UNIVERSIDADE DE SÃO PAULO
Faculdade de Medicina de Ribeirão Preto
Departamento de Genética

Flávia Sacilotto Donaires

**Utilização de análise bioinformática e validação por PCR
quantitativa em tempo real na identificação da indução
transcricional dos fatores de transcrição *E2F1* e *E2F4* em
linhagens de glioblastoma**

Ribeirão Preto
2011

Flávia Sacilotto Donaires

**Utilização de análise bioinformática e validação por PCR
quantitativa em tempo real na identificação da indução
transcricional dos fatores de transcrição *E2F1* e *E2F4* em
linhagens de glioblastoma**

Dissertação apresentada à Faculdade
de Medicina de Ribeirão Preto da
Universidade de São Paulo para
obtenção do título de Mestre em
Ciências, área de concentração em
Genética

Orientadora: **Elza Tiemi Sakamoto Hojo**

Ribeirão Preto

2011

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

Donaires, Flávia Sacilotto

Utilização de análise bioinformática e validação por PCR quantitativa em tempo real na identificação da indução transcricional dos fatores de transcrição *E2F1* e *E2F4* em linhagens de glioblastoma – Ribeirão Preto, 2011.

91 p. : il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Genética.

Orientadora: Hojo, Elza Tiemi Sakamoto.

1. Glioblastoma. 2. Microarranjos. 3. Fatores de transcrição. 4. *E2F1*. 5. *E2F4*.

APOIO E SUPORTE FINANCEIRO

Este trabalho foi realizado com o apoio financeiro das seguintes entidades e instituições:

- Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP
- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES
(Bolsa de mestrado – março/2009 a novembro/2010)
- Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
(Bolsa de mestrado – dezembro/2010 a fevereiro/2011)
- Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP – USP)
- Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo (FFCLRP – USP)
- Fundação de Apoio ao Ensino, Pesquisa e Assistência do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FAEPA – HCFMRP – USP)

Tu perguntas

o que uma lagosta tece lá embaixo com seus pés dourados?

Respondo que o oceano sabe.

E por quem a medusa espera em sua veste transparente?

Está esperando pelo tempo, como tu.

'Quem as algas apertam em seu abraço...', perguntas

'mais firme que uma hora e um mar certos?' Eu sei.

Perguntas sobre a presa branca do narval

e eu respondo contando como o unicórnio do mar,

arpado, morre.

Perguntas sobre as plumas do rei-pescador

que vibram nas puras primaveras dos mares do sul.

Quero te contar que o oceano sabe isto:

que a vida, em seus estojos de jóias,

é infinita como a areia incontável, pura;

e o tempo, entre uvas cor de sangue

tornou a pedra dura e lisa encheu a água-viva de luz,

desfez o seu nó, soltou seus fios musicais

de uma cornicópia feita de infinita madreperola.

Sou só a rede vazia diante dos olhos humanos na escuridão

e de dedos habituados à longitude

do tímido globo de uma laranja.

Caminho como tu, investigando a estrela sem fim

e em minha rede, durante a noite, acordo nu.

A única coisa capturada é um peixe preso dentro do vento.

Pablo Neruda

AGRADECIMENTOS

À **Profa. Dra. Elza Tiemi Sakamoto Hojo** pela orientação deste trabalho e também por ter proporcionado as condições para o desenvolvimento da pesquisa. Agradeço por todo o conhecimento fornecido e pela sua colaboração em minha formação.

À **Profa. Dra. Catarina Satie Takahashi** por todo o incentivo e apoio, pela simpatia e pelos momentos de descontração. E também não poderia deixar de agradecê-la pelos tomates com os quais sempre fui presenteada. Salada de tomate é uma das minhas preferências gastronômicas!

Ao **Prof. Dr. Ademilson Espencer E. Soares** (coordenador do programa de pós-graduação), pelo apoio em diversos momentos ao longo do mestrado.

À **Profa. Dra. Zilá Luz Paulino Simões**, por gentilmente ceder alguns aparelhos de seu laboratório.

Às secretárias do Departamento de Genética, especialmente à **Susie** e à **Sílvia**, pelo apoio constante.

Aos membros da **banca examinadora** por terem aceitado participar da avaliação deste trabalho, contribuindo para sua melhoria.

Às meninas do laboratório de Imunogenética, especialmente à **Dri** e à **Ernna**, que me auxiliaram em alguns experimentos. E também ao **Prof. Dr. Geraldo A. S. Passos**, que sempre me ouvia quando eu aparecia meio perdida nos corredores do seu laboratório.

A todos do laboratório de Bioinformática, especialmente ao **Daniel** e à **Profa. Dra. Silvana Giuliatti**, pela amizade e apoio.

Ao pessoal do laboratório de Genética e Biologia Molecular de Fungos, principalmente à **Nalu** e ao **Tiago**, pela atenção e paciência ao ajudar na padronização da técnica de qRT-PCR em tempo real.

Ao “ex-amigo de laboratório”, porém não ex-amigo, **Danillo (Bollor)**, por ter me ajudado muito (muito mesmo) na fase mais crítica do meu trabalho, quando nada dava certo. Além disso, agradeço pela amizade, pelas dicas, pelos conselhos e pelos divertidos momentos de conversa!

À minha querida xará (**Flávia**) do bloco A, que apesar de ser uma abelhudinha (no bom sentido!) me ajudou muito ao longo de todo meu trabalho, dando dicas, sugestões e também me ouvindo nos momentos em que nada dava certo. Além disso, me ajudou muito na correção deste trabalho, a qual me foi muito importante. Sua amizade também é parte dos meus “resultados”, aliás, um dos melhores!

A todas as minhas grandes amigas do laboratório, cultivadas cuidadosamente durante anos de convívio! Às vezes cultivar amigas é mais difícil do que cultivar células: células não têm dias bons ou ruins, não têm variações de humor ou problemas (se tiverem, a gente joga fora). Também não tem a capacidade de nos ajudar, de nos ouvir ou dar conselhos. Admito que nessa complicada tarefa de cultivar amigas, todos foram impecáveis.

À **Ana Paula**, pela ótima ajuda na correção. Ao **Danilo (Xitão)**: encontrei alguém pior que eu quando o quesito é “deixar pra última hora”. Ao **Felipe (Magal)** pelas aulas de violino, por me acompanhar nos assovios, por gostar de animais e plantas (alguns biólogos não gostam!) e por me ensinar vários conceitos da biologia. À **Fernanda**, que me deu dicas muito boas na hora da escrita. À **Férrnanda**, que carregou um fardo pesado por minha causa, literalmente (duas resmas e um cartucho!). Ao **Gustavo (Polvilho)** pela amizade e companhia nos cafés da tarde. Ao **Léo**, o cara do nanotubo, pelas risadas na cozinha e pela companhia na espera do *bus*. À **Pati**, que não está mais no laboratório, mas deixou ótimas lembranças. Ao **Paulo (Cop, ou Lordinho, se preferirem)** por me ajudar grandemente nesse trabalho, me ensinando a cultivar células e extrair RNA em pleno sábado à noite! Quem mais faria isso por mim? À **Paulinha**, que veio especialmente dos Estados Unidos para compor o “trio parada droga”! Em pouco tempo de convivência, você já nos conquistou (o Xitão que o diga!). À **Raquel**, por sempre me dar ótimos conselhos. Ao **Tiago**, o mais novo integrante do laboratório! À **Gi** (minha *druginha*) por sua grande amizade e companheirismo. Pouquíssimas pessoas têm essa capacidade de encantar os outros. Também por me ouvir rir, chorar, desabafar, cantar paródias e falar abobrinhas, além das caronas para casa, sempre muito bem vindas. Você foi o apoio que eu precisei em muitos momentos de “crise laboratorial”.

À **Vera**, que cuida com muito zelo da limpeza do laboratório, e também do nosso almoço, sempre fazendo um suco de limão ou trazendo um docinho para a sobremesa.

Ao **Júnior (Juco)** e **Sueli (Ruivão)**, não apenas pelo suporte técnico, mas também pelos ótimos momentos e pelas boas risadas.

Ao meu querido namorado, **Felipe (Mimi)**! Talvez ele tenha sido a maior “vítima” dos meus maus momentos, mas sempre esteve pronto a me ouvir e me acalmar. Obrigada por toda paciência e compreensão. Você não é só um namorado, mas um grande amigo. Espero ter você sempre ao meu lado!

A toda minha família, que sempre me apoiou e esteve ao meu lado, mesmo nos momentos mais difíceis. Agradeço especialmente à minha mãe, sem a qual não teria conseguido chegar até aqui! Obrigada por me aturar, ouvir, compreender e direcionar, mesmo quando as palavras não eram suficientes para me confortar!

SUMÁRIO

RESUMO	i
ABSTRACT	iii
1. INTRODUÇÃO	13
1.1 Microarranjos de cDNA e sua aplicação no estudo do câncer	13
1.2 Glioblastoma: necessidade de novas terapias mais eficazes	14
1.3 Bancos de dados de microarranjos e a necessidade de análises integrativas	15
1.4 Alvos transcricionais e “assinaturas de câncer”	18
1.5 Fatores de transcrição e sua importância na regulação gênica	18
1.6 Potenciais alvos moleculares e a necessidade de validação	19
2. OBJETIVOS	21
2.1 Objetivo geral	21
2.2 Objetivos específicos	21
3. MATERIAL E MÉTODOS	22
3.1 Análise <i>in silico</i> de microarranjos de cDNA	22
3.1.1 Conjunto de dados de microarranjos de cDNA	22
3.1.2 Forma de análise dos dados de microarranjos de cDNA	22
3.1.2.1 Análise estatística e agrupamentos gênicos	23
3.1.2.2 Análise de fatores de transcrição	24
3.1.2.3 Análise de associação a assinaturas de expressão gênica	25
3.1.2.4 Análise de vias de sinalização celular	27
3.2 Linhagens celulares	27
3.3 Cultivo celular	28
3.4 Extração de RNA total	28
3.5 Grupo controle	29
3.6 Amostras de RNA da linhagem U87 tratada com drogas antitumorais	29
3.7 Transcrição reversa (RT-PCR) para realização da qRT-PCR em tempo real	30
3.8 Avaliação de expressão gênica por RT-PCR quantitativa (qRT-PCR) em tempo real	30
3.8.1 Desenho de iniciadores (<i>primers</i>)	30
3.8.2 Padronização da técnica para os FTs <i>E2F1</i> e <i>E2F4</i> e para os genes de referência <i>B2M</i> , <i>GUSB</i> , <i>HPRT1</i> e <i>TBP</i>	33
3.8.3 Montagem da reação	35
3.8.4 Análise dos resultados	36
3.8.4.1 Método de escolha dos genes de referência	36
3.8.4.2 Análise da expressão dos FTs <i>E2F1</i> e <i>E2F4</i>	37

4. RESULTADOS	38
4.1 Análise <i>in silico</i> de microarranjos de cDNA	38
4.1.1 Análise estatística e agrupamentos gênicos	38
4.1.2 Análise de fatores de transcrição	40
4.1.3 Análise de associação a assinaturas de expressão gênica	40
4.1.4 Comparação entre a análise de fatores de transcrição e a análise de associação a assinaturas de expressão gênica	42
4.1.5 Análise de vias de sinalização celular	43
4.2 Avaliação de expressão gênica por RT-PCR quantitativa (qRT-PCR) em tempo real	48
4.2.1 Padronização da técnica para os FTs <i>E2F1</i> e <i>E2F4</i> e para os genes de referência <i>B2M</i> , <i>GUSB</i> , <i>HPRT1</i> e <i>TBP</i>	48
4.2.2 Escolha de genes de referência	53
4.2.3 Expressão de <i>E2F1</i> e <i>E2F4</i> em linhagens de glioblastoma	54
4.2.3.1 Expressão de <i>E2F1</i> e <i>E2F4</i> nas linhagens T98, U251, U138, U87, U343, MO59J e MO59K	55
4.2.3.2 Expressão de <i>E2F1</i> e <i>E2F4</i> na linhagem U87 tratada com agentes quimioterápicos	58
5. DISCUSSÃO	61
6. CONCLUSÕES	78
REFERÊNCIAS BIBLIOGRÁFICAS	79

RESUMO

O emprego da metodologia de microarranjos no estudo do câncer tem permitido a identificação de genes com alterações em seus perfis de expressão, os quais estão direta ou indiretamente envolvidos na etiologia dessa doença. O crescente número de publicações de experimentos de expressão gênica em repositórios de dados de microarranjos demonstra que, em geral, a limitação não está na quantidade ou qualidade desses experimentos, mas no processamento desses dados. Dessa forma, há a necessidade de desenvolver metodologias de bioinformática que sejam capazes de analisar tais perfis de forma integrada, incluindo no contexto da análise, dados provenientes de outros experimentos.

O desenvolvimento do câncer tem sido associado principalmente a distúrbios nos mecanismos de controle do ciclo celular, cujas vias são dependentes de uma maquinaria transcricional e de seus elementos regulatórios; entre estes últimos, os fatores de transcrição (FTs) têm sido estudados como potenciais alvos para a terapia molecular.

No presente trabalho, um conjunto de dados de microarranjos realizados a partir de amostras de glioblastoma foi obtido em repositórios públicos (*GEO* e *ArrayExpress*). O teste estatístico SAM foi aplicado aos dados e os genes diferencialmente expressos ($FDR \leq 0,05$), induzidos em glioblastoma (1.830 genes), foram submetidos a uma análise de associação a FTs (programa FatiGO+), a qual associou os FTs *HNF1A*, *IRF7* e *E2F* ($p \leq 0,05$) aos genes induzidos. Adicionalmente, a lista de genes foi submetida a uma análise de associação a um banco de dados de assinaturas transcricionais do *software TBrowser*, extraídas de diversos experimentos de microarranjos do *GEO*. Como resultado, 7.910 assinaturas foram associadas ($p \leq 0,01$), sendo que as cem primeiras também foram relacionadas a FTs por meio de ferramentas disponibilizadas no próprio *TBrowser*. Os FTs *E2F1* e *E2F4* foram associados a mais de 80% das assinaturas analisadas. Dessa forma, ambas as análises apontaram o FT *E2F*, mais especificamente *E2F1* e *E2F4*, como associados à lista inicial de genes fornecida pelo SAM. A expressão de *E2F1* e *E2F4* foi avaliada por meio da técnica de RT-PCR quantitativa em tempo real em amostras de RNA de sete linhagens de glioblastoma (T98, U251, U138, U87, U343, MO59J e MO59K). Interessantemente, ambos os genes foram apontados como superexpressos em todas as linhagens, a maioria com significância estatística.

A família de FTs *E2F* apresenta papéis importantes no controle da proliferação celular e apoptose. Alguns membros atuam como oncogenes, e outros, como genes supressores de tumor, dependendo do contexto molecular nos quais se encontram. Muitos trabalhos da literatura têm apontado a importância desses FTs, especialmente *E2F1*, no processo de

tumorigênese. Com base nos resultados obtidos no presente trabalho, o *status* de expressão (indução) de *E2F1* e *E2F4* é provavelmente associado ao glioblastoma. Esses FTs podem influenciar a expressão de uma série de genes cruciais, frequentemente alterados nessa doença, o que ainda requer um estudo mais detalhado, visando à validação desses FTs como biomarcadores, ou até mesmo como alvos moleculares que possam ser empregados no estabelecimento de novas modalidades de terapias.

ABSTRACT

The application of DNA microarrays in cancer research has provided the identification of abnormal expression profiles of a series of genes, which may be directly involved in the etiology of such disease. The increasing number of publications related to gene expression profile in microarray repositories has demonstrated that the limitation of such experiments is not associated with the quantity neither the quality of such experiments, but with data processing. Therefore, there is a great interest in the development of methodologies in bioinformatics that allow the integrated analysis of such profiles, including data set from other experiments.

Cancer development has been associated mostly with changes in cell cycle control, whose intracellular pathways are dependent on the transcriptional machinery. Regulatory elements involved in transcription, which include transcriptional factors (TFs), can be potential targets for molecular therapy.

In the present study, a set of microarray data achieved from glioblastoma samples were obtained from public repositories (*GEO* and *ArrayExpress*). Data were submitted to statistical analysis using SAM, and 1,830 up-regulated genes ($FDR \leq 0.05$) were submitted to TFs association analysis (FatiGO+ program). Those genes were associated to the TFs *HNF1A*, *IRF7* and *E2F* ($p \leq 0.05$). Moreover, the set of genes was submitted to a transcription signature bank data association analysis using the *software TBrowser*, extracted from *GEO*, which provided a wide range of data sets from microarray experiments. The analysis led to 7,910 associated signatures ($p \leq 0.01$); out of them, the first 100 were related to TFs according to the tools available in *TBrowser*. The TFs *E2F1* and *E2F4* were associated to more than 80% from the analyzed signatures. Therefore, both analysis suggested that the TF *E2F*, specifically *E2F1* and *E2F4*, were associated to the initial gene set obtained by the SAM analysis. The expression of *E2F1* and *E2F4* was evaluated by quantitative real-time PCR using RNA samples from seven glioblastoma cell lines (T98, U251, U138, U87, U343, MO59J and MO59K). Interestingly, both genes were found up-regulated in all cell lines.

The *E2F* TFs family members present important roles in cell proliferation control and apoptosis. Depending on the molecular context involved, some members act as oncogenes and others as tumor suppressor genes. A great number of studies has suggested the importance of such TFs, specially *E2F1*, in the tumorigenesis process. According to the present results, the expression status (induced) of *E2F1* and *E2F4* may be associated with glioblastoma. These TFs may influence the abnormal expression of important genes in cancer, even though

detailed studies should be necessary in order to validate these TFs as biomarkers or molecular targets for therapeutical intervention.

1. INTRODUÇÃO

1.1 Microarranjos de cDNA e sua aplicação no estudo do câncer

Alterações nos perfis de expressão dos genes podem gerar respostas celulares não convencionais, as quais são passíveis de desencadear doenças. A metodologia de microarranjos tornou possível o estudo de expressão gênica em larga escala de maneira eficaz, permitindo analisar simultaneamente os perfis de expressão de milhares de genes em uma situação biológica de interesse. O emprego dessa metodologia no estudo de doenças, por exemplo, tem permitido a identificação de genes com alterações em seus perfis de expressão, os quais estão direta ou indiretamente envolvidos na etiologia dessas enfermidades (De Magalhães *et al.*, 2009).

Em aproximadamente duas décadas de uso da técnica de microarranjos de cDNA, centenas de experimentos têm sido realizados, gerando perfis globais e quantitativos de expressão gênica em diversas situações experimentais (Dudley *et al.*, 2009), sendo que muitos trabalhos investigam doenças como o câncer (Dopazo *et al.*, 2001; Rhodes e Chinnaiyan, 2005; Rhodes *et al.*, 2007).

O câncer é uma das principais causas de morte em todo o mundo. Estima-se que aproximadamente 7,4 milhões de pessoas morrem por ano vítimas de algum tipo de câncer, perfazendo 13% das mortes em todo o mundo (Rosser e Goodison, 2010). Apesar de extensa pesquisa no campo da oncologia e do desenvolvimento de algumas estratégias bem sucedidas na prevenção, diagnóstico e tratamento, as causas e desenvolvimento do câncer ainda não são completamente conhecidos. As perspectivas são de aumento no número de mortes pela doença, projetadas para 9,2 milhões em 2015 e 12 milhões em 2030 (Calvert e Frucht, 2002). Dessa maneira, tratamentos mais eficazes, assim como programas de prevenção e inovações nas metodologias de diagnóstico devem ser desenvolvidos para que os números reais de mortes devidas ao câncer, no futuro, subvertam as atuais perspectivas.

O desenvolvimento do câncer tem sido principalmente associado a distúrbios no controle do ciclo celular, caracterizando-se como uma instabilidade genômica gerada por falhas nesses mecanismos (Hartwell e Kastan, 1994; Kinzler e Vogelstein, 1997; Bartek *et al.*, 1999; Rosser e Goodison, 2010). O papel dos defeitos na maquinaria do ciclo celular no desenvolvimento e progressão de neoplasias malignas tornou-se cada vez mais evidente, segundo uma extensa literatura (Hall e Peters, 1996; Carnero, 2002; Hahn e Weinberg, 2002; Park e Lee, 2003; Tessema *et al.*, 2004), mas as pesquisas desenvolvidas em escala genômica têm demonstrado

que muitas outras classes de genes são também alteradas, tais como: controle metabólico, resposta ao estresse, reparo do DNA, adesão, apoptose, controle da transcrição e tradução, dentre outras (Bassi *et al.*, 2008).

Por meio de experimentos de microarranjos, tipos e subtipos de câncer têm sido distinguidos com base em seus padrões de expressão gênica. Além disso, novos subtipos moleculares de câncer têm sido descobertos associando-se as propriedades tumorais ao paciente, indicando, por exemplo, propensão à metástase, e sensibilidade ou resistência a terapias específicas (Alizadeh *et al.*, 2000; Alizadeh *et al.*, 2001; Sorlie *et al.*, 2003; Sotiriou *et al.*, 2003; Rhodes e Chinnaiyan, 2005; Goodison *et al.*, 2010). Assim, a técnica de microarranjos tem auxiliado, mesmo que de forma indireta, na descoberta de novas modalidades de tratamento.

Estudos clínicos evidenciam a importância dos perfis de expressão gênica obtidos nos experimentos de microarranjos, demonstrando que esses perfis podem influenciar a tomada de decisão na prática clínica. Essa metodologia já é utilizada em alguns centros de pesquisa em pacientes com câncer de mama e linfoma (Van De Vijver *et al.*, 2002; Dave *et al.*, 2004).

Há perspectivas de utilização dos perfis de expressão gênica em larga escala no diagnóstico e prognóstico, bem como na definição de estratégias de tratamento mais eficazes. Mas antes que tais aplicações se tornem realidade, faz-se necessária a realização de validações cuidadosas dessas assinaturas de expressão gênica (Michiels *et al.*, 2005).

1.2 Glioblastoma: necessidade de novas terapias mais eficazes

A incidência de alguns tumores, como os tumores cerebrais primários, aumentou drasticamente ao longo das últimas décadas (Brandes, 2003). Esses tumores cerebrais (excluindo meningiomas) constituem aproximadamente 2% de todos os tumores primários (Reifenberger e Collins, 2004). Em todo o mundo, aproximadamente 176.000 casos de câncer de sistema nervoso central (SNC) são diagnosticados por ano, com uma mortalidade anual de 128.000 pessoas (Parkin *et al.*, 2001).

Os gliomas malignos são os tumores primários mais comuns do SNC em adultos (Grossman e Batará, 2004). Esses tumores são nomeados de acordo com o tipo celular com o qual mais se assemelham e que provavelmente os originou. Os principais subtipos histológicos de glioma são astrocitoma, oligodendroglioma e ependimoma. Astrocitomas são classificados em graus de I a IV, de acordo com a WHO – *World Health Organization* (Louis *et al.*, 2007): os graus I e II constituem os astrocitomas de crescimento lento, o grau III

corresponde a astrocitomas anaplásicos e o grau IV consiste no glioblastoma multiforme, o mais comum (65%) e maligno dentre os tumores cerebrais (Bai *et al.*, 2011).

O tratamento de pacientes com glioblastoma envolve terapias agressivas, consistindo na combinação de cirurgia, radioterapia e quimioterapia adjuvante paralela com temozolomida, um agente metilante que atua no DNA formando aductos (Stupp *et al.*, 2001). Entretanto, tais terapias agressivas não garantem a cura e estão associadas à significativa morbidade (Ramos *et al.*, 2010).

Outras modalidades de tratamento também vêm sendo testadas recentemente no tratamento do glioblastoma, como o uso de adenovírus recombinante na transfecção de genes que podem ter função terapêutica (Dent *et al.*, 2008), inibição de genes e vias específicas, como fatores de crescimento, seus receptores e seus efetores (Sathornsumetee *et al.*, 2007), inibição de fatores de transcrição (Mees *et al.*, 2009) e inibição de genes de reparo do DNA (Belzile *et al.*, 2006; Madhusudan e Middleton, 2005).

Apesar do avanço ocorrido na área clínica, a sobrevida média dos pacientes com essa doença é de aproximadamente um ano a partir do diagnóstico (Avgeropoulos e Batchelor, 1999; Aoki *et al.*, 2007).

Um dos maiores desafios no estudo de tumores cerebrais é identificar biomarcadores e/ou assinaturas de expressão gênica que complementem, especifiquem e aumentem a sensibilidade e especificidade das atuais análises clinicopatológicas (Ziobar *et al.*, 2010). A busca por novas estratégias de tratamento, como a terapia molecular, pode ser bastante interessante e promissora, visto que os tratamentos atuais geralmente são agressivos para os pacientes e, em muitos casos, há o problema de resistência aos agentes antitumorais utilizados (Rosser e Goodison, 2010; Bai *et al.*, 2011). Dessa forma, pesquisas focadas no desenvolvimento de novas drogas, assim como de novos alvos e abordagens são necessárias (Reardon e Wen, 2006).

1.3 Bancos de dados de microarranjos e a necessidade de análises integrativas

Muitos repositórios públicos de dados de microarranjos surgiram como consequência da grande quantidade de experimentos realizados. Além disso, a maior parte das revistas científicas passou a requerer a disponibilização dos dados de microarranjos em bases públicas (Anonymous, 2002), tornando exponencial o crescimento da quantidade desses dados em repositórios internacionais (Parkinson *et al.*, 2009).

Nesse contexto, duas bases de dados se destacaram (Anonymous, 2002): o *GEO* (*Gene Expression Omnibus* – (Barrett *et al.*, 2005; Barrett *et al.*, 2011), base de dados do NCBI (*National Center for Biotechnology Information*), e o *ArrayExpress* (Brazma *et al.*, 2003; Parkinson *et al.*, 2005; Parkinson *et al.*, 2009; Parkinson *et al.*, 2011), base de dados do EBI (*European Bioinformatics Institute*). Ambas utilizam o padrão de anotação MIAME (*Minimum Information About a Microarray Experiment*) (Brazma *et al.*, 2001), o que contribuiu para o seu sucesso no meio científico. A padronização MIAME descreve quais informações de um experimento de microarranjos precisam ser publicadas para assegurar que a interpretação dos resultados não seja ambígua. A adoção dessas orientações para publicação no meio científico tem garantido a reprodutibilidade das análises e dos resultados encontrados pelos pesquisadores que publicaram os dados, além de torná-los úteis para novas análises (Brazma, 2009).

Usando-se as palavras “câncer” e “humano” como chaves de busca em bancos de dados de experimentos de microarranjos, listas imensas de resultados são geradas. Até 2005, por exemplo, na base de dados de publicações da *Affymetrix* havia 646 artigos relacionados com câncer, sendo 453 publicados apenas entre 2004 e 2005. No banco de dados *GEO*, já havia 84 conjuntos de dados de estudos em câncer. Já a base de dados *Oncomine* (Rhodes *et al.*, 2004), específica de microarranjos de câncer, possuía 114 conjuntos de dados, totalizando mais de 8.000 experimentos (Rhodes e Chinnaian, 2005).

Atualmente, a base de dados de publicações da *Affymetrix* conta com 23.067 publicações. O *GEO* possui 2.721 conjuntos de dados, totalizando 541.888 amostras e o *Oncomine* totaliza 534 conjuntos de dados e 43.021 amostras. O *ArrayExpress* conta com 20.982 experimentos. Essas estatísticas demonstram o crescente número de experimentos realizados ao longo dos últimos anos. Normalmente, tais experimentos são analisados individualmente e publicados nas bases de dados, sem que haja uma preocupação em realizar análises expandidas, nas quais vários conjuntos de dados sejam englobados (Moreau *et al.*, 2003).

Considerando a enorme quantidade de experimentos de microarranjos realizados e de informações geradas, as análises integradas, as quais avaliam esses resultados de transcriptoma do câncer no contexto de outras fontes de dados são, em geral, pouco abordadas (Golub *et al.*, 1999; Cui *et al.*, 2011). A *análise integrativa* apresenta-se como uma nova forma de análise, cujo objetivo é examinar os perfis de expressão gênica de um experimento (como microarranjos em câncer), adicionando ao contexto da análise dados oriundos de outros experimentos. Esses dados podem ser provenientes de experimentos nos quais as mesmas técnicas foram aplicadas, porém implementadas por outros grupos de pesquisa, ou podem

representar dados genômicos obtidos por meio de outras técnicas (Rhodes e Chinnaiyan, 2005). Dessa forma, as análises integrativas possibilitam a geração de conclusões mais “profundas” e menos “óbvias” a partir dos dados gerados.

Abordagens integrativas são capazes de simplificar perfis de expressão gênica complexos de tumores, transformando uma assinatura de câncer unidimensional em redes de interação multidimensional e extraíndo mecanismos regulatórios intrínsecos a partir dos níveis de expressão dos genes. Essas informações intrínsecas normalmente não são detectadas nas análises convencionais, como análise de expressão diferencial baseada em estatística e aplicação de algoritmos de agrupamento hierárquico, por exemplo (Tomlins *et al.*, 2006).

Atualmente, existem várias abordagens analíticas e computacionalmente integrativas: meta-análise, análise de enriquecimento funcional, análise de interactoma, análise de redes transcricionais e análise integrativa de sistemas modelo, dentre outras (Segal *et al.*, 2005).

No contexto do estudo do câncer, as abordagens integrativas podem ser aplicadas de diversas formas: métodos de meta-análise propõem integração de dados na busca de assinaturas de câncer; análises de melhoramento buscam identificar módulos gênicos funcionais regulados coordenadamente; redes de interação de proteínas procuram detectar complexos de interação desregulados em câncer; redes transcricionais visam inferir mecanismos regulatórios em câncer; análises de perfis de expressão em sistemas modelo integram dados obtidos *in vitro*, como perfis de expressão de tumores humanos, com experimentos *in vivo* para inferir vias oncogênicas (Rhodes e Chinnaiyan, 2005).

Com o crescimento exponencial dos dados em escala genômica, abordagens de bioinformática integrativa serão decisivas para a inferência de elementos regulatórios e de redes de interação entre genes e células, possibilitando melhor entendimento das disfunções gênicas que levam ao desenvolvimento de doenças, incluindo-se o câncer.

Não obstante, por causa da heterogeneidade existente entre os conjuntos de dados, já que há diferentes plataformas, inúmeros protocolos e delineamentos experimentais distintos, análises integrativas tornam-se complicadas no campo dos dados de microarranjos (Warnat *et al.*, 2005). Assim, diferentes conjuntos de dados nem sempre são diretamente comparáveis (Kuo *et al.*, 2002; Irizarry *et al.*, 2005), fazendo-se estritamente necessária a aplicação de métodos estatísticos para normalizar os dados (Jiang *et al.*, 2004). Dessa forma, é importante definir padrões para tornar possível a integração entre diversos conjuntos de dados de microarranjos. Atualmente, não há consenso ou diretrizes claras que delimitem a melhor maneira de realizar a transformação dos dados de modo a torná-los diretamente comparáveis (Fishel *et al.*, 2007).

1.4 Alvos transcricionais e “assinaturas de câncer”

Muitas “assinaturas de câncer” (conjunto de genes diferencialmente expressos na doença) são obtidas em condições específicas, retratando apenas a expressão gênica instantânea. Isso dificulta o entendimento e a diferenciação entre a relação de causa e efeito das assinaturas de expressão gênica. Por exemplo, *gene A* ativa *gene B*, *gene B* ativa *gene A* ou ambos os *genes A* e *B* são ativados pelo *gene C*?

Com o intuito de auxiliar no estabelecimento de relações de causa e efeito, algumas abordagens integrativas preocupam-se em inferir causalidade e definir vias direcionais ativadas em assinaturas de câncer. Um exemplo dessas abordagens é a associação das assinaturas de expressão a sítios de ligação de fatores de transcrição (inferência de alvos *downstream*) (Elkon *et al.*, 2003).

A identificação de genes regulados pelos mesmos fatores de transcrição é muito promissora no estudo de doenças relacionadas ao ciclo celular, como o câncer, pois permite gerar estratégias mais eficazes para o tratamento do câncer. Um exemplo seria o uso desses fatores de transcrição como alvos de drogas (Mees *et al.*, 2009).

Dessa forma, abordagens de inferência de causalidade permitem reduzir assinaturas de câncer complexas em um pequeno número de elementos transcricionais ativados, auxiliando na compreensão dos mecanismos de desenvolvimento neoplásico. Além disso, essas abordagens podem fornecer informações valiosas sobre biomarcadores e potenciais alvos para intervenção terapêutica (Rhodes e Chinnaiyan, 2004).

1.5 Fatores de transcrição e sua importância na regulação gênica

Fatores de transcrição (FTs) são elementos endógenos, usualmente proteínas, os quais são requeridos na iniciação, estimulação ou terminação do processo de transcrição gênica. Essas proteínas apresentam domínios de ligação ao DNA, por meio dos quais interagem com essa molécula, promovendo sua regulação. A *RNA polimerase II* (enzima que realiza a transcrição) requer FTs para iniciar o processo transcricional do DNA nos eucariotos. Os FTs ajudam a posicionar a *RNA polimerase* corretamente na região promotora, auxiliando na separação da dupla fita do DNA para permitir o início da transcrição (Alberts *et al.*, 2002).

Em vários organismos estudados, os genomas contêm uma grande quantidade de DNA cis-regulatório, cuja função é direcionar padrões de expressão gênica em função do tempo e do espaço (por exemplo, em tecidos), como resposta a estímulos metabólicos ou externos (Lemon e Tjian, 2000). Dessa forma, FTs e elementos cis-regulatórios correspondentes,

também chamados de motivos de ligação de FTs, são considerados componentes chave na regulação da expressão gênica (Choi *et al.*, 2009).

Genes de organismos multicelulares contêm módulos proximais (como os elementos promotor-proximal) e distais (*enhancers* e silenciadores), assim como um aparato basal da transcrição (por exemplo, *TATA box* e outros elementos responsáveis pelo posicionamento da *RNA polimerase II*). Os módulos distais podem se situar a milhares de bases de uma região codificadora, assim como posicionados em regiões intrônicas. Cada módulo pode conter múltiplos sítios de ligação que interagem com uma combinação específica de FTs (Davidson, 2001). A caracterização de tais regiões é um passo fundamental para a compreensão de vias de regulação gênica ainda pouco exploradas ou conhecidas.

A maquinaria transcricional, assim como seus elementos transcricionais regulatórios são potenciais alvos para a terapia molecular. O aumento do conhecimento sobre mecanismos de regulação gênica e redes transcricionais pode levar a um melhor entendimento das funções críticas dos FTs. O recente desenvolvimento dos mecanismos de transfecção (lipossomos e nanopartículas), assim como o uso de moléculas efetoras mais precisas e potentes (shRNA, siRNA, ribozimas, oligonucleotídeos antisense, pequenas moléculas inibidoras e “*zinc fingers*”) podem ser utilizadas futuramente na prática clínica, por meio da inibição de FTs alvo (Mees *et al.*, 2009).

1.6 Potenciais alvos moleculares e a necessidade de validação

Na área do câncer, a maioria dos estudos de expressão gênica tem identificado grande quantidade de potenciais marcadores para a doença, tornando-se necessária a realização de validações experimentais independentes antes que um gene, ou um conjunto de genes, seja considerado como um biomarcador válido.

A validação é estritamente necessária, pois a metodologia de microarranjos pode gerar uma parcela de resultados falso positivos devido a uma série de razões, entre as quais incluem-se: viés amostral, acaso, problemas no processo de hibridação e de análise dos resultados finais. Técnicas como PCR quantitativa em tempo real (qRT-PCR em tempo real), *Northern blot* ou microarranjos de tecidos são indicadas para validar os resultados obtidos em experimentos de microarranjo. A técnica de qRT-PCR em tempo real tem se consolidado como uma das mais poderosas ferramentas com esse propósito, graças à sua sensibilidade na detecção e quantificação da amplificação do gene de interesse. Entretanto, algumas padronizações devem

ser realizadas, dentre essas, a escolha de genes de referência adequados, visando garantir a qualidade e a confiabilidade dos resultados obtidos (Valente *et al.*, 2009).

2. OBJETIVOS

2.1 Objetivo geral

O objetivo geral deste trabalho foi realizar análises *in silico* a partir de um conjunto de dados de glioblastoma depositado em bases de dados públicas de experimentos de microarranjos com o intuito de buscar genes diferencialmente expressos que pudessem ser associados a assinaturas transcricionais e a fatores de transcrição (FTs). Além disso, foi também proposto aplicar técnicas laboratoriais *in vitro* para validação dos FTs resultantes das análises *in silico*.

2.2 Objetivos específicos

- ✓ Realizar análise estatística (SAM) em um conjunto de dados de glioblastoma obtido em bases de dados públicas de experimentos de microarranjos (*GEO* e *ArrayExpress*);
- ✓ Associar os genes diferencialmente expressos obtidos na análise pelo SAM ($FDR \leq 0,05$) a FTs por meio da ferramenta FatiGO+;
- ✓ Realizar análise integrativa por meio de comparação entre os genes diferencialmente expressos (análise estatística) e um banco de dados de assinaturas transcricionais do *software TranscriptomeBrowser (TBrowser)*;
- ✓ Associar as assinaturas transcricionais a FTs por meio do *software TBrowser*;
- ✓ Utilizar a técnica de qRT-PCR em tempo real para quantificar a expressão dos principais FTs obtidos nas análises *in silico* em linhagens de glioblastoma cultivadas;
 - Avaliar a expressão dos FTs em sete linhagens de glioblastoma (T98, U251, U138, U87, U343, MO59J e MO59K) não submetidas a qualquer tipo de tratamento;
 - Avaliar a expressão dos FTs na linhagem U87 tratada com drogas antitumorais (cisplatina, doxorubicina e temozolomida).

3. MATERIAL E MÉTODOS

3.1 Análise *in silico* de microarranjos de cDNA

3.1.1 Conjunto de dados de microarranjos de cDNA

Um conjunto de dados de expressão gênica foi utilizado como ponto de partida para a obtenção de uma lista de genes diferencialmente expressos (assinatura de expressão). O conjunto de dados foi obtido em bases públicas de experimentos de microarranjos, no *GEO* e no *ArrayExpress*.

Este se refere aos resultados de um trabalho publicado com amostras de glioma humano (Sun *et al.*, 2006). Nesse trabalho, foi utilizada a plataforma *GeneChip® Human Genome U133 Plus 2.0 Arrays* (Affymetrix, Inc., Santa Clara, USA), compreendendo a expressão de todo o genoma humano (total de 54.675 transcritos). As lâminas foram hibridadas com RNA de amostras de glioma (graus II, III e IV, incluindo oligodendroglioma, astrocitoma e glioblastoma) e amostras de tecido cerebral não tumoral de pacientes com epilepsia (representando os controles experimentais). As amostras de tecidos foram retiradas de pacientes não submetidos a qualquer tipo de tratamento, sendo congeladas e utilizadas para a extração de RNA.

Nesse trabalho, foram analisadas 178 amostras, sendo 50 amostras de oligodendroglioma, 25 de astrocitoma, 81 de glioblastoma e 23 amostras de pacientes com epilepsia. Apenas os dados referentes às amostras de glioblastoma e epilepsia foram utilizados na presente análise (104 amostras), por se tratarem dos dados de interesse para esse trabalho.

Os identificadores de acesso para os dados no *GEO* e no *ArrayExpress* são *GSE4290* e *E-GEOD-4290*, respectivamente. Os identificadores correspondem ao mesmo conjunto de dados, porém nas diferentes bases de dados.

3.1.2 Forma de análise dos dados de microarranjos de cDNA

Os dados gerados pelo método de microarranjos de cDNA são analisados por meio de diversas ferramentas de bioinformática. A série sequencial de análise envolve basicamente duas etapas: o pré-processamento dos dados (quantificação das imagens das lâminas dos microarranjos, filtragem dos pontos perdidos e normalização) e a análise da matriz de expressão gênica gerada na etapa de pré-processamento. Os dados de microarranjos obtidos

nas bases de dados já estavam pré-processados. O arquivo de texto tabulado contendo genes e seus respectivos valores de expressão foi submetido à análise estatística, sendo que os genes diferencialmente expressos resultantes foram agrupados por meio de algoritmos de agrupamento hierárquico.

3.1.2.1 Análise estatística e agrupamentos gênicos

O aplicativo *TIGR Multiexperiment Viewer* (MeV) (Saeed *et al.*, 2003) – versão 4.5.1 – foi utilizado para realizar os testes estatísticos e a análise de agrupamento gênico. O algoritmo SAM (*Significance Analysis of Microarrays*) (Tusher *et al.*, 2001) foi aplicado aos dados para a análise estatística. O SAM pode ser usado para classificar genes significativamente modulados baseado na expressão diferencial entre conjuntos de amostras. Uma característica valiosa do SAM é que este estima a taxa de falsas descobertas (FDR – *False Discovery Rate*) (Benjamini, 1995), que corresponde à proporção de genes possivelmente identificados erroneamente como diferencialmente expressos. Além disso, o SAM é um algoritmo muito interativo, visto que permite ao usuário visualizar a distribuição do teste estatístico e então determinar limites de significância (pelo ajuste do parâmetro *delta*) após a análise da distribuição.

O arquivo de texto tabulado contendo a matriz de valores de expressão foi carregado no aplicativo MeV. Alguns ajustes foram realizados nos dados por meio de ferramentas disponibilizadas no próprio aplicativo. Esses ajustes modificam a matriz de expressão, a base de toda a análise.

Um dos ajustes aplicados foi o *filtro de corte pela porcentagem de valores válidos*. Estes são considerados válidos quando são maiores que zero e diferentes de “NA” (valor inexistente). A subtração do *background* e a filtragem, realizadas no pré-processamento, podem levar à perda de valores de expressão na matriz de expressão gênica. A aplicação do filtro é essencial para a qualidade dos resultados da análise, pois mantém apenas as linhas (genes) da matriz que apresentam uma quantidade de valores válidos (maiores que zero e diferentes de NA) acima da porcentagem. Essa porcentagem é definida pelo usuário. O valor de porcentagem foi estabelecido em 80%, ou seja, genes com menos de 80% de valores válidos foram ignorados.

Outro ajuste realizado foi o *ajuste pela mediana dos genes (linhas)*, que substituiu cada valor de expressão de um gene pelo próprio valor subtraído da mediana da linha ao qual esse valor pertencia (ou seja, a mediana foi calculada para os valores de expressão de um mesmo gene, porém em diferentes amostras).

O teste SAM foi aplicado para duas classes não pareadas (amostras de glioblastoma e de cérebro de pacientes com epilepsia), com número de permutações ajustado para mil. Os genes escolhidos foram aqueles com valor de FDR igual ou inferior a 0,05.

Após a análise estatística, a lista de genes diferencialmente expressos foi submetida à análise de agrupamento hierárquico (HCL), a qual agrupa os genes com base nas similaridades de seus níveis de expressão. O agrupamento foi realizado tanto por genes (linhas) quanto por amostras (colunas da matriz de expressão), o que caracteriza o *biclustering*. A distância métrica selecionada foi a Pearson não-centrada (*Pearson uncentered*); o método de agrupamento utilizado foi o de ligação média (*average linkage*).

3.1.2.2 Análise de fatores de transcrição

Adicionalmente, foi realizado um estudo de fatores de transcrição (FTs) relacionados aos genes significativamente modulados segundo a análise pelo SAM. A escolha dos melhores agrupamentos para o estudo de FTs foi realizada por meio de inspeção visual.

Primeiramente, foi necessário observar se havia um bom agrupamento das amostras (colunas), ou seja, se havia uma boa separação entre as amostras controles e as amostras de glioblastoma. Como houve uma boa separação visual entre as amostras de glioblastoma e os controles, os agrupamentos gênicos (linhas) foram escolhidos quando havia uma diferença de expressão (indução – representada pela cor vermelha, repressão – representada pela cor verde) entre os grupos amostrais. Houve basicamente dois conjuntos de agrupamentos, um compreendendo os genes induzidos em glioblastoma e reprimidos nos controles (epilepsia) e outro representando a situação inversa (repressão em glioblastoma e indução nos controles). Foi escolhido o grupo de genes induzidos em glioblastoma. Dessa forma, os genes presentes no agrupamento escolhido foram submetidos ao programa de análise de FTs.

O programa utilizado para a análise de FTs foi o FatiGO+ (Al-Shahrour *et al.*, 2005), um aplicativo baseado na web capaz de associar fatores de transcrição comuns a um conjunto de genes usados como parâmetros para o programa. Esse programa utiliza a base de dados *TRANSFAC* (Wingender *et al.*, 2000) de fatores de transcrição, incluindo seus respectivos sítios de ligação e genes regulados. Os sítios de ligação desses FTs podem ser estudados em diferentes localizações dos promotores (nas primeiras 1 kb, em 5 kb ou acima de 10 kb).

O FatiGO+ requer dois parâmetros de entrada: o primeiro é a lista de genes presentes em cada agrupamento; o segundo parâmetro requerido é a lista de todos os genes utilizados no experimento de microarranjos do qual foi selecionado o grupo gênico. No caso do

agrupamento gênico escolhido para esta análise, foi utilizado como segundo parâmetro a lista dos 54.675 transcritos do conjunto de dados original, sem qualquer tipo de análise estatística.

A finalidade da utilização desses dois parâmetros é verificar se há diferenças significativas quanto às características funcionais dos genes selecionados em uma das listas relativas à outra lista. Essa operação representa a comparação dos genes modulados estatisticamente (SAM) em relação aos genes de referência (usualmente todos os genes envolvidos no experimento) (Al-Shahrour *et al.*, 2007).

Os valores p obtidos na análise de elementos regulatórios foram ajustados pelo programa utilizando o teste exato de Fisher bicaudal. Os FTs escolhidos como estatisticamente significativos foram aqueles cujo valor p estava abaixo de 0,05. Os genes foram submetidos ao FatiGO+, sendo utilizado o identificador *Ensembl* (<http://www.ensembl.org/index.html>) dos genes. A distância gênica escolhida para a análise dos FTs foi de 5 kb (5000 bases).

Após a escolha dos FTs relacionados aos genes modulados estatisticamente (SAM), uma busca foi realizada no *PubMed* (<http://www.ncbi.nlm.nih.gov/sites/entrez/>) utilizando o nome de cada FT juntamente com a palavra “câncer”. Essa busca foi realizada com o intuito de verificar se os FTs escolhidos foram anteriormente associados a algum tipo de câncer na literatura.

3.1.2.3 Análise de associação a assinaturas de expressão gênica

A lista de genes diferencialmente expressos, os quais foram induzidos em glioblastoma, resultado da análise pelo SAM, foi utilizada como parâmetro de entrada de uma análise comparativa entre diferentes assinaturas de expressão.

Por meio do teste exato de Fisher, a lista de genes foi comparada a um banco de dados de assinaturas de expressão gênica. Essas assinaturas transcricionais foram extraídas em análises de diversos conjuntos de dados de microarranjos recuperados a partir da base de dados *GEO*, utilizando experimentos realizados em diferentes plataformas da empresa *Affymetrix*. Tais assinaturas de expressão são descritas por Lopez *et al.* (2008) como parte dos resultados do desenvolvimento do *software TranscriptomeBrowser (TBrowser)*. A descrição da metodologia para obtenção das assinaturas é apresentada na **Figura 3.1**.

Figura 3.1: Pipeline utilizado na obtenção de assinaturas transcricionais a partir de dados de microarranjos armazenados no GEO. As assinaturas fazem parte da base de dados do *software TBrowser*, por meio do qual é possível realizar análises de enriquecimento funcional. Adaptado de Lopez *et al.* (2008).

As assinaturas transcricionais da base de dados do *TBrowser* foram associadas à lista de genes diferencialmente expressos induzidos em glioblastoma, gerando um valor p . As assinaturas com $p \leq 0,05$ foram separadas para análise de enriquecimento funcional no *TBrowser*. Nessa análise, as cem primeiras assinaturas foram anotadas utilizando o *CisRed* (Robertson *et al.*, 2006), base de dados de elementos regulatórios conservados, e o *ECRbase* (Loots e Ovcharenko, 2007), base de dados de regiões conservadas, promotores e sítios de ligação de FTs. Essas bases de dados fazem parte das ferramentas disponíveis no *TBrowser* para anotação das assinaturas. Além disso, foi utilizado o *TFBSconserved*, um *plugin* do *TBrowser* que relaciona cada assinatura de expressão a sítios de ligação de FTs. Dentre essas cem assinaturas analisadas, também foi verificada a plataforma *Affymetrix* utilizada nos experimentos que geraram cada assinatura, bem como a situação biológica abordada nesses experimentos.

A análise (mais especificamente a associação da lista de genes inicial com o banco de dados de assinaturas do *TBrowser*) foi realizada sob a supervisão do Dr. Denis Puthier

(Inserm U928, TAGC, *Parc Scientifique de Luminy* e ESIL, *Université de Provence et de la Méditerranée*, Marseille, França), um dos responsáveis pelo desenvolvimento do *software TBrowser*.

3.1.2.4 Análise de vias de sinalização celular

A lista de 1.830 genes induzidos em glioblastoma também foi utilizada para a análise de vias de sinalização celular. Nessa etapa, foi utilizada a ferramenta de anotação funcional *DAVID* (Dennis *et al.*, 2003; Huang *et al.*, 2009), disponível no endereço <http://david.abcc.ncifcrf.gov/>.

Essa ferramenta realiza a associação de uma dada lista de genes a vias celulares, apontando quais genes da lista inicial estão presentes em determinadas vias. Um valor *p* é gerado para cada via de sinalização associada. Foram escolhidas as vias obtidas a partir da base de dados do *KEGG* (Kanehisa *et al.*, 2010).

3.2 Linhagens celulares

Sete linhagens celulares de glioblastoma foram utilizadas:

- ✓ Linhagens T98G (T98), U251MG (U251) e U138MG (U138): linhagens de glioblastoma mutantes para os genes *TP53*, *p16*, *p14ARF* e *PTEN*
- ✓ Linhagens U343MG-a (U343) e U87MG (U87): linhagens de glioblastoma mutantes para os genes *p16*, *p14ARF* e *PTEN*, mas proficientes para *TP53*
- ✓ Linhagem MO59J: linhagem de glioblastoma mutante para o gene *DNA-PK*
- ✓ Linhagem MO59K: linhagem de glioblastoma proficiente para o gene *DNA-PK*

A linhagem U343 foi gentilmente cedida pelo Prof. James T. Rutka (*The Arthur and Sonia Labatt Brain Tumour Research Centre*, Canadá) e repassada pelo Prof. Dr. Carlos Gilberto Carlotti Júnior (Departamento de Cirurgia e Anatomia do Hospital das Clínicas – FMRP – USP), enquanto que a linhagem T98G foi gentilmente cedida pela Prof. Dra. Mari Cleide Sogayar (Departamento de Bioquímica – Instituto de Química – USP, SP). As linhagens U87, U251 e U138 foram gentilmente cedidas pelo Prof. Dr. Guido Lenz (Departamento de Biofísica – UFRGS, RS). As linhagens MO58J e MO59K foram gentilmente cedidas pelo Dr. Adayabalam S. Balajee (*Center for Radiological Research – Columbia University Medical Center*, New York, USA).

3.3 Cultivo celular

As linhagens celulares foram mantidas em nitrogênio líquido (-195°C) em alíquotas de 1×10^6 células/ml em solução de congelamento (90% soro bovino fetal e 10% glicerol). Para a realização dos experimentos, as células foram descongeladas e cultivadas em monocamada, em frascos de cultura de 25 cm^2 (TPP, Trasadigen, Suíça) com 10 ml de meio de cultura (DEM + F10 – Sigma, St Louis, USA) suplementado com 10% de soro bovino fetal (Cultilab, Campinas, Brasil) e incubadas em estufa a 37°C e 5% de CO_2 , até atingirem o estado de confluência ($\sim 5 \times 10^6$ células), no qual necessitam de subcultivo. Foram cultivadas três alíquotas de cada linhagem, provenientes de descongelamentos diferentes (triplicata biológica), as quais foram semeadas (um milhão de células) e incubadas por 48 h, sendo posteriormente submetidas à extração de RNA. Todos os experimentos foram realizados com as células cultivadas entre a quarta e a décima passagem após seu descongelamento.

3.4 Extração de RNA total

Para a extração de RNA total de todas as células cultivadas foi utilizado o *kit illustra RNAspin Mini RNA Isolation* (GE Healthcare, Buckinghamshire, UK) conforme instruções do fabricante. Após a extração, as amostras de RNA foram estocadas a -80°C até a sua utilização.

As amostras foram quantificadas no espectrofotômetro NanoDrop ND1000 (NanoDrop Technologies, Wilmington, USA) nos comprimentos de onda 230nm, 260nm e 280nm. A pureza do RNA foi avaliada pela razão A_{260}/A_{280} , na qual o ideal é obter valores próximos de 2,0 (entre 1,8 e 2,1), demonstrando ausência de contaminação por proteínas e DNA. Também foi avaliada a razão A_{260}/A_{230} , na qual valores inferiores a 1,7 ou superiores a 2,2 demonstram contaminação por compostos orgânicos, tais como o fenol. Os valores obtidos para as razões A_{260}/A_{280} e A_{260}/A_{230} variaram entre 2,02 e 2,11, e entre 1,81 e 2,18, respectivamente, descartando a hipótese de contaminação das amostras de RNA.

A integridade do RNA foi avaliada por meio de eletroforese em gel de agarose (1,5%) com formaldeído, corado com GelRed (Biotium, Hayward, USA) e visualizado por transiluminador UV. A formação de duas bandas (18S e 28S) foi detectada nitidamente para cada amostra, sem a presença de “rastros” entre elas no gel, que são indicativos de degradação do RNA.

3.5 Grupo controle

Como grupo controle das linhagens celulares de glioblastoma foi adquirido um RNA de referência de cérebro humano (*FirstChoice® Human Brain Reference RNA*, número de catálogo AM6050, *Ambion®*, Foster City, USA). Esse RNA caracteriza-se como de alto padrão de qualidade e foi utilizado no projeto MAQC (*MicroArray Quality Control* – (Shi *et al.*, 2006; Shi *et al.*, 2010). O RNA foi obtido a partir de múltiplos doadores (23 pessoas, 13 homens e 10 mulheres, na faixa etária de 23 a 86 anos) e de diversas regiões do cérebro, garantindo uma cobertura imparcial. Embora sua qualidade fosse garantida pelo fabricante, esse RNA também foi submetido aos processos de quantificação no espectrofotômetro e avaliação da integridade por meio de eletroforese em gel, como já descrito acima, antes da realização da RT-PCR.

3.6 Amostras de RNA da linhagem U87 tratada com drogas antitumorais

Amostras de RNA da linhagem U87 tratada com drogas antitumorais, que haviam sido utilizadas em projetos anteriores no laboratório de pesquisa, também foram utilizadas no presente trabalho.

As células da linhagem U87 foram semeadas e cultivadas. Tais células foram tratadas separadamente por 24 h com três diferentes drogas antitumorais:

- ✓ cisplatina (*Sigma-Aldrich*, St Louis, USA)
- ✓ doxorubicina (cloridrato de doxorubicina, medicamento Doxolem® – *Zodiac*)
- ✓ temozolomida (princípio extraído do medicamento Temodal® – *Schering Plough*, Kenilworth, USA)

As drogas foram administradas nas seguintes concentrações: 25 μ M de cisplatina, 0,15 μ g/ml de doxorubicina e 200 μ M de temozolomida. As células foram colhidas nos tempos de 6 e 24 h após os tratamentos. Os controles experimentais caracterizaram-se como a linhagem U87 sem nenhum tratamento, também cultivada e colhida nos tempos de 6 e 24 h. Dois experimentos independentes foram realizados, sendo cada um deles em triplicata. O RNA foi extraído, quantificado, e sua qualidade foi averiguada.

3.7 Transcrição reversa (RT-PCR) para realização da qRT-PCR em tempo real

Inicialmente, utilizou-se o *kit Deoxyribonuclease I, Amplification Grade* (Invitrogen, Carlsbad, USA) para a remoção de possíveis moléculas de DNA nas amostras de RNA. Para cada amostra, utilizou-se 1 µg de RNA total, 1 µL de *10x DNase I Reaction Buffer* e 1 µL de *DNase I Ampgrade* (1U/µL). O volume final foi completado para 10 µL com H₂O DEPC. As amostras foram incubadas por 15 minutos à temperatura ambiente. Em seguida, a DNase foi inativada com adição de 1 µL de *EDTA* (25 mM) e as amostras foram aquecidas por 10 minutos a 65°C, estando, assim, prontas para a reação de RT-PCR.

Na reação de RT-PCR foi utilizado o *kit SuperScript® III Reverse Transcriptase* (Invitrogen, Carlsbad, USA), adicionando-se 1 µL de iniciadores randômicos (*Random Hexamers*, 50 µM – *Applied Biosystems*, Foster City, USA) e 1 µL de *dNTP* (10 mM) aos RNAs provenientes da reação com DNase. As amostras foram incubadas por 5 minutos a 65°C e imediatamente transferidas para o gelo por 2 minutos. Após essa etapa, as amostras foram centrifugadas e acrescentou-se 4 µL de *5x First-Strand Buffer*, 1 µL de *DTT* (0,1 M) e 1 µL de *SuperScript® III RT* (200 U/µL) em cada amostra. As amostras foram divulsionadas e incubadas a 25°C por 5 minutos e a 50°C por 50 minutos. A inativação da reação foi permitida incubando-se as amostras a 70°C por 15 minutos. As amostras foram estocadas a –20°C até sua utilização nos experimentos de RT-PCR quantitativa em tempo real.

A confirmação do sucesso da reação de transcrição reversa foi verificada por meio da realização de uma PCR convencional com as amostras de cDNA utilizando-se os iniciadores do *B2M* (gene de referência). Logo após, foi realizada uma eletroforese em gel de agarose (2%), na qual foi verificada a formação de apenas uma banda na faixa entre 100 e 200 pb (que corresponde ao tamanho do produto de PCR para os iniciadores do *B2M*), confirmando, assim, o sucesso na síntese do cDNA e sua integridade.

3.8 Avaliação de expressão gênica por RT-PCR quantitativa (qRT-PCR) em tempo real

3.8.1 Desenho de iniciadores (*primers*)

As sequências dos iniciadores dos genes utilizados na técnica de PCR quantitativa em tempo real foram obtidas por meio do seguinte protocolo de desenho de iniciadores, desenvolvido no laboratório de pesquisa: inicialmente, a sequência em formato FASTA do mRNA de cada gene (em humanos) foi adquirida no *site* do NCBI (<http://www.ncbi.>

nml.nih.gov/). A sequência de nucleotídeos foi utilizada como entrada para o programa de alinhamento BLAST, que permitiu a realização do alinhamento da sequência inicial com um banco de dados escolhido (*Celera cWGA Scaffolds*), utilizando-se o algoritmo específico para alinhamento de sequências de nucleotídeos (BLASTN). O programa foi utilizado via *web*, no próprio *site* do NCBI (<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606>). Dentre os alinhamentos resultantes, foi escolhido o de menor valor *E* (o valor ideal é 0.0, indicando que o alinhamento entre a sequência de entrada e a região cromossômica descrita no banco de dados não foi meramente ao acaso).

Por meio das coordenadas das sequências *Subject* e *Query*, encontradas na página do alinhamento, foi possível determinar a posição de cada éxon dentro da região cromossômica. Subsequentemente, pode ser estabelecida a ordem dos éxons no gene, além de ter sido verificado se haviam sobreposições de nucleotídeos nas fronteiras dos éxons adjacentes. A sequência da região cromossômica do alinhamento escolhido foi salva em formato FASTA. Essa sequência foi utilizada como entrada no *software Artemis* (versão 11.4.1) (Rutherford *et al.*, 2000), o qual foi útil na tarefa de “acertar” as fronteiras dos éxons, permitindo eliminar manualmente possíveis sobreposições de nucleotídeos. Para tal, as coordenadas do alinhamento foram inseridas no *Artemis*, ou seja, foram inseridos os pares de coordenadas (coordenadas *Subject* do início e fim) de cada éxon do gene.

Depois de tudo anotado, as fronteiras *éxon – íntron* foram ajustadas manualmente de acordo com o esquema da **Figura 3.2**.

Figura 3.2: Nucleotídeos finalizadores e iniciadores característicos das fronteiras *éxon – íntron*, de acordo com a direção da sequência (5' – 3' ou 3' – 5'). Os retângulos acinzentados representam os éxons.

Dois éxons adjacentes com a maior região intrônica foram selecionados. Tais éxons foram concatenados e alinhados com o genoma humano por meio do BLAST (base de dados *genome (all assemblies)*) – <http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606>. Essa etapa permitiu verificar a especificidade dos éxons escolhidos por meio do valor *E* e dos

identificadores das regiões alinhadas (NT, NM, NW, etc), que deveriam corresponder aos identificadores do gene de interesse.

Após essa verificação, a sequência dos éxons concatenados foi submetida ao programa de desenho de iniciadores, o *Primer3*, disponível na web pelo endereço eletrônico <http://frodo.wi.mit.edu/primer3/>. Os parâmetros especificados foram: tamanho do produto entre 100 e 120 bases, tamanho do iniciador entre 18 e 22 bases e temperatura de *melting* entre 58° e 62°. Dentre os iniciadores desenhados pelo programa, foi verificado qual par “*left primer*” e “*right primer*” seguia as seguintes especificações: um dos iniciadores desenhados deveria estar na região de intersecção entre as duas seqüências de éxons, ou, caso não fosse possível obter um dos iniciadores na intersecção, ao menos cada iniciador deveria estar localizado em éxons diferentes. Essa especificação é necessária para assegurar que, em caso de contaminação com DNA genômico, o par de iniciadores não se anele e nem amplifique seqüências genômicas, mas apenas mRNA.

O par “*left primer*” e “*right primer*” que seguiu tais especificações teve sua especificidade (em relação ao gene de interesse) verificada por meio de alinhamento no BLAST (*Basic BLAST – Nucleotide BLAST* no endereço http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome). Foram verificados os valores *E* dos alinhamentos e os identificadores das regiões alinhadas, os quais deviam corresponder aos identificadores do gene de interesse. Além disso, um *site* específico para testar iniciadores (disponibilizado pelo NCBI) foi utilizado (<http://www.ncbi.nlm.nih.gov/projects/e-pcr/reverse.cgi?orgdb=0&taxid=&type=table&unistslist=&margin=50&mism=0&gaps=0>). O par de iniciadores foi comparado a duas bases de dados de humanos, a primeira do *genoma* e a segunda do *transcriptoma*. No primeiro caso, os iniciadores não deveriam se alinhar com nenhuma sequência (Hits: 0), demonstrando que o par não se anelava a seqüências genômicas. Já no segundo caso, os iniciadores deveriam se alinhar somente ao gene de interesse, assegurando sua especificidade (Hits > 0).

Iniciadores dos genes *E2F1* e *E2F4* (FTs preditos nas análises *in silico*) foram desenhados, além dos genes *B2M*, *GUSB*, *HPRT1* e *TBP*, genes de referência utilizados na normalização dos dados de expressão gênica de cada linhagem e do controle (**Tabela 3.1**). Os iniciadores foram confeccionados pela IDT (*Integrated DNA Technologies*, Coralville, USA).

Tabela 3.1: Combinação de iniciadores utilizados para a análise de expressão gênica pelo método de PCR quantitativa em tempo real.

Iniciador	Sequência	Tamanho do produto de RT-PCR (pb)
<i>E2F1 – forward</i>	5' – CACAGATCCCAGCCAGTCTCTA – 3'	119
<i>E2F1 – reverse</i>	5' – GAGAAGTCCTCCCGCACATG – 3'	
<i>E2F4 – forward</i>	5' – CCCCAGAGAGCTGTCAGAAA – 3'	114
<i>E2F4 – reverse</i>	5' – GGTGGAGAAAGACGAAGCAG – 3'	
<i>B2M – forward</i>	5' – AGGCTATCCAGCGTACTCCA – 3'	112
<i>B2M – reverse</i>	5' – TCAATGTCGGATGGATGAAA – 3'	
<i>GUSB – forward</i>	5' – CACCAGGATCCACCTCTGAT – 3'	115
<i>GUSB – reverse</i>	5' – TCCAAATGAGCTCTCCAACC – 3'	
<i>HPRT1 – forward</i>	5' – TCATTATGCTGAGGATTTGGA – 3'	104
<i>HPRT1 – reverse</i>	5' – GATGGCCTCCCATCTCCTT – 3'	
<i>TBP – forward</i>	5' – AGGAGCCAAGAGTGAAGAACAG – 3'	117
<i>TBP – reverse</i>	5' – CTCCCCACCATGTTCTGAAT – 3'	

Dessa forma, dois FTs e quatro genes de referência (**Tabela 3.2**) foram avaliados quanto à expressão gênica transcricional, comparando-se tal expressão nas linhagens de glioblastoma e no controle (RNA de referência de cérebro humano).

Tabela 3.2: Genes estudados para avaliação de expressão transcricional por qRT-PCR em tempo real. Os quatro últimos genes foram testados para a escolha dos genes de referência mais adequados para realizar a normalização dos dados.

Símbolo	Nome
<i>E2F1</i>	E2F transcription factor 1
<i>E2F4</i>	E2F transcription factor 4, p107/p130-binding
<i>B2M</i>	Beta-2-microglobulin
<i>GUSB</i>	Glucuronidase beta
<i>HPRT1</i>	Hypoxanthine phosphoribosyltransferase 1
<i>TBP</i>	TATA box binding protein

3.8.2 Padronização da técnica para os FTs *E2F1* e *E2F4* e para os genes de referência *B2M*, *GUSB*, *HPRT1* e *TBP*

Segundo protocolo do *kit SYBR® Green PCR Master Mix* (Applied Biosystems, Foster City, USA), utilizado nos experimentos de qRT-PCR em tempo real, alguns testes podem ser realizados para padronização dos experimentos, a fim de otimizar os resultados. Dentre esses testes incluem-se os de otimização da concentração de iniciadores e também da concentração

de cDNA. A finalidade da otimização da concentração de iniciadores é determinar a concentração de iniciadores mínima que resulta no menor valor de C_T (*Cycle threshold*), minimizando amplificações não específicas. A concentração de cDNA é otimizada no sentido de padronizar a quantidade de cDNA a ser utilizada nos experimentos, evitando “enviesar” resultados devido ao uso de diferentes concentrações de cDNA.

Dessa forma, com o intuito de otimizar a concentração de iniciadores, várias concentrações foram testadas. As razões de concentração “*primer forward* / *primer reverse*” (em nM) testadas foram as seguintes: 100 / 100, 200 / 200, 300 / 300 e 400 / 400. Também foi utilizado um NTC (*No Template Control*), controle negativo da reação, com razão de concentração extrapolada (800 / 800) para detectar a possível ocorrência de “*primer dimer*” (anelamento dos *primers forward* e *reverse* consigo mesmos). A quantidade de cDNA utilizada em todos esses testes foi fixada em 80 ng/μL.

Para o teste de concentração de cDNA, as variações de quantidade utilizadas (em ng/μL) foram as seguintes: 40, 80, 160 e 320. A razão de concentração de iniciadores foi fixada em 200 / 200 (em nM).

A partir desses testes, foram escolhidas as concentrações padrão de iniciadores e cDNA com base na análise dos valores de C_T , observando-se os menores valores de C_T com as menores variações entre as triplicatas de mesma concentração, dando preferência àquelas em que havia ausência de amplificações não específicas (como *primer dimer*, por exemplo).

Além disso, diferentes diluições de um *pool* de cDNA foram testadas para cada gene estudado, com o intuito de obter a “curva padrão” (método de quantificação absoluta) gerada pelos iniciadores de cada gene. A partir dessa curva obtém-se o valor de *slope*, o qual é utilizado para o cálculo da eficiência (E) dos iniciadores, segundo a equação:

$$E = 10^{-1/slope} - 1 \text{ (Bustin, 2000)}$$

O valor de eficiência E é utilizado posteriormente na análise dos resultados da qRT-PCR em tempo real.

As diluições do *pool* de cDNA utilizadas para construção da “curva padrão” foram as seguintes: 1, 1/5, 1/25, 1/125 e 1/625 (fator de diluição de 1/5).

3.8.3 Montagem da reação

A realização da qRT-PCR em tempo real foi possível por meio da utilização do *kit SYBR® Green PCR Master Mix* (Applied Biosystems, Foster City, USA). As reações foram realizadas em volume final de 12,5 µL. Para esse volume, o *mix* foi preparado de acordo com a seguinte proporção: 6,25 µL de *SYBR Green PCR Master Mix*, 1 µL do *primer forward* e 1 µL do *primer reverse* (ambos com solução estoque de 10 µM), 3,25 µL de água livre de RNase e 1 µL de cDNA das amostras obtido na reação de transcrição reversa. As amostras de cDNA foram previamente quantificadas no espectrofotômetro NanoDrop ND1000 e diluídas para a obtenção de uma concentração final de 80 ng/µL. As razões de concentração “*primer forward* / *primer reverse*” (em nM) utilizadas para cada gene são descritas na **Tabela 3.3**.

Tabela 3.3: Razões de concentração “*primer forward* / *primer reverse*” (em nM) utilizadas nas reações de PCR quantitativa em tempo real.

Gene	Razão de concentração “ <i>primer forward</i> / <i>primer reverse</i> ” (em nM)
<i>E2F1</i>	200 / 200
<i>E2F4</i>	200 / 200
<i>B2M</i>	200 / 200
<i>GUSB</i>	100 / 100
<i>HPRT1</i>	400 / 400
<i>TBP</i>	100 / 100

As concentrações otimizadas de iniciadores e cDNA foram obtidas em testes prévios, como descrito acima.

As reações foram montadas em placas de 96 poços (*MicroAmp® Optical 96 Well Reaction Plate* – Applied Biosystems, Foster City, USA), onde foram adicionadas o *mix* da reação e as amostras de cDNA. Após a montagem das reações de PCR, as placas foram devidamente seladas com o *MicroAmp® Optical Adhesive Covers* (Applied Biosystems, Foster City, USA). Foram realizadas a incubação e a ciclagem, utilizando-se o seguinte programa: incubação (2 min a 50°C), ativação da enzima *Taq polimerase* (10 min a 95°C), seguido por quarenta ciclos de denaturação do cDNA (15 segundos a 95°C), anelamento dos iniciadores (1 min a 60°C), finalizando com o resfriamento (10 min a 4°C).

O aparelho utilizado para amplificação e detecção foi o *ABI PRISM® 7000 Sequence Detection System* (Applied Biosystems, Foster City, USA).

3.8.4 Análise dos resultados

A comparação da expressão gênica entre as diferentes linhagens de glioblastoma (sem e com tratamentos) e seus respectivos controles, assim como a escolha dos melhores genes de referência para a normalização dos dados, foi possível com a obtenção dos valores de C_T (*Cycle threshold*) fornecidos pelo *software ABI PRISM® 7000 Sequence Detection System* (Applied Biosystems, Foster City, USA). Os valores de C_T provenientes do *software* foram exportados para planilhas eletrônicas (*Excel – MS Office 2007*), a partir das quais foram iniciadas as análises.

Para cada uma das sete linhagens de glioblastoma (e para a linhagem com tratamentos) foi feita triplicata biológica, e para cada triplicata biológica, foi feita triplicata experimental. Já para o controle, caracterizado como um *pool* de doadores, foi feita apenas triplicata experimental.

Os C_{Ts} das triplicatas experimentais de cada amostra foram analisados e pontos em que havia variação de mais de 0,5 C_T quando comparados entre os outros pontos da triplicata foram excluídos (exclusão de no máximo um ponto por triplicata experimental). A média entre os pontos restantes de cada triplicata foi calculada para cada amostra.

3.8.4.1 Método de escolha dos genes de referência

Na escolha dos genes de referência mais apropriados para a normalização dos dados de PCR em tempo real, quatro genes (*B2M*, *GUSB*, *HPRT1* e *TBP*) foram testados. Os genes foram amplificados por qRT-PCR em tempo real em todas as amostras (triplicatas biológicas de cada linhagem e triplicata do controle).

O *software geNorm* (Vandesompele *et al.*, 2002) foi utilizado para a escolha dos dois melhores genes de referência, ou seja, dos dois genes com as menores variações de expressão entre as amostras. O *geNorm* é um algoritmo que determina os genes de referência mais estáveis dentre os testados em um dado conjunto de amostras. Para isso, um fator de normalização de expressão gênica é calculado para cada amostra baseado na média geométrica do número de genes de referência definido pelo usuário. Além disso, o *software* calcula uma medida de estabilidade de expressão gênica (M) para cada endógeno. Passos sucessivos de exclusão do gene com maior valor M (menos estável) e recálculo dos valores M permitem classificar os de referência testados de acordo com seus valores de estabilidade de expressão. Dessa forma, os dois genes com os menores valores M foram escolhidos como referência.

3.8.4.2 Análise da expressão dos FTs *E2F1* e *E2F4*

Para a análise da expressão gênica dos FTs *E2F1* e *E2F4*, foram realizadas comparações entre cada uma das sete linhagens (sem tratamento) *versus* o controle (*pool* de doadores e de regiões do cérebro). No caso da linhagem U87 tratada com drogas (cisplatina, doxorrubicina e temozolomida) nos tempos de 6 e 24 h, cada tratamento foi comparado ao controle (U87 sem tratamento – 6 e 24 h), nos respectivos tempos, separadamente.

O *software REST* (*Relative Expression Software Tool* – (Pfaffl *et al.*, 2002) foi utilizado nas comparações dos grupos e na análise estatística dos resultados de expressão relativa. A quantificação da expressão gênica relativa utiliza a expressão dos genes de referência para normalizar os níveis de expressão dos genes de interesse nas diferentes amostras. Esse método permite ajustar os dados para compensar variações devido a diferenças de processamento das amostras.

O *REST* utiliza um modelo matemático que leva em consideração as diferentes eficiências dos iniciadores dos genes analisados. Essas eficiências (*E*) são calculadas a partir do *slope*, como explicado anteriormente. Além disso, o *REST* provê informações estatísticas para a comparação dos níveis de expressão nos grupos (não tratado *versus* tratado, por exemplo). O *software* integra métodos de randomização e *bootstrapping*, testando a significância estatística das razões de expressão calculadas. A variação de expressão de cada gene de interesse é visualizada em gráficos do tipo *boxplot*.

As médias dos C_{Ts} das triplicatas experimentais das amostras foram os parâmetros de entrada do *REST*, a partir dos quais todos os cálculos foram feitos.

4. RESULTADOS

4.1 Análise *in silico* de microarranjos de cDNA

O conjunto de dados de microarranjos de cDNA, o qual foi obtido em bases de dados públicas (*GEO* e *ArrayExpress*) compreendendo o estudo de expressão gênica de diferentes tipos de glioma, foi utilizado para a obtenção de uma lista de genes significativamente modulados. Para isso, a matriz de valores de expressão (obtida no formato de texto tabulado) foi editada de forma a excluir as amostras de oligodendroglioma e astrocitoma, restando apenas amostras de glioblastoma e tecido não tumoral retirado de pacientes com epilepsia (controles). Os dados relativos às amostras de oligodendroglioma e astrocitoma foram excluídos, uma vez que os resultados obtidos nas análises *in silico* foram testadas *in vitro* apenas em linhagens caracterizadas como sendo de glioblastoma.

4.1.1 Análise estatística e agrupamentos gênicos

O arquivo contendo a matriz de valores de expressão foi carregado no *software* MeV, por meio do qual foi aplicado o teste estatístico SAM. A análise resultou em dois grupos de genes significativamente expressos ($FDR \leq 0,05$): um grupo contendo 1.830 genes significativamente induzidos nas amostras de glioblastoma e outro grupo contendo 2.190 genes significativamente reprimidos nessas mesmas amostras, quando comparados aos níveis de expressão dos controles. Apenas os genes induzidos em glioblastoma foram utilizados nas análises posteriores, uma vez que um dos objetivos primordiais do trabalho foi o de apontar potenciais alvos para terapia molecular e, nesse caso, a inibição de alvos é mais promissora (Rhodes e Chinnaiyan, 2004).

A lista dos 1.830 genes significativamente induzidos em glioblastoma foi submetida à análise de agrupamento hierárquico, também no MeV. Houve agrupamento visualmente distinto entre os dois tipos amostrais (glioblastoma e controle), ou seja, as amostras controles e de glioblastoma foram agrupadas separadamente (**Figura 4.1**). Isso demonstra que os genes significativamente modulados apresentaram padrões de expressão similar quando comparados com os mesmos tipos amostrais.

Figura 4.1: Comparação dos perfis de expressão gênica entre as amostras de glioblastoma e controles (epilepsia) após a aplicação do algoritmo de agrupamento hierárquico. (A) Agrupamento da matriz de expressão dos 1.830 genes selecionados pela análise do SAM (induzidos em glioblastoma). (B) Aumento, para melhor visualização, do dendrograma de agrupamento das amostras. (C) Aumento, para melhor visualização, de uma região do agrupamento. O valor de expressão normalizada para cada gene (de acordo com a escala acima) é representado pelas cores: vermelho indica níveis de expressão maiores que a mediana; verde indica níveis de expressão menores que a mediana ($-3,0$ a $3,0$ em $\log_2$); preto indica níveis de expressão iguais à mediana.

4.1.2 Análise de fatores de transcrição

Após a aplicação do algoritmo SAM e a realização do agrupamento hierárquico, a lista dos 1.830 genes diferencialmente expressos (induzidos em glioblastoma) foi submetida no programa FatiGO+ para associação a fatores de transcrição em comum. A **Tabela 4.1** apresenta os resultados obtidos. Três FTs foram relacionados à lista de genes com significância estatística ($p \leq 0,05$). As porcentagens apresentadas correspondem à quantidade de genes da lista que foi relacionada a cada FT. Trabalhos que já relacionaram esses FTs a câncer foram pesquisados no *PubMed*.

Tabela 4.1: Fatores de transcrição relacionados à lista de genes induzidos em glioblastoma (SAM – FDR $\leq 0,05$).

Fator de transcrição	Porcentagem de genes	Valor p	Já relacionado a câncer
<i>HNF1</i>	50,97	5,88 E–3	Ishigaki <i>et al.</i> , 2002; Laurent-Puig <i>et al.</i> , 2003; Rebouissou <i>et al.</i> , 2004; Jeannot <i>et al.</i> , 2006; Piessen <i>et al.</i> , 2007
<i>IRF7</i>	54,36	4,06 E–2	Não
<i>E2F</i>	55,42	4,57 E–2	Lin <i>et al.</i> , 2001; Gorgoulis <i>et al.</i> , 2002; Macaluso <i>et al.</i> , 2003; DuPree <i>et al.</i> , 2004; Imai <i>et al.</i> , 2004; Onda <i>et al.</i> , 2004; Nelson <i>et al.</i> , 2006; Reimer <i>et al.</i> , 2006; Roberts, 2006; Salon <i>et al.</i> , 2006; dentre outros

4.1.3 Análise de associação a assinaturas de expressão gênica

A lista de genes diferencialmente expressos induzidos em glioblastoma, resultado da análise pelo SAM, além de ser utilizada para a análise de FTs no programa FatiGO+, também foi utilizada como parâmetro de entrada para uma análise comparativa com diferentes assinaturas transcricionais.

O teste exato de Fisher foi aplicado nessa comparação entre a lista de genes e o banco de dados de assinaturas transcricionais do programa *TBrowser*, gerando valores p para a comparação entre a lista inicial e cada assinatura. Assinaturas transcricionais com $p \leq 0,05$

foram selecionadas. No total, 7.910 assinaturas seguiram esse critério ($4,31 \text{ E-}214 \leq p \leq 9,69 \text{ E-}3$). Estas foram classificadas por ordem crescente de valor p .

Visto que o número de assinaturas associadas foi demasiado grande, inviabilizando a análise individual de cada uma por meio do *software TBrowser*, apenas as cem primeiras foram analisadas. Dentre essas cem primeiras, 70% correspondiam a assinaturas provenientes de experimentos de microarranjos relacionados com algum tipo de neoplasia, sendo que grande parte das neoplasias correspondiam a diferentes tipos de glioma (38% dessas assinaturas).

As duas primeiras assinaturas transcricionais associadas à lista inicial de genes ($p = 4,31 \text{ E-}214$ e $p = 4,82 \text{ E-}211$, respectivamente) foram extraídas a partir do mesmo conjunto de dados de microarranjos utilizado na análise pelo SAM (*GSE4290*).

Algumas ferramentas de análise de enriquecimento funcional disponibilizadas pelo *TBrowser*, cujas anotações eram relacionadas a FTs e a sítios de ligação destes, foram utilizadas para complementar a análise das assinaturas de expressão gênica. Novamente, apenas as cem primeiras assinaturas foram analisadas.

Na anotação baseada no banco de dados *CisRed*, 92,98% das assinaturas foram relacionadas ao FT *E2F1* ($6,39 \text{ E-}15 \leq q \leq 3,44 \text{ E-}3$) e também 92,98% foram relacionadas ao FT *E2F4* ($3,84 \text{ E-}15 \leq q \leq 1,69 \text{ E-}2$). Esses dois FTs (*E2F1* e *E2F4*) estavam sempre presentes no topo da lista de FTs relacionados a cada assinatura (menores valores q).

Na anotação baseada no banco de dados *ECRbase*, 81,54% das assinaturas foram relacionadas ao FT *E2F* com os menores valores q dentre todos os FTs associados ($4,08 \text{ E-}17 \leq q \leq 2,88 \text{ E-}3$).

O *plugin TFBSconserved*, implementado no *TBrowser*, também foi utilizado. Verificou-se que 76,05% das assinaturas também foram associadas ao FT *E2F* com os menores valores q ($6,71 \text{ E-}49 \leq q \leq 3,08 \text{ E-}17$).

Além de *E2F1* e *E2F4*, outros FTs também foram relacionados às assinaturas transcricionais analisadas, porém com menor frequência (**Tabela 4.2**).

Tabela 4.2: Outros fatores de transcrição relacionados às cem assinaturas transcricionais analisadas no software *TBrowser*, de acordo com as três anotações utilizadas (*CisRed*, *ECRbase* e *TFBSconserved*).

Fator de transcrição	Porcentagem de assinaturas	Valor q
<i>CisRed</i>		
<i>CBF</i>	8,57	$9,58 \text{ E-}3 \leq q \leq 3,12 \text{ E-}2$
<i>Elk-1</i>	4,28	$2,00 \text{ E-}6 \leq q \leq 1,88 \text{ E-}2$
<i>IRF2</i>	4,28	$8,18 \text{ E-}5 \leq q \leq 5,59 \text{ E-}2$
<i>NRF</i>	20,00	$3,67 \text{ E-}10 \leq q \leq 2,45 \text{ E-}2$
<i>ECRbase</i>		
<i>Elk-1</i>	2,86	$1,90 \text{ E-}3 \leq q \leq 3,72 \text{ E-}2$
<i>NRF</i>	8,57	$2,87 \text{ E-}11 \leq q \leq 2,65 \text{ E-}2$
<i>ZF5</i>	28,57	$1,81 \text{ E-}15 \leq q \leq 4,17 \text{ E-}2$
<i>TFBSconserved</i>		
<i>CREBP1</i>	11,43	$4,01 \text{ E-}5 \leq q \leq 3,95 \text{ E-}2$
<i>ELF1</i>	5,71	$7,36 \text{ E-}5 \leq q \leq 1,39 \text{ E-}2$
<i>Elk-1</i>	11,43	$4,83 \text{ E-}5 \leq q \leq 1,62 \text{ E-}3$
<i>IRF2</i>	8,57	$1,68 \text{ E-}14 \leq q \leq 1,07 \text{ E-}2$
<i>MycMax</i>	25,71	$3,69 \text{ E-}7 \leq q \leq 3,07 \text{ E-}2$
<i>NRF2</i>	32,86	$5,36 \text{ E-}13 \leq q \leq 3,86 \text{ E-}2$
<i>PAX3</i>	4,28	$4,71 \text{ E-}3 \leq q \leq 1,96 \text{ E-}2$

Algumas assinaturas transcricionais, além das cem primeiras já descritas, foram escolhidas aleatoriamente dentre as 7.810 assinaturas restantes e foram analisadas no *TBrowser* com o intuito de verificar se os resultados encontrados na análise das cem primeiras se mantinha ao longo das assinaturas restantes. Também foi observada a frequência de *E2F1* e *E2F4* (menores valores q) nessas assinaturas, demonstrando que esses FTs provavelmente foram os mais frequentes na maioria das assinaturas transcricionais associadas à lista inicial de genes.

4.1.4 Comparação entre a análise de fatores de transcrição e a análise de associação a assinaturas de expressão gênica

Uma vez que houve duas formas de análise de FTs partindo-se da mesma lista inicial de genes, ou seja, da lista contendo 1.830 genes significativamente induzidos em glioblastoma, foi possível fazer uma comparação direta entre as duas formas de análise (análise de FTs realizada a partir do programa FatiGO+ e análise de assinaturas transcricionais realizada a partir do software *TBrowser*), como pode ser observado na **Figura 4.2**.

Figura 4.2: Diagrama de Venn representando os FTs encontrados nas duas análises realizadas: (A) análise de FTs pelo programa FatiGO+ e (B) análise de associação a assinaturas transcricionais pelo programa *TBrowser*.

A intersecção dos resultados obtidos nas duas análises de FTs interessantemente apontou o *E2F* como significativamente relacionado à lista inicial de genes e às assinaturas transcricionais analisadas. De forma mais específica, *E2F1* e *E2F4* foram encontrados na análise de assinaturas. Ainda, vale destacar que estes FTs se apresentaram no topo das listas (menores valores *q*) nas análises de enriquecimento funcional de cada assinatura, para todas as anotações utilizadas no *TBrowser* (*CisRed*, *ECRbase* e *TFBSconserved*).

4.1.5 Análise de vias de sinalização celular

A análise de vias de sinalização celular foi realizada a partir da lista de genes induzidos com o intuito de complementar os resultados já obtidos. Essa análise auxiliou na visualização geral dos genes induzidos, enfocando suas interrelações, já que não seria possível analisar a lista contendo 1.830 genes “manualmente”.

Como resultado, três vias celulares associadas aos genes foram escolhidas (**Tabela 4.3**) por apresentarem importante relação com a abordagem do presente trabalho, pelos baixos valores *p* e também por apresentarem os maiores números de genes associados.

Tabela 4.3: Vias de sinalização celular associadas à lista de genes induzidos em glioblastoma. A análise foi realizada por meio da ferramenta *DAVID*. “Genes” representa a quantidade de genes da lista inicial associada a cada via; os respectivos valores p ($p \leq 0,01$) também são apresentados.

Via de sinalização celular	Genes	Valor p
Ciclo celular	32	1,21 E-9
Sinalização da p53	17	2,83 E-5
Câncer	42	6,32 E-4

As três vias selecionadas permitem compreender, mesmo que de maneira simplificada, a relação entre os genes e os principais processos biológicos nos quais estão envolvidos (**Figuras 4.3 a 4.5**).

Figura 4.3: Via de sinalização do ciclo celular (*KEGG*), associada à lista de genes induzidos em glioblastoma, por meio do *DAVID* ($p = 1,21 \text{ E-}9$). No total, 32 genes dessa lista foram associados à via. Tais genes são identificados por meio de estrelas vermelhas posicionadas no canto superior esquerdo do retângulo verde representante de cada gene.

Figura 4.4: Via de sinalização da p53 (KEGG), associada à lista de genes induzidos em glioblastoma, por meio do DAVID ($p = 2,83 \text{ E-}5$). No total, 17 genes dessa lista foram associados à via. Tais genes são identificados por meio de estrelas vermelhas posicionadas no canto superior esquerdo do retângulo verde representante de cada gene.

4.2 Avaliação de expressão gênica por RT-PCR quantitativa (qRT-PCR) em tempo real

4.2.1 Padronização da técnica para os FTs *E2F1* e *E2F4* e para os genes de referência *B2M*, *GUSB*, *HPRT1* e *TBP*

Seguindo-se as especificações do protocolo do kit *SYBR® Green PCR Master Mix* (*Applied Biosystems*, Foster City, USA), utilizado nos experimentos de qRT-PCR em tempo real, alguns testes foram realizados para padronização dos experimentos, a fim de otimizar os resultados.

Dentre esses, foram realizados testes de otimização da concentração de iniciadores e também da concentração de cDNA. Para a obtenção da concentração ideal de iniciadores, várias razões de concentração “*primer forward / primer reverse*” foram testadas. A finalidade foi construir uma curva de dissociação para cada um dos genes testados (*E2F1*, *E2F4*, *B2M*, *GUSB*, *HPRT1* e *TBP*). Os dados dessa curva para cada gene, primeira derivativa da fluorescência em relação à temperatura, foram representados graficamente. Apenas como exemplo, essa representação é mostrada para o gene *E2F4* (**Figura 4.6**).

As combinações de iniciadores foram eficientes na amplificação dos fragmentos, sob temperaturas bastante próximas de *melting* (entre 78 e 88°C, dependendo do gene). Não houve a formação de dímeros, o que pode ser observado pela presença de somente um pico de fluorescência e pela ausência de sinal nas reações sem amostra (NTC), representadas pelas três linhas na derivativa zero.

Figura 4.6: Curvas de dissociação de uma série variável de diluições de iniciadores do gene *E2F4* (razões de concentração 100 / 100, 200 / 200, 300 / 300 e 400 / 400, em nM). O gráfico indica que todas as concentrações de iniciadores foram eficientes na amplificação do fragmento desejado sem a formação de dímeros (*primer dimer*).

Para a determinação da concentração (de iniciadores) mais eficiente na reação, curvas de amplificação foram desenhadas para cada gene testado. Essas curvas indicaram que todas as concentrações de iniciadores foram eficientes na amplificação do fragmento desejado, com valores de C_T pouco variáveis. Novamente, apenas o gráfico do gene *E2F4* é mostrado, como exemplo (**Figura 4.7**).

Assim, foram escolhidas as razões de concentração (em nM) de 100 / 100 (*GUSB* e *TBP*), 200 / 200 (*B2M*, *E2F1* e *E2F4*) e 400 / 400 (*HPRT1*), por representarem as razões nas quais as triplicatas experimentais mostraram-se mais semelhantes e os valores de C_T foram intermediários, não ultrapassaram os limites de detecção ótima do aparelho.

Figura 4.7: Curvas de amplificação a partir de concentrações variáveis de iniciadores do gene *E2F4*, em reações com concentrações de cDNA constantes (80 ng/μL). Em todas as concentrações de iniciadores houve amplificação em C_T semelhante (entre 23,5 e 24,5).

Além do teste de concentração de iniciadores, diferentes concentrações de cDNA também foram testadas (40, 80, 160 e 320 ng/μL) para todos os genes. Gráficos representativos da curva de dissociação (derivativa da fluorescência *versus* temperatura) foram obtidos, sendo que o gráfico para o *E2F4* é apresentado como exemplo (**Figura 4.8**). Todas as quantidades de cDNA testadas foram suficientes para a amplificação dos fragmentos, em temperaturas semelhantes de *melting* ($\approx 81^\circ\text{C}$ no caso do *E2F4*, por exemplo).

Curvas de amplificação também foram desenhadas para os genes testados. Essas curvas indicaram que todas as concentrações de cDNA foram suficientes para a amplificação dos fragmentos, com valores de C_T pouco variáveis (**Figuras 4.9**, exemplificando para o *E2F4*). Assim, foi escolhida a concentração de 80 ng/μL para todos os genes, por representar a concentração cujas triplicatas experimentais mostraram-se mais semelhantes e os valores de C_T foram intermediários, não ultrapassaram os limites de detecção ótima do aparelho.

Figura 4.8: Curvas de dissociação de diferentes concentrações de cDNA (40, 80, 160 e 320 ng/μL) testadas com quantidade fixa de iniciadores do gene *E2F4* (200 / 200 nM). O gráfico indica que todas as concentrações de cDNA foram suficientes para a amplificação do fragmento desejado sem a formação de dímeros (*primer dimer*).

Figura 4.9: Curvas de amplificação obtidas a partir de concentrações variáveis de cDNA para o gene *E2F4*, em reações com concentrações de iniciadores constantes (200 / 200 nM). Em todas as diferentes concentrações de cDNA houve amplificação em C_T pouco variáveis (entre 21,5 e 23,5).

Após a escolha das concentrações de iniciadores e cDNA para a realização da reação de qRT-PCR em tempo real, foi avaliada a eficiência de amplificação dos iniciadores de todos os seis genes, utilizando-se diluições seriadas de um *pool* de cDNA. As diluições foram utilizadas nas seguintes razões: 1, 1/5, 1/25, 1/125 e 1/625. As curvas de dissociação foram representadas graficamente (**Figura 4.10**, exemplificando para o *E2F4*). Também foram obtidas curvas padrão para cada um dos genes (**Figura 4.11**, exemplificando para o *E2F4*).

Todos os iniciadores testados apresentaram valores de *slope* entre $-2,79$ e $-3,44$, indicando alta eficiência de amplificação para todos os genes (**Tabela 4.4**), já que porcentagens ótimas de eficiência variam entre 90 e 110 (Bustin, 2000).

Assim, a quantidades de iniciadores e cDNA foram padronizadas com sucesso e as curvas padrão obtidas confirmaram a eficiência de amplificação das regiões dos genes de interesse, graças à boa qualidade dos iniciadores utilizados. Por meio desses resultados também foi possível constatar a eficiência do protocolo de desenho de iniciadores padronizado no laboratório de pesquisa.

Figura 4.10: Curva de amplificação de uma série de diluições de um *pool* de cDNA (1, 1/5, 1/25, 1/125 e 1/625) do gene *E2F4*. Cada diluição de cDNA corresponde a um C_T diferente (23,5 a 33).

Figura 4.11: Curva padrão construída a partir de cinco diluições de um *pool* de cDNA utilizando iniciadores para o gene *E2F4*.

Tabela 4.4: Valores de *slope* e de eficiência (*E*, em porcentagem) para os iniciadores de cada um dos genes testados, indicando boa eficiência de amplificação para todos os genes.

Gene	Valor de <i>slope</i>	Eficiência (<i>E</i>) – %
<i>E2F1</i>	–2,7871	128,45
<i>E2F4</i>	–3,1663	106,93
<i>B2M</i>	–3,4190	96,10
<i>GUSB</i>	–3,0154	114,60
<i>HPRT1</i>	–3,4401	95,30
<i>TBP</i>	–3,4367	95,42

4.2.2 Escolha de genes de referência

Quatro genes de referência (*B2M*, *GUSB*, *HPRT1* e *TBP*) foram testados para a escolha dos mais apropriados para a normalização dos dados de PCR em tempo real. Tais genes foram amplificados por qRT-PCR em tempo real em todas as amostras e os valores de C_T obtidos foram submetidos ao *software geNorm*. O *software* calculou uma medida de estabilidade de expressão gênica (*M*) para cada endógeno (**Figura 4.12**).

Figura 4.12: Valores de média de estabilidade de expressão (M) dos genes de referência testados ($B2M$, $GUSB$, $HPRT1$ e TBP). Os valores M e a representação gráfica foram gerados pelo *software* *geNorm*.

Como resultado da análise realizada no *geNorm*, os genes $GUSB$ e $HPRT1$ foram escolhidos como controles de referência, pois apresentaram a menor variação de expressão entre todas as amostras (menores valores de M), garantindo que as diferenças entre as amostras não interferissem na expressão dos genes escolhidos, caracterizando-os como controles de referência adequados para a normalização dos dados de expressão gênica avaliados por PCR em tempo real.

4.2.3 Expressão de $E2F1$ e $E2F4$ em linhagens de glioblastoma

O *software* *REST* foi utilizado nas comparações dos grupos e na análise estatística dos resultados de expressão relativa.

Os genes $E2F1$ e $E2F4$ foram amplificados por qRT-PCR em tempo real para todas as amostras de linhagens de glioblastoma (na ausência de qualquer tipo de tratamento e na presença de tratamentos com agentes quimioterápicos), assim como os genes de referência $GUSB$ e $HPRT1$, escolhidos como normalizadores no cálculo da expressão relativa. As médias dos C_{Ts} obtidas de triplicatas experimentais das amostras foram os parâmetros de entrada do *REST*, a partir dos quais todos os cálculos foram realizados. O *software* também

levou em consideração os valores de eficiência dos iniciadores de cada gene nas análises estatísticas.

4.2.3.1 Expressão de *E2F1* e *E2F4* nas linhagens T98, U251, U138, U87, U343, MO59J e MO59K

Para a análise da expressão gênica de *E2F1* e *E2F4* nas sete linhagens de glioblastoma (T98, U251, U138, U87, U343, MO59J e MO59K) na ausência de qualquer tipo de tratamento, foram realizadas comparações separadas: cada linhagem *versus* o controle (*pool* de RNA de tecido cerebral de vários doadores).

Houve um padrão de indução nos perfis de expressão das linhagens quando comparadas ao controle. Ambos os FTs (*E2F1* e *E2F4*) foram induzidos em todas as linhagens e a maior parte das linhagens mostrou indução com diferença estatisticamente significativa ($P(H_1) \leq 0,037$). Apenas as linhagens U343 e MO59K não apresentaram diferença significativa quanto à expressão de *E2F4* quando comparadas ao controle (**Tabela 4.5**).

Tabela 4.5: Expressão de *E2F1* e *E2F4* em cada linhagem, e seus respectivos valores estatísticos. $P(H_1)$ representa a probabilidade da hipótese alternativa de que a diferença entre os grupos (linhagens e controle) seja apenas devida ao acaso. Em *Resultado*, *UP* representa os genes cuja expressão obteve diferença estatística (indução), e “–” representa ausência de diferença estatística. Os dados foram gerados na análise efetuada pelo *software REST*.

Linhagem	Gene	Nível de Expressão	Erro padrão	Intervalo de Confiança (95%)	$P(H_1)$	Resultado
T98	<i>E2F1</i>	9,769	9,011 – 10,781	8,464 – 11,336	0,035	UP
	<i>E2F4</i>	2,409	2,065 – 2,826	2,004 – 2,923	0,032	UP
U251	<i>E2F1</i>	9,489	8,572 – 10,659	8,265 – 11,225	0,000	UP
	<i>E2F4</i>	2,148	1,944 – 2,400	1,886 – 2,482	0,035	UP
U138	<i>E2F1</i>	6,813	5,488 – 9,176	5,071 – 9,663	0,032	UP
	<i>E2F4</i>	1,952	1,843 – 2,079	1,790 – 2,150	0,032	UP
U87	<i>E2F1</i>	10,221	7,749 – 13,637	7,160 – 14,360	0,000	UP
	<i>E2F4</i>	2,259	1,740 – 3,479	1,688 – 3,598	0,037	UP
U343	<i>E2F1</i>	6,260	5,450 – 7,584	5,243 – 7,986	0,000	UP
	<i>E2F4</i>	1,152	0,951 – 1,384	0,923 – 1,432	0,368	–
MO59J	<i>E2F1</i>	5,912	5,022 – 7,195	4,641 – 7,577	0,000	UP
	<i>E2F4</i>	2,512	2,309 – 2,813	2,240 – 2,909	0,000	UP
MO59K	<i>E2F1</i>	6,177	3,749 – 10,425	3,464 – 10,978	0,000	UP
	<i>E2F4</i>	1,182	1,006 – 1,528	0,976 – 1,580	0,191	–

Os valores de expressão gerados pelo *REST* foram transformados em valores de $\log_2$ e representados graficamente (**Figura 4.13**).

Figura 4.13: Perfis de expressão gênica transcricional obtidos por qRT-PCR em tempo real (valores relativos de expressão em $\log_2$) na comparação isolada entre sete linhagens de glioblastoma (T98, U251, U138, U87, U343, MO59J e MO59K) e o controle. (*) Houve diferença estatisticamente significativa ($P(H_1) \leq 0,037$).

Os valores de expressão gênica transcricional resultantes da análise relativa (*REST*) também foram disponibilizados em um gráfico do tipo *boxplot* (**Figura 4.14**). Este mostra a distribuição dos valores de expressão para ambos os FTs em cada linhagem analisada.

Figura 4.14: Distribuição dos valores de expressão de *E2F1* e *E2F4* nas sete linhagens de glioblastoma (T98, U251, U138, U87, U343, MO59J e MO59K) comparadas ao controle, separadamente. Valores de expressão foram obtidos por qRT-PCR em tempo real (valores relativos) por meio da análise do *software* REST.

4.2.3.2 Expressão de *E2F1* e *E2F4* na linhagem U87 tratada com agentes quimioterápicos

Foi analisada a expressão de *E2F1* e *E2F4* na linhagem U87 tratada com drogas (cisplatina, doxorrubicina e temozolomida) nos tempos de 6 e 24 h, sendo que cada tratamento foi comparado ao controle (U87 sem tratamento; 6 e 24 h), nos respectivos tempos, separadamente.

Não houve diferença nos níveis de expressão dos FTs na linhagem quando comparada ao controle nos tratamentos com cisplatina e doxorrubicina, em ambos os tempos (6 e 24 h). Entretanto, no tratamento com temozolomida houve diferença nos níveis de expressão dos FTs para ambos os tempos (indução de *E2F1* e repressão de *E2F4*), mas não houve diferença estatisticamente significativa. Diferenças significativas foram observadas no tratamento com cisplatina nos tempos de 6 h (indução de *E2F1*) e 24 h (indução de *E2F4*). Também foi observada repressão significativa de *E2F1* no tratamento com doxorrubicina no tempo de 24 h (Tabela 4.6).

Tabela 4.6: Expressão de *E2F1* e *E2F4* obtida para a linhagem U87 nos diferentes tratamentos e tempos. $P(H_1)$ representa a probabilidade da hipótese alternativa de que a diferença entre os grupos (tratamentos e controle) seja apenas devida ao acaso. Em *Resultado*, *UP* e *DOWN* representam os genes cuja expressão mostra diferença significativa (indução e repressão, respectivamente), e “–” representa ausência de diferença estatística. Os dados foram gerados na análise efetuada pelo *software REST*.

Tratamento	Gene	Nível de Expressão	Erro padrão	Intervalo de Confiança (95%)	$P(H_1)$	Resultado
Cisplatina 6 h	<i>E2F1</i>	1,566	1,191 – 2,089	1,076 – 2,290	0,000	<i>UP</i>
	<i>E2F4</i>	0,840	0,671 – 1,053	0,648 – 1,090	0,663	–
Cisplatina 24 h	<i>E2F1</i>	0,557	0,468 – 0,662	0,465 – 0,667	0,161	–
	<i>E2F4</i>	1,088	1,028 – 1,153	1,024 – 1,157	0,000	<i>UP</i>
Doxorrubicina 6 h	<i>E2F1</i>	1,137	0,665 – 2,096	0,519 – 2,543	0,663	–
	<i>E2F4</i>	1,056	0,909 – 1,229	0,877 – 1,271	0,663	–
Doxorrubicina 24 h	<i>E2F1</i>	0,758	0,657 – 0,874	0,653 – 0,879	0,000	<i>DOWN</i>
	<i>E2F4</i>	1,183	0,878 – 1,600	0,838 – 1,673	0,663	–
Temozolomida 6 h	<i>E2F1</i>	1,150	0,875 – 1,536	0,784 – 1,695	0,663	–
	<i>E2F4</i>	0,855	0,706 – 1,038	0,682 – 1,074	0,481	–
Temozolomida 24 h	<i>E2F1</i>	1,268	1,162 – 1,385	1,154 – 1,394	0,161	–
	<i>E2F4</i>	0,854	0,735 – 0,995	0,701 – 1,040	0,320	–

Os valores de expressão gerados pelo *REST* foram transformados em valores de $\log_2$ e representados graficamente (**Figura 4.15**), permitindo a visualização dos padrões de expressão (indução e repressão) dos FTs nas condições de tratamentos realizados em ambos os tempos de 6 e 24 h.

Figura 4.15: Perfis de expressão gênica transcricional obtidos por qRT-PCR em tempo real (valores relativos de expressão em $\log_2$) para a linhagem U87 na comparação isolada entre três diferentes tratamentos com drogas (*Cis* – cisplatina, *Doxo* – doxorubicina, *Tmz* – temozolomida) nos tempos de 6 e 24 h e seus respectivos controles (U87 sem tratamento; 6 e 24 h). (*) Houve diferença estatisticamente significativa ($P(H_1) = 0,000$).

Os valores de expressão gênica transcricional resultantes da análise relativa realizada pelo *REST* também são mostrados em um gráfico do tipo *boxplot* (**Figura 4.16**). Este apresenta a distribuição dos valores de expressão para ambos os FTs em cada tratamento, nos dois tempos analisados (6 e 24 h).

Figura 4.16: Distribuição dos valores de expressão de *E2F1* e *E2F4* na linhagem de glioblastoma U87 tratada com diferentes drogas (*Cis* – cisplatina, *Doxo* – doxorubicina, *Tmz* – temozolomida) nos tempos de 6 e 24 h, comparativamente aos controles (U87 sem tratamento – 6 e 24 h). Valores de expressão foram obtidos por qRT-PCR em tempo real (valores relativos) na análise pelo *software* REST.

5. DISCUSSÃO

O aumento crescente do número de metodologias de análise de expressão gênica em larga escala, além da publicação de grandes quantidades de dados gerados pelas mesmas em bases de dados públicas, indicam que abordagens integrativas de análise são cada vez mais cruciais para integrar, validar e extrair o máximo de informações biológicas a partir de conjuntos de dados em escala genômica (Choi *et al.*, 2007).

Abordagens integrativas são propostas de análise em que vários conjuntos de dados são utilizados, realizando-se a associação destes com os resultados analisados. Essas abordagens são capazes de extrair informações biológicas mais “profundas”, diminuindo a possibilidade de resultados falsos, uma vez que há a combinação de dados e resultados provenientes de múltiplos estudos (Rhodes e Chinnaiyan, 2005). Entretanto, não há metodologias definidas para a realização de análises integrativas aplicadas a resultados de microarranjos (Fishel *et al.*, 2007), fato este que confere certa flexibilidade na metodologia adotada para a realização desse tipo de análise.

Algumas análises integrativas objetivam o estabelecimento de relações entre os genes diferencialmente expressos, associando-os, por exemplo, a fatores de transcrição (FTs) que são reguladores comuns desses genes.

O estudo de FTs comuns a um conjunto de genes é interessante no sentido de buscar marcadores que possam ser utilizados futuramente no estabelecimento de estratégias terapêuticas mais eficazes em pacientes acometidos pelo câncer, como o glioblastoma (no contexto deste trabalho), visto que os FTs apresentam potencial para serem utilizados como alvo específico de drogas antitumorais (Mees *et al.*, 2009). FTs e elementos cis-regulatórios correspondentes são considerados componentes chave na regulação gênica (Choi *et al.*, 2009). Assim, determinados FTs constituem potenciais alvos moleculares, que em conjunto com o tratamento quimio ou radioterápico, por exemplo, podem ser promissores em termos de estratégias para o tratamento do câncer.

No presente trabalho, foram utilizados dados de microarranjos provenientes de um estudo que utilizou diversas amostras de gliomas, sendo que tais dados foram obtidos no *GEO*, além de estarem também disponíveis no *ArrayExpress*; ambas as bases públicas de dados de microarranjos têm sido extensamente utilizadas pela comunidade científica (Anonymous, 2002).

A escolha dos dados de microarranjos do trabalho de Sun e colaboradores (2006) levou em consideração a qualidade desses dados, a utilização de uma boa plataforma de microarranjos

(a *HG-U133 Plus 2.0*, da *Affymetrix*, a qual engloba todo o transcriptoma), além da grande quantidade de amostras testadas, compreendendo um grupo controle (23 amostras) e um grupo de glioblastoma (81 amostras). As amostras foram obtidas por remoção cirúrgica em pacientes, o que significa que não houve um tratamento específico empregado nas amostras de glioblastoma. Essa característica é importante, já que, a princípio, não se objetivou estudar os efeitos de um tratamento específico. Além disso, alguns testes *in vitro*, posteriores às análises *in silico*, foram realizados em linhagens de glioblastoma na ausência de tratamento químico- ou radioterápico.

A matriz de expressão gênica obtida na base de dados foi submetida à análise estatística (SAM – $FDR \leq 0,05$). Os genes foram divididos em dois grupos principais: induzidos e reprimidos em glioblastoma.

Os resultados obtidos na análise realizada pelo SAM foram visualmente disponibilizados em um *heatmap* após a aplicação do algoritmo de agrupamento hierárquico. Métodos de agrupamento (*clustering*) de dados de expressão gênica em larga escala são caracterizados pelo uso de algoritmos estatísticos padronizados para agrupar genes de acordo com as similaridades nos seus perfis de expressão (Eisen *et al.*, 1998).

Por meio do *heatmap* (agrupamento hierárquico), mais especificamente por meio do dendograma, foi possível verificar que os dois tipos de amostras em estudo foram separados em dois grupos distintos (um grupo contendo as amostras de glioblastoma e outro, as amostras de cérebro de pacientes com epilepsia). Isso demonstra que houve um padrão distinto de expressão entre os grupos amostrais e, além disso, que os genes apontados pela análise realmente são bons indicadores (quanto à expressão diferencial) de ambas as condições: tumoral e não tumoral.

Após a análise estatística, os genes obtidos nessa etapa foram submetidos à análise de FTs. Eisen e colaboradores (1998) observaram que os genes agregados em um mesmo agrupamento apresentavam funções em comum, podendo também estar sob o controle dos mesmos FTs. O fato de que genes co-expressos tendem a desempenhar papéis comuns nas células é amplamente aceito na literatura (Stuart *et al.*, 2003; Lee *et al.*, 2004), o que reforça a hipótese de que genes co-expressos podem ser regulados por FTs em comum.

Um dos objetivos do presente trabalho foi promover a associação dos genes diferencialmente expressos (apontados pelo SAM) a FTs que poderiam ser reguladores comuns de um grande número desses genes. Vale ressaltar que a lista de genes submetida ao programa FatiGO+ (para a análise de FTs) e também utilizada na análise de assinaturas transcricionais corresponde apenas à lista dos genes significativamente expressos induzidos em glioblastoma, comparados aos controles (epilepsia). A lista dos genes significativamente

reprimidos em glioblastoma não foi utilizada, uma vez que o intuito principal desse trabalho foi encontrar genes (FTs, mais especificamente) que pudessem ser apontados como biomarcadores em glioblastoma e também como potenciais alvos para intervenção molecular, por meio de métodos de inibição (RNA interferente, por exemplo).

Genes são considerados potenciais marcadores quando são diferencialmente *induzidos* em um tipo particular de câncer e suas funções moleculares ou localização sugerem que os respectivos produtos gênicos possam ser passíveis de detecção no plasma ou tecidos quando traduzidos (Rhodes e Chinnaiyan, 2004).

Welsh e colaboradores (2001) realizaram uma análise comparativa entre tecido tumoral e normal de ovário utilizando dados de microarranjos e ferramentas de anotação da base de dados *Oncomine*. Interessantemente, o quinto gene mais superexpresso na análise, o *PRSS8* (*Protease Serine 8*), já havia sido descrito em outros estudos como novo biomarcador para câncer de ovário (Mok *et al.*, 2001). Dessa forma, supõe-se que outros genes que se apresentaram significativamente superexpressos possam também servir como biomarcadores, o que deverá ser testado.

Outro estudo do mesmo grupo de pesquisa (Welsh *et al.*, 2003) utilizou uma abordagem similar para a busca de marcadores a partir de dados de microarranjos de diferentes tipos de câncer. Nessa análise foram apontados 74 genes significativamente induzidos como potenciais marcadores de câncer. Alguns produtos de marcadores identificados têm sido detectados no plasma de pacientes com câncer, incluindo membros da família *KLK* (*kallikrein*) em carcinoma de ovário, próstata, mama e rim (White *et al.*, 2010), *GAST* (*gastrin*) em carcinoma de pulmão (Zhao e Wang, 2011) e *AFP* (*alpha-fetoprotein*) em carcinoma de fígado (Yuen e Lai, 2005).

Dessa forma, genes candidatos a biomarcadores correspondem àqueles apontados como superexpressos em doenças, uma vez que esses genes podem levar à tradução de um número elevado (em relação ao normal) de proteínas, por meio das quais é possível realizar medições e quantificações das mesmas no plasma, por exemplo, possibilitando novas estratégias de diagnóstico. Além disso, tais genes são potenciais alvos para inibição por meio de drogas. Assim, apenas a lista de genes identificados como induzidos em glioblastoma (análise pelo SAM) foi utilizada nas análises de associação a assinaturas transcricionais e a FTs.

Essa lista foi utilizada como parâmetro de entrada no programa FatiGO+. Como resultado da análise, três FTs (*HNF1*, *IRF7* e *E2F*) foram associados estatisticamente ($5,88 \text{ E-}3 \leq p \leq 4,57 \text{ E-}2$) a um grande número de genes. Pouco mais da metade dos genes, e não necessariamente os mesmos genes, foi associada a cada um dos três FTs.

O *HNF1*, também conhecido por *HNF1A* (*Hepatocyte Nuclear Factor 1 – Alpha*) e *TCF-1* (*Transcription Factor 1*) codifica um FT requerido para a expressão de diversos genes específicos do fígado. Esse gene tem sido relacionado a doenças (segundo uma extensa literatura) como diabetes, incluindo diabetes *mellitus* tipo I e II, diabetes gestacional e *MODY3 – Maturity Onset Diabetes of the Young type 3* – (Harries *et al.*, 2006; Narayana *et al.*, 2006; Rowley *et al.*, 2006; Maraschin *et al.*, 2008; Umeyama *et al.*, 2009; Pal *et al.*, 2010; Pinés Corrales *et al.*, 2010), doenças renais (Malecki *et al.*, 2005), adenoma (Bluteau *et al.*, 2002; Laurent-Puig *et al.*, 2003), câncer de mama, de endométrio e de ovário (Rebouissou *et al.*, 2004), neoplasias colorretais (Jeannot *et al.*, 2006; Bélanger *et al.*, 2010) e neoplasias de esôfago (Piessen *et al.*, 2007).

O *HNF1* está envolvido em diversos processos biológicos, dentre eles a regulação da via de sinalização WNT (Thierry-Mieg e Thierry-Mieg, 2006). Esta via exerce papel fundamental em vários processos do desenvolvimento normal em metazoários, tais como proliferação celular, diferenciação e migração. O mediador central da sinalização WNT canônica é a β -catenina (*CTNNB1*), a qual interage no núcleo com membros da família de FTs *TCF* e *LEF* (*Lymphoid Enhancer Factor*) para induzir genes alvos para WNT. Entre esses genes estão incluídos reguladores de progressão do ciclo celular (Wodarz e Nusse, 1998; Fogarty *et al.*, 2005). Como o complexo β -catenina/*HNF1*/*LEF* liga-se ao DNA para promover a progressão do ciclo, a formação desse complexo tem sido considerada uma etapa primordial na tumorigênese induzida por WNT (Clevers, 2006; Klaus e Birchmeier, 2008). A inibição de alguns genes atuantes na via WNT tem demonstrado estagnação no crescimento de tumores (Dealmeida *et al.*, 2007).

A ativação constitutiva da via WNT/ β -catenina (via canônica) tem sido apontada como a principal causa do câncer de cólon, sendo que o *HNF1A* é um dos principais transdutores dessa via no intestino (Najdi *et al.*, 2009). Além disso, Roth e colaboradores (2000) demonstraram que a via WNT se apresenta funcional em células de glioma, inibindo a motilidade e promovendo a proliferação das células tumorais. Dessa forma, a indução do FT *HNF1* pode estar associada ao disparo da via WNT e à proliferação tumoral em glioblastoma.

Ademais, a via WNT está relacionada à expressão de genes de sinalização, como o *VEGF* (*Vascular Endothelial Growth Factor*) (Klaus e Birchmeier, 2008; Dilek *et al.*, 2010). O crescimento de tumores sólidos requer suprimento de sangue que é obtido por meio de neoangiogênese, a qual é controlada por diversos fatores de crescimento, como o *VEGF* (Shao *et al.*, 2005). Esse gene tem sido extensamente relacionado à vascularização em tumores, inclusive em gliomas (Dunn *et al.*, 2000; Goldbrunner *et al.*, 2000; Lamszus *et al.*, 2004; Sun

et al., 2006). Interessantemente, o *VEGF* foi identificado na lista de genes induzidos nas amostras de glioblastoma.

Na análise das vias de sinalização celular obtidas pelo *DAVID* a partir da mesma lista de genes induzidos, também foi possível detectar vários genes da via WNT (via representada como parte da via do câncer). Dentre esses, membros da família de genes *WNT* (*Wingless-type MMTV integration site family*), da família de receptores transmembrana *Frizzled* (genes *FZD1-10*), além do gene codificador da proteína *Survivin* (*BIRC5 – Baculoviral Inhibitor of apoptosis Repeat-Containing 5*), membro da família de inibidores de apoptose (Sah *et al.*, 2006), e o *c-Myc* (ou *MYC*, *v-myc avian myelocytomatosis viral oncogene homolog*), os quais desencadeiam proliferação celular e evasão da apoptose. Além disso, também foram apontados alguns genes pertencentes à via de sinalização VEGF (também representada como parte da via do câncer): o gene *VEGF*, como já descrito, o *HIF- α* (*Hypoxia Inducible Factor 1, Alpha subunit*) e o *TGF- β* (*Transforming Growth Factor Beta*), todos participantes do processo de angiogênese.

Outro FT apontado na análise pelo FatiGO+ foi o *IRF7* (*Interferon Regulatory Factor 7*). A ativação de fatores regulatórios de *interferons*, como o *IRF7*, é essencial para a indução de *interferons* do tipo I (*IFN*) e respostas antivirais inatas (Bentz *et al.*, 2010). Esse FT é normalmente associado a doenças virais (Zhang e Pagano, 1997; Wathelet *et al.*, 1998; Tsitoura *et al.*, 2009; Oldak *et al.*, 2011; Sathish *et al.*, 2011). Já é estabelecido que o sistema imune também tem a capacidade de atacar células tumorais, já que durante a transformação de normais em malignas, essas células adquirem inúmeras alterações moleculares e bioquímicas que as tornam potencialmente vulneráveis às células do sistema imune (Steer *et al.*, 2010). Alguns pesquisadores ainda sugerem que alterações no sistema imune podem ser utilizadas como ferramentas para detecção de câncer (Hanash, 2011).

Assim, o *IRF7*, selecionado na análise pelo FatiGO+, também foi apontado na lista de genes induzidos em glioblastoma.

Além da análise realizada pelo FatiGO+, a mesma lista de genes induzidos foi comparada a um banco de dados de assinaturas transcricionais, compreendendo um total de 18.250 assinaturas. Estas foram extraídas de diversos conjuntos de dados de microarranjos disponibilizados no *GEO* e fazem parte do banco de dados de assinaturas transcricionais do *software Transcriptome Browser* (*TBrowser – Lopez et al.*, 2008).

Os valores *p* associados às assinaturas e obtidos nessa comparação foram extremamente baixos (entre 4,31 E-214 e 9,69 E-3), indicando a confiabilidade dos resultados encontrados e a semelhança entre as assinaturas transcricionais e a lista de genes utilizada como parâmetro. Além disso, outra evidência interessante é a de que as duas primeiras assinaturas relacionadas

foram assinaturas obtidas a partir do mesmo conjunto de dados utilizado na análise inicial pelo SAM (*GSE4290*). A maioria das demais assinaturas (70%, dentre as analisadas) também eram assinaturas provenientes de estudos do câncer; dentre estas, 30% foram obtidas em experimentos específicos de glioma, incluindo alguns estudos específicos com glioblastoma. Esses resultados demonstram de maneira interessante que pode haver uma relação constante entre os genes diferencialmente expressos em diferentes experimentos que abordem as mesmas doenças.

Dudley e colaboradores (2009) realizaram a meta-análise mais extensa (em termos de volume de dados) a partir de resultados de microarranjos. O grupo avaliou 429 experimentos de microarranjos, representando 238 doenças e 122 tecidos analisados a partir de um total de 8.435 lâminas. Segundo os autores, houve maior concordância nos perfis de expressão gênica entre doenças do que entre tecidos. Em outras palavras, a expressão dos genes foi afetada de forma semelhante em tecidos diferentes com a mesma doença. Já era esperado que a expressão apresentasse concordância entre os mesmos tecidos com a mesma doença, mas essa análise mostrou que uma mesma doença tende a afetar diferentes tecidos de uma mesma forma. Por outro lado, não houve concordância na expressão de genes num mesmo tecido quando o mesmo foi afetado por doenças distintas. Os resultados encontrados no presente trabalho, embora em menores proporções, corroboram os descritos por Dudley e seus colaboradores, uma vez que uma grande quantidade das assinaturas transcricionais relacionadas foi obtida a partir de diferentes tecidos caracterizados como neoplasia.

As cem primeiras assinaturas analisadas no *software* *TBrowser* apresentaram resultados interessantes, devido à presença constante de alguns FTs em todas as assinaturas analisadas. Além disso, foi observado que essa constância prevaleceu também nas outras assinaturas escolhidas aleatoriamente para análise (à parte das cem primeiras). Dessa forma, empiricamente pode-se afirmar que houve consistência nos resultados observados, sendo que alguns FTs (*E2F*) foram apontados como relacionados com a maioria das assinaturas analisadas nas três diferentes anotações utilizadas (*CisRed*, *ECRbase* e *TFBSconserved*).

O principal FT apontado na análise das assinaturas transcricionais foi o *E2F* (*E2F1* e *E2F4*, especificamente), também associado à lista inicial de genes na análise do FatiGO+. Esse FT será discutido adiante.

Outros FTs também foram associados aos genes das assinaturas, porém com frequência menor que o *E2F*. Dentre esses FTs, destacaram-se *NRF2* (*Nuclear Respiratory Factor 2*), *ZF5* (*Zinc Finger Protein 5 Homolog*), o complexo *MycMax* (*MAX – MYC Associated Factor X*), *Elk-1* (*ETS-like gene 1*) e *CREBP1* (ou *ATF2*, *Activating Transcription Factor 2*), por estarem associados às maiores porcentagens de assinaturas transcricionais, depois de *E2F1,4*.

Elk-1 é membro da família de FTs *ETS*, os quais são oncogenes. Alguns trabalhos relataram a associação desse FT ao câncer (Xiao *et al.*, 2002; Chen *et al.*, 2006). *CREBP1* também foi relacionado com câncer em alguns trabalhos, incluindo câncer de pulmão, de mama e neuroblastoma (Woo *et al.*, 2002).

A intersecção dos resultados obtidos nas duas análises de FTs (programa FatiGO+ e análise de associação a assinaturas transcricionais) interessantemente apontou o *E2F*. De forma mais específica, os FTs *E2F1* e *E2F4* foram encontrados na análise de assinaturas. Em face a essa relação observada na comparação entre as análises, uma atenção especial foi dada ao *E2F1* e *E2F4*, no sentido de validar com experimentos laboratoriais *in vitro* tais resultados obtidos nas análises *in silico*.

Com esse intuito, optou-se pela utilização da técnica de PCR quantitativa em tempo real, pelo fato desta ser capaz de detectar e medir em pequenas frações de tempo as quantidades de ácidos nucleicos em uma ampla gama de amostras de numerosas fontes. Atualmente, tem sido extensamente utilizada com sucesso para diagnóstico molecular em diversas áreas como, por exemplo, em pesquisa básica e aplicada à medicina (Bustin, 2000; Kubista *et al.*, 2006). É uma técnica que apresenta simplicidade conceitual e prática, combinada à rapidez, sensibilidade e especificidade na quantificação de ácidos nucleicos (Bernard e Wittwer, 2002).

Com o intuito de otimizar os resultados obtidos pelo método de qRT-PCR, algumas padronizações foram realizadas para cada gene estudado (FTs *E2F1,4* e genes de referência *B2M*, *GUSB*, *HPRT1* e *TBP*).

Quatro genes de referência foram escolhidos para a normalização dos dados de expressão gênica por PCR quantitativa em tempo real (*B2M*, *GUSB*, *HPRT1* e *TBP*). Tais genes têm sido extensamente utilizados para essa finalidade em diversos tipos celulares e teciduais (Cicinnati *et al.*, 2008; Lyng *et al.*, 2008; Gubern *et al.*, 2009; Rho *et al.*, 2010; Manzano *et al.*, 2011). Entretanto, nem todos os genes caracterizados como referência são necessariamente estáveis e de expressão constante em todos os tipos celulares e teciduais (normal e patológico) e sob condições diferentes de tratamento (drogas e radiação, por exemplo) (Romanowski *et al.*, 2007). Genes de referência ideais para qRT-PCR em tempo real devem ser aqueles cuja expressão do mRNA é constante (mesmo nível) em todas as amostras sob investigação, independentemente do tipo de tecido, estado da doença, medicação ou condições experimentais. Além disso, os níveis de expressão devem ser comparáveis aos do gene alvo (Suzuki *et al.*, 2000; Coulson *et al.*, 2008).

A expressão gênica quantitativa de tecidos e linhagens celulares tumorais é frequentemente comparada a tecidos ou linhagens não-tumorais para a seleção de genes marcadores e para a

avaliação do grau de diferenciação transcricional em câncer. A seleção dos genes de referência apropriados para cada caso, os quais variam em diferentes tipos de tumores, subtipos moleculares e tecidos normais, é essencial para a quantificação exata da expressão gênica (Gur-Dedeoglu *et al.*, 2009), evitando eventuais imprecisões decorrentes da utilização de genes de referência inadequados (Dheda *et al.*, 2004; Meldgaard *et al.*, 2006; Derks *et al.*, 2008).

Valente e colaboradores (2009) testaram sete genes de referência comumente utilizados como controle interno (*ACTB*, *GAPDH*, *GUSB*, *HMBS*, *HPRT1*, *TBP* e *18S rRNA*). O objetivo do trabalho foi avaliar a expressão destes em glioblastoma (30 amostras de glioblastoma e 9 de substância branca provenientes de disseções de diferentes indivíduos) e apontar os melhores normalizadores para esse tipo de tumor. Segundo os autores, 80% dos trabalhos publicados nos últimos dois anos utilizaram os genes *ACTB*, *GAPDH* e *18S rRNA* como controles de referência em análises de expressão de tumores ou linhagens celulares de glioma. *ACTB*, *GAPDH* e *18S rRNA* são os genes de referência mais utilizados, não só em gliomas. Entretanto, diversos estudos têm mostrado sólida evidência de que seus níveis de expressão variam significativamente entre diferentes indivíduos, tipos celulares, estágios de desenvolvimento e/ou condições experimentais (Stamova *et al.*, 2009; Tong *et al.*, 2009).

Os resultados obtidos por Valente e colaboradores (2009) corroboraram esses estudos, já que apontaram a combinação dos genes *TBP* e *HPRT1* como adequada para a normalização de dados de qRT-PCR em tempo real no estudo de glioblastoma.

No presente trabalho, a análise realizada pelo *software* *geNorm* apontou os genes de referência *GUSB* e *HPRT1* como a combinação mais adequada para a normalização dos dados de expressão obtidos em sete linhagens de glioblastoma e o grupo controle, já que esses genes obtiveram os menores valores de estabilidade de expressão *M*, o que representa genes mais estáveis. Tais genes obtiveram $M = 0,739$, valor inferior ao valor de corte (1,5) sugerido pelos autores do *geNorm* (Vandesompele *et al.*, 2002). Esses resultados não diferiram muito dos obtidos por Valente e colaboradores (2009), visto que o *HPRT1* foi selecionado em ambos os trabalhos.

No *software* *geNorm*, ao invés do uso de um único gene de referência, Vandesompele e colaboradores (2002) propuseram a seleção de mais de um gene de referência válido para a normalização dos dados. A utilização de dois ou mais genes de referência devidamente validados pode melhorar a confiabilidade dos resultados (Huggett *et al.*, 2005; Hellemans *et al.*, 2007).

O *software* *REST* foi utilizado na quantificação da expressão gênica relativa, utilizando a expressão dos genes *GUSB* e *HPRT1* para normalizar os níveis de expressão dos FTs *E2F1* e

E2F4 nas diferentes amostras. Uma das vantagens do *REST*, com relação a outras metodologias, é justamente o fato de permitir o uso de mais de um gene endógeno no cálculo da expressão gênica relativa (Pfaffl *et al.*, 2002). Além disso, o *REST* leva em consideração eventuais diferenças na eficiência da reação de PCR, diferente de outras metodologias amplamente utilizadas, como o método de cálculo do $2^{-\Delta\Delta C_t}$ (Livak e Schmittgen, 2001), o qual normalmente assume todas as eficiências como sendo 100% (Pfaffl, 2001; Bustin *et al.*, 2009).

Na análise de expressão quantitativa por PCR em tempo real, foram utilizados cDNAs obtidos a partir de sete diferentes linhagens de glioblastoma (células cultivadas). Como resultado das análises dos dados de PCR em tempo real, foi observada a superexpressão de ambos os FTs, sendo que houve maior diferença de expressão de *E2F1* do que de *E2F4*, com relação ao controle.

E2F1 e *E2F4* pertencem à família de FTs *E2F*, composta por oito membros (*E2F1-8*) em mamíferos (Zalmas *et al.*, 2008). As proteínas codificadas por esses genes são participantes de importantes processos do ciclo celular, tais como progressão, transição G1/S, segregação e condensação dos cromossomos, replicação e reparo do DNA, diferenciação e apoptose (Dimova e Dyson, 2005). Assim, essas proteínas desempenham papel crucial no controle do ciclo celular; além disso, apresentam ação sobre proteínas supressoras de tumor. As proteínas *E2F* possuem múltiplos domínios conservados evolutivamente, os quais são encontrados na maioria dos membros dessa família, em diversos organismos, de plantas (Ramírez-Parra *et al.*, 1999; Albani *et al.*, 2000; Guo *et al.*, 2007) a eucariotos superiores, como mamíferos (Reichel *et al.*, 1987; Sardet *et al.*, 1995; Cartwright *et al.*, 1998). A família *E2F* é composta, basicamente, por dois tipos de proteínas: *E2F* típicos (*E2F1-6*) e *E2F* atípicos (*E2F7,8*). A diferença entre os dois grupos de proteínas está na quantidade e tipo dos domínios. Nos *E2F* típicos, há um domínio de ligação ao DNA responsável pela interação da proteína com os genes alvo, um domínio de heterodimerização responsável pela dimerização com a proteína *DP* (*DRTF Polypeptide*), uma região para interação proteína-proteína, conferindo especificidade à resposta transcricional, e um domínio de ligação à proteína do Retinoblastoma (pRB). Por outro lado, nos *E2F* atípicos, há apenas dois domínios descritos, ambos com finalidade de ligação ao DNA (Lammens *et al.*, 2009).

As proteínas *E2F* típicas formam heterodímeros com outro FT, a proteína *DP*. Em mamíferos, duas proteínas *DP* (*DP1,2*) já foram caracterizadas (Zalmas *et al.*, 2008). Uma vez que tanto *E2F* quanto *DP* possuem um domínio de ligação ao DNA, o heterodímero formado possui dois sítios de ligação ao DNA, por meio dos quais ocorre a ativação da transcrição. A dimerização entre essas duas proteínas é um pré-requisito para alta afinidade e ligação

sequência-específica das proteínas E2F aos elementos *cis*-regulatórios de genes alvo (Van Den Heuvel e Dyson, 2008). A via *E2F-DP* é crucial na regulação da duplicação do DNA, apoptose e diferenciação celular (Degregori e Johnson, 2006), enquanto que a desregulação de sua atividade apresenta alto impacto na relação saúde-doença (Tsuge *et al.*, 2005; Höglinger *et al.*, 2007; Opavsky *et al.*, 2007).

Além disso, existe outra importante via de regulação da transcrição pelas proteínas E2F típicas, a qual é controlada pela proteína do Retinoblastoma (pRB), uma proteína supressora de tumor. Já foi demonstrado em experimentos *in vitro* que existe alta afinidade entre o complexo E2F/DP e pRB (Giacinti e Giordano, 2006). A via *pRB/E2F/DP* é conservada em mamíferos e plantas. Em mamíferos, a família pRB possui três membros: pRB/p105, p107 e pRB2/p130 (Dimova e Dyson, 2005).

De acordo com a atividade positiva ou negativa na regulação da transcrição gênica, *E2Fs* são divididos em dois grupos: ativadores (*E2F1-3*) e supressores (*E2F4-6*) transcricionais (Helin, 1998). As proteínas pRB são as responsáveis pelas atividades antagônicas dos FTs *E2F*, pois são responsáveis por “mascarar” o domínio de transativação das proteínas E2F e recrutar proteínas modificadoras de cromatina (Brehm *et al.*, 1998). Em animais, pRB liga-se ao heterodímero E2F/DP no domínio de transativação C-terminal, inibindo o heterodímero e, conseqüentemente, a transcrição. A hiperfosforilação da pRB por quinases dependentes de ciclina causa a dissociação dessa proteína do heterodímero E2F/DP, liberando o domínio de transativação e permitindo a transcrição de genes alvo do *E2F*. Em alguns casos, a associação de pRB ao complexo E2F/DP não só promove inativação, mas repressão na transcrição de genes alvo de *E2F*, como ocorre com *E2F4,5* (Giacinti e Giordano, 2006).

E2F1, assim como E2F2 e E2F3, possui um domínio adicional de ligação a ciclinas. Essa proteína liga-se preferencialmente à pRB/p105, podendo mediar tanto a proliferação celular quanto apoptose dependente ou independente de p53 (Dimova e Dyson, 2005).

Quanto à E2F4, esta se liga a todas as três proteínas da família pRB, mas com maior afinidade à p107 e p130. Em situações normais, esse FT possui um importante papel na supressão de genes associados à proliferação. Mutações no gene codificador desse FT e expressão aumentada têm sido associadas ao câncer humano (Rakha *et al.*, 2004).

Ishii e colaboradores (1999) realizaram um trabalho com o intuito de estabelecer o *status* de diversos genes supressores tumorais em trinta e quatro linhagens celulares de glioma humano selecionadas aleatoriamente. Dentre os resultados, foi observada alta incidência de alterações na via celular da pRB (64,7%). Essa alta incidência demonstra a relação existente entre a proliferação tumoral e mutações nos genes codificadores de proteínas dessa importante via do ciclo celular. As linhagens de glioblastoma analisadas nesse trabalho podem conter

genes mutados participantes dessa via, corroborando com a expressão diferencial de alguns genes, incluindo os FTs *E2F1* e *E2F4*. Em células com pRB não funcional, há aumento da atividade dos FTs *E2F* (Prost *et al.*, 2007), promovendo desregulação do ciclo celular e aumento na probabilidade de ocorrência de fenótipos patológicos (Giacinti e Giordano, 2006).

O gene codificador da pRB (*RBI*) é funcionalmente inativado na maioria das neoplasias humanas, seja diretamente por mutação/deleção (como ocorre no retinoblastoma, osteossarcoma e carcinoma pulmonar) ou indiretamente por alterações na expressão ou atividade de genes regulatórios (Liu *et al.*, 2004). Assim, pRB apresenta papel fundamental no controle negativo do ciclo celular e sua inativação pode contribuir para a progressão tumoral.

No final da fase G1, ciclinas E e A formam complexos com CDK2, os quais especificamente se ligam às pRB (De Luca *et al.*, 1997). Esse passo é crucial para a inativação dos membros da família pRB e a liberação dos E2F (Giacinti e Giordano, 2006).

Interessantemente, na via do ciclo celular, *p107* foi apontado como induzido, assim como *E2F5*. Isso demonstraria a associação da *p107* ao *E2F5*, culminando na inativação da ação desse FT. Entretanto, também aparecem induzidos nessa via, tanto a ciclina A quanto CDK2, as quais podem participar na formação de complexos inibidores de p107, também superexpressa. Essa hipótese é sustentada pelo fato de que *c-Myc*, normalmente reprimido pelo complexo *p107/E2F5*, apresenta-se induzido na mesma via. A superexpressão de *c-Myc* tem sido associada a diversas neoplasias, inclusive glioblastoma (Walter *et al.*, 2002; Nilsson e Cleveland, 2003; Dang *et al.*, 2005).

Alguns trabalhos têm demonstrado que pRB/E2F1 integra resposta proliferativa com as vias de sobrevivência celular por meio da indução de *TP53* e do acúmulo de p53 (Qin *et al.*, 1994; Shan e Lee, 1994; Wu e Levine, 1994; Kowalik *et al.*, 1995; Kowalik *et al.*, 1998).

O *TP53*, embora seja considerado um gene supressor tumoral, quando sofre alguns tipos de mutações, pode exercer um efeito negativo dominante, já que a maioria das mutações leva à perda de função. Dessa forma, o produto de um único alelo mutado interage e inativa o produto do alelo normal, induzindo o câncer e atuando, portanto, como um oncogene (Levine e Vosburgh, 2008). Ishii e colaboradores (1999) demonstraram que a incidência de mutação ou deleção no gene codificador da p53 alcança cerca de 80% em gliomas; ainda, os autores demonstraram que há alta incidência de alterações na via celular de p53 (94,1%).

Alguns estudos indicaram uma ativação preferencial da via p53 em glioblastomas pediátricos (Litofsky *et al.*, 1994; Pollack *et al.*, 1997; Pollack *et al.*, 2001). Além disso, Ganigi e colaboradores (2005) associaram a expressão elevada desta proteína aos pacientes com pior prognóstico de glioblastoma. Diversos genes da via p53 foram apontados como induzidos na análise pelo SAM, inclusive o próprio *TP53*. As vias de ciclo celular, de

sinalização da p53 e de progressão tumoral, obtidas na análise pelo *DAVID*, ilustram esses resultados.

Dentre as linhagens de glioblastoma utilizadas no presente trabalho, algumas são mutantes (T98, U251 e U138) e outras proficientes (U87, U343, MO59J e MO59K) para o *TP53*. Essa diferença não influenciou na expressão dos FTs *E2F1,4*, já que foram superexpressos de maneira semelhante em todas as linhagens. Tal observação demonstra que, provavelmente, não só mutações no próprio *TP53*, mas também em outros genes de sua via, corroboram para a desregulação da mesma e para o processo de progressão tumoral.

A proteína c-Myc (cujo gene apresentou-se induzido na análise pelo SAM), assim como *E2F1*, também apresenta papel crítico no controle da proliferação celular e está relacionada com a resposta antagônica de indução de *TP53* e progressão tumoral (Hermeking e Eick, 1994; Wagner *et al.*, 1994). O produto do gene *Myc* associa-se ao produto de *Max* formando heterodímeros protéicos com função de FT que regulam genes de controle da proliferação celular (Blackwell *et al.*, 1990; Blackwood *et al.*, 1991). O complexo MycMax foi associado a 25,71% das assinaturas transcricionais pelo *TFBSconserved* na análise do *TBrowser*.

Assim como *c-Myc*, na via de ciclo celular do *KEGG*, quatro genes foram apontados como induzidos, dentre os seis existentes na família do *MCM* (*Mini Chromosome Maintenance*): *MCM2,3,5* e *7*. As proteínas MCM são requeridas na iniciação da replicação do DNA de eucariotos. A família é constituída por seis proteínas, MCM2 a MCM7. As proteínas MCM associam-se formando um grande complexo, o qual, em humanos, tem sido observado *in vitro* com atividade de helicase (Bryant e Aves, 2011). Além disso, foi descrito que MCM7 interage com *Cdc45* (outro gene apontado como induzido na via do ciclo celular) no complexo MCM, também apresentando atividade de helicase, permitindo a separação da dupla fita de DNA, imprescindível para o início da transcrição (Moyer *et al.*, 2006; Pacek *et al.*, 2006).

Embora *E2F1* e *E2F4* não estivessem presentes na lista dos 1.830 genes induzidos na análise conduzida pelo SAM e, conseqüentemente, não aparecessem nas vias do *KEGG*, os experimentos realizados *in vitro* confirmaram a hipótese de que esses FTs estavam superexpressos em glioblastoma, possivelmente regulando outros genes de maneira a induzir sua expressão, conforme indicado nas análises *in silico* dos dados de microarranjos, da associação de FTs por meio do FatiGO+ e da análise de assinaturas transcricionais. Nessas análises somente foram utilizados os genes induzidos em glioblastoma. Dessa forma, possivelmente esses FTs induzidos desencadeiam a indução de outros genes presentes nas diversas vias celulares, como ciclo celular, progressão tumoral e sinalização de p53. Assim, é possível supor que a expressão desregulada desses FTs, em conjunto com outros genes, pode

contribuir significativamente para a tumorigênese e progressão da doença. Essa suposição também é reforçada pelo fato de que diversos estudos têm demonstrado a associação de membros da família *E2F* ao câncer (**Tabela 4.1 – Resultados**).

E2F1 atua num ponto crítico da transição G1/S (Wyllie, 2002), sendo que as células transfectadas com *E2F1* e DP1 induziram tumores em camundongos (Johnson *et al.*, 1994), e a atividade aumentada do gene *E2F1* pode induzir tumores de pele em camundongos deficientes para o *TP53* (Pierce *et al.*, 1998). A expressão aumentada de *E2F1* (amplificação de mRNA de 8 a 12 vezes maior que o normal) foi detectada na linhagem HEL de eritroleucemia, sugerindo que a superexpressão de *E2F1* confere uma vantagem de crescimento nessas células tumorais (Saito *et al.*, 1995). Onda e colaboradores (2004) verificaram a superexpressão de *E2F1* como um evento específico em tumores de tireóide. Dessa forma, *E2F1* é requerido para a proliferação eficiente de células tumorais (Pan *et al.*, 1998). Zhang e colaboradores (2000) também propuseram a expressão aumentada de *E2F1* como um indicador de proliferação em neoplasias de mama.

No desempenho de suas funções normais, *E2F1* também possui atividade pró-apoptótica, atividade essa exclusiva desse FT dentre todos os outros membros da família *E2F* (Degregori *et al.*, 1997; Kowalik *et al.*, 1998), o que lhe confere, paradoxalmente, as funções tanto de proliferação quanto de morte celular. No entanto, a razão por trás dessa conexão é que a célula, durante o curso do seu ciclo celular, possui a capacidade de acionar um programa de autodestruição no caso de danos irreversíveis, especialmente durante a síntese de DNA. Essa tarefa é mais eficaz se ambas as funções mencionadas são coordenadas pelo mesmo fator (Tsantoulis e Gorgoulis, 2005). Um dos mecanismos pelos quais *E2F1* desempenha sua função pró-apoptótica é por meio da p53. Se o gene *TP53* estiver mutado e apresentar função de oncogene, há propensão de aumento na proliferação celular, ao invés de apoptose. Também tem sido proposto que *E2F1* pode conferir resistência a estímulos apoptóticos, possivelmente por inibição de genes pró-apoptóticos; entretanto, o mecanismo pelo qual *E2F1* inibe a apoptose ainda não foi descrito.

Em diversos estudos na área do câncer, a indução de *E2F1* é frequentemente associada à desregulação da via p53 (Gorgoulis *et al.*, 2002). Johnson e DeGregori (2006) propuseram que a combinação de superexpressão de *E2F1* com a mutação de *TP53* potencializa a tumorigênese.

Além do *TP53*, desregulação do gene codificador da pRB também pode ser crucial para a proliferação tumoral. Tsantoulis e Gorgoulis (2005) observaram que em carcinoma pulmonar os níveis protéicos de *E2F1* são aumentados em relação ao tecido normal e que há alta concentração de pRB fosforilada, contribuindo para a forma livre ativa do *E2F1*

superexpresso. Além disso, os autores relataram que níveis elevados de *E2F1* foram associados ao *status* de pRB anormal, demonstrando que *E2F1* foi desassociado do efeito inibitório de pRB. Análises da cinética tumoral revelaram que a superexpressão de *E2F1* foi positivamente relacionada ao índice de crescimento tumoral. Assim, a ligação entre *E2F1* e apoptose, nesse tipo de câncer, é rompida.

Em gliomas, especificamente, um trabalho de revisão de Alonso e colaboradores (2008) apontou que o *E2F1* desempenha a função de oncogene nesse tipo de tumor.

Embora o *E2F1* seja o mais estudado, o *E2F4* é o mais abundante dentre os FTs da família *E2F*, compondo a maior porção de proteínas E2F na maioria das células (Ikeda *et al.*, 1996; Moberg *et al.*, 1996). A superexpressão desse FT foi relatada como responsável pela indução da entrada na fase S em fibroblastos de embrião de rato, e consequente falha na indução de apoptose nas células afetadas (Lukas *et al.*, 1996; Degregori *et al.*, 1997). Rakha e colaboradores (2004) apontaram o *E2F4* como importante atuante na progressão do câncer de mama, sendo que a expressão aumentada no núcleo foi associada a tumores em estágios mais avançados de progressão e com baixa resposta a tratamentos. Verificou-se que a infecção lentiviral por shRNA para supressão da expressão de *E2F4* retardou a transição G1/S e a proliferação, tanto de células epiteliais intestinais normais quanto de células de câncer de cólon. Também foi observada a repressão de *CDK2*, ciclinas A e D1, *Cdc25A* e *c-Myc* (Garneau *et al.*, 2009). Os autores também relataram que *E2F4* e um de seus alvos, a ciclina A, estavam superexpressos, sobretudo no núcleo, em células de tumor colorretal humano quando comparadas ao epitélio normal correspondente.

No presente trabalho, a superexpressão de *E2F1* e *E2F4* pode levantar algumas hipóteses: caso *E2F4* não esteja mutado e esteja desempenhando uma de suas funções normais (a repressão do ciclo), o maior aumento de expressão de *E2F1* em relação a *E2F4* pode estar sobrepujando a ação de *E2F4* e promovendo a progressão no ciclo e, consequentemente, a proliferação tumoral; caso esteja mutado, a superexpressão desse FT pode contribuir para a progressão do glioblastoma, como já relatado em outros tipos de câncer.

Além disso, com relação a *E2F1*, esse FT liga-se seletivamente a pRB, e não a p107 ou p130 (Stevaux e Dyson, 2002). A suposta falta de indução de *RBI* (inferida pela ausência desse gene na lista de genes induzidos) pode estar associada à baixa taxa de formação de complexos pRB/E2F1/DP e, consequentemente, à falha na inativação da progressão em G1/S. De maneira semelhante, *E2F4* (embora induzido nos resultados de PCR em tempo real) pode não estar sendo recrutado por p107 (a principal proteína da família pRB que se liga a *E2F4*) e, assim, não estar exercendo sua função de repressão na progressão do ciclo celular na fase S (Giacinti e Giordano, 2006).

A ação conjunta de todos os membros da família *E2F*, assim como as funções que irão desempenhar nas células (que, muitas vezes, são antagônicas), podem ser compreendidas apenas quando os contextos celulares específicos são levados em consideração (Tsantoulis e Gorgoulis, 2005; Degregori e Johnson, 2006; Johnson e Degregori, 2006). Por exemplo, expressão desregulada de *E2F1* pode tanto promover ou inibir a tumorigênese, dependendo da natureza das mutações em outros genes envolvidos em suas interações (Alonso *et al.*, 2008).

Além da avaliação da expressão dos FTs *E2F1* e *E2F4* nas sete linhagens de glioblastoma sem qualquer tipo de tratamento, a expressão destes também foi avaliada por PCR em tempo real em células da linhagem U87 submetidas a tratamento com drogas (cisplatina, doxorrubicina e temozolomida), analisadas em dois tempos (6 e 24 h), em comparação ao controle na ausência de tratamento.

Nas últimas décadas, muitos estudos têm visado à elucidação dos mecanismos celulares desencadeados em células tumorais na presença de drogas. Esses processos são considerados primordiais para o entendimento dos mecanismos de tumorigênese e progressão tumoral, além de possibilitar o estabelecimento de novas estratégias terapêuticas. Em glioblastoma, as células apresentam acúmulo de diversas alterações genéticas (Ishii *et al.*, 1999), dentre as quais as mais frequentes são as mutações em genes relacionados ao controle do ciclo celular, reparo e apoptose. A expressão diferencial de diversos genes relacionados a essas vias foi observada no presente trabalho. Combinações diferentes de genes mutados também podem resultar em diferentes respostas a tratamento com drogas.

Além disso, com relação aos perfis de expressão gênica em glioblastoma, observados após tratamentos quimioterápicos, as diferenças de modulação na expressão de muitos genes provavelmente são devidas ao mecanismo de ação peculiar de cada droga. Os mecanismos de ação de drogas antitumorais envolvem complexas vias de resposta aos danos no DNA, as quais ainda são pouco compreendidas, apesar de uma extensa literatura. A diversidade de resposta dos tumores, além da resistência adquirida ou intrínseca a agentes citotóxicos continua sendo o principal obstáculo no tratamento do câncer (Sarkaria *et al.*, 2008).

Como resultado da análise dos dados de PCR em tempo real, os FTs não apresentaram um padrão de expressão diferencial, nem em relação ao tratamento, nem em relação ao tempo, exceto no tratamento com temozolomida, em que houve indução de *E2F1* e repressão de *E2F4* nos dois tempos, porém sem diferença estatisticamente significativa.

Diferenças significativas foram observadas no tratamento com cisplatina em 6 h (indução de *E2F1*) e 24 h (indução de *E2F4*), bem como no tratamento com doxorrubicina em 24 h (repressão de *E2F1*). As funções desempenhadas pelos produtos desses genes

diferencialmente expressos podem ser antagônicas e contribuir tanto para apoptose quanto para progressão, também dependendo do contexto celular.

As drogas utilizadas nos tratamentos das células U87 causam danos no DNA (por exemplo, formação de aductos), os quais desencadeiam cascatas de sinalização que podem ou não culminar em apoptose. Entretanto, como os genes dessas vias provavelmente não estão desempenhando suas funções normais, por se tratar de células cancerosas, os resultados pós-tratamento podem não ser totalmente previsíveis.

Com base nos resultados obtidos, principalmente nas linhagens de glioblastoma sem tratamentos, esse trabalho pode apontar o *status* de expressão dos FTs *E2F1* e *E2F4* como associado ao glioblastoma. Essa hipótese é apoiada pela evidente superexpressão do mRNA desses FTs em todas as sete linhagens de glioblastoma analisadas (T98, U251, U138, U87, U343, MO59J e MO59K), quando comparada à expressão dos mesmos FTs no mRNA de um *pool* de regiões cerebrais não-neoplásicas de diversos doadores. Uma vez que a superexpressão foi observada em todas as linhagens, também pode ser demonstrado que a expressão diferencial desses FTs é constante, independentemente das diferenças inerentes a cada linhagem (devido ao acúmulo de diferentes mutações).

Diversos trabalhos na literatura têm demonstrado a importância da desregulação dos FTs da família *E2F* em muitos tipos de câncer. De maneira complementar, alguns trabalhos, como o de Kaelin (2003), propuseram o *E2F1* como potencial alvo para terapia molecular em câncer.

Dessa forma, *E2F1* e *E2F4* podem representar genes cruciais alterados em gliomas, sendo que os mesmos requerem um estudo mais detalhado, visando à obtenção de dados que sejam importantes ou promissores para o desenvolvimento de terapia molecular, além de validações para verificar a possibilidade de tais genes serem utilizados como biomarcadores dessa doença.

Biomarcadores moleculares podem ser úteis, não somente como marcadores para diagnóstico, mas também para prognóstico, além da possibilidade de buscar alvos para novas estratégias terapêuticas (Cooper *et al.*, 2007).

Entretanto, para estabelecer genes como biomarcadores, validações cuidadosas devem ser realizadas. Estas geralmente envolvem a confirmação da expressão do produto protéico de um gene altamente expresso em nível transcricional. Isso é necessário uma vez que a maioria dos testes clínicos envolve a quantificação dos níveis protéicos, seja por imuno-histoquímica (no caso de biomarcadores em tecidos) ou por ELISA (no caso de biomarcadores no soro sanguíneo). Além disso, a validação deve ser realizada em uma grande amostra definida para

assegurar que os resultados possam ser generalizados para a população de pacientes (Rhodes e Chinnaiyan, 2004).

Sendo assim, futuras validações devem ser realizadas com o intuito de confirmar a relação entre *E2F1,4* e glioblastoma. Dentre estas, pretende-se realizar a análise proteica por *Western blot*, silenciamento desses FTs por meio de RNAi, bem como a verificação da expressão e possível mutação de outros genes também importantes na via *E2F*, como o *RBI*, o qual codifica a proteína pRB.

Em conjunto, tais informações podem ser valiosas no árduo processo de caracterização e elucidação das vias desreguladas no glioblastoma.

6. CONCLUSÕES

O presente trabalho gerou novas informações a partir de análises efetuadas sobre um conjunto de dados de glioblastoma, fornecido por bases públicas de dados de microarranjos (*GEO* e *ArrayExpress*).

Os resultados observados nas análises *in silico* foram interessantes, principalmente no que diz respeito à comparação dos FTs associados pelo programa FatiGO+ com os FTs obtidos nas anotações das assinaturas transcricionais. Nessa comparação, *E2F1* e *E2F4* destacaram-se sobremaneira.

As análises *in vitro* realizadas utilizando-se a técnica de qRT-PCR em tempo real confirmaram a expressão diferencial dos FTs *E2F1* e *E2F4*. Estes se apresentaram superexpressos em todas as sete linhagens de glioblastoma sem tratamento (T98, U251, U138, U87, U343, MO59J e MO59K), quando comparadas ao controle. Quanto aos resultados de expressão dos FTs nas células da linhagem U87 submetidas a tratamentos com drogas (cisplatina, doxorrubicina e temozolomida), não foi verificado um padrão comum de expressão, havendo resultados diferentes para cada droga testada. Diferenças significativas foram observadas no tratamento com cisplatina em 6 h (indução de *E2F1*) e 24 h (indução de *E2F4*), bem como no tratamento com doxorrubicina em 24 h (repressão de *E2F1*). Esse fato se deve provavelmente aos mecanismos peculiares de ação de cada droga e às complexas vias desencadeadas a partir desses tratamentos.

Com base nos resultados obtidos, os dados do presente trabalho podem apontar o *status* de expressão dos FTs *E2F1* e *E2F4* como possivelmente associado ao glioblastoma. Esses FTs podem representar genes cruciais alterados nessa doença, sendo que os mesmos ainda requerem estudo mais detalhado, visando à validação como biomarcadores, ou até mesmo como alvos moleculares que possam ser empregados no estabelecimento de novas modalidades de terapias.

REFERÊNCIAS BIBLIOGRÁFICAS

AL-SHAHROUR, F. *et al.* FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res* [S.I.], v. 35, n. Web Server issue, p. W91-6, Jul 2007.

_____. BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res* [S.I.], v. 33, n. Web Server issue, p. W460-4, Jul 1 2005.

ALBANI, D. *et al.* DcE2F, a functional plant E2F-like transcriptional activator from *Daucus carota*. *J Biol Chem* [S.I.], v. 275, n. 25, p. 19258-67, Jun 2000.

ALBERTS, B. *et al.* *Molecular Biology of the Cell*. 4th. ed. New York: Garland Science, 2002.

ALIZADEH, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* [S.I.], v. 403, n. 6769, p. 503-11, Feb 2000.

_____. Towards a novel classification of human malignancies based on gene expression patterns. *J Pathol* [S.I.], v. 195, n. 1, p. 41-52, Sep 2001.

ALONSO, M. M. *et al.* E2F1 in gliomas: a paradigm of oncogene addiction. *Cancer Lett* [S.I.], v. 263, n. 2, p. 157-63, May 2008.

ANONYMOUS. Microarray standards at last. v. 419: *Nature*, 2002. p. 323.

AOKI, T. *et al.* Management of glioblastoma. *Expert Opin Pharmacother* [S.I.], v. 8, n. 18, p. 3133-46, Dec 2007.

AVGEROPOULOS, N. G.; BATCHELOR, T. T. New treatment strategies for malignant gliomas. *Oncologist* [S.I.], v. 4, n. 3, p. 209-24, 1999.

BAI, R. Y. *et al.* Molecular targeting of glioblastoma: Drug discovery and therapies. *Trends Mol Med* [S.I.], Mar 2011.

BARRETT, T. *et al.* NCBI GEO: mining millions of expression profiles--database and tools. *Nucleic Acids Res* [S.I.], v. 33, n. Database issue, p. D562-6, Jan 2005.

_____. NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res* [S.I.], v. 39, n. Database issue, p. D1005-10, Jan 2011.

BARTEK, J. *et al.* Perspective: defects in cell cycle control and cancer. *J Pathol* [S.I.], v. 187, n. 1, p. 95-9, Jan 1999.

BASSI, C. *et al.* Transcriptional changes in U343 MG-a glioblastoma cell line exposed to ionizing radiation. *Hum Exp Toxicol* [S.I.], v. 27, n. 12, p. 919-29, Dec 2008.

BELZILE, J. P. *et al.* Targeting DNA repair proteins: a promising avenue for cancer gene therapy. *Curr Gene Ther* [S.I.], v. 6, n. 1, p. 111-23, Feb 2006.

BENJAMINI, Y. H., Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B.* [S.I.], v. 57, p. 289-300, 1995.

- BENTZ, G. L. *et al.* Epstein-Barr virus BRLF1 inhibits transcription of IRF3 and IRF7 and suppresses induction of interferon-beta. *Virology* [S.I.], v. 402, n. 1, p. 121-8, Jun 2010.
- BERNARD, P. S.; WITTEWER, C. T. Real-time PCR technology for cancer diagnostics. *Clin Chem* [S.I.], v. 48, n. 8, p. 1178-85, Aug 2002.
- BLACKWELL, T. K. *et al.* Sequence-specific DNA binding by the c-Myc protein. *Science* [S.I.], v. 250, n. 4984, p. 1149-51, Nov 1990.
- BLACKWOOD, E. M. *et al.* The Myc:Max protein complex and cell growth regulation. *Cold Spring Harb Symp Quant Biol* [S.I.], v. 56, p. 109-17, 1991.
- BLUTEAU, O. *et al.* Bi-allelic inactivation of TCF1 in hepatic adenomas. *Nat Genet* [S.I.], v. 32, n. 2, p. 312-5, Oct 2002.
- BRANDES, A. A. State-of-the-art treatment of high-grade brain tumors. *Semin Oncol* [S.I.], v. 30, n. 6 Suppl 19, p. 4-9, Dec 2003.
- BRAZMA, A. Minimum Information About a Microarray Experiment (MIAME)--successes, failures, challenges. *ScientificWorldJournal* [S.I.], v. 9, p. 420-3, 2009.
- BRAZMA, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* [S.I.], v. 29, n. 4, p. 365-71, Dec 2001.
- _____. ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* [S.I.], v. 31, n. 1, p. 68-71, Jan 2003.
- BREHM, A. *et al.* Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature* [S.I.], v. 391, n. 6667, p. 597-601, Feb 1998.
- BRYANT, J. A.; AVES, S. J. Initiation of DNA replication: functional and evolutionary aspects. *Ann Bot* [S.I.], Apr 2011.
- BUSTIN, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* [S.I.], v. 25, n. 2, p. 169-93, Oct 2000.
- BUSTIN, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* [S.I.], v. 55, n. 4, p. 611-22, Apr 2009.
- BÉLANGER, A. S. *et al.* Regulation of UGT1A1 and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells. *BMC Mol Biol* [S.I.], v. 11, p. 9, 2010.
- CALVERT, P. M.; FRUCHT, H. The genetics of colorectal cancer. *Ann Intern Med* [S.I.], v. 137, n. 7, p. 603-12, Oct 2002.
- CARNERO, A. Targeting the cell cycle for cancer therapy. *Br J Cancer* [S.I.], v. 87, n. 2, p. 129-33, Jul 2002.
- CARTWRIGHT, P. *et al.* E2F-6: a novel member of the E2F family is an inhibitor of E2F-dependent transcription. *Oncogene* [S.I.], v. 17, n. 5, p. 611-23, Aug 1998.
- CHEN, A. G. *et al.* Overexpression of Ets-like protein 1 in human esophageal squamous cell carcinoma. *World J Gastroenterol* [S.I.], v. 12, n. 48, p. 7859-63, Dec 2006.
- CHIANG, S. Y. *et al.* A comparison of DNA-binding drugs as inhibitors of E2F1- and Sp1-DNA complexes and associated gene expression. *Biochemistry* [S.I.], v. 37, n. 9, p. 3109-15, Mar 1998.

- CHOI, D. *et al.* Application of Biostatistics and Bioinformatics Tools to Identify Putative Transcription Factor-Gene Regulatory Network of Ankylosing Spondylitis and Sarcoidosis. *Commun Stat Theory Methods* [S.I.], v. 38, n. 18, p. 3326-3338, Jan 2009.
- CHOI, H. *et al.* A latent variable approach for meta-analysis of gene expression data from multiple microarray experiments. *BMC Bioinformatics* [S.I.], v. 8, p. 364, 2007.
- CICINNATI, V. R. *et al.* Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR. *BMC Cancer* [S.I.], v. 8, p. 350, 2008.
- CLEVERS, H. Wnt/beta-catenin signaling in development and disease. *Cell* [S.I.], v. 127, n. 3, p. 469-80, Nov 2006.
- COOPER, C. S. *et al.* Mechanisms of Disease: biomarkers and molecular targets from microarray gene expression studies in prostate cancer. *Nat Clin Pract Urol* [S.I.], v. 4, n. 12, p. 677-87, Dec 2007.
- COULSON, D. T. *et al.* Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue. *BMC Mol Biol* [S.I.], v. 9, p. 46, 2008.
- CUI, J. *et al.* Gene-expression signatures can distinguish gastric cancer grades and stages. *PLoS One* [S.I.], v. 6, n. 3, p. e17819, 2011.
- DANG, C. V. *et al.* The great MYC escape in tumorigenesis. *Cancer Cell* [S.I.], v. 8, n. 3, p. 177-8, Sep 2005.
- DAVE, S. S. *et al.* Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med* [S.I.], v. 351, n. 21, p. 2159-69, Nov 2004.
- DAVIDSON, E. *Genomic Regulatory Systems: Development and Evolution*. 1st. ed. San Diego: Academic Press, 2001.
- DE LUCA, A. *et al.* A unique domain of pRb2/p130 acts as an inhibitor of Cdk2 kinase activity. *J Biol Chem* [S.I.], v. 272, n. 34, p. 20971-4, Aug 1997.
- DE MAGALHÃES, J. P. *et al.* Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* [S.I.], v. 25, n. 7, p. 875-81, Apr 2009.
- DEALMEIDA, V. I. *et al.* The soluble wnt receptor Frizzled8CRD-hFc inhibits the growth of teratocarcinomas in vivo. *Cancer Res* [S.I.], v. 67, n. 11, p. 5371-9, Jun 2007.
- DEGREGORI, J.; JOHNSON, D. G. Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis. *Curr Mol Med* [S.I.], v. 6, n. 7, p. 739-48, Nov 2006.
- DEGREGORI, J. *et al.* Distinct roles for E2F proteins in cell growth control and apoptosis. *Proc Natl Acad Sci U S A* [S.I.], v. 94, n. 14, p. 7245-50, Jul 1997.
- DENNIS, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* [S.I.], v. 4, n. 5, p. P3, 2003.
- DENT, P. *et al.* Searching for a cure: gene therapy for glioblastoma. *Cancer Biol Ther* [S.I.], v. 7, n. 9, p. 1335-40, Sep 2008.
- DERKS, N. M. *et al.* Housekeeping genes revisited: different expressions depending on gender, brain area and stressor. *Neuroscience* [S.I.], v. 156, n. 2, p. 305-9, Oct 2008.
- DHEDA, K. *et al.* Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* [S.I.], v. 37, n. 1, p. 112-4, 116, 118-9, Jul 2004.

- DILEK, F. H. *et al.* β -Catenin and its relation to VEGF and cyclin D1 expression in pT3 rectosigmoid cancers. *Turk J Gastroenterol* [S.I.], v. 21, n. 4, p. 365-71, Dec 2010.
- DIMOVA, D. K.; DYSON, N. J. The E2F transcriptional network: old acquaintances with new faces. *Oncogene* [S.I.], v. 24, n. 17, p. 2810-26, Apr 2005.
- DOPAZO, J. *et al.* Methods and approaches in the analysis of gene expression data. *J Immunol Methods* [S.I.], v. 250, n. 1-2, p. 93-112, Apr 2001.
- DUDLEY, J. T. *et al.* Disease signatures are robust across tissues and experiments. *Mol Syst Biol* [S.I.], v. 5, p. 307, 2009.
- DUNN, I. F. *et al.* Growth factors in glioma angiogenesis: FGFs, PDGF, EGF, and TGFs. *J Neurooncol* [S.I.], v. 50, n. 1-2, p. 121-37, 2000 Oct-Nov 2000.
- DUPREE, E. L. *et al.* Genotoxic stress induces expression of E2F4, leading to its association with p130 in prostate carcinoma cells. *Cancer Res* [S.I.], v. 64, n. 13, p. 4390-3, Jul 2004.
- EISEN, M. B. *et al.* Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* [S.I.], v. 95, n. 25, p. 14863-8, Dec 1998.
- ELKON, R. *et al.* Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res* [S.I.], v. 13, n. 5, p. 773-80, May 2003.
- FISHEL, I. *et al.* Meta-analysis of gene expression data: a predictor-based approach. *Bioinformatics* [S.I.], v. 23, n. 13, p. 1599-606, Jul 2007.
- FOGARTY, M. P. *et al.* Morphing into cancer: the role of developmental signaling pathways in brain tumor formation. *J Neurobiol* [S.I.], v. 64, n. 4, p. 458-75, Sep 2005.
- GANIGI, P. M. *et al.* Expression of p53, EGFR, pRb and bcl-2 proteins in pediatric glioblastoma multiforme: a study of 54 patients. *Pediatr Neurosurg* [S.I.], v. 41, n. 6, p. 292-9, 2005 Nov-Dec 2005.
- GARNEAU, H. *et al.* E2F4 expression is required for cell cycle progression of normal intestinal crypt cells and colorectal cancer cells. *J Cell Physiol* [S.I.], v. 221, n. 2, p. 350-8, Nov 2009.
- GIACINTI, C.; GIORDANO, A. RB and cell cycle progression. *Oncogene* [S.I.], v. 25, n. 38, p. 5220-7, Aug 2006.
- GOLDBRUNNER, R. H. *et al.* Vascular endothelial growth factor-driven glioma growth and vascularization in an orthotopic rat model monitored by magnetic resonance imaging. *Neurosurgery* [S.I.], v. 47, n. 4, p. 921-9; discussion 929-30, Oct 2000.
- GOLUB, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* [S.I.], v. 286, n. 5439, p. 531-7, Oct 1999.
- GOODISON, S. *et al.* Derivation of cancer diagnostic and prognostic signatures from gene expression data. *Bioanalysis* [S.I.], v. 2, n. 5, p. 855-62, May 2010.
- GORGOLIS, V. G. *et al.* Transcription factor E2F-1 acts as a growth-promoting factor and is associated with adverse prognosis in non-small cell lung carcinomas. *J Pathol* [S.I.], v. 198, n. 2, p. 142-56, Oct 2002.
- GROSSMAN, S. A.; BATARA, J. F. Current management of glioblastoma multiforme. *Semin Oncol* [S.I.], v. 31, n. 5, p. 635-44, Oct 2004.

- GUBERN, C. *et al.* Validation of housekeeping genes for quantitative real-time PCR in in-vivo and in-vitro models of cerebral ischaemia. *BMC Mol Biol* [S.I.], v. 10, p. 57, 2009.
- GUO, J. *et al.* Genome-wide identification and expression analysis of rice cell cycle genes. *Plant Mol Biol* [S.I.], v. 64, n. 4, p. 349-60, Jul 2007.
- GUR-DEDEOGLU, B. *et al.* Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues. *Oncol Res* [S.I.], v. 17, n. 8, p. 353-65, 2009.
- HAHN, W. C.; WEINBERG, R. A. Modelling the molecular circuitry of cancer. *Nat Rev Cancer* [S.I.], v. 2, n. 5, p. 331-41, May 2002.
- HALL, M.; PETERS, G. Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer. *Adv Cancer Res* [S.I.], v. 68, p. 67-108, 1996.
- HANASH, S. Harnessing the immune response for cancer detection. *Cancer Epidemiol Biomarkers Prev* [S.I.], v. 20, n. 4, p. 569-70, Apr 2011.
- HARRIES, L. W. *et al.* Isomers of the TCF1 gene encoding hepatocyte nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes. *Hum Mol Genet* [S.I.], v. 15, n. 14, p. 2216-24, Jul 2006.
- HARTWELL, L. H.; KASTAN, M. B. Cell cycle control and cancer. *Science* [S.I.], v. 266, n. 5192, p. 1821-8, Dec 1994.
- HELIN, K. Regulation of cell proliferation by the E2F transcription factors. *Curr Opin Genet Dev* [S.I.], v. 8, n. 1, p. 28-35, Feb 1998.
- HELLEMANS, J. *et al.* qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* [S.I.], v. 8, n. 2, p. R19, 2007.
- HERMEKING, H.; EICK, D. Mediation of c-Myc-induced apoptosis by p53. *Science* [S.I.], v. 265, n. 5181, p. 2091-3, Sep 1994.
- HUANG, D. W. *et al.* Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* [S.I.], v. 4, n. 1, p. 44-57, 2009.
- HUGGETT, J. *et al.* Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* [S.I.], v. 6, n. 4, p. 279-84, Jun 2005.
- HÖGLINGER, G. U. *et al.* The pRb/E2F cell-cycle pathway mediates cell death in Parkinson's disease. *Proc Natl Acad Sci U S A* [S.I.], v. 104, n. 9, p. 3585-90, Feb 2007.
- IKEDA, M. A. *et al.* A unique role for the Rb protein in controlling E2F accumulation during cell growth and differentiation. *Proc Natl Acad Sci U S A* [S.I.], v. 93, n. 8, p. 3215-20, Apr 1996.
- IMAI, M. A. *et al.* Overexpression of E2F1 associated with LOH at RB locus and hyperphosphorylation of RB in non-small cell lung carcinoma. *J Cancer Res Clin Oncol* [S.I.], v. 130, n. 6, p. 320-6, Jun 2004.
- IRIZARRY, R. A. *et al.* Multiple-laboratory comparison of microarray platforms. *Nat Methods* [S.I.], v. 2, n. 5, p. 345-50, May 2005.
- ISHIGAKI, K. *et al.* Aberrant localization of beta-catenin correlates with overexpression of its target gene in human papillary thyroid cancer. *J Clin Endocrinol Metab* [S.I.], v. 87, n. 7, p. 3433-40, Jul 2002.

- ISHII, N. *et al.* Frequent co-alterations of TP53, p16/CDKN2A, p14ARF, PTEN tumor suppressor genes in human glioma cell lines. *Brain Pathol* [S.I.], v. 9, n. 3, p. 469-79, Jul 1999.
- JEANNOT, E. *et al.* Hepatocellular adenoma displaying a HNF1alpha inactivation in a patient with familial adenomatous polyposis coli. *J Hepatol* [S.I.], v. 45, n. 6, p. 883-6, Dec 2006.
- JIANG, H. *et al.* Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinformatics* [S.I.], v. 5, p. 81, Jun 2004.
- JOHNSON, D. G. *et al.* Oncogenic capacity of the E2F1 gene. *Proc Natl Acad Sci U S A* [S.I.], v. 91, n. 26, p. 12823-7, Dec 1994.
- JOHNSON, D. G.; DEGREGORI, J. Putting the Oncogenic and Tumor Suppressive Activities of E2F into Context. *Curr Mol Med* [S.I.], v. 6, n. 7, p. 731-8, Nov 2006.
- KAELIN, W. G. E2F1 as a target: promoter-driven suicide and small molecule modulators. *Cancer Biol Ther* [S.I.], v. 2, n. 4 Suppl 1, p. S48-54, 2003 Jul-Aug 2003.
- KANEHISA, M. *et al.* KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* [S.I.], v. 38, n. Database issue, p. D355-60, Jan 2010.
- KINZLER, K. W.; VOGELSTEIN, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* [S.I.], v. 386, n. 6627, p. 761, 763, Apr 1997.
- KLAUS, A.; BIRCHMEIER, W. Wnt signalling and its impact on development and cancer. *Nat Rev Cancer* [S.I.], v. 8, n. 5, p. 387-98, May 2008.
- KOWALIK, T. F. *et al.* E2F1-specific induction of apoptosis and p53 accumulation, which is blocked by Mdm2. *Cell Growth Differ* [S.I.], v. 9, n. 2, p. 113-8, Feb 1998.
- _____. E2F1 overexpression in quiescent fibroblasts leads to induction of cellular DNA synthesis and apoptosis. *J Virol* [S.I.], v. 69, n. 4, p. 2491-500, Apr 1995.
- KUBISTA, M. *et al.* The real-time polymerase chain reaction. *Mol Aspects Med* [S.I.], v. 27, n. 2-3, p. 95-125, 2006 Apr-Jun 2006.
- KUO, W. P. *et al.* Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* [S.I.], v. 18, n. 3, p. 405-12, Mar 2002.
- LAMMENS, T. *et al.* Atypical E2Fs: new players in the E2F transcription factor family. *Trends Cell Biol* [S.I.], v. 19, n. 3, p. 111-8, Mar 2009.
- LAMSZUS, K. *et al.* Angiogenesis-related growth factors in brain tumors. *Cancer Treat Res* [S.I.], v. 117, p. 169-90, 2004.
- LAURENT-PUIG, P. *et al.* Frequent mutations of hepatocyte nuclear factor 1 in colorectal cancer with microsatellite instability. *Gastroenterology* [S.I.], v. 124, n. 5, p. 1311-4, May 2003.
- LEE, H. K. *et al.* Coexpression analysis of human genes across many microarray data sets. *Genome Res* [S.I.], v. 14, n. 6, p. 1085-94, Jun 2004.
- LEMON, B.; TJIAN, R. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* [S.I.], v. 14, n. 20, p. 2551-69, Oct 2000.
- LEVINE, A. J.; VOSBURGH, E. P53 mutations in lymphomas: position matters. *Blood* [S.I.], v. 112, n. 8, p. 2997-8, Oct 2008.

- LIN, W. C. *et al.* Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation. *Genes Dev* [S.I.], v. 15, n. 14, p. 1833-44, Jul 2001.
- LITOFISKY, N. S. *et al.* The lack of a role for p53 in astrocytomas in pediatric patients. *Neurosurgery* [S.I.], v. 34, n. 6, p. 967-72; discussion 972-3, Jun 1994.
- LIU, H. *et al.* New roles for the RB tumor suppressor protein. *Curr Opin Genet Dev* [S.I.], v. 14, n. 1, p. 55-64, Feb 2004.
- LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* [S.I.], v. 25, n. 4, p. 402-8, Dec 2001.
- LOOTS, G.; OVCHARENKO, I. ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics* [S.I.], v. 23, n. 1, p. 122-4, Jan 2007.
- LOPEZ, F. *et al.* TranscriptomeBrowser: a powerful and flexible toolbox to explore productively the transcriptional landscape of the Gene Expression Omnibus database. *PLoS One* [S.I.], v. 3, n. 12, p. e4001, 2008.
- LOUIS, D. N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* [S.I.], v. 114, n. 2, p. 97-109, Aug 2007.
- LUKAS, J. *et al.* Deregulated expression of E2F family members induces S-phase entry and overcomes p16INK4A-mediated growth suppression. *Mol Cell Biol* [S.I.], v. 16, n. 3, p. 1047-57, Mar 1996.
- LYNG, M. B. *et al.* Identification of genes for normalization of real-time RT-PCR data in breast carcinomas. *BMC Cancer* [S.I.], v. 8, p. 20, 2008.
- MACALUSO, M. *et al.* pRb2/p130-E2F4/5-HDAC1-SUV39H1-p300 and pRb2/p130-E2F4/5-HDAC1-SUV39H1-DNMT1 multimolecular complexes mediate the transcription of estrogen receptor-alpha in breast cancer. *Oncogene* [S.I.], v. 22, n. 23, p. 3511-7, Jun 2003.
- MADHUSUDAN, S.; MIDDLETON, M. R. The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer. *Cancer Treat Rev* [S.I.], v. 31, n. 8, p. 603-17, Dec 2005.
- MALECKI, M. T. *et al.* Renal malformations may be linked to mutations in the hepatocyte nuclear factor-1alpha (MODY3) gene. *Diabetes Care* [S.I.], v. 28, n. 11, p. 2774-6, Nov 2005.
- MALLUCCI, L. *et al.* Turning cell cycle controller genes into cancer drugs. A role for an antiproliferative cytokine (betaGBP). *Biochem Pharmacol* [S.I.], v. 66, n. 8, p. 1563-9, Oct 2003.
- MANZANO, R. *et al.* Housekeeping gene expression in myogenic cell cultures from neurodegeneration and denervation animal models. *Biochem Biophys Res Commun* [S.I.], Mar 2011.
- MARASCHIN, J. E. F. *et al.* HNF1alpha mutations are present in half of clinically defined MODY patients in South-Brazilian individuals. *Arq Bras Endocrinol Metabol* [S.I.], v. 52, n. 8, p. 1326-31, Nov 2008.
- MEES, C. *et al.* Transcription factors: their potential as targets for an individualized therapeutic approach to cancer. *Cancer Gene Ther* [S.I.], v. 16, n. 2, p. 103-12, Feb 2009.
- MELDGAARD, M. *et al.* Validation of two reference genes for mRNA level studies of murine disease models in neurobiology. *J Neurosci Methods* [S.I.], v. 156, n. 1-2, p. 101-10, Sep 2006.

- MICHIELS, S. *et al.* Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* [S.I.], v. 365, n. 9458, p. 488-92, 2005 Feb 5-11 2005.
- MOBERG, K. *et al.* E2F-4 switches from p130 to p107 and pRB in response to cell cycle reentry. *Mol Cell Biol* [S.I.], v. 16, n. 4, p. 1436-49, Apr 1996.
- MOK, S. C. *et al.* Prostatin, a potential serum marker for ovarian cancer: identification through microarray technology. *J Natl Cancer Inst* [S.I.], v. 93, n. 19, p. 1458-64, Oct 2001.
- MOREAU, Y. *et al.* Comparison and meta-analysis of microarray data: from the bench to the computer desk. *Trends Genet* [S.I.], v. 19, n. 10, p. 570-7, Oct 2003.
- MOYER, S. E. *et al.* Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A* [S.I.], v. 103, n. 27, p. 10236-41, Jul 2006.
- NAJDI, R. *et al.* A Wnt kinase network alters nuclear localization of TCF-1 in colon cancer. *Oncogene* [S.I.], v. 28, n. 47, p. 4133-46, Nov 2009.
- NARAYANA, N. *et al.* Diabetes mellitus due to misfolding of a beta-cell transcription factor: stereospecific frustration of a Schellman motif in HNF-1alpha. *J Mol Biol* [S.I.], v. 362, n. 3, p. 414-29, Sep 2006.
- NELSON, M. A. *et al.* Increased gene copy number of the transcription factor E2F1 in malignant melanoma. *Cancer Biol Ther* [S.I.], v. 5, n. 4, p. 407-12, Apr 2006.
- NILSSON, J. A.; CLEVELAND, J. L. Myc pathways provoking cell suicide and cancer. *Oncogene* [S.I.], v. 22, n. 56, p. 9007-21, Dec 2003.
- OLDAK, M. *et al.* Differential regulation of human papillomavirus type 8 by interferon regulatory factors 3 and 7. *J Virol* [S.I.], v. 85, n. 1, p. 178-88, Jan 2011.
- ONDA, M. *et al.* Up-regulation of transcriptional factor E2F1 in papillary and anaplastic thyroid cancers. *J Hum Genet* [S.I.], v. 49, n. 6, p. 312-8, 2004.
- OPAVSKY, R. *et al.* Specific tumor suppressor function for E2F2 in Myc-induced T cell lymphomagenesis. *Proc Natl Acad Sci U S A* [S.I.], v. 104, n. 39, p. 15400-5, Sep 2007.
- PACEK, M. *et al.* Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* [S.I.], v. 21, n. 4, p. 581-7, Feb 2006.
- PAL, A. *et al.* Evaluation of serum 1,5 anhydroglucitol levels as a clinical test to differentiate subtypes of diabetes. *Diabetes Care* [S.I.], v. 33, n. 2, p. 252-7, Feb 2010.
- PAN, H. *et al.* Key roles for E2F1 in signaling p53-dependent apoptosis and in cell division within developing tumors. *Mol Cell* [S.I.], v. 2, n. 3, p. 283-92, Sep 1998.
- PARK, M. T.; LEE, S. J. Cell cycle and cancer. *J Biochem Mol Biol* [S.I.], v. 36, n. 1, p. 60-5, Jan 2003.
- PARKIN, D. M. *et al.* Estimating the world cancer burden: Globocan 2000. *Int J Cancer* [S.I.], v. 94, n. 2, p. 153-6, Oct 15 2001.
- PARKINSON, H. *et al.* ArrayExpress update--from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res* [S.I.], v. 37, n. Database issue, p. D868-72, Jan 2009.

_____. ArrayExpress update--an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res* [S.I.], v. 39, n. Database issue, p. D1002-4, Jan 2011.

_____. ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* [S.I.], v. 33, n. Database issue, p. D553-5, Jan 2005.

PFAFFL, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* [S.I.], v. 29, n. 9, p. e45, May 2001.

PFAFFL, M. W. *et al.* Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* [S.I.], v. 30, n. 9, p. e36, May 2002.

PIERCE, A. M. *et al.* Increased E2F1 activity induces skin tumors in mice heterozygous and nullizygous for p53. *Proc Natl Acad Sci U S A* [S.I.], v. 95, n. 15, p. 8858-63, Jul 1998.

PIESSEN, G. *et al.* Regulation of the human mucin MUC4 by taurodeoxycholic and taurochenodeoxycholic bile acids in oesophageal cancer cells is mediated by hepatocyte nuclear factor 1alpha. *Biochem J* [S.I.], v. 402, n. 1, p. 81-91, Feb 2007.

PINÉS CORRALES, P. J. *et al.* Clinical differences between patients with MODY-3, MODY-2 and type 2 diabetes mellitus with I27L polymorphism in the HNF1alpha gene. *Endocrinol Nutr* [S.I.], v. 57, n. 1, p. 4-8, Jan 2010.

POLLACK, I. F. *et al.* Age and TP53 mutation frequency in childhood malignant gliomas: results in a multi-institutional cohort. *Cancer Res* [S.I.], v. 61, n. 20, p. 7404-7, Oct 2001.

_____. The relationship between TP53 mutations and overexpression of p53 and prognosis in malignant gliomas of childhood. *Cancer Res* [S.I.], v. 57, n. 2, p. 304-9, Jan 1997.

PROST, S. *et al.* E2F regulates DDB2: consequences for DNA repair in Rb-deficient cells. *Oncogene* [S.I.], v. 26, n. 24, p. 3572-81, May 2007.

QIN, X. Q. *et al.* Deregulated transcription factor E2F-1 expression leads to S-phase entry and p53-mediated apoptosis. *Proc Natl Acad Sci U S A* [S.I.], v. 91, n. 23, p. 10918-22, Nov 1994.

RAKHA, E. A. *et al.* Expression of E2F-4 in invasive breast carcinomas is associated with poor prognosis. *J Pathol* [S.I.], v. 203, n. 3, p. 754-61, Jul 2004.

RAMOS, M. *et al.* Management of squamous cell carcinoma of the head and neck: updated European treatment recommendations. *Expert Rev Anticancer Ther* [S.I.], v. 10, n. 3, p. 339-44, Mar 2010.

RAMÍREZ-PARRA, E. *et al.* The cloning of plant E2F, a retinoblastoma-binding protein, reveals unique and conserved features with animal G(1)/S regulators. *Nucleic Acids Res* [S.I.], v. 27, n. 17, p. 3527-33, Sep 1999.

REARDON, D. A.; WEN, P. Y. Therapeutic advances in the treatment of glioblastoma: rationale and potential role of targeted agents. *Oncologist* [S.I.], v. 11, n. 2, p. 152-64, Feb 2006.

REBOUISSOU, S. *et al.* Mutation of TCF1 encoding hepatocyte nuclear factor 1alpha in gynecological cancer. *Oncogene* [S.I.], v. 23, n. 45, p. 7588-92, Sep 2004.

REICHEL, R. *et al.* Developmental control of a promoter-specific factor that is also regulated by the E1A gene product. *Cell* [S.I.], v. 48, n. 3, p. 501-6, Feb 1987.

- REIFENBERGER, G.; COLLINS, V. P. Pathology and molecular genetics of astrocytic gliomas. *J Mol Med* [S.I.], v. 82, n. 10, p. 656-70, Oct 2004.
- REIMER, D. *et al.* Expression of the E2F family of transcription factors and its clinical relevance in ovarian cancer. *Ann N Y Acad Sci* [S.I.], v. 1091, p. 270-81, Dec 2006.
- RHO, H. W. *et al.* Identification of valid reference genes for gene expression studies of human stomach cancer by reverse transcription-qPCR. *BMC Cancer* [S.I.], v. 10, p. 240, 2010.
- RHODES, D. R.; CHINNAIYAN, A. M. Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. *Ann N Y Acad Sci* [S.I.], v. 1020, p. 32-40, May 2004.
- _____. Integrative analysis of the cancer transcriptome. *Nat Genet* [S.I.], v. 37 Suppl, p. S31-7, Jun 2005.
- RHODES, D. R. *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* [S.I.], v. 9, n. 2, p. 166-80, Feb 2007.
- _____. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* [S.I.], v. 6, n. 1, p. 1-6, 2004 Jan-Feb 2004.
- ROBERTS, J. D. E2F1 amplification and genetic heterogeneity in melanoma. *Cancer Biol Ther* [S.I.], v. 5, n. 6, p. 691-2, Jun 2006.
- ROBERTSON, G. *et al.* cisRED: a database system for genome-scale computational discovery of regulatory elements. *Nucleic Acids Res* [S.I.], v. 34, n. Database issue, p. D68-73, Jan 2006.
- ROMANOWSKI, T. *et al.* [Housekeeping genes as a reference in quantitative real-time RT-PCR]. *Postepy Hig Med Dosw (Online)* [S.I.], v. 61, p. 500-10, 2007.
- ROSSER, C. J.; GOODISON, S. Today's discoveries to tomorrow's care: cancer biomarkers revisited. Foreword. *Biomark Med* [S.I.], v. 4, n. 4, p. 491-3, Aug 2010.
- ROTH, W. *et al.* Secreted Frizzled-related proteins inhibit motility and promote growth of human malignant glioma cells. *Oncogene* [S.I.], v. 19, n. 37, p. 4210-20, Aug 2000.
- ROWLEY, C. W. *et al.* Mechanisms of mutual functional interactions between HNF-4alpha and HNF-1alpha revealed by mutations that cause maturity onset diabetes of the young. *Am J Physiol Gastrointest Liver Physiol* [S.I.], v. 290, n. 3, p. G466-75, Mar 2006.
- RUTHERFORD, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* [S.I.], v. 16, n. 10, p. 944-5, Oct 2000.
- SAEED, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* [S.I.], v. 34, n. 2, p. 374-8, Feb 2003.
- SAH, N. K. *et al.* Structural, functional and therapeutic biology of survivin. *Cancer Lett* [S.I.], v. 244, n. 2, p. 164-71, Dec 2006.
- SAITO, M. *et al.* Amplification of the E2F1 transcription factor gene in the HEL erythroleukemia cell line. *Genomics* [S.I.], v. 25, n. 1, p. 130-8, Jan 1995.
- SALON, C. *et al.* E2F1 induces apoptosis and sensitizes human lung adenocarcinoma cells to death-receptor-mediated apoptosis through specific downregulation of c-FLIP(short). *Cell Death Differ* [S.I.], v. 13, n. 2, p. 260-72, Feb 2006.

- SARDET, C. *et al.* E2F-4 and E2F-5, two members of the E2F family, are expressed in the early phases of the cell cycle. *Proc Natl Acad Sci U S A* [S.I.], v. 92, n. 6, p. 2403-7, Mar 1995.
- SARKARIA, J. N. *et al.* Mechanisms of chemoresistance to alkylating agents in malignant glioma. *Clin Cancer Res* [S.I.], v. 14, n. 10, p. 2900-8, May 2008.
- SATHISH, N. *et al.* Mechanisms of autoinhibition of IRF-7 and a probable model for inactivation of IRF-7 by Kaposi's sarcoma-associated herpesvirus protein ORF45. *J Biol Chem* [S.I.], v. 286, n. 1, p. 746-56, Jan 2011.
- SATHORNSUMETEE, S. *et al.* Molecularly targeted therapy for malignant glioma. *Cancer* [S.I.], v. 110, n. 1, p. 13-24, Jul 2007.
- SEGAL, E. *et al.* From signatures to models: understanding cancer using microarrays. *Nat Genet* [S.I.], v. 37 Suppl, p. S38-45, Jun 2005.
- SHAN, B.; LEE, W. H. Deregulated expression of E2F-1 induces S-phase entry and leads to apoptosis. *Mol Cell Biol* [S.I.], v. 14, n. 12, p. 8166-73, Dec 1994.
- SHAO, J. *et al.* Prostaglandin E2 Stimulates the beta-catenin/T cell factor-dependent transcription in colon cancer. *J Biol Chem* [S.I.], v. 280, n. 28, p. 26565-72, Jul 2005.
- SHI, L. *et al.* The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* [S.I.], v. 28, n. 8, p. 827-38, Aug 2010.
- _____. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* [S.I.], v. 24, n. 9, p. 1151-61, Sep 2006.
- SORLIE, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* [S.I.], v. 100, n. 14, p. 8418-23, Jul 2003.
- SOTIRIOU, C. *et al.* Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* [S.I.], v. 100, n. 18, p. 10393-8, Sep 2003.
- STAMOVA, B. S. *et al.* Identification and validation of suitable endogenous reference genes for gene expression studies in human peripheral blood. *BMC Med Genomics* [S.I.], v. 2, p. 49, 2009.
- STEER, H. J. *et al.* Harnessing the immune response to treat cancer. *Oncogene* [S.I.], v. 29, n. 48, p. 6301-13, Dec 2010.
- STEVAUX, O.; DYSON, N. J. A revised picture of the E2F transcriptional network and RB function. *Curr Opin Cell Biol* [S.I.], v. 14, n. 6, p. 684-91, Dec 2002.
- STUART, J. M. *et al.* A gene-coexpression network for global discovery of conserved genetic modules. *Science* [S.I.], v. 302, n. 5643, p. 249-55, Oct 2003.
- STUPP, R. *et al.* Current and future developments in the use of temozolomide for the treatment of brain tumours. *Lancet Oncol* [S.I.], v. 2, n. 9, p. 552-60, Sep 2001.
- SUN, L. *et al.* Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* [S.I.], v. 9, n. 4, p. 287-300, Apr 2006.
- SUZUKI, T. *et al.* Control selection for RNA quantitation. *Biotechniques* [S.I.], v. 29, n. 2, p. 332-7, Aug 2000.

- TESSEMA, M. *et al.* Cell cycle and no end. *Virchows Arch* [S.I.], v. 444, n. 4, p. 313-23, Apr 2004.
- THIERRY-MIEG, D.; THIERRY-MIEG, J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* [S.I.], v. 7 Suppl 1, p. S12.1-14, 2006.
- TOMLINS, S. A. *et al.* Integrative biology of prostate cancer progression. *Annu Rev Pathol* [S.I.], v. 1, p. 243-71, 2006.
- TONG, Z. *et al.* Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Mol Biol* [S.I.], v. 10, p. 71, 2009.
- TSANTOULIS, P. K.; GORGOULIS, V. G. Involvement of E2F transcription factor family in cancer. *Eur J Cancer* [S.I.], v. 41, n. 16, p. 2403-14, Nov 2005.
- TSITOURA, E. *et al.* Infection with herpes simplex type 1-based amplicon vectors results in an IRF3/7-dependent, TLR-independent activation of the innate antiviral response in primary human fibroblasts. *J Gen Virol* [S.I.], v. 90, n. Pt 9, p. 2209-20, Sep 2009.
- TSUGE, M. *et al.* A variable number of tandem repeats polymorphism in an E2F-1 binding element in the 5' flanking region of SMYD3 is a risk factor for human cancers. *Nat Genet* [S.I.], v. 37, n. 10, p. 1104-7, Oct 2005.
- TUSHER, V. G. *et al.* Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* [S.I.], v. 98, n. 9, p. 5116-21, Apr 24 2001.
- UMEYAMA, K. *et al.* Dominant-negative mutant hepatocyte nuclear factor 1alpha induces diabetes in transgenic-cloned pigs. *Transgenic Res* [S.I.], v. 18, n. 5, p. 697-706, Oct 2009.
- VALENTE, V. *et al.* Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. *BMC Mol Biol* [S.I.], v. 10, p. 17, 2009.
- VAN DE VIJVER, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* [S.I.], v. 347, n. 25, p. 1999-2009, Dec 2002.
- VAN DEN HEUVEL, S.; DYSON, N. J. Conserved functions of the pRB and E2F families. *Nat Rev Mol Cell Biol* [S.I.], v. 9, n. 9, p. 713-24, Sep 2008.
- VANDESOMPELE, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* [S.I.], v. 3, n. 7, p. RESEARCH0034, Jun 2002.
- WAGNER, A. J. *et al.* Myc-mediated apoptosis requires wild-type p53 in a manner independent of cell cycle arrest and the ability of p53 to induce p21waf1/cip1. *Genes Dev* [S.I.], v. 8, n. 23, p. 2817-30, Dec 1994.
- WALTER, K. A. *et al.* Scatter factor/hepatocyte growth factor stimulation of glioblastoma cell cycle progression through G(1) is c-Myc dependent and independent of p27 suppression, Cdk2 activation, or E2F1-dependent transcription. *Mol Cell Biol* [S.I.], v. 22, n. 8, p. 2703-15, Apr 2002.
- WARNAT, P. *et al.* Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC Bioinformatics* [S.I.], v. 6, p. 265, 2005.
- WATHELET, M. G. *et al.* Virus infection induces the assembly of coordinately activated transcription factors on the IFN-beta enhancer in vivo. *Mol Cell* [S.I.], v. 1, n. 4, p. 507-18, Mar 1998.
- WELSH, J. B. *et al.* Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc Natl Acad Sci U S A* [S.I.], v. 100, n. 6, p. 3410-5, Mar 2003.

_____. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci U S A* [S.I.], v. 98, n. 3, p. 1176-81, Jan 2001.

WHITE, N. M. *et al.* Dysregulation of kallikrein-related peptidases in renal cell carcinoma: potential targets of miRNAs. *Biol Chem* [S.I.], v. 391, n. 4, p. 411-23, Apr 2010.

WINGENDER, E. *et al.* TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* [S.I.], v. 28, n. 1, p. 316-9, Jan 1 2000.

WODARZ, A.; NUSSE, R. Mechanisms of Wnt signaling in development. *Annu Rev Cell Dev Biol* [S.I.], v. 14, p. 59-88, 1998.

WOO, I. S. *et al.* Infrequent mutations of the activating transcription factor-2 gene in human lung cancer, neuroblastoma and breast cancer. *Int J Oncol* [S.I.], v. 20, n. 3, p. 527-31, Mar 2002.

WU, X.; LEVINE, A. J. p53 and E2F-1 cooperate to mediate apoptosis. *Proc Natl Acad Sci U S A* [S.I.], v. 91, n. 9, p. 3602-6, Apr 1994.

WYLLIE, A. H. E2F1 selects tumour cells for both life and death. *J Pathol* [S.I.], v. 198, n. 2, p. 139-41, Oct 2002.

XIAO, D. *et al.* GRP receptor-mediated immediate early gene expression and transcription factor Elk-1 activation in prostate cancer cells. *Regul Pept* [S.I.], v. 109, n. 1-3, p. 141-8, Nov 2002.

YUEN, M. F.; LAI, C. L. Serological markers of liver cancer. *Best Pract Res Clin Gastroenterol* [S.I.], v. 19, n. 1, p. 91-9, Feb 2005.

ZALMAS, L. P. *et al.* DNA-damage response control of E2F7 and E2F8. *EMBO Rep* [S.I.], v. 9, n. 3, p. 252-9, Mar 2008.

ZHANG, L.; PAGANO, J. S. IRF-7, a new interferon regulatory factor associated with Epstein-Barr virus latency. *Mol Cell Biol* [S.I.], v. 17, n. 10, p. 5748-57, Oct 1997.

ZHANG, S. Y. *et al.* E2F-1: a proliferative marker of breast neoplasia. *Cancer Epidemiol Biomarkers Prev* [S.I.], v. 9, n. 4, p. 395-401, Apr 2000.

ZHAO, X.; WANG, M. [Clinical Utility of Serum Tumor Markers in Lung Cancer.]. *Zhongguo Fei Ai Za Zhi* [S.I.], v. 14, n. 3, p. 286-291, Mar 2011.

ZIOBER, A. F. *et al.* Is gene expression profiling of head and neck cancers ready for the clinic? *Biomark Med* [S.I.], v. 4, n. 4, p. 571-80, Aug 2010.