressão destes genes eram preditores de um prognóstico mais favorável.

A predição de prognóstico através da expressão gênica é uma tarefa bastante complicada em tumores do sistema nervoso, pois são diversos os fatores que influenciam no prognóstico (como a localização ou tamanho do tumor, a idade do paciente e a taxa de crescimento da massa tumoral). Entretanto, a expressão do gene *MYC* foi unânime na sua relação com prognóstico, nas duas coortes avaliadas, conforme observado nas curvas de sobrevivência na figura 19A.

Analisando estas mesmas coortes sob distinção de faixas de idade (separando-as entre <50 anos e >50 anos), verificamos que, mesmo havendo uma associação para a coorte toda, a associação de *MYC* com o prognóstico era um fator dependente de idade, e não ocorria em pacientes com mais de 50 anos (figura 19B). Portanto a relação de *MYC* com o prognóstico mostrou-se significativa apenas em pacientes com menos de 50 anos de idade.

A alta instabilidade genética de tumores do tipo GBM pode estar associada com a aquisição do que se chama “fenótipo mutador”. A aquisição deste fenótipo é

postulada como um importante passo na evolução de tumores a formas mais agressivas (Loeb, 1991; Carter et al, 2006), e é um potencial responsável pela geração de uma maior diversidade de fenótipos celulares, aumentando a probabilidade da existência de células capazes de transpor as barreiras anti-tumorais do organismo (Cahill, 1999). No entanto, esta mesma instabilidade genômica é limitada pelos mecanismos de seleção natural, onde células em demasiado aberrantes acabam por colapsar devido ao comprometimento de suas funções (Cahill, 1999). Ao aprofundar os estudos a respeito do gene *MYC* foi verificado que a expressão deste gene em tumores encontra-se associada com altas taxas de instabilidade genômica (Kuzyk & Mai, 2014), no entanto, mais estudos são necessários para compreender a dinâmica de *MYC* em glioblastomas, bem como o impacto de mudanças na sua expressão na instabilidade genômica e no fluxo de informação de rede nestes tumores.

## Capítulo 2 – Conclusão

Este trabalho avaliou a entropia local de rede em coortes de pacientes acometidos por tumores do sistema nervoso, comparando a entropia local da rede de GBM com a de tecido nervoso sadio e com a de outros tumores do sistema nervoso, verificando que a entropia local de rede encontra-se aumentada em tumores do sistema nervoso, e principalmente em tumores do tipo glioblastoma multiforme (GBM). Este aumento no desequilíbrio de fluxo de informação de rede pode estar atribuído à alta instabilidade genômica destes tumores.

Foi verificado também que, dentro de uma mesma rede de interação proteica, as mudanças de entropia não ocorrem de maneira igual para todos os genes, havendo genes que apresentam um aumento em sua entropia local de rede (S), bem como genes que praticamente não sofrem alterações em seus valores de S e genes que inclusive apresentam uma redução nos seus valores de S, sendo isso resultante das diferentes taxas de expressão dos genes correspondentes, possivelmente ocasionada por alterações nos componentes dos mecanismos regulatórios destes genes.

A análise de enriquecimento funcional dos (5%) genes de maior ganho de entropia identificou uma série de rotas possivelmente afetadas pela mudança de entropia de seus genes. Estas rotas pertencem a processos biológicos relativos: ao controle do destino celular (ciclo celular e apoptose), a rotas inflamatórias ou processos do sistema imunitário, a manutenção do genoma, a processos

específicos do sistema nervoso central, controle da adesão celular e vias de transdução de sinal.

Avaliando o perfil da expressão gênica das rotas identificadas pela análise funcional, verificamos a presença de genes diferencialmente expressos em cada uma das rotas, reforçando a ideia da existência de alterações neste processo, apontadas pela análise de entropia. Dado que a análise de expressão gênica traz uma noção bastante superficial sobre o status de cada um dos processos biológicos analisados, sendo necessárias análises experimentais para compreender de fato a dimensão destas alterações em cada um dos processos, as análises de entropia de redes, combinadas com a avaliação de rotas de sinalização foram bastante efetiva na elucidação de processos biológicos relevantes para a fisiopatologia de GBM.

A medida de entropia local de rede é uma métrica que explora a topologia de uma rede de interação proteica integrada a dados de expressão gênica. Dado que experimentos que analisam a expressão gênica (como *RNAseq*<sup>®</sup> ou *Microarrays*) geralmente necessitam de algumas centenas de milhares de células para obter mRNA suficiente para a análise, o que acontece é que a diversidade de padrões de expressão é muito maior em um tumor, em comparação a um tecido sadio. Em outras palavras, a análise da expressão gênica de uma massa tumoral tende a ter bastante ruído, em função de contemplar a expressão gênica de diversos programas celulares (oriundos de uma população diversa de células), ao passo que o funcionamento de células (e, portanto a expressão gênica) em um tecido sadio seria muito mais homogêneo. No entanto, ainda assim é possível identificar programas celulares comuns dentre a população de células tumorais. Estes programas celulares se

traduzem em rotas biológicas as quais o sofrem uma reprogramação do seu fluxo de informação de rede, no processo de transformação celular. A identificação destes processos através de diferentes genes nas diferentes coortes reforça a importância destes processos para o fenótipo tumoral.

As alterações genéticas características de glioblastomas constituem-se de mutações, amplificações ou superexpressão de genes como *PTEN*, *EGFR*, *PDGFR-a*, *TP53*, e *mTOR*, no entanto, estas alterações estão mais relacionadas com a gênese destes tumores, sendo por esta razão, incapazes de predizer a sobrevida do paciente. Ao analisar a correlação de expressão gênica de genes alterados (a nível de expressão gênica) com a sobrevida de pacientes, verificamos que a expressão de alguns destes genes, dentre eles o *MYC* possuíam correlação direta com a sobrevida de pacientes de glioblastoma com idade inferior a 50 anos. O gene *MYC* codifica um fator de transcrição que atua na regulação de um grande número de genes do genoma (bem como de genes presentes em grande parte das rotas analisadas), as alterações genéticas neste e em outros fatores de transcrição repercutem nas alterações das redes regulatórias que implicariam na reprogramação da expressão gênica necessária para a progressão de um tumor.

A compreensão da interação entre os mecanismos seletivos (barreiras antitumorais) do organismo e a dinâmica de alterações genômicas em células tumorais- através do estudo nos fluxos de informações em redes de interação proteica - é um potencial caminho para prospecção de rotas a serem testadas enquanto alvos terapêuticos, utilizando-se, por exemplo, inibidores farmacológicos de enzimas-chave das rotas identificadas por análise de biologia de sistemas.

## Referências Bibliográficas:

- Bar-Yam Y, Minai AA,(eds), **Unifying themes in complex systems**, Volume IIIB, Proceedings of the third international conference on complex systems, Springer, ISBN-3-540-35864-1 Springer Berlin Heidelberg New York. 2013.
- Barabasi, A. L. & Oltvai, Z. N. **Network biology: understanding the cell's functional organization**. Nat Rev Genet. 2004.
- Bozdag S, Li A, Baysan M, Fine HA.: **Master regulators, regulatory networks, and pathways of glioblastoma subtypes**. Cancer Inform. 2014
- Brennan C, Momota H, Hambardzumyan D, Ozawa T, Tandon A, Pedraza A, et al. **Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations**. PLoS ONE 2009
- Brown KR, Jurisica I: **Online predicted human interaction database**. Bioinformatics. 2005.
- Cahill DP, Kinzler KW, Vogelstein B, e Lengauer C: **Genetic instability and darwinian selection in tumours**. Trends in Cell Biology - Millenium issue. 1999.
- Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z: **A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers**. Nat Genet. 2006.
- Cenci T, Martini M, Montano N, D'Alessandris QG, Falchetti ML, Annibaldi D, Savino M, Bianchi F, Pierconti F, Nasi S, Pallini R, Larocca LM.: **Prognostic relevance of**

- c-Myc and BMI1 expression in patients with glioblastoma.** Am J Clin Pathol. 2012.
- Choi JW, Schroeder MA, Sarkaria JN, Bram RJ: **Cyclophilin B supports Myc and mutant p53-dependent survival of glioblastoma multiforme cells.** Cancer Res. 2014.
- Edgar R, Domrachev M, Lash AE. **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** Nucleic Acids Res. 2002.
- Gevaert O, Plevritis S.: **Identifying master regulators of cancer and their downstream targets by integrating genomic and epigenomic features.** Pac Symp Biocomput. 2013.
- Goswami CP and Nakshatri H. **PROGgene: gene expression based survival analysis web application for multiple cancers.** J Clin Bioinforma. 2013.
- Hornberg JJ, Bruggeman FJ, Westerhoff HV, Lankelma J: **Cancer: a systems biology disease.** Biosystems. 2006.
- Huang DW, Sherman BT, Lempicki RA. **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** Nucleic Acids. 2009.
- Jonsson PF, Bates PA: **Global topological features of cancer proteins in the human interactome.** Bioinformatics. 2006.
- Klein CA: **Random mutations, selected mutations: A PIN opens the**

**door to new genetic landscapes.** PNAS. 2006

Kolesnikov N. et al.: **ArrayExpress update-simplifying data submissions.** Nucleic Acids Res. 2015.

Kuzyk A, Mai S.: **c-MYC-induced genomic instability.** Cold Spring Harb Perspect Med. 2014.

Loeb LA: **A Mutator Phenotype in Cancer**<sup>1</sup>. Cancer Research. 2001.

Loeb, LA, **Mutator phenotype may be required for multistage carcinogenesis.** Cancer Res. 1991.

Marashi SA, Tefagh M, **A mathematical approach to emergent properties of metabolic networks: Partial coupling relations, hyperarcs and flux ratios,** Journal of Theoretical Biology. 2014.

Mast FD, Ratushny VA, Aitchison JD: **Systems cell biology.** JCB, 2014.

Milinkovic V, Bankovic J, Rakic M, Milosevic N, Stankovic T, Jokovic M, Milosevic Z, Skender-Gazibara M, Podolski-Renic A, Pesic M, Ruzdijic S, Tanic N. **Genomic instability and p53 alterations in patients with malignant glioma.** Exp Mol Pathol. 2012.

Nibbe, R. K., Koyutrk, M. & Chance, M. R.: **An integrative -omics approach to identify functional sub-networks in human colorectal cancer.** PLoS Comput Biol. 2010.

- Platzer A, Perco P, Lukas A, Mayer B: **Characterization of proteininteraction networks in tumors.** BMC Bioinformatics. 2007.
- Prasad TS, Kandasamy K, Pandey A: **Human protein reference database and human proteinpedia as discovery tools for systems biology.** Methods Mol Biol. 2009.
- Prasad, T. S. K. et al.: **Human Protein Reference Database - 2009 Update.** Nucleic Acids Research. 2009.
- Rajagopalan V, Vaidyanathan M, Janardhanam VA, Bradner JE.: **Pre-clinical analysis of changes in intra-cellular biochemistry of glioblastoma multiforme (GBM) cells due to c-Myc silencing.** Cell Mol Neurobiol. 2014.
- Soni D, King JA, Kaye AH, Hovens CM. **Genetics of glioblastoma multiforme: mitogenic signaling and cell cycle pathways converge.** J Clin Neurosci 2005.
- Stelzl U, Wanker EE: **The value of high quality protein-protein interaction networks for systems biology.** Curr Opin Chem Biol. 2006.
- Stewart LA. **Chemotherapy in adult high-grade glioma: a systematic review and meta-analysis of individual patient data from 12 randomised trials.** Lancet 2002.
- Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, et al: **Dynamic modularity in protein interaction networks predicts breast cancer outcome.** Nat Biotechnol. 2009.

Vineis P, Berwick M: **The population dynamics of cancer: a Darwinian perspective.** International Journal of Epidemiology. 2006.

Yu C, Friday BB, Yang L, Atadja P, Wigle D, Sarkaria J, et al.: **Mitochondrial Bax translocation partially mediates synergistic cytotoxicity between histone deacetylase inhibitors and proteasome inhibitors in glioma cells.** Neuro Oncol 2008;10(3):309–19.

## Anexo: 1

### Colaborações em Outros Trabalhos

1 – Artigo publicado no periódico *Molecular Neurobiology* em 2014

(primeira página em anexo):

***Gene Expression Profile of NF- $\kappa$ B, Nrf2, Glycolytic, and p53 Pathways  
During the SH-SY5Y Neuronal Differentiation Mediated by Retinoic Acid.***

Matheus Augusto de Bittencourt Pasquali, Vitor Miranda de Ramos, Ricardo D'Oliveira Albanus, Alice Kunzler, Luis Henrique Trentin de Souza, Rodrigo Juliani Siqueira Dalmolin, Daniel Pens Gelain, Leila Ribeiro, Luigi Carro, José Cláudio Fonseca Moreira

2 – Artigo submetido para publicação no periódico *Journal of Experimental Medicine*, em 2015:

***BRCA1 balances homologous recombination and NRF2 pathway activities to promote alkylation survival***

Alfeu Zanotto-Filho, Pragathi Masamsetti, Eva Loranc, Sonal Tonapi, Aparna Gorthi, Xavier Bernard, Ravi Dashnamoorthy, Luis Henrique Souza, Jose Claudio Moreira, Uthra Suresh, Yidong Chen, and Alexander Bishop,

## Gene Expression Profile of NF- $\kappa$ B, Nrf2, Glycolytic, and p53 Pathways During the SH-SY5Y Neuronal Differentiation Mediated by Retinoic Acid

Matheus Augusto de Bittencourt Pasquali · Vitor Miranda de Ramos · Ricardo D'Oliveira Albanus · Alice Kunzler · Luis Henrique Trentin de Souza · Rodrigo Juliani Siqueira Dalmolin · Daniel Pens Gelain · Leila Ribeiro · Luigi Carro · José Cláudio Fonseca Moreira

Received: 21 August 2014 / Accepted: 12 November 2014  
 © Springer Science+Business Media New York 2014

**Abstract** SH-SY5Y cells, a neuroblastoma cell line that is a well-established model system to study the initial phases of neuronal differentiation, have been used in studies to elucidate the mechanisms of neuronal differentiation. In the present study, we investigated alterations of gene expression in SH-SY5Y cells during neuronal differentiation mediated by retinoic acid (RA) treatment. We evaluated important pathways involving nuclear factor kappa B (NF- $\kappa$ B), nuclear E2-related factor 2 (Nrf2), glycolytic, and p53 during neuronal differentiation. We also investigated the involvement of reactive oxygen species (ROS) in modulating the gene expression profile of those pathways by antioxidant co-treatment with Trolox<sup>®</sup>, a hydrophilic analogue of  $\alpha$ -tocopherol. We found that RA treatment increases levels of gene expression of NF- $\kappa$ B, glycolytic, and antioxidant pathway genes during neuronal differentiation of SH-SY5Y cells. We also found that ROS production induced by RA treatment in SH-SY5Y

cells is involved in gene expression profile alterations, chiefly in NF- $\kappa$ B, and glycolytic pathways. Antioxidant co-treatment with Trolox<sup>®</sup> reversed the effects mediated by RA NF- $\kappa$ B, and glycolytic pathways gene expression. Interestingly, co-treatment with Trolox<sup>®</sup> did not reverse the effects in antioxidant gene expression mediated by RA in SH-SY5Y. To confirm neuronal differentiation, we quantified endogenous levels of tyrosine hydroxylase, a recognized marker of neuronal differentiation. Our data suggest that during neuronal differentiation mediated by RA, changes in profile gene expression of important pathways occur. These alterations are in part mediated by ROS production. Therefore, our results reinforce the importance in understanding the mechanism by which RA induces neuronal differentiation in SH-SY5Y cells, principally due this model being commonly used as a neuronal cell model in studies of neuronal pathologies.

**Keywords** Neuronal differentiation · Neuroblastoma · Retinoic acid · SH-SY5Y cells · Oxidative stress

M. A. de Bittencourt Pasquali (✉) · V. M. de Ramos · R. D. Albanus · A. Kunzler · L. H. T. de Souza · D. P. Gelain · J. C. F. Moreira  
 Departamento de Bioquímica, Centro de Estudos em Estresse Oxidativo, Universidade Federal do Rio Grande do Sul—UFRGS, Rua Ramiro Barcelos 2600, Cep 91035-003 Porto Alegre, Rio Grande do Sul, Brazil  
 e-mail: 00124262@ufrgs.br

M. A. de Bittencourt Pasquali · R. J. S. Dalmolin  
 Departamento de Bioquímica, Instituto de Medicina Tropical, Universidade Federal do Rio Grande do Norte—UFRN, Avenida Senador Salgado Filho 3000, Cep 59078-900 Natal, Rio Grande do Norte, Brazil

L. Ribeiro · L. Carro  
 Departamento de Informática Aplicada, Universidade Federal do Rio Grande do Sul—UFRGS, Rua Bento Gonçalves 9500, Cep 91501-970 Porto Alegre, Rio Grande do Sul, Brazil

### Introduction

The brain is a unique and complex structure. In order to understand the processes leading to neural development, one must rely both on neuroscience and on developmental biology for describing the cellular and molecular mechanisms by which the emergence of the complex nervous system during the embryonic development and throughout life. Additionally, the evolutionary processes of neuronal differentiation are tightly regulated and remain largely unknown. The regulation of gene expression during these events is governed by different mechanisms at transcriptional level. Proteins and