Ulises Maximiliano Mancini Villagra

Análise de *splicing* alternativo utilizando dados de
sequências expressas

São Paulo
2009

Ulises Maximiliano Mancini Villagra

Análise de *splicing* alternativo utilizando dados de sequências expressas

Tese Apresentada ao Instituto de
Biociências da Universidade de São Paulo,
para obtenção de Título de Doutor em
Ciências, na Área de Genética.

Orientadora: Eloiza Helena Tajara da Silva

São Paulo
2009

Mancini Villagra, Ulises Maximiliano
Análise de *splicing* alternativo utilizando dados de sequências
expressas
102pp

Tese (Doutorado) - Instituto de Biociências da Universidade de São
Paulo.

Departamento de Genética e Evolução
1- *Splicing* alternativo. 2- Transcritoma

Universidade de São Paulo. Instituto de Biociências.
Departamento de Genética e Evolução

Comissão Julgadora:

Prof(a) Dr(a)

Prof(a) Dr(a)

Prof(a) Dr(a)

Prof(a) Dr(a)

Profa. Dra.
Orientadora

À Deus
À minha família

AGRADECIMENTOS

À minha orientadora, Eloiza Helena Tajara pela sua sabedoria, pela sua inesgotável paciência e pelos seus incansáveis conselhos afetuosos.

Ao Instituto de Biociências da Universidade de São Paulo - IB/USP, por ter-me acolhido e oferecido um lugar sonhado para o desenvolvimento do meu doutorado.

Aos coordenadores, professores, alunos e funcionários do programa de Pós-Graduação em Genética do Instituto de Biociências da Universidade de São Paulo - IB/USP

À Faculdade de Medicina de São José do Rio Preto - FAMERP, por ter-me brindado o espaço físico para o desenvolvimento das minhas pesquisas durante o meu doutorado

À Fundação de Amparo À Pesquisa do Estado de São Paulo - FAPESP, pela bolsa de doutorado com a que fui deferido, sem a qual este trabalho e minha permanência no Brasil não teriam sido possíveis e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelos auxílios financeiros proporcionados ao nosso laboratório.

Aos Profs. Drs. Emmanuel Dias Neto, Elida Ojopi, Sandro de Souza, Paulo Oliveira, Andréia Leopoldino e Carlos Tomich pela sua valiosa participação no meu trabalho de doutorado.

Aos grupos de pesquisadores do *Head and Neck Genome Project* - GENCAPO e *Head and Neck Transcriptome Initiative*.

Aos meus pais Sara e Giovanni e ao meu irmão Fernando pelo seu incondicional apoio, carinho, amor, dedicação e paciência e à Sandra, pela sua serenidade, orientação, carinho, cuidado e amor, e pelos nossos projetos, pelos nossos planos e pelos nossos sonhos. Te Amo.

Aos meus colegas do Laboratório de Marcadores Moleculares e Bioinformática Médica pela sua amizade, carinho, conselhos, cumplicidade e paciência. A todos eles, àqueles que estão e àqueles que já não estão, e que sempre formarão parte das minhas lembranças.

A Rhamses e à Margarita porque são parte essencial da minha vida, do meu dia a dia. Eles não falam com palavras, mas se comunicam com um olhar cheio de alegria e amor. Amo vocês, para todo e sempre.

ÍNDICE

Capitulo 1.	
Introdução Geral	1
<i>Splicing</i> do transcrito primário	1
<i>Splicing</i> alternativo	3
<i>Splicing</i> alternativo e doenças humanas	5
Objetivos	8
Capitulo 2.	9
Introdução	9
Materiais e métodos	9
Resultados e Discussão	9
Trabalho <i>Cancer Research</i> 2005	11
Capitulo 3.	
Introdução	18
Materiais e métodos	18
Resultados e Discussão	19
Trabalho <i>Genome</i> 2007	21
Capitulo 4.	
Introdução	33
Materiais e métodos	34
Resultados e Discussão	35
Trabalho <i>RNA</i> 2009 (em submissão)	36
Capitulo 5.	
Introdução	62
Materiais e métodos	63
Resultados e Discussão	72
Abstract	93
Ressumo	94
Conclusões	95
Referencias Bibliográficas	96

Capítulo 1

I - Introdução Geral

Os transcritos primários, ou pré-RNAs mensageiros (mRNAs), da maioria dos eucariotos são processados após sua síntese. Esse processamento, que gera um transcrito maduro funcional, inclui adição de uma 7-metilguanossina na extremidade 5' da molécula, clivagem e poliadenilação da extremidade 3' e desaminação de nucleotídeos específicos. As sequências não codificadoras (íntrons) são também removidas e as sequências codificadoras (exons) tornam-se contíguas por um mecanismo conhecido como *splicing*.

Desde sua síntese até seu destino no citoplasma, os pré-mRNAs e os mRNAs são associados a proteínas nucleares abundantes, denominadas ribonucleoproteínas heterogêneas nucleares (hnRNPs), formando complexos ribonucleoprotéicos (RNPs). As hnRNPs atuam em todos os aspectos da biologia dos RNAs, incluindo transcrição, *splicing*, poliadenilação transporte, localização, tradução e estabilidade (Dreyfuss *et al.*, 1993).

A maioria das hnRNP e de outras proteínas ligadoras de RNA possui estrutura modular e apresenta um ou mais domínios de reconhecimento e ligação ao RNA (Varani & Nagai, 1998). Entre eles, estão o RBD (*RNA-binding domain*, também conhecido como RRM ou *RNA recognition motif*), o domínio KH *K-homology*) e o box RGG (Arg-Gly-Gly) revisto por (Glisovic *et al.*, 2008).

***Splicing* do Transcrito Primário**

Aproximadamente 95% dos genes humanos contêm exons separados por íntrons. A maioria dos exons apresenta uma centena de nucleotídeos enquanto os íntrons exibem em média mais de mil pares de bases (pb) e alguns podem atingir até 500 mil pb (Rowen *et al.*, 2002). Os íntrons possuem sequências consenso curtas fundamentais para o processamento do RNA. As mais conservadas estão presentes nas suas extremidades 5' (ou sítio doador) e 3' (ou sítio aceptor) e geralmente compreendem um par de nucleotídeos GU e AG (Goldstrohm *et al.*, 2001). Outras duas regiões próximas ao sítio

aceptor também são importantes: uma sequência de aproximadamente 15 nucleotídeos rica em pirimidinas (PTT) e uma sequência chamada sequência de ramificação composta de sete a oito bases na qual a penúltima delas, nos vertebrados, é uma adenina (Aebi *et al.*, 1986). Nos casos em que a sequência de ramificação está ausente, a maquinaria de *splicing* busca resíduos de adenina na região 3' do íntron e, eventualmente, utiliza sítios doadores e aceptores alternativos (Lodish *et al.*, 2000; Lewin 2000).

Os íntrons de transcritos primários de eucariotos superiores, ao contrário daqueles observados em genomas de organelas de plantas e de algumas bactérias, necessitam de ATP para *splicing*. Também são necessários RNAs nucleares pequenos (snRNAs), que se associam a proteínas e formam as ribonucleoproteínas pequenas nucleares (snRNPs). Os componentes protéicos das snRNPs interagem entre si durante o processamento do RNA, modulando e estabilizando a estrutura tridimensional dos complexos formados (Kohtz *et al.*, 1994). São conhecidas cinco snRNPs (U1, U2, U4, U5 e U6) e cerca de 300 fatores de *splicing* não-snRNPs (Kramer, 1996; Patel & Bellini, 2008).

Durante a fase inicial do *splicing*, as snRNPs hibridam com as sequências consenso do transcrito levando à formação de um complexo ribonucleoprotéico denominado *spliceosome*. Esse grande complexo ribonucleoprotéico cataliza duas reações de transesterificação envolvendo grupos funcionais de três regiões reativas do pré-mRNA. Inicialmente, a hidroxila 2' da adenosina da sequência de ramificação intrônica forma um enlace 2',5' fosfodiéster com uma guanina do sítio doador 5'. Como consequência, é gerada uma extremidade 3' livre do exon e o íntron adota uma estrutura de laço ou *lariat* entre a sua extremidade 5' e a adenosina da sequência de ramificação. Nesse momento, a hidroxila 3' da extremidade livre do exon atua como nucleófilo e liga-se por meio de outra reação de transesterificação ao nucleotídeo 5' do exon seguinte, resultando na excisão do íntron (van Alphen *et al.*, 2009).

É o *spliceosome* o responsável por mudanças estruturais do transcrito que permitem o correto reconhecimento e o pareamento de sítios de *splicing* com diferentes ribonucleoproteínas, bem como o posicionamento desses sítios de modo a favorecer as reações de transesterificação.

As snRNPs U1, U2, U4/U6 e U5 são as unidades funcionais principais do *spliceosome*, cuja montagem tem início com a formação do complexo E (*early complex*), comumente chamado de complexo de compromisso (CC) porque sua presença identifica

um pré-mRNA comprometido com o *splicing*. Nessa etapa, a snRNP U1 liga-se por pareamento de bases ao sítio doador 5'. Tal interação é estabilizada por membros da família de proteínas ricas em serina e arginina (SR) e pelas proteínas presentes em U1. Esta fase inicial do *spliceosome* também envolve a ligação da proteína SF1/BBP e do fator auxiliar U2 (U2AF) à sequência de ramificação e à sequência adjacente rica em pirimidinas (PTT), respectivamente. Essas proteínas interagem cooperativamente por meio da extremidade C-terminal de SF1/BBP com a subunidade de 65 kDa de U2AF. A subunidade de 35 kDa de U2AF liga-se também aos dinucleotídeos AG do sítio acceptor 3' (revisto por Wahl *et al.*, 2009).

Em seguida, a snRNP U2 inicia, de forma ATP-dependente, o pareamento de bases com a sequência de ramificação, que é estabilizado pelas subunidades de U2, SF3a e SF3b (Gozani *et al.*, 1996), e pelo domínio rico em serinas e argininas da proteína U2AF (Valcarcel, *et al.*, 1996). A associação de U2 (Complexo A) promove a substituição de SF1/BBP pela proteína SF3b na sequência de ramificação (Will *et al.*, 2001).

Após o pareamento de U2, as snRNPs U4/U6 e U5 são recrutadas para o local e formam o complexo B. A saída de U1 e U4 e o pareamento de U6 com o sítio doador 5' e com U2, ativam o complexo B, que promove a primeira reação catalítica do *splicing* e determina a formação do complexo C (Konarska *et al.*, 2006). Após a segunda reação catalítica, as snRNPs U2, U5 e U6 e o íntron são liberados, sendo as primeiras reutilizadas em novos eventos de processamento de RNA e o íntron é degradado por nucleases (Nelson & Cox, 2000).

Splicing Alternativo

Além do aumento no número de genes, a complexidade e a diversidade do genoma podem ser ampliadas durante a evolução pela utilização de mecanismos que geram diversidade protéica. Entre os principais mecanismos que geram essa diversidade nos eucariotos, estão aqueles relacionados com o processamento do RNA, como a utilização de diferentes sítios de início de transcrição e de poliadenilação, as modificações no padrão de edição do mRNA e o *splicing* alternativo. Este último é certamente um dos mais importantes geradores de diversidade protéica e controle de níveis de expressão gênica em vertebrados (Black, 2000; Ben-Dov *et al.*, 2008; Wang *et al.*, 2008; Villate *et al.*, 2008), que permite o enriquecimento do transcrito e,

portanto, do proteoma de um organismo, sem a necessidade de expansão do genoma. O *splicing* alternativo pode ser regulado diferencialmente dependendo do tipo celular, do estágio do desenvolvimento ou por diferentes padrões de sinalização (Cooper *et al.*, 2009) e mostra um interessante nível de conservação, o que o torna uma poderosa ferramenta de predição de ancestrais eucarióticos comuns (Irimia *et al.*, 2007; Kim *et al.*, 2007; Roy & Penni, 2007; Irimia & Roy, 2008; Irimia *et al.*, 2009). Além disso, é um processo que explica a disparidade entre o pequeno número de genes na espécie humana sugerido pelo seqüenciamento do seu genoma (Venter *et al.*, 2001; Lander *et al.*, 2001) e a complexidade observada em relação a outros organismos com número similar de genes, como o *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium**, 1998).

Os dados de algumas análises recentes sugerem que cerca de 90% dos genes humanos exibem *splicing* alternativo (Kampa *et al.*, 2004; Pan *et al.*, 2008; Wang *et al.*, 2008). Existem exemplos nos quais as possibilidades de *splicing* são enormes, como os casos dos genes *TTN* (Titin) com 316 exons, *RYR1* (*Rainodyn Receptor 1*) com 106 exons, *DMD* (distrofin) com 79 exons e *NF1* (neurofibromin 1) com 58 exons, o que reforça o potencial desse mecanismo na variabilidade genética constitucional e somática.

No processo de *splicing*, diferentes íntrons ou exons (ou partes de íntrons e exons) podem ser alternativamente removidos ou incluídos, e os sítios 5' e 3' de um dado íntron podem ser diferencialmente selecionados levando a mudanças no comprimento do exon adjacente (Goldstrohm *et al.*, 2001).

A seleção do sítio de *splicing* nos eucariotos superiores é determinada por múltiplos fatores. Sítios que apresentam grande afinidade por U1 ou U2AF, em geral, são mais eficientemente reconhecidos pela maquinaria celular. Entretanto, como muitas sequência consenso são relativamente degeneradas, a composição dos sítios de *splicing* não é capaz sozinha de direcionar a eficiência do *spliceosome*. Na verdade, a seleção e o reconhecimento desses sítios são na maioria dos casos influenciados por sequências reguladoras presentes no pré-mRNA, conhecidas como estimuladores ou silenciadores (*enhancers* ou *silencers*) intrônicos (ISEs e ISSs, respectivamente) ou exônicos (ESEs e ESSs) (Graveley, 2000; Cartegni *et al.*, 2002; Singh & Valcarcel, 2005). Essas sequências *cis* atuam como regiões de ligação de elementos reguladores *trans* que, por sua vez,

recrutam as snRNPs ou, nos casos de reguladores negativos, previnem sua associação ao pré-RNA.

Muitas das escolhas dos sítios de *splicing* são feitas durante as etapas iniciais da formação do *spliceosome*. Como referido acima, essas escolhas são determinadas pela ação combinada de fatores reguladores positivos e negativos (Smith & Valcarcel, 2000). No entanto, estudos recentes indicam que essas decisões também podem ser efetuadas em estágios posteriores do processo, incluindo a transição do complexo A para o complexo B (Bonnal *et al.*, 2008; Sharma *et al.*, 2008), ou mesmo depois da primeira reação de transesterificação (Lallena *et al.*, 2002).

***Splicing* alternativo e doenças humanas**

Aproximadamente 75% dos exons alternativos codificam partes importantes da proteína (Blencowe, 2006) e possuem potencial para gerar múltiplas proteínas a partir de um único gene. Propriedades biológicas novas podem ser adquiridas e envolver mudanças em interações proteína-proteína, localização celular ou habilidades catalíticas. O mapeamento de regiões alternativamente processadas de proteínas com estruturas conhecidas sugere que muitas dessas regiões são intrinsicamente desordenadas e, portanto, devem permitir diversidade funcional sem levar a grandes problemas estruturais da molécula (Romero *et al.*, 2006). Entretanto, muitos exons alternativos introduzem *stop* códons prematuros em seus mRNAs, gerando proteínas truncadas ou degradação do transcrito, e estão potencialmente relacionadas com o desenvolvimento de doenças. Em relação à maquinaria de *splicing* propriamente dita, existe um número menor de relatos sobre mutações patogênicas em seus elementos centrais, sugerindo que tais alterações nem sempre são compatíveis com a vida.

Apesar do potencial patogênico, o *splicing* alternativo é bem tolerado pelos organismos, mas muitas vezes é causa de doenças ou é um indicador de defeito prévio. A primeira descrição de mudanças na seleção dos sítios de *splicing* foi publicada por Mardon e colaboradores em 1987 (Mardon *et al.*, 1987). Dez anos depois, uma revisão sobre o impacto das mutações exônicas nesses sítios postulava que, embora sem modificar o código genético presente no DNA, elas podem provocar doenças (Cooper & Mattox, 1997). Nos últimos anos, um grande número de dados foi obtido e permitiu uma melhor compreensão dos mecanismos biológicos envolvidos e seu efeito na saúde

humana. Um exemplo de doença causada por mutação de ponto em elemento regulador é a atrofia muscular espinhal (SMA). A SMA é causada por perda ou mutação dos genes *SMN1* e *SMN2*, este último com *splicing* alternativo do exon 7 (Wirth *et al.*, 2006). Os produtos de *SMN1* e *SMN2* estão envolvidos na agregação de snRNPs e, portanto, podem afetar o processamento de pré-RNAs de forma específica ou generalizada (Burghes & Beattie, 2009) e provocar a morte seletiva de neurônios (Zhang *et al.*, 2008). Outros exemplos de doenças causadas por fatores reguladores de *splicing* incluem formas autosômicas dominantes de retinite pigmentosa que possuem mutação no fator PRPF31/U4 (Vithana *et al.*, 2001; Wilkie *et al.*, 2008) e no componente PRP8 da snRNP U5 (Boon *et al.*, 2007).

Mudanças na proporção de isoformas protéicas geradas por *splicing* alternativo ocorrem em doenças do sistema nervoso central conhecidas como tauopatias. As tauopatias mostram acúmulo intracelular de filamentos que contêm a proteína tau. O gene que codifica essa proteína, *MAPT*, apresenta *splicing* alternativo regulado temporal e espacialmente em oito dos seus 16 exons. A proteína tau liga-se a microtúbulos por regiões repetidas da molécula, uma delas codificada pelo exon alternativo 10. Quando presente, o transcrito exibe quatro repetições (4R) e, na ausência do exon 10, o transcrito apresenta três repetições (3R) (Tazi *et al.*, 2009). Mutações em elementos reguladores de *splicing* alteram a taxa de inclusão do exon 10 no transcrito e podem levar a doenças neurodegenerativas e demência (Andreadis, 2006; Gallo *et al.*, 2007).

A utilização de sítios de *splicing* crípticos também é um fator causal de doenças humanas, como ocorre na síndrome de envelhecimento prematuro Hutchinson-Gilford, que apresenta síntese de produto truncado do gene da lamina nuclear A/C, causada por provável alteração de sequência ESE. Mutações em polimorfismos de base única, bem como em sítios doadores e presença de elementos repetitivos também podem alterar o padrão de *splicing* e serem fatores de predisposição a doenças como Alzheimer, doença coronariana e várias condições hereditárias (Tazi *et al.*, 2009).

Em câncer, os exemplos são numerosos, mas ainda não está completamente esclarecido se os produtos alternativos contribuem para iniciação e progressão neoplásica ou se compreendem apenas uma consequência do fenótipo maligno. Uma busca recente de artigos científicos catalogados no MedLine utilizando o PubMed e as palavras *cancer AND alternative splicing* mostrou 1452 resultados, dos quais 371

referem-se a trabalhos de revisão. Utilizando *breast cancer* ou *oral cancer AND alternative splicing*, foram extraídos 529 e 37 artigos, respectivamente. São citados como afetados por processamento alterado os genes *BRCA1*, *PR* ou do receptor de progesterona, *LKB1*, *KLF6*, *KIT*, *APC*, *TP53*, entre outros, relacionados com carcinomas de mama, de próstata, tumores orais, gastrointestinais e familiares (Srebrow *et al.*, 2006; Bourdon, 2007; Cork *et al.*, 2008; Tazi *et al.*, 2009).

A caracterização funcional das proteínas alternativas e a definição dos sítios e reguladores *cis/trans* de *splicing* fornecem novas opções para interferência terapêutica. Algumas mutações podem ter valor preditivo; as isoformas, por outro lado, são alvos potenciais de drogas, estratégias anti-sense ou de terapia radioisotópica com peptídeo-receptor-específico (van Alphen *et al.*, 2009; Tazi *et al.*, 2009).

II - Objetivos

O presente trabalho teve como objetivo geral estudar o *splicing* alternativo em células normais e neoplásicas utilizando abordagens moleculares e computacionais. Os objetivos específicos do trabalho compreenderam:

1. Avaliar detalhadamente *in silico* 788 genes com formas novas de *splicing* sugeridas pelo projeto *Head and Neck Transcriptome Initiative* (FAPESP / Instituto Ludwig) bem como *contigs* identificados como quimeras, sítios doadores e aceptores de íntrons envolvidos em *splicing* alternativo potencial, elementos reguladores *cis* e alteração da matriz de leitura e de domínios protéicos causados por introdução ou deleção de segmentos.
2. Validar formas novas de *splicing* em amostras de tecidos normais obtidos de autópsia e de portadores de câncer, e em linhagens celulares submetidas ou não a condições de estresse, utilizando a reação em cadeia da polimerase via transcriptase reversa (RT-PCR), com oligonucleotídeos flanqueando ou presentes na região de *splicing*, e sequenciamento dos produtos obtidos.
3. Analisar, utilizando PCR quantitativa e semi-quantitativa, o nível de expressão de isoformas validadas previamente, em amostras de tecidos normais obtidos de autópsia e de portadores de câncer, e em linhagens celulares submetidas ou não a condições de estresse.

A metodologia utilizada foi, em essência, a mesma em todos os trabalhos relatados nos diferentes capítulos do presente trabalho.

Os estudos foram aprovados pelos comitês de ética das instituições envolvidas e pelo Comitê Nacional de Ética em Pesquisa (CONEP 1763/05, 18/05/2005).

Capítulo 2

Introdução

Os tumores epidermóides de cabeça e pescoço são neoplasmas comuns e agressivos, com uma incidência anual estimada de mais de 600 mil casos novos no mundo e uma taxa crescente de incidência em muitos países. O câncer de tiróide é a neoplasia mais comum do sistema endócrino e possui uma incidência anual de cerca de 120 mil casos no mundo. A maioria das variações de frequência desses tumores em diferentes regiões geográficas é provavelmente decorrente de fatores tais como radiação, hábitos alimentares ou estilo de vida.

O fenótipo maligno é caracterizado pelo acúmulo de mutações e eventos epigenéticos que modificam a função da proteína e/ou causam alterações nos padrões transcricionais. Tais eventos geralmente conferem vantagens às células e podem levar à malignização do tecido. A análise em larga escala do transcritoma é uma ferramenta importante na avaliação de alguns dos mecanismos envolvidos no processo neoplásico.

Materiais e Métodos

Para gerar um banco de dados do perfil transcricional derivado de tecidos tumorais e não tumorais, foi desenvolvido um projeto de sequenciamento de cDNA, com base em mini-bibliotecas de cDNA normalizadas, geradas pela metodologia de ORESTES (*open reading frame expressed sequence tag*), o que resultou em um conjunto de mais de um milhão de sequências expressas (ESTs) de diversos tecidos humanos.

As sequências ORESTES são predominantemente derivadas da região codificadora central dos genes, o que favorece a identificação da função gênica com base na similaridade protéica. Além disso, essa metodologia normaliza parcialmente a população de sequências, permitindo o estudo de transcritos raros.

Resultados e Discussão

O presente trabalho descreve a análise detalhada de 134.495 ORESTES derivadas de câncer de cabeça e pescoço e 79.141 ORESTES derivadas de tecidos de tiróide,

seguida de validação experimental de novos transcritos e novas isoformas derivadas de *splicing* alternativo, bem como de avaliação de marcadores potenciais para esses tumores.

Os transcritos com sequências homólogas a genes humanos conhecidos foram identificados e novas variantes de *splicing* alternativo foram submetidas à análise manual, resultando em 788 novas isoformas potenciais, a maioria resultante de eventos de inserção. Um conjunto de 34 novas isoformas alternativas foi selecionado e 23 isoformas foram confirmadas pela reação em cadeia da polimerase via transcriptase reversa (RT-PCR) e por sequenciamento de DNA.

Novos genes potenciais foram revelados, incluindo seis transcritos mapeados em cromossomos bem estudados, como o 22, bem como transcritos mapeados em 253 regiões intergênicas. Além disso, 2.252 RNAs intrônicos não codificadores, eventualmente envolvidos em regulação transcricional, foram detectados. Um conjunto de 250 marcadores candidatos foi selecionado após a identificação de transcritos mapeados em regiões genômicas previamente conhecidas como amplificadas ou deletadas em tumores de cabeça e pescoço e de tireóide. Três desses marcadores foram avaliados por PCR quantitativa em um conjunto diferente de amostras. Juntamente com detalhada informação clínica, os dados estão disponíveis publicamente (http://verjo19.iq.usp.br/java/jsp/head_neck/) como fonte de informação para investigações posteriores.

Esta primeira reconstrução *in silico* dos transcritomas de cabeça e pescoço e de tireóide revela novos marcadores potenciais que podem ser utilizados para estudos futuros sobre a base molecular desses tumores.

Capítulo 3

Introdução

As ribonucleoproteínas heterogêneas nucleares (hnRNPs) compreendem uma grande família de proteínas com cerca de 30 membros que exibem domínios estruturais comuns. Um desses domínios é o RBD (*RNA-binding domain*), também denominado RRM (*RNA recognition motif*). As hnRNPs desempenham importantes papéis na biogênese de telômeros, reparo de DNA, sinalização celular e na regulação da expressão gênica, tanto no nível transcricional como traducional. Tais funções podem estar relacionadas com a transformação maligna por meio de DNA não reparado, instabilidade cromossômica, apoptose, angiogênese e invasão tecidual.

Vinte e dois genes que codificam as hnRNPs têm sido mapeados em diferentes cromossomos humanos e vários deles geram múltiplas isoformas por *splicing* alternativo. Um dos membros mais estudado dessa família, hnRNPK, tem sido detectado no núcleo, bem como no citoplasma e na mitocôndria, interagindo com transdutores de sinais, outras proteínas, RNA e também com elementos da região promotora de genes como *c-myc*, *c-src*, *BRCA1* and *thymidine kinase*. A hnRNPK já foi associada a uma variedade de processos, incluindo remodelamento da cromatina, transcrição, eventos de *splicing* e tradução.

O objetivo do presente trabalho foi investigar a expressão de sequências homólogas do gene humano *HNRPK* e analisar a evolução de membros da família hnRNP. Considerando que nenhum modelo molecular de isoformas *HNRPK* está disponível, foi investigado se diferenças no terminal C das variantes **a** e **b**, derivadas de *splicing* alternativo, são importantes para sua estrutura.

Materiais e Métodos

Com o objetivo de identificar sequências homólogas ao gene *HNRPK*, foi utilizada a sequência da proteína hnRNPK e a ferramenta *Map Viewer* (NCBI) e o *Genome Browser* da Universidade da Califórnia, Santa Cruz (UCSC). A ferramenta BLASTN foi aplicada e somente sequências com >70% de identidade sequencial foram

selecionadas. As análises de elementos repetidos foram realizadas utilizando o *RepeatMasker* (<http://www.repeatmasker.org>). Para cada sequência, um TBLASTN foi realizado.

A sequência codificadora e as sequências homólogas do gene *HNRPK* foram comparadas utilizando o MultiAlin (*Multiple sequence alignment with hierarchical clustering*) e o algoritmo Clustal. A análise de ORFs empregou o ORF Finder (<http://www.icb.ufmg.br/~infobio/sms/>) e a de domínios conservados da proteína empregou o *software* ScanProsite (<http://us.expasy.org/prosite>). Essa pesquisa foi realizada contra o banco de dados PROSITE. As predições de fosforilação foram feitas com o servidor NetPhos 2.0 (<http://www.cbs.dtu.dk/services/NetPhos/>).

A expressão das sequências de interesse foi investigada em amostras de tecidos normais e linhagens celulares por meio da reação em cadeia da polimerase via transcrição reversa (RT-PCR).

A modelagem por homologia do primeiro domínio da hnRNP utilizou o programa MODELLER. O alinhamento de sequências foi realizado com o pacote AMPS (*Alignment of Multiple Protein Sequences*). As análises filogenéticas e de evolução molecular utilizaram a versão MEGA 3.0 version 3.1. Tanto os alinhamentos de proteínas como de nucleotídeos foram obtidos com o *Profile Alignment Mode in Multialin*. As árvores filogenéticas foram construídas empregando os métodos UPGMA e *neighbor-joining*.

Resultados e Discussão

No presente estudo, foram analisados os pseudogenes processados *HNRPK* ($\psi 1$ - $\psi 4$ *HNRPK*) e as sequências codificadoras. Os pseudogenes *HNRPKs* são aparentemente não funcionais e o $\psi 1$ pode corresponder a transcritos de um gene ancestral. As análises filogenéticas e de sequência sugerem que os genes *HNRP* surgiram por duplicação e, posteriormente, novas características estruturais ampliaram as funções das hnRNPs.

Utilizando modelagem molecular, foram obtidos modelos dos domínios KH1 e KH3 que revelaram importantes resíduos envolvidos na interação DNA-proteína e evidenciaram a inexistência de diferenças estruturais entre as isoformas a e b da hnRNP.

Este é o primeiro trabalho filogenético catalogado pelo Medline que descreve uma análise detalhada dos genes *HNRNPs* e dos pseudogenes *HNRNPKs* e compara os domínios KH das isoformas da proteína hnRNPk. Os dados apresentados são importantes porque essa proteína representa um alvo potencial para intervenções farmacológicas em replicações virais e câncer.

Capítulo 4

Introdução

As proteínas RIP (*receptor-interacting proteins*) compreendem uma família de quinases de serina/treonina que integram sinais de estresse extra e intracelulares causados por diferentes fatores, incluindo infecções, inflamação e lesões no DNA. Respostas pró-sobrevivência, inflamatórias e imune, bem como padrões de sinalização indutores de morte celular podem ser iniciados pelas quinases RIP.

A proteína RIP2 (também conhecida como RICK ou CARDIAK), assim como outras RIPs, ativa o NF- κ B (*nuclear factor κ B*) interagindo com os fatores associados ao receptor do fator de necrose tumoral. A RIP2 mostra isoformas geradas por *splicing* alternativo, cujas funções e significados ainda não foram determinados. Esta proteína interage especificamente com o domínio CARD da caspase-1, uma caspase pró-inflamatória, e ativa diretamente o padrão de sinalização ERK 1 / 2. Alguns estudos referem a participação de RIP2 em imunidade inata e adaptativa e em sinais iniciados por receptores de ambas as cascatas imunes: receptores *toll-like* (TLRs) e receptores intracelulares NOD (*nucleotide-binding and oligomerization domain*) para imunidade inata, e receptores antígeno-específicos para imunidade adaptativa.

Nós identificamos previamente uma variante do gene *RIPK2* gerada por *splicing* alternativo (AY562996), combinando análise por ferramentas de bioinformática e pela reação em cadeia da polimerase via transcriptase reversa (RT-PCR) de sequências ORESTES (*open reading frame expressed sequence tags*). Essa nova variante apresenta remoção do exon 2, que resulta em um produto predito truncado com o domínio quinase parcialmente deletado.

A utilização de sítios de *splicing* alternativo difere entre tipos celulares, tecidos e fases do desenvolvimento e parece ser afetada por condições fisiológicas e estresse, tais como choque térmico, oxidativo e osmótico. Dado que as quinases RIP integram sinais de estresse extra e intracelulares e iniciam diferentes cascatas de sinalização que resultam tanto em proliferação como em morte celular, mudanças ambientais podem afetar sua expressão e ter importantes implicações.

Materiais e Métodos

No presente estudo, a modelagem molecular das duas isoformas RIP2 derivadas de *splicing* alternativo foi realizada com o *software* MODELLER. O alinhamento de sequências foi realizado com o pacote AMPS (*Alignment of Multiple Protein Sequences*). Para as análises estruturais, o programa INSIGHT II foi empregado. Além disso, a expressão das duas isoformas foi investigada em células normais humanas procedentes de autópsia e em amostras de carcinoma epidermóide de cavidade oral. Para avaliar se a variante 2 é conservada entre espécies, o cDNA de tumores de mama caninos foi amplificado utilizando condições de baixa estringência.

Com a finalidade de determinar se ambas as variantes são reguladas no nível pré-traducional por condições de estresse, sua expressão foi analisada sob exposição ao frio e ao calor e sob choque osmótico e ácido em três linhagens celulares humanas (FaDu/HTB-43, SiHa/HTB-35 and HeLa/CCL-2).

As linhagens celulares foram descongeladas e colocadas em meio MEM EARLE suplementado com soro fetal bovino, aminoácidos não essenciais, piruvato de sódio, penicilina e estreptomicina, sendo mantidas a 37°C em atmosfera de 5% de CO₂. Quando a densidade celular mostrava-se próxima ao estado de confluência, o material dos frascos foi tripsinizado e subdividido em réplicas. Para choque ácido, as células foram submetidas, imediatamente após a troca do meio, a uma atmosfera com CO₂ 10% por 24h e 72h, e mantidas a 37°C em atmosfera de CO₂ 5% por 24h e 72h. O pH do meio foi rigorosa e repetidamente checado a cada 6h. Para choque osmótico, as culturas foram alimentadas com meio MEM completo contendo glicerol 0,4M e 0,8M (soluto eletricamente neutro) enquanto os respectivos controles foram mantidos em ausência desse soluto. O bloqueio das culturas foi efetuado duas ou cinco horas após iniciado o experimento. Para choque térmico, as linhagens celulares foram submetidas, após a troca do meio, a temperaturas de 17°C ou 5°C por períodos de 64 h, ou a 39-40°C por aproximadamente 3h. Os controles foram mantidos por períodos idênticos a 37°C. Um segundo experimento de choque térmico foi realizado com dez réplicas, duas para cada condição.

Resultados e Discussão

Foi observado que as duas variantes RIP2 são expressas em todos os tecidos analisados e que a razão variante 1 / variante 2 é regulada em resposta à temperatura. A variante 2 também foi detectada nas amostras de tecidos caninos, evidenciando sua conservação entre espécies relativamente distantes. Apesar da baixa similaridade de sequência entre as estruturas das proteínas RIP2 homólogas, os sítios ativos dessas isoformas são estruturalmente similares e razoavelmente bem conservados.

Como as quinase RIP possuem um papel crítico na integração de sinais de estresse, a elucidação dos fatores que atuam na regulação dessas proteínas é importante. A mudança de expressão das isoformas RIP pode levar a respostas opostas - proliferação ou morte celular - com consequências importantes em diferentes condições patológicas. Seria interessante determinar o papel fisiológico de cada isoforma na adaptação ao estresse em células normais e tumorais. Entretanto, o desafio será integrar os achados *in vitro* e *in vivo*.

TEMPERATURE REGULATES ALTERNATIVE SPLICING IN *RIPK2* GENE

Mancini UM^{1,2}, Silva CHTP³, Rodrigues RV^{1,2}, Souza CF^{1,2}, Polachini GM¹, Feitosa OA³,
Head and Neck Genome Project GENCAPO⁴, Tajara EH^{1,2}

¹Departamento de Biologia Molecular, Faculdade de Medicina de São José do Rio Preto, FAMERP, São José do Rio Preto, SP, Brazil.

²Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, USP, São Paulo, SP, Brazil

³Departamento de Ciências Farmacêuticas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil

⁴<http://ctc.fmrp.usp.br/clinicalgenomics/cp/group.asp>. Complete authors list and addresses are also presented in the Appendix.

⁵Corresponding author: Eloiza Helena Tajara, PhD, Department of Molecular Biology, School of Medicine/FAMERP, São José do Rio Preto, CEP 15090-000, SP, Brazil. Phone: +55 17 3201-5737, e-mail: tajara@famerp.br

RUNNING TITLE

Alternative splicing in *RIPK2* gene

KEY WORDS

Alternative splicing, *RIPK2*, temperature regulation, gene expression.

ABSTRACT

Receptor-interacting proteins (RIP) are a family of serine/threonine kinases which integrate extra- and intracellular stress signals caused by different factors, including infections, inflammation and DNA damage. Pro-survival, inflammatory and immune responses as well as death-inducing signaling pathways can all be initiated by the RIP kinases.

RIP2 is an important component of the nuclear factor κ B (NF- κ B) activation by tumor necrosis factors and shows isoforms generated by alternative splicing, whose precise functions and significance remain to be determined. In the present study, homology modeling of two RIP2 isoforms was performed using the MODELLER software. Pairwise sequence alignment and assessment of statistical significance, multiple alignment and additional functions were carried out using the AMPS (Alignment of Multiple Protein Sequences) package, and structural analyses were performed with the INSIGHT II program. In addition, the expression of *RIPK2* transcripts was investigated in normal human tissues and in human and canine tumor samples. In order to determine if both variants are regulated at the pre-translational level in response to stress conditions, we analyzed their expression upon heat/cold exposure and acidic and osmotic shock in three human tumor cell lines.

We observed that both variants are expressed in all tissues analyzed and demonstrated that the variant 1/variant 2 ratio is regulated in response to temperature stress. Despite overall low sequence similarity among the complex structures of the homologue RIP2 proteins, the active sites of both isoforms are structurally similar and reasonably well conserved.

INTRODUCTION

Receptor-interacting proteins (RIP) are a family of serine/threonine kinases, which integrate extra and intracellular stress signals caused by different factors, including infections, inflammation, cellular differentiation and DNA damage. Pro-survival, inflammatory and immune responses as well as death-inducing signaling pathways can all be initiated by the RIP proteins (Meylan e Tschopp, 2005).

RIP1, the best-characterized member of the RIP family, is an important component of the nuclear factor κ B (NF- κ B) activation by tumor necrosis factors (TNFs). TNFs and related cytokines initiate different signaling cascades that can result in opposite responses, both cell proliferation and cell death. Binding of TNF to its receptor leads to the recruitment and assembling of a signaling complex containing death domain proteins and the TNF receptor-associated factor 2 (TRAF2), triggering pro-apoptotic responses. The additional recruitment of RIP1 to this complex as well as its polyubiquitination through the lysine at position 63 (Lys63) of ubiquitin activate the mitogen-activated protein kinases (MAPKs) and the I κ B kinase (IKK), which phosphorylates NF- κ B inhibitors, targeting these proteins for proteasomal degradation. NF- κ B transcription factors are thereby released and may induce expression of proinflammatory, prosurvival and immunomodulatory genes (Zhang, Kovalenko *et al.*, 2000; Ea, Deng *et al.*, 2006; O'donnell, Legarda-Addison *et al.*, 2007).

Besides its pro-survival role, RIP1 can act as a pro-death molecule, forming a complex with caspase-8 to initiate apoptosis (Lin, Devin *et al.*, 1999; Kim, Choi *et al.*, 2000; Martinon, Holler *et al.*, 2000; O'donnell, Legarda-Addison *et al.*, 2007).

RIP1 interacts with RIP3, another member of the RIP family, which promotes apoptosis in various cell types when overexpressed (Sun, Yin *et al.*, 2002).

RIP4, unlike RIP3, induces the activation of NF- κ B and AP-1 pathways when overexpressed. This member of the RIP kinase family is a modulator of keratinocyte differentiation and, as RIP5, is characterized by the presence of ankyrin repeats in their C-terminal domain (Meylan, Martinon *et al.*, 2002; Zha, Zhou *et al.*, 2004; Meylan e Tschopp, 2005; Adams, Pankow *et al.*, 2007). RIP6 (or LRRK1) and RIP7 (or LRRK2) are additional RIP kinases that have a leucine-rich repeat (LRR) motif and may be involved in

pathogen-, damage- or stress-associated molecular patterns (Festjens, Vanden Berghe *et al.*, 2007).

RIP2 (also called RICK or CARDIAK), like other RIPs, activates NF- κ B, interacting with the TNFR-associated factors. This RIP specifically interacts with the CARD of caspase-1, a pro-inflammatory caspase, and directly activates ERK1/2 signaling (Inohara, Del Peso *et al.*, 1998; Thome, Hofmann *et al.*, 1998; Navas, Baldwin *et al.*, 1999). Some reports claim the participation of RIP2 in innate and adaptive immunity, and in transducing signals from receptors of both immune cascades: Toll-like receptors (TLRs) and nucleotide-binding and oligomerization domain (NOD) intracellular receptors for innate immunity and antigen-specific receptors for adaptive immunity (Chin, Dempsey *et al.*, 2002; Kobayashi, Inohara *et al.*, 2002). Otherwise, the studies of Park *et al.* (Park, Kim *et al.*, 2007) and Hall *et al.* (Hall, Wilhelm *et al.*, 2008) indicated that the absence of RIP2 causes reduced cytokine production and are defective in their responses to NOD but does not affect T cell proliferation, T helper differentiation or TLR responses.

Similar to RIP1 in the TNF signaling cascade, Lys63-linked polyubiquitination of RIP2 induces NF- κ B activation after NOD stimulation. The polyubiquitin chains are linked at lysine 209 in RIP2 kinase domain (Hasegawa, Fujimoto *et al.*, 2008) whereas, in RIP1, are linked at lysine 377, in an intermediate region between the N-terminal kinase and C-terminal death domain (Ea, Deng *et al.*, 2006; Li, Kobayashi *et al.*, 2006). In NOD signaling pathway, the interaction of RIP2 and NEMO, a regulatory subunit of IKK, is also essential for NF- κ B activation (Inohara, Koseki *et al.*, 2000). Recently, Clark and collaborators (Clark, Marinis *et al.*, 2008) demonstrated that MAP3K4, a MAPK kinase kinase that is a major mediator of environmental stress, sequesters RIP2 from the NOD signaling inhibiting NF- κ B cascades.

We previously described a new alternative splicing variant of the *RIPK2* gene (AY562996), using a combined bioinformatic and reverse transcription polymerase chain reaction (RT-PCR) analysis of open reading frame expressed sequence tags (ORESTES) (Mancini e Tajara). This new variant presents skipping of exon 2 resulting in a predicted shorter truncated product with a partially deleted kinase domain.

The use of alternative splice sites may differ among cell types, tissues and phases of development (Clarke, Jordan *et al.*, 2000; Sun, Yin *et al.*, 2002; Zeng, Sharpe *et al.*, 2006) and has been shown to be affected by physiologic and stress conditions, such as heat

shock, oxidative and osmotic stress (Shen, Beall *et al.*, 1993; Takechi, Hosokawa *et al.*, 1994; Tomasini, Samir *et al.*, 2001; Hashimoto, Zhang *et al.*, 2002; Gray, Iglesias *et al.*, 2003; Matsuzaki, Manabe *et al.*, 2004; Nurmi, Puolakkainen *et al.*, 2005; Wakasugi, Nakano *et al.*, 2005; Bond, Patel *et al.*, 2006; Kim e Gladyshev, 2006; Fiol, Mak *et al.*, 2007). For example, Fujikake and collaborators (Fujikake, Nagai *et al.*, 2005) identified three new variants of the *Drosophila* heat shock transcription factor gene (*Hsf*) regulated in response to heat/cold stress. In mouse, one alternatively spliced mRNA of the *HSP47* (heat shock protein 47) gene is temperature-induced (Takechi, Hosokawa *et al.*, 1994). A similar event was observed for the “readthrough” transcript (type R) of the acetylcholinesterase gene (*Ache*) in mouse neuroblastoma N18TG2 cells (Perrier, Salani *et al.*, 2005).

In humans, an interesting case of temperature-dependent alternative splicing was observed in fibroblasts of a patient with Ehlers–Danlos Syndrome type VII (Weil, D'alessio *et al.*, 1989). The patient showed a mutation in the last nucleotide of exon 6 of the *COL1A1* gene, which produced an aberrant spliced transcript not detectable at 31°C but present at 39°C. The insertion of a cryptic exon in the neurofibromatosis type 1 mRNA induced by cold-shock conditions was also observed by Ars and collaborators (Ars, Serra *et al.*, 2000).

RIPK1, 2, 3 and 5 genes show variants generated by alternative splicing of the primary transcripts whose precise functions and significance remain to be determined. Since RIP kinases integrate extra and intracellular stress signals and initiate different signaling cascades that can result both in cell proliferation and cell death, environmental changes may affect their expression with important implications.

In the present study, homology modeling of two RIP2 isoforms derived from alternative splicing was performed. Furthermore, the expression of two *RIPK2* isoforms was investigated in normal and tumor cells. In order to determine if both variants are regulated at the pre-translational level in response to stress conditions, we analyzed their expression upon heat/cold exposure and acidic and osmotic shock in human tumor cell lines.

MATERIALS AND METHODS

Samples

Nine samples of normal human tissues removed at autopsy (brain, testis, heart, lung, stomach, kidney, larynx, liver and tongue) and samples from 32 patients with squamous cell carcinoma of the oral cavity and 4 dogs (*Canis lupus familiaris*) with mammary tumors (matched tumor/surgical margins) were used to evaluate the expression of two splicing variants of the *RIPK2* gene using RT-PCR and quantitative PCR. Proteins from another set of 15 patients with squamous cell carcinoma of the oral cavity were analyzed by Western Blot.

For the stress experiments, we utilized the human ATCC cell lines HeLa/CCL-2, (adenocarcinoma of the cervix), FaDu/HTB-43 (squamous cell carcinoma of the pharynx), and SiHa/HTB-35 (squamous cell carcinoma of the cervix). All three cell lines were cultured in Modified Eagle's Medium (Cultilab, Campinas, Brazil) supplemented with 10% fetal bovine serum (Cultilab), 10mM of non-essential amino acids (Invitrogen Corp., Carlsbad, CA, USA), 0.11g/L of sodium pyruvate, 100 units/ml penicillin (Cultilab), and 90µg/ml streptomycin (Cultilab), in a humidified atmosphere with 5% CO₂ at 37°C.

The study protocol was approved by the National Committee of Ethics in Research (CONEP 1763/05, 18/05/2005) and informed consent was obtained from all patients enrolled.

Heat, osmotic and acidic shock

Prior to stress treatment, replicas from each cell line were allowed to grow to 80-90% confluence. The medium was replaced with 15mL of fresh medium at 0 hour. For the heat shock experiments, cultures were maintained at 40°C for 3 h, and at 17°C or 5°C for 64 h. Osmotic shock was achieved by incubation of cell line cultures in medium containing 0.4M or 0.8M ultrapure glycerol (hypertonic stress) for 2 or 5 h at 37°C. Acidic shock was performed by maintaining the cultures in an atmosphere with elevated CO₂ (10% CO₂) for 24 h or 72 h. Four replicas were cultured under each stress condition, and four control replicas were cultured without addition of glycerol, in a humidified atmosphere with 5% CO₂ at 37°C for the same time periods.

Cells were immediately lysed after the incubation period by adding TRIzol (Invitrogen), and stored at -80°C until RNA extraction.

RNA Extraction and Reverse Transcription Polymerase Chain Reaction (RT-PCR)

Total RNAs from tissue samples and cell lines were obtained using TRIzol reagent and treated with DNase. One microgram of total RNA was converted to cDNA using the ThermoScript™ RT-PCR System or SuperScript™ III Reverse Transcriptase (Invitrogen), according to the manufacturer's instructions. Integrity of the RNA was confirmed by agarose gel electrophoresis, and the concentration was determined by obtaining the ratio of absorbance values at 260 and 280 nm, following standard procedures.

Detection of RIPK2 variants by semiquantitative PCR and sequencing

The PCR amplification of both longer and shorter *RIPK2* variants was performed using the oligonucleotides 5'-CGCCTCTGGCACTGTGTCGT-3' (sense) and 3'-CGTGACTGTGAGAGGGACAT-3' (anti-sense). The total reaction volume was 25 µl and contained 1X PCR buffer, 1 mM MgCl₂, 2 µM of each *RIPK2* primer, 2 µM of *GAPDH* primers, 5 mM dNTPs mix, 1U Taq DNA polymerase (Invitrogen) and 50ng of cDNA. After pre-incubation for 5 min at 94°C (initial denaturation), amplification was carried out through 35, 28 or 21 cycles at 94°C for 50 s, 58°C for 40 s, 72°C for 50 s, and 72°C for 10 min, using a thermal cycler (9700 GeneAmp PCR System – Applied Biosystems, Foster City, CA, USA). PCR amplification of the endogenous control gene, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), was carried out using sense (5'-CTGCACCACCAACTGCTTA-3' or 5'-ACCCACTCCTCCACCTTTGA-3') and anti-sense (5'-CTAGACGGCAGGTCAGGTC-3' or 5'-CTGTTGCTGTAGCCAAATTCGT-3') primers. The PCR products from *RIPK2* variant 1 (457 bp), variant 2 (303 bp), and *GAPDH* (296 bp and 101 bp) were separated on 2% agarose gels in the presence of ethidium bromide and quantified by an imager system (EDAS 290, Kodak, New Haven, CT). The PCR bands were cut out of the agarose gel and directly sequenced in both directions with sense and antisense primers, using a 377 ABI Prism Sequencer (Applied Biosystems). The sequences were analyzed using BLAST similarity search against a Nucleotide Database ([Www.Ncbi.Nlm.Nih.Gov/Blast/](http://www.Ncbi.Nlm.Nih.Gov/Blast/)).

Splice variant-specific real time PCR

The expression of *RIPK2* variant 1 and variant 2 following stress treatment of FaDu cells was investigated by real time PCR. The reaction was performed in quadruplicate using an ABI Prism 7500 Sequence Detection System (Applied Biosystems). The primers were manually designed using the following parameters: 19-24 bp length, 30-70% GC content and a short amplicon size (66-104 bp), as follows, *RIPK2* variant 1, sense (5'-AGAAGCTGAAATTTACACAAAGC-3') and antisense (5'-CCATTTGGCATGTATTTCAGTAAC-3'); *RIPK2* variant 2, sense (5'-TGCTCGACAGAAAAGTGAATATC-3') and antisense (5'-AAGGAGGAGTCATATTGTGCAG-3'); *GAPDH*, sense (5'-ACCCACTCCTCCACCTTTGA-3') and antisense (5'-CTGTTGCTGTAGCCAAATTCGT-3'); *TUB6A*, sense (5'-TCAACACCTTCTTCAGTGAAACG-3') and antisense (5'-AGTGCCAGTGCGAACTTCATC-3'); *ACTB*, sense (5'-GGCACCCAGCACAATGAAG-3') and antisense (5'-CCGATCCACACGGAGTACTTG-3'). All primers were purchased from Invitrogen. Briefly, the reactions were carried out in a total volume of 20 µl, with 10 µl of *Power SYBR Green PCR Master Mix* (Applied Biosystems), 250 nM of each primer and 10 ng cDNA. The PCR conditions were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec, 58°C for 10 sec, 60°C for 1 min. Following the PCR, dissociation curve analysis was performed to confirm the desired single gene product.

The gene expression stability over different samples was analyzed using the geNorm software (<http://medgen.ugent.be/~jvdesomp/genorm>). Each transcript level was normalized by division with the expression values of *GAPDH*, *ACTB* and *TUBA6*, used as internal controls.

The transcript level was calculated using the $2^{-\Delta\Delta C_t}$ method (Livak e Schmittgen, 2001). The fold difference (relative abundance) was calculated using the formula $2^{-\Delta\Delta C_t}$ and was plotted as means $\pm$ SD, with $n=4$ technical replicates. Fold differences of >3 were considered to be significant.

All reactions were performed with a negative control. One sample without treatment was chosen to be used as the calibrator control for the reactions. Experiments were repeated whenever the coefficient of variation was higher than 5%. After each reaction, the products were analyzed on a 2% agarose gel stained with ethidium bromide.

Sequence alignment and homology modeling procedures

Homology modeling of two RIP2 isoforms was carried out using the MODELLER software that performs modeling by satisfaction of spatial restraints (Sali e Blundell, 1990). Six homologues with solved structures in the Protein Data Bank (codes 2GSF, 1JPA, 1K2P, 1U59, 1UWH, 2EVA) used as templates were selected through a non-redundant BLASTp search. Two putative conserved domains with statistical significance were detected: TyrKc and S_TKc, which correspond, respectively, to the catalytic domains of tyrosine and serine/threonine protein kinases, and include the Leu10-Thr296 sequence. These six templates share sequence identities of 27.1% (Eph receptor tyrosine kinase, PDB code 1JPA) to 30% [Transforming growth factor-beta (TGF-beta)-activated kinase 1 - TAK1, PDB code 2EVA] with RIP2. Analyses were performed using pairwise alignments via the AMPS (Alignment of Multiple Protein Sequences) package (Barton e Sternberg, 1987). Previous to the modeling, a final multiple alignment was obtained by analyzing the superposition of the six structures regarding the alpha-carbons of the residues, using the INSIGHT II program, version 2005 (Accelrys, San Diego, CA), which allowed refine the previous alignment obtained from the AMPS MULTALIGN module. The information of secondary structure in the template sequence was incorporated into this previous alignment using the MULTALIGN module of the AMPS package, with the restriction that all insertions and deletions were limited to regions outside the common core of alpha-helices and beta-sheets. A gap penalty of 1000 was fixed to any deletion or insertion inside a secondary structure element. The alignment obtained was edited, investigating and considering aligned the residues which were close in the space, as visualized in the structural superposition. This procedure results in a final alignment that is different from the one based on the Dayhoff matrix used in AMPS.

NetPhosK 1.0 server (Blom, Sicheritz-Ponten *et al.*, 2004) was used for phosphorylation site analyses. The algorithm produces neural network predictions of kinase-specific eukaryotic protein phosphorylation sites. Currently, NetPhosK covers the following kinases: PKA, PKC, PKG, CKII, Cdc2, CaM-II, ATM, DNA PK, Cdk5, p38 MAPK, GSK3, CKI, PKB, RSK, INSR, EGFR, and Src.

Western blot

For Western blot analysis, the antibodies used were polyclonal anti-RICK diluted 1:100 (C-19, sc-8611, Santa Cruz Biotechnology Inc, Santa Cruz, CA, USA), which targets a peptide near the C-terminus of human RIP2, and anti beta-actin diluted 1:5000 (Sigma-Aldrich, St Louis, MO, USA). In brief, protein samples (10 ug) were loaded onto 12% resolving gel with 5% stacking gel (SDS-PAGE) in denaturing conditions at 130V for 90 min. The molecular weight ladders used were the PageRuler™ Prestained Protein Ladder (SM0671; Fermentas Life Sciences, Burlington, Ontario, Canada) and the See Blue Plus2 Pre- Stained Standard (Invitrogen).

The proteins were then transferred electrophoretically (325 mA per blot 70 min; Mini Protean 3 Cell, BioRad, Hercules, CA, USA) to polyvinylidene fluoride (PVDF) paper (Immobilon-P, Millipore, Billerica, MA, USA) soaked in transfer buffer (25 mM Tris, 0.2 M glycine, 20% v/v methanol; Merck, Campinas, SP, Brazil). The PVDF membranes were submitted to chromogenic staining using the Western Breeze kit (Invitrogen). The blots were then scanned and analyzed (Gel Logic HP 2200 imaging system; Carestream Health Inc./Kodak Health Group, Rochester, NY, USA).

RESULTS

Identification and Expression Patterns of RIPK2 variants

The open reading frame expressed sequence tags (ORESTES, (Dias Neto, Correa *et al.*, 2000) generated by the Human Cancer Genome Project (HCGP); (Camargo, Samaia *et al.*, 2001; Brentani, Caballero *et al.*, 2003); were matched to the RefSeq and mRNA public data sets (Www.Ncbi.Nlm.Nih.Gov). Genomic mapping of the ORESTES sequences revealed the existence of a putative new *RIPK2* splice variant. RT-PCR and cDNA sequencing were performed to confirm the splicing event, using normal and tumor tissues. This approach revealed the expression of both longer *RIPK2* variant 1 (NM_003821) and shorter variant 2 (AY562996) in different samples. The latter is generated by skipping of exon 2 (154bp), which alters the reading frame, producing several premature stop codons.

In the shorter *RIPK2* isoform, a potential translation initiation codon AUG (nucleotides 85-87 of exon 3) is in-frame with the downstream *RIPK2* sequence. Translation from this codon may give rise to a predicted amino-terminal truncated

protein (403 residues, predicted molecular weight of 45.59 kDa) lacking the first 137 amino acids of RIP2, ([Http://Ca.Expasy.Org/Prosite/](http://Ca.Expasy.Org/Prosite/), ; [Http://Www.Ncbi.Nlm.Nih.Gov/Structure/Cdd/Wrpsb.Cgi](http://Www.Ncbi.Nlm.Nih.Gov/Structure/Cdd/Wrpsb.Cgi)), (Figure 1)

Normal tissue samples from brain, testis, heart, lung, stomach, kidney, larynx, liver and tongue, and matched tumor/surgical margins samples from 32 patients with oral cancer co-expressed *RIPK2* variants 1 and 2 (Figure 2A and B). The co-expression of both isoforms was reinforced by database analysis (UCSC Genome Browser - <http://genome.ucsc.edu/>) of ESTs covering the exon 2 of *RIPK2*. This analysis showed the skipping of exon 2 in 33 of the 80 ESTs covering this transcript region. To investigate whether splice variant 2 is conserved across species, we amplified cDNA from canine mammary tumors by using low stringent conditions. Again, a putative splice variant was detected (Figure 2C).

Splicing patterns of RIPK2 under stress conditions

To investigate whether alternative splicing of *RIPK2* is induced or inhibited by stress conditions, three cell lines (FaDu, SiHa and HeLa) were exposed to severe heat shock (40°C, 17°C or 5°C), osmotic shock (hypertonic medium) or acidic shock (atmosphere of 10% CO₂). Only heat shock altered the splicing pattern. The semi-quantitative RT-PCR results showed an increase in the transcription of variant 1 at 40°C and an increase in the transcription of variant 2 at 17°C compared to that at 37°C (Figure 3 B). In all replicas of the Fadu cell line at 40°C, the result was confirmed by quantitative PCR, and in most replicas at 17°C, variant 2 showed higher expression than variant 1 (Figure 3 A). These results suggest that alternatively spliced exon 2-containing mRNA may be transcribed more efficiently at higher temperatures than the variant without exon 2.

Sequence alignment and homology modeling procedures

The model built for the catalytic domain of the isoform 1 (Figure 4) was obtained as described in the previous section, showing good stereochemical quality as indicated by the general G-factor (– 0.29) that resumes the overall quality, and reasonable atomic contact quality (– 2.21), above the expected for a crystal structure of the template with 2.95 Å resolution (PDB code 2EVA). Good stereochemistry was also evidenced by 80.9 % residues present in the most favored region and 14.7 % in additional allowed regions.

Despite overall low sequence identity among the complex structures of the homologue RIP2 proteins, the active sites are structurally similar and reasonably well conserved.

The model obtained for the catalytic domain of the isoform 2 (Figure 4) also showed good stereochemical quality, evidenced by a G-Factor of -0.19 and 81.2 % residues present in the most favored region and 16.1 % in additional allowed regions. Reasonable atomic contact quality (-2.1) was obtained for this model, considering the threshold of -3.0 , where values above this index indicate good models or well-refined structures. The model built for the isoform 1 is 137 residues larger than the one obtained for the isoform 2, and includes Leu10 to Thr296 of the overall sequence.

Regarding the phosphorylation sites predicted for the isoform 1, the NetPhosK 1.0 server pointed out 10 sites, which are absent in the isoform 2 sequence: Ser8, Thr12, Ser25, Ser29, Thr31, Ser33, Ser58, Ser76, Thr95 and Ser102. The study of Dorsch et al. 2006 (Dorsch, Wang *et al.*, 2006) suggested that Ser176 is an important autophosphorylation site for RIP2, and this phosphorylation can be used to monitor the activation state of RIP2. In Figure 4, we show the superposition of the two models, as well as the localization of Ser176, which seems to be conveniently accessible to the solvent and to phosphorylation. The additional 10 phosphorylation sites predicted for the isoform 1 are also shown in Figure 4, as well as the Lys47 in the conserved ATP-binding site and the critical polyubiquitination site Lys209.

Western blot

A subset of 23 samples (13 oral carcinomas and 10 surgical margins) was analyzed by Western blot (Figure 2 D). The blots showed a band of ~ 64 kDa in all samples, which corresponds approximately to the predicted molecular weight of isoform 1 (~ 61 kDa). Another band around 55 kDa was also observed in all samples.

DISCUSSION

In the present study, we detected the expression of two *RIPK2* variants in normal tissues and tumor cells. We found evidence that transcript 2 is conserved across species, since it was also detected in canine mammary tissue. At the protein level, a western blot band, which may correspond to isoform 2, was detected in the samples analyzed.

We observed a shift in the expression of *RIPK2* variants 1 and 2 as a response to temperature stress *in vitro*. Hyperthermia accompanies some pathophysiological processes, such as infection, injury and inflammation, and may lead to enhancement of immune responses to the causal stimuli. A beneficial effect of hyperthermia has also been observed when used therapeutically together with cancer treatment (Wust, Hildebrandt *et al.*, 2002; Yoo e Lee, 2007). However, the molecular mechanisms that mediate the effects of hyperthermia are not completely known so far.

Yan and collaborators (Yan, Xiu *et al.*, 2007) found evidence that hyperthermia induces TLR expression and TLR signaling-mediated activation of NF- κ B and MAPK pathways, resulting in increased synthesis of pro- and anti-inflammatory cytokines. Zhao et al (Zhao, An *et al.*, 2007) also observed TLR activation and significantly increased cytokine production when TLR stimulation was accompanied by a hyperthermia pretreatment. These data suggest that fever may modulate innate immune responses by TLR pathway and, although the role of RIP2 in this signaling remains controversial (Chin, Dempsey *et al.*, 2002; Kobayashi, Inohara *et al.*, 2002; Park, Kim *et al.*, 2007; Hall, Wilhelm *et al.*, 2008), provide a possible link to *RIPK2* expression changes depending on the temperature variation.

RIP2 integrates extra and intracellular stress signals and is an important component of the NF- κ B signaling cascade, transducing signals from innate immune system proteins, including cytoplasmic NOD proteins, which are involved in detecting pathogens in intracellular compartments. In addition, RIP2 has been suggested to be required for T-cell-receptor signaling and therefore may also participate of the adaptive immune response (Kobayashi, Inohara *et al.*, 2002).

Hasegawa and collaborators (Hasegawa, Fujimoto *et al.*, 2008) demonstrated that Lys63-linked polyubiquitination at Lys209 within the kinase domain of RIP2 is essential for NOD-mediated NF- κ B activation. The authors also showed that the interaction between NEMO and the intermediate region spanning residues 293–319 of RIP2 is required for NOD signaling. Otherwise, the residue critical for kinase activity of RIP2 (Lys47) is important but not essential for NF- κ B activation (Inohara, Del Peso *et al.*, 1998; Inohara, Koseki *et al.*, 2000) and apparently does not participate in polyubiquitination and interaction with NEMO. The kinase domain may mediate other functions, such as regulation of ERKs, p38 kinases, and own degradation (Navas, Baldwin *et al.*, 1999; Chin,

Dempsey *et al.*, 2002; Hasegawa, Fujimoto *et al.*, 2008). Considering these data, it is reasonable to suppose that isoform 2 of RIP2 may exhibit some impaired functions, other than polyubiquitination and interaction with NEMO, since its kinase domain is partially deleted.

In the present study, *RIPK2* variants 1 and 2 were observed in all normal and tumor tissues investigated, including human and canine samples. Variant 2 showed a tendency to have increased expression at low temperatures than variant 1, whereas the opposite occurred at 40°C.

This variability – a shift in the expression of *RIPK2* variants 1 and 2 as a response to temperature changes – suggests that stress signals caused by conditions such as infections and inflammation may affect the expression and splicing of *RIPK2*. We hypothesize that, under stress conditions, a putative mechanism would induce higher or lower expression of the exon 2-containing *RIPK2* product, which clearly affects the sequence of the encoded protein. If isoform 2 is a non-functional kinase but still mediates NF-κB activation, the balance of isoforms 1 and 2 would change, with consequences on signaling pathways related to kinase domain of RIP2. For instance, hyperthermia would increase the turnover rate of RIP2 or affect ERK pathways, whereas hypothermia might trigger proapoptotic responses. This hypothesis obviously requires experimental confirmation.

An important point is that, because the skipping of exon 2 causes a shift of reading frame, introducing several premature stop codons, a putative in-frame AUG codon downstream of the first AUG may be used for translation of the short variant. Alternative mechanisms of translation initiation are described following cell stress. One of these mechanisms is the internal ribosome entry sites (IRES)-mediated translation, which is observed in cells under stress conditions and may represent a regulatory process (reviewed by (Vagner, Galy *et al.*, 2001; Spriggs, Stoneley *et al.*, 2008).

Recently, Dasgupta and collaborators (Dasgupta, Agarwal *et al.*, 2008) described a C-terminal fragment of another member of RIP family, RIP1, which can activate signaling events, including NF-κB and TNF pathways, and consequently, induce survival and apoptotic responses. They concluded that short RIP1 affects the long isoform levels and may represent a new regulation mechanism. Albeit exhibiting some differences, RIP1 and 2 participate in the same regulatory net and are probably derived from a common

ancestry. Therefore, it is tempting to hypothesize that short RIP2 also has a similar function and may modulate long RIP2 under different conditions.

As RIP kinases play a critical role in integrating stress signals, the elucidation of the factors that take part in the regulation of these proteins is of major importance. The change in expression of RIP isoforms may lead to opposite responses - cell proliferation or cell death - with important consequences in different pathological conditions. It would be interesting to determine the physiological roles of each isoform for stress adaptation of normal and cancer cells. However, the challenge will be to link in vitro and in vivo findings.

ACKNOWLEDGMENT

The authors acknowledge support from FAPESP/Fundação de Amparo à Pesquisa do Estado de São Paulo (Grants 04/12054-9 and 03/07802-3), Conselho Nacional de Desenvolvimento Científico e Tecnológico/CNPq (Grant 401043/05-3) and LICR/Ludwig Institute for Cancer Research.

APPENDIX

The GENCAPO (Head and Neck Genome Project) authors are the following:

Alves JL^h, Amar A^h, Arap SS^f, Araújo-Filho V^f, Barbieri RB^h, Brandão LG^f, Brandão RM^k, Canto AL^d, Carmona-Raphe J^b, Carvalho A^h, Carvalho-Neto PB^h, Cerione M^e, Cernea CR^f, Chedid H^h, Chiappini PBO^h, Cominato ML^e, Correa PMS^d, Correia LA^h, Costa A^l, Curioni OA^h, Cury PM^g, de Carvalho MB^h, Dias Neto E^c, Dias THG^c, Durazzo M^f, Dutra RL^h, Ferraz AR^f, Figueiredo DLAⁱ, Figueiredo RO^l, Fortes CS^l, Franzi SA^h, Fukuyama EE^e, Gazito D^h, Góis-Filho JF^e, Gutierrez AP^h, Henriques S^h, Inamine R^l, Kaneto CM^k, Lehn CN^h, Leopoldino AM^o, Lisoni FCCR^b, López RVM^l, Macarenco R^d, Magalhães RP^f, Mamede RCMⁱ, Mancini UM^b, Martins GS^e, Meneses C^d, Mercante AMC^h, Michaluart-Junior P^f, Montenegro FLM^f, Moreira-Filho CA^a, Mota AJ^d, Moyses RA^f, Nóbrega FG^d, Nóbrega MP^d, Nunes FD^m, Ojopi EPB^c, Okamoto OKⁿ, Oliveira FB^g, Oliveira MIS^h, Paiva R^e, Paraventi C^f, Pinheiro DG^k, Polachini GM^b, Ramos O^f, Rapoport A^h, Rodini CO^m, Rodrigues AN^l, Rodrigues RV^b, Santos ARD^k, Santos M^h, Santos VPP^h, Secco L^e, Serafini LN^j, Settani F^e, Severino P^a, Silva AMA^h, Silva AMN^h, Silva C^f, Silva IT^k, Silva MJ^e, Silva-Filho GB^f, Silva-Junior WA^k, Silveira NJF^p, Smith RB^f, Souza SCOM^m, Souza TB^h, Stabenow E^f, Tajara EH^b,

Takamori JT^h, Tarlá MVC^k, Tavares MR^f, Teixeira AP^b, Turcano R^f, Valentim PJ^e, Vidotto A^b, Volpi EM^f, Wanderley AC^h, Wünsch-Filho V^l, Xavier FCA^m, Yamagushi F^e, Zago MA^q.

Affiliations: ^aInstituto de Ensino e Pesquisa Albert Einstein, São Paulo, SP; ^bDepartamento de Biologia Molecular, Faculdade de Medicina de São José do Rio Preto, SP; ^cInstituto de Psiquiatria, Faculdade de Medicina, USP, São Paulo, SP; ^dInstituto de Pesquisa e Desenvolvimento, UNIVAP, São José dos Campos, SP; ^eServiço de Cirurgia de Cabeça e Pescoço, Instituto do Câncer Arnaldo Vieira de Carvalho, São Paulo, SP; ^fDepartamento de Cirurgia de Cabeça e Pescoço, Faculdade de Medicina, USP, São Paulo, SP; ^gDepartamento de Patologia, Faculdade de Medicina de São José do Rio Preto, SP; ^hServiço de Cirurgia de Cabeça e Pescoço, Hospital Heliópolis, São Paulo, SP; ⁱServiço de Cirurgia de Cabeça e Pescoço, Faculdade de Medicina de Ribeirão Preto, USP, SP; ^jDepartamento de Patologia, Faculdade de Medicina de Ribeirão Preto, USP, SP; ^kDepartamento de Genética, Faculdade de Medicina de Ribeirão Preto, USP, SP; ^lDepartamento de Epidemiologia, Faculdade de Saúde Pública, USP, São Paulo, SP; ^mDepartamento de Estomatologia, Faculdade de Odontologia da USP, São Paulo, SP; ⁿDepartamento de Neurologia/Neurocirurgia, UNIFESP, São Paulo, SP; ^oDepartamento de Análises Clínicas, Toxicológicas e Bromatológicas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, USP, SP; ^pCiências da Computação, UNIVAP, São José dos Campos, SP; ^qDepartamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, USP, SP; Brasil.

FIGURE LEGENDS

Figure 1. Diagrams of *RIPK2* splice variants. The variant 1 encodes the longer isoform and the variant 2 presents skipping of exon 2. Arrowheads indicate the positions of the primers used for RT-PCR expression analysis. TyrKc=tyrosine kinase, catalytic domain; S_TKc=serine/threonine protein kinases, catalytic domain; CARD=caspase recruitment domain.

Figure 2. Expression of *RIPK2* mRNA and RIP2 protein in normal and tumor tissues. RT-PCR products from *RIPK2* variant 1 (456bp), variant 2 (302bp), and *GAPDH* (305 bp) in (A) normal human tissues, (B) samples from patients with oral cancer, and (C) canine mammary tumor samples. (D) Representative Western blots illustrating the RIP2 expression in a subset of oral squamous cell carcinoma and surgical margin samples by using anti-RIP2 and anti beta actin antibodies. T=tumor; M=surgical margin; 1=brain; 2=testis; 3=heart; 4=lung; 5=stomach; 6=kidney; 7=larynx; 8=liver; 9=tongue; L=100 bp fragment size marker; MW=PageRuler Prestained Protein Ladder. *GAPDH* mRNA was included as a control.

Figure 3. A. Splicing patterns of *RIPK2* under temperature stress. *RIPK2* splice variants in FaDu cell line cultures maintained at 40°C for 3 h and at 17°C or 5°C for 64 h. **A.** Quantification of *RIPK2* splice variant 1 (white) and 2 (black) expression by real time PCR. **B.** RT-PCR products from *RIPK2* variant 1 (456 bp), variant 2 (302 bp), and *GAPDH* (101 bp). Arrows indicate shifted bands.

Figure 4. Homology modeling of the catalytic domain of RIP2 isoforms. Superposition of the models built for the RIP2 isoforms 1 and 2 (ribbons representation in magenta and green, respectively). Isoform 2 is 137 residues larger than isoform 1, and the fold of this domain is surrounded by a dashed square. In blue and cyan, the additional 10 phosphorylation sites (serine and threonine residues, respectively) predicted for isoform 1 and absent in isoform 2 sequence are indicated. In red, other predicted sites for isoform 2 are represented. Ser176 is shown in yellow stick, and is indicated to be conveniently accessible to the solvent and to phosphorylation. Lysine 47 and 209 are also shown in sticks colored by atoms.

Figure 1

Figure 2

Figure 3

Figure 4

REFERENCES

- Adams, S., S. Pankow, *et al.* Regulation of NF-kappaB activity and keratinocyte differentiation by the RIP4 protein: implications for cutaneous wound repair. J Invest Dermatol, v.127, n.3, Mar, p.538-44. 2007.
- Ars, E., E. Serra, *et al.* Cold shock induces the insertion of a cryptic exon in the neurofibromatosis type 1 (NF1) mRNA. Nucleic Acids Res, v.28, n.6, Mar 15, p.1307-12. 2000.
- Barton, G. J. e M. J. Sternberg. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. J Mol Biol, v.198, n.2, Nov 20, p.327-37. 1987.
- Blom, N., T. Sicheritz-Ponten, *et al.* Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics, v.4, n.6, Jun, p.1633-49. 2004.
- Bond, C. E., P. Patel, *et al.* Astroglia up-regulate transcription and secretion of 'readthrough' acetylcholinesterase following oxidative stress. Eur J Neurosci, v.24, n.2, Jul, p.381-6. 2006.
- Brentani, H., O. L. Caballero, *et al.* The generation and utilization of a cancer-oriented representation of the human transcriptome by using expressed sequence tags. Proc Natl Acad Sci U S A, v.100, n.23, Nov 11, p.13418-23. 2003.
- Camargo, A. A., H. P. Samaia, *et al.* The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. Proc Natl Acad Sci U S A, v.98, n.21, Oct 9, p.12103-8. 2001.
- Chin, A. I., P. W. Dempsey, *et al.* Involvement of receptor-interacting protein 2 in innate and adaptive immune responses. Nature, v.416, n.6877, Mar 14, p.190-4. 2002.
- Clark, N. M., J. M. Marinis, *et al.* MEKK4 sequesters RIP2 to dictate NOD2 signal specificity. Curr Biol, v.18, n.18, Sep 23, p.1402-8. 2008.
- Clarke, L. A., P. Jordan, *et al.* Cell type specificity in alternative splicing of the human mismatch repair gene hMSH2. Eur J Hum Genet, v.8, n.5, May, p.347-52. 2000.
- Dasgupta, M., M. K. Agarwal, *et al.* Transposon-based mutagenesis identifies short RIP1 as an activator of NFkappaB. Cell Cycle, v.7, n.14, Jul 15, p.2249-56. 2008.
- Dias Neto, E., R. G. Correa, *et al.* Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proc Natl Acad Sci U S A, v.97, n.7, Mar 28, p.3491-6. 2000.

Dorsch, M., A. Wang, *et al.* Identification of a regulatory autophosphorylation site in the serine-threonine kinase RIP2. Cell Signal, v.18, n.12, Dec, p.2223-9. 2006.

Ea, C. K., L. Deng, *et al.* Activation of IKK by TNF α requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. Mol Cell, v.22, n.2, Apr 21, p.245-57. 2006.

Festjens, N., T. Vanden Berghe, *et al.* RIP1, a kinase on the crossroads of a cell's decision to live or die. Cell Death Differ, v.14, n.3, Mar, p.400-10. 2007.

Fiol, D. F., S. K. Mak, *et al.* Specific TSC22 domain transcripts are hypertonically induced and alternatively spliced to protect mouse kidney cells during osmotic stress. FEBS J, v.274, n.1, Jan, p.109-24. 2007.

Fujikake, N., Y. Nagai, *et al.* Alternative splicing regulates the transcriptional activity of Drosophila heat shock transcription factor in response to heat/cold stress. FEBS Lett, v.579, n.17, Jul 4, p.3842-8. 2005.

Gray, S. G., A. H. Iglesias, *et al.* Modulation of splicing events in histone deacetylase 3 by various extracellular and signal transduction pathways. Gene Expr, v.11, n.1, p.13-21. 2003.

Hall, H. T., M. T. Wilhelm, *et al.* RIP2 contributes to Nod signaling but is not essential for T cell proliferation, T helper differentiation or TLR responses. Eur J Immunol, v.38, n.1, Jan, p.64-72. 2008.

Hasegawa, M., Y. Fujimoto, *et al.* A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF-kappaB activation. EMBO J, v.27, n.2, Jan 23, p.373-83. 2008.

Hashimoto, Y., C. Zhang, *et al.* An alternatively spliced isoform of transcriptional repressor ATF3 and its induction by stress stimuli. Nucleic Acids Res, v.30, n.11, Jun 1, p.2398-406. 2002.

[Http://Ca.Expasy.Org/Prosite/](http://Ca.Expasy.Org/Prosite/). <http://ca.expasy.org/prosite/>
[Http://Www.Ncbi.Nlm.Nih.Gov/Structure/Cdd/Wrpsb.Cgi](http://Www.Ncbi.Nlm.Nih.Gov/Structure/Cdd/Wrpsb.Cgi).
<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>.

Inohara, N., L. Del Peso, *et al.* RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. J Biol Chem, v.273, n.20, May 15, p.12296-300. 1998.

Inohara, N., T. Koseki, *et al.* An induced proximity model for NF-kappa B activation in the Nod1/RICK and RIP signaling pathways. J Biol Chem, v.275, n.36, Sep 8, p.27823-31. 2000.

Kim, H. Y. e V. N. Gladyshev. Alternative first exon splicing regulates subcellular distribution of methionine sulfoxide reductases. BMC Mol Biol, v.7, p.11. 2006.

- Kim, J. W., E. J. Choi, *et al.* Activation of death-inducing signaling complex (DISC) by pro-apoptotic C-terminal fragment of RIP. Oncogene, v.19, n.39, Sep 14, p.4491-9. 2000.
- Kobayashi, K., N. Inohara, *et al.* RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. Nature, v.416, n.6877, Mar 14, p.194-9. 2002.
- Li, H., M. Kobayashi, *et al.* Ubiquitination of RIP is required for tumor necrosis factor alpha-induced NF-kappaB activation. J Biol Chem, v.281, n.19, May 12, p.13636-43. 2006.
- Lin, Y., A. Devin, *et al.* Cleavage of the death domain kinase RIP by caspase-8 prompts TNF-induced apoptosis. Genes Dev, v.13, n.19, Oct 1, p.2514-26. 1999.
- Livak, K. J. e T. D. Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, v.25, n.4, Dec, p.402-8. 2001.
- Mancini, U. e E. H. Tajara. www.ncbi.nlm.gov/nucore/45685662.
- Martinon, F., N. Holler, *et al.* Activation of a pro-apoptotic amplification loop through inhibition of NF-kappaB-dependent survival signals by caspase-mediated inactivation of RIP. FEBS Lett, v.468, n.2-3, Feb 25, p.134-6. 2000.
- Matsuzaki, S., T. Manabe, *et al.* Metals accelerate production of the aberrant splicing isoform of the presenilin-2. J Neurochem, v.88, n.6, Mar, p.1345-51. 2004.
- Meylan, E., F. Martinon, *et al.* RIP4 (DIK/PKK), a novel member of the RIP kinase family, activates NF-kappa B and is processed during apoptosis. EMBO Rep, v.3, n.12, Dec, p.1201-8. 2002.
- Meylan, E. e J. Tschopp. The RIP kinases: crucial integrators of cellular stress. Trends Biochem Sci, v.30, n.3, Mar, p.151-9. 2005.
- Navas, T. A., D. T. Baldwin, *et al.* RIP2 is a Raf1-activated mitogen-activated protein kinase kinase. J Biol Chem, v.274, n.47, Nov 19, p.33684-90. 1999.
- Nurmi, J. T., P. A. Puolakkainen, *et al.* Intron 1 retaining cyclooxygenase 1 splice variant is induced by osmotic stress in human intestinal epithelial cells. Prostaglandins Leukot Essent Fatty Acids, v.73, n.5, Nov, p.343-50. 2005.
- O'donnell, M. A., D. Legarda-Addison, *et al.* Ubiquitination of RIP1 regulates an NF-kappaB-independent cell-death switch in TNF signaling. Curr Biol, v.17, n.5, Mar 6, p.418-24. 2007.

Park, J. H., Y. G. Kim, *et al.* RICK/RIP2 mediates innate immune responses induced through Nod1 and Nod2 but not TLRs. J Immunol, v.178, n.4, Feb 15, p.2380-6. 2007.

Perrier, N. A., M. Salani, *et al.* The readthrough variant of acetylcholinesterase remains very minor after heat shock, organophosphate inhibition and stress, in cell culture and in vivo. J Neurochem, v.94, n.3, Aug, p.629-38. 2005.

Sali, A. e T. L. Blundell. Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. J Mol Biol, v.212, n.2, Mar 20, p.403-28. 1990.

Shen, J., C. J. Beall, *et al.* Tissue-specific alternative splicing of the Drosophila dopa decarboxylase gene is affected by heat shock. Mol Cell Biol, v.13, n.8, Aug, p.4549-55. 1993.

Spriggs, K. A., M. Stoneley, *et al.* Re-programming of translation following cell stress allows IRES-mediated translation to predominate. Biol Cell, v.100, n.1, Jan, p.27-38. 2008.

Sun, X., J. Yin, *et al.* Identification of a novel homotypic interaction motif required for the phosphorylation of receptor-interacting protein (RIP) by RIP3. J Biol Chem, v.277, n.11, Mar 15, p.9505-11. 2002.

Takechi, H., N. Hosokawa, *et al.* Alternative 5' splice site selection induced by heat shock. Mol Cell Biol, v.14, n.1, Jan, p.567-75. 1994.

Thome, M., K. Hofmann, *et al.* Identification of CARDIAK, a RIP-like kinase that associates with caspase-1. Curr Biol, v.8, n.15, Jul 16, p.885-8. 1998.

Tomasini, R., A. A. Samir, *et al.* Molecular and functional characterization of the stress-induced protein (SIP) gene and its two transcripts generated by alternative splicing. SIP induced by stress and promotes cell death. J Biol Chem, v.276, n.47, Nov 23, p.44185-92. 2001.

Vagner, S., B. Galy, *et al.* Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. EMBO Rep, v.2, n.10, Oct, p.893-8. 2001.

Wakasugi, K., T. Nakano, *et al.* Oxidative stress-responsive intracellular regulation specific for the angiostatic form of human tryptophanyl-tRNA synthetase. Biochemistry, v.44, n.1, Jan 11, p.225-32. 2005.

Weil, D., M. D'alessio, *et al.* Temperature-dependent expression of a collagen splicing defect in the fibroblasts of a patient with Ehlers-Danlos syndrome type VII. J Biol Chem, v.264, n.28, Oct 5, p.16804-9. 1989.

Wust, P., B. Hildebrandt, *et al.* Hyperthermia in combined treatment of cancer. Lancet Oncol, v.3, n.8, Aug, p.487-97. 2002.

Www.Ncbi.Nlm.Nih.Gov. www.ncbi.nlm.nih.gov.

Www.Ncbi.Nlm.Nih.Gov/Blast/. www.ncbi.nlm.nih.gov/BLAST/.

Yan, X., F. Xiu, *et al.* Fever range temperature promotes TLR4 expression and signaling in dendritic cells. Life Sci, v.80, n.4, Jan 2, p.307-13. 2007.

Yoo, J. e Y. J. Lee. Effect of hyperthermia on TRAIL-induced apoptotic death in human colon cancer cells: development of a novel strategy for regional therapy. J Cell Biochem, v.101, n.3, Jun 1, p.619-30. 2007.

Zeng, Z., C. R. Sharpe, *et al.* The expression and alternative splicing of alpha-neurexins during *Xenopus* development. Int J Dev Biol, v.50, n.1, p.39-46. 2006.

Zha, J., Q. Zhou, *et al.* RIP5 is a RIP-homologous inducer of cell death. Biochem Biophys Res Commun, v.319, n.2, Jun 25, p.298-303. 2004.

Zhang, S. Q., A. Kovalenko, *et al.* Recruitment of the IKK signalosome to the p55 TNF receptor: RIP and A20 bind to NEMO (IKKgamma) upon receptor stimulation. Immunity, v.12, n.3, Mar, p.301-11. 2000.

Zhao, W., H. An, *et al.* Hyperthermia differentially regulates TLR4 and TLR2-mediated innate immune response. Immunol Lett, v.108, n.2, Feb 15, p.137-42. 2007.

Capítulo 5

Introdução

O *splicing* alternativo compreende um dos principais mecanismos geradores de diversidade protéica em vertebrados. Um grupo importante de proteínas, como as SR, e a própria maquinaria de transcrição e a estrutura do promotor parecem modular esse processo. Estudos recentes indicam que esse mecanismo pode afetar a função do produto gênico e, em alguns casos, pode levar à transformação maligna. Mutações localizadas nas regiões intrônicas, tais como as que alteram o sítio de ramificação, os sítios doador e acceptor ou a sequência polipirimidínica, e os polimorfismos localizados em regiões codificadoras devem ter importantes consequências, criando ou rompendo sequências consenso e estimuladores ou acentuadores do processamento do RNA. Em função disso, foi proposto o presente trabalho cujo objetivo geral é estudar o *splicing* alternativo em células normais e neoplásicas utilizando abordagens moleculares e computacionais. Os objetivos específicos do trabalho compreenderam:

1. Avaliar detalhadamente *in silico* 788 genes com formas novas de *splicing* sugeridas pelo projeto *Head and Neck Transcriptome Initiative* (FAPESP / Instituto Ludwig) bem como *contigs* identificados como quimeras, sítios doadores e aceptores de íntrons envolvidos em *splicing* alternativo potencial, elementos reguladores *cis* e alteração da matriz de leitura e de domínios protéicos causados por introdução ou deleção de segmentos.

2. Validar formas novas de *splicing* em amostras de tecidos normais obtidos de autópsia e de portadores de câncer, e em linhagens celulares, utilizando a reação em cadeia da polimerase via transcriptase reversa (RT-PCR), com oligonucleotídeos flanqueando ou presentes na região de *splicing*, e sequenciamento dos produtos obtidos.

3. Analisar, utilizando PCR quantitativa, o nível de expressão de isoformas validadas previamente, em amostras de tecidos normais obtidos de autópsia e de portadores de câncer, e em linhagens celulares.

Materiais e Métodos

MATERIAIS

Das 788 sequências com sugestão de novos eventos de *splicing* identificadas pelo projeto *Head and Neck Transcriptome Initiative* (Reis EM, Ojopi EP, Alberto FL, Rahal P, Tsukumo F, Mancini UM, *Head and Neck Annotation Consortium. Cancer Res.*, 65:1693-9, 2005), 148 (18,8%) foram excluídas porque as informações mais recentes disponíveis em bancos de dados públicos não confirmaram a presença de *splicing* alternativo. Portanto, o total em análise pelo presente projeto somou 640 (81,2% de 788) sequências.

Com a utilização de ferramentas públicas de bioinformática, foram analisadas, portanto, 640 sequências com formas novas de *splicing*. Dessas sequências, 197 corresponderam ao grupo exibindo *splicing* do tipo perda e ganho de exon, perda de regiões internas do exon ou ganho de uma região intrônica completa (Classe A), 203 corresponderam ao grupo exibindo *splicing* do tipo extensão e deleção da região 5' do exon (Classe B) e 240 corresponderam ao grupo exibindo *splicing* do tipo extensão e deleção da região 3' do exon (Classe C).

Para validação das novas formas de *splicing*, foram utilizadas 17 amostras de tecidos normais de autópsias (cérebro, testículo, fígado, laringe, pulmão, rim, estômago e coração), três linhagens celulares ATCC (Manassas, Virgínia), uma de carcinoma espinocelular de faringe (FaDu/HTB-43), uma de adenocarcinoma cervical (HeLa/CCL-2) e uma de carcinoma de células escamosas de colo de útero (SiHa/HTB-35) e 75 amostras de tecidos normais e neoplásicos (tumor, margem cirúrgica e/ou linfonodo) de 29 pacientes com carcinoma epidermóide de assoalho da boca e de língua (exceto base de língua) selecionados pelo projeto temático “Busca de marcadores de agressividade em tumores de cabeça e pescoço” (FAPESP 04/12054-9). Segundo os critérios adotados nesse projeto temático, as amostras de pacientes foram classificadas em agressivas e menos agressivas de acordo com os parâmetros da *Union Internationale contre le Cancer* (UICC) e do *American Joint Committee on Cancer* (AJCC) (Ambrosch *et al.* 1998). Foram considerados agressivos os casos apresentando nódulos comprometidos por células neoplásicas (N+, que corresponde aos casos N1 a N3) e tamanho T1 ou T2, portanto pequenos. Os menos

agressivos corresponderam aos tumores que apresentaram nódulos não comprometidos (N0) e tamanho geralmente maior (T2 e T3).

Em função do número de experimentos de validação, as amostras de tumores foram também reunidas em pools, três contendo casos previamente classificados como agressivos (cada um deles referentes a margem cirúrgica, tumor e linfonodo) e outros três com casos menos agressivos, igualmente subdivididos como nos grupos agressivos.

As amostras foram obtidas após consentimento livre e esclarecido, segundo protocolos aprovados pelo CONEP (parecer 1763/2005).

MÉTODOS

Análises *in silico*

As análises incluíram: tipo e tamanho de *splicing* (inserção / deleção), localização do evento de *splicing* (CDS ou UTR), regra GT/AG das extremidades dos íntrons, matriz de leitura (*ORF*), presença de códon de terminação (*stop* códon) e de sequências repetitivas (Alu, Line, Mir, etc), efeito do *splicing* sobre o domínio protéico, sobre o produto ou sobre ambos, sítios potenciais de glicosilação e fosforilação, domínios transmembrânicos e função do produto gênico e processo biológico envolvido. Além disso, foi avaliada a presença de elementos reguladores de *splicing* (estimuladores e inibidores), tanto no exon afetado como nos exons vizinhos.

Para confirmação e classificação do evento de *splicing*, foram utilizadas informações do NCBI (<http://www.ncbi.nlm.nih.gov/>), ExPASy (<http://us.expasy.org/>), além de ferramentas disponíveis nesses bancos de dados e em outras páginas como BLAT (<http://genome.ucsc.edu/>), BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>), *The Sequence Manipulation Suite* da UFMG (<http://www.icb.ufmg.br/~infobio/sms/>), *CBS Prediction Servers* (NetOGlyc, NetPhos, TMHMM) e SIM4. Para avaliação de elementos reguladores, foram utilizadas as ferramentas BLAST e BLAT, além do RESCUE-ESSE (<http://genes.mit.edu/burgelab/rescue-ese/>), que é um banco de dados obtidos por um método híbrido computacional/experimental, e desenvolvido para identificar seqüências com atividade ESE.

A análise de elementos reguladores de *splicing* foi também realizada por programas específicos de busca de elementos reguladores *cis* (estimuladores e silenciadores), potenciais ou já experimentalmente testados, tanto no exon envolvido no

evento de *splicing* como nos exons vizinhos. Essa análise foi desenvolvida no laboratório do Prof. Dr. Sandro de Souza (Instituto Ludwig, SP). A sequência dos exons de interesse foi escaneada em busca de elementos reguladores, escorregando um nucleotídeo por vez. Todas as ocorrências foram anotadas em suas categorias bem como suas posições (Sakabe e de Souza, 2007). A organização, a análise e a interpretação dos resultados obtidos foram realizadas manualmente. Apenas os elementos reguladores potenciais ou conhecidos, com número de ocorrências maior ou igual a três, foram selecionados. Os demais foram excluídos. Tanto os elementos ESE (*exonic splicing enhancers* ou estimuladores de *splicing* exônicos) como os PESE (*putative exonic splicing enhancers* ou estimuladores de *splicing* exônicos putativos) foram posicionados na sequência do exon para facilitar a análise. Nessa etapa, foi feita a busca de padrões específicos de posição ou de presença/ausência de elementos estimuladores ou silenciadores em relação às características do exon alternativo, como perda ou ganho de seqüências, ou em relação às características do exon vizinho (5' ou 3').

Cultivo celular

O cultivo celular foi realizado em etapas anteriores ao presente trabalho e seguiu protocolo estabelecido pelo grupo. Em resumo, as linhagens celulares foram descongeladas e colocadas em meio MEM EARLE (Cultilab), suplementado com 10% de soro fetal bovino, 10mM de aminoácidos não essenciais (Invitrogen), 100mM de piruvato de sódio, 100 unidades/mL de penicilina e 100µg/mL de estreptomicina (Cultilab), sendo mantidas a 37°C em atmosfera de 5% de CO₂. Após um período de 24h, que permite a fixação das células na garrafa de cultura, o meio foi trocado para retirada completa do DMSO presente na solução de congelamento. O meio foi, então, substituído a cada três dias e o crescimento e a morfologia celular foram avaliados diariamente em microscópio invertido.

Extração de RNA total e síntese de cDNA

O processamento para obtenção do RNA seguiu protocolo descrito por Sambrook & Russell (2001). Após sua extração, o material foi quantificado em espectrofotômetro, aplicado (1µg) em gel de agarose 1% TAE e corado com brometo de etídio para confirmação das duas bandas ribossômicas 18S e 28S. Uma vez aprovada a qualidade do

RNA, o mesmo foi transcrito a cDNA utilizando-se os kits comerciais *High Capacity cDNA Archive Kit* (Applied Biosystems) e *SuperScript™ III Reverse Transcriptase* (Invitrogen Life Technologies), seguindo protocolo do fabricante.

Amplificação do cDNA por PCR não quantitativa

Os oligonucleotídeos foram selecionados considerando o tamanho do produto esperado (entre aproximadamente 70pb a 180pb), o conteúdo em dinucleotídeos GC (~60%), a temperatura de hibridação ($T_m^\circ = 60^\circ\text{C}$) e a ausência de homologia com sequências repetitivas no genoma. Todos os oligonucleotídeos desenhados não apresentaram, *in silico*, homologia ou similaridade significativa a outras sequências que pudesse interferir nos resultados.

Para cada sequência de interesse, foram utilizados dois pares de oligonucleotídeos diferentes, um para amplificar uma região constitutiva e um outro para amplificar a região alternativa (Tabela 1 Figura 1), utilizando a ferramenta Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). O desenho de oligonucleotídeos levou em consideração a obtenção, por PCR, de fragmentos com tamanhos diferentes correspondentes aos tipos de *splicing* constitutivo e alternativo esperados para o gene em questão. Em todos os casos, era esperada a presença de produto.

A reação em cadeia da polimerase (PCR) não quantitativa foi realizada em volume final de 25µL, contendo 100ng de cDNA, Tris-HCl-KCl 10mM (pH 8.3), MgCl₂ 2mM, dNTPs 200µM (dATP, dCTP, dGTP, dTTP), 25pmol de cada oligonucleotídeo e 1U de Taq DNA polimerase. Todas as reações de PCR foram efetuadas com um controle de contaminação (mix sem amostra). Finalmente, os produtos foram separados em géis de agarose 2%. Para purificação do produto e seu posterior sequenciamento, a PCR foi desenvolvida em volume final de 50µL.

Devido à padronização dos oligonucleotídeos quanto ao tamanho, conteúdo em GC e T_m° , o programa de PCR consenso compreendeu uma primeira etapa de 5min a 94°C e 35 ciclos de 50s a 94°C para desnaturação do DNA, 40s para hibridação dos oligonucleotídeos a 60°C e 50s a 72°C para extensão das cadeias, além de 10min a 72°C para extensão final.

Eletroforese em gel de Agarose

A separação dos produtos de PCR foi efetuada por eletroforese em gel de agarose 1% a temperaturas de 15°C a 20°C. As respectivas bandas foram evidenciadas por coloração com brometo de etídio. A fonte eletroforética foi regulada para uma corrida de 80 volts. O marcador molecular utilizado foi de 100pb (*Ladder Gibco DNA 100bp*) para produtos de *splicing* e de 1Kb (*Ladder Novagen Perfect DNA 1Kb*) para produtos de minissatélite.

PCR em tempo real

As reações de PCR em tempo real foram realizadas pelo método *CYBR Green*, no aparelho *Real-Time PCR System 7500* (Applied Biosystems). Os oligonucleotídeos utilizados foram os mesmos empregados para a PCR não quantitativa, exceto para os genes *RQCD1* e *SEC13L1*. Para esses dois genes, foram utilizados outros oligonucleotídeos, que geraram produtos entre 100 e 200pb.

As condições de termociclagem compreenderam uma incubação de 2min a 50°C, seguida por uma desnaturação inicial de 10min a 95°C. Em seguida, o programa continuou com 40 ciclos de 15s a 95°C, 1min a 60°C para hibridação de oligonucleotídeos e extensão das cadeias, seguido de 35 seg a 65°C para coleta do sinal. A curva de dissociação foi gerada após amplificação e compreendeu um passo de 15s a 95°C, seguido de 1min a 60°C.

Todas as reações foram preparadas em triplicatas e/ou quadruplicatas, incluindo os controles internos (controles endógenos) e processadas em volume final de 20µL contendo *Power SYBR® Green PCR Master Mix* e 100nM de cada oligonucleotídeo, segundo protocolo do fabricante. As reações foram repetidas até três vezes para cada par de oligonucleotídeo, para minimizar qualquer desvio por causas externas.

As réplicas com diferenças de C_t maiores que 2%, foram reanalisadas utilizando sempre o mesmo cDNA mãe (*Cycle threshold* ou C_t é o ciclo no qual as amostras atingem o ponto de detecção da fluorescência pelo equipamento). A partir dos valores de C_t de cada amostra, foram calculadas as médias das duplicatas. Posteriormente, foi calculado o ΔC_t a partir da subtração das médias obtidas para a sequência de interesse pela média do controle interno (*ACTB*, *GAPDH* e *TUBA6*). Para o cálculo do $\Delta -\Delta C_t$, foi escolhido o tecido normal como calibrador e o seu valor foi calculado a partir das diferenças dos valores de

ΔC_t de cada tecido em relação ao ΔC_t do calibrador. Em seguida, foi calculado o $2^{-\Delta-\Delta C_t}$, ou seja, o $(\Delta-\Delta C_t)^2$. Finalmente, para uma melhor representação gráfica, os resultados foram colocados em uma escala logarítmica de base 3 ($\log_3$).

Controle de transcritos anti-senses

Como controle de eventuais transcritos anti-senses (RNAs não codificadores/ncRNA), foi realizada a síntese de cDNA ‘dirigida’ com os oligonucleotídeos das sequências de interesse separadamente (o oligonucleotídeo anti-sense como controle de expressão de transcritos sense e o oligonucleotídeo sense como controle de expressão de eventuais transcritos anti-senses não codificadores), a partir de seis pools de RNA, três contendo amostras previamente classificadas como agressivas (normal, tumor, linfonodo) e outros três classificadas como menos agressivas (normal, tumor, linfonodo).

A síntese de cDNA ‘dirigida’ foi realizada de acordo as especificações do fabricante do kit comercial *SuperScript™ III One-Step RT-PCR System with Platinum® Taq DNA Polymerase* (Invitrogen Life Technologies).

O programa de síntese de cDNA e amplificação de produtos compreendeu um ciclo de 30min a 55°C para hibridação do oligonucleotídeo (sense ou anti-sense) e síntese da primeira fita de cDNA. Posteriormente, o programa continuou com um ciclo de 2min a 94°C para desnaturação do DNA. Nesse passo, o programa foi pausado para adicionar às amostras o outro oligonucleotídeo e continuar com a amplificação normal do produto. Essa etapa foi imediatamente seguida de 40 ciclos de 15s a 94°C para desnaturação do DNA, 30s a 60°C para hibridação dos oligonucleotídeos e 20s a 68°C para extensão das cadeias (o fabricante recomenda 68°C 1min/Kb), além de 5min a 68°C para extensão final.

Sequenciamento

Os produtos de PCR contendo volume final de 50µL foram transferidos para tubos de 1,5mL e receberam 5µL de acetato de sódio 3M, 1µL de *Ultra Pure™ Glycogen* (Invitrogen Life Technologies) e 150µL de etanol absoluto. Em seguida, foram estocados por 12h a -20°C. A purificação dos produtos continuou com centrifugação a 14000 rpm

por 20 min. O sobrenadante foi descartado e o sedimento recebeu 150µL de etanol 70%. Após centrifugação a 14000 rpm por 20 min, o sobrenadante foi novamente descartado e os tubos foram mantidos em centrífuga a vácuo a 60°C por 30 min. Finalmente, foram acrescentados 20µL de água DEPC em cada tubo.

Após purificação, as amostras foram submetidas à reação de amplificação com 2µL de *Big Dye Terminator* e 3,2 pmol de oligonucleotídeo. A reação compreendeu uma primeira etapa de 5min a 95°C para desnaturação inicial das cadeias de DNA, seguida de 30 ciclos de 30s a 95°C, 20 seg a 60°C para hibridação e 4min a 60°C para extensão das cadeias. Para cada amostra, foi realizada uma reação de seqüenciamento com o oligonucleotídeo sense e uma reação com o anti-sense.

As amostras foram precipitadas com 80µL de solução de isopropanol 75%. O material foi transferido para tubos de 1,5mL, agitado, incubado a 4°C por 15min e centrifugado a 13000 rpm por 20 min. Após o descarte do isopropanol, foi adicionado 1mL de etanol 70% e o material foi centrifugado a 13000 rpm por 15 min. O sedimento foi seco em centrífuga a vácuo a 60°C por 30min e mantido 12h em lugar seco e protegido da luz.

Posteriormente, foram adicionados 2µL de tampão de corrida (formamida e *blue dextran*) a cada amostra. Após incubação por 2min a 95°C para desnaturação, o material foi colocado rapidamente em gelo e aplicado no gel de seqüenciamento.

Tabela 1. Sequência de oligonucleotídeos constitutivos e alternativos dos 43 genes estudados no presente trabalho.

Gene	Oligonucleotídeos Constitutivos	Oligonucleotídeos Alternativos
<i>ACTN1</i>	S:CGGTCAAAGTGTTGAAGG/A:CCAGGACCATCAATGAGGTAG	S: GGAAATCATCCGTGTCCATC/A:CCAGGACCATCAATGAGGTAG
<i>BAT3</i>	S:AGCTGAGGACTCGACATTGG/A:ATGACAGGAAATGGGACTCG	S: GCCTGATGAGTGATCTGGTG/A:ACACAGCCTGGTGGTGTTC
<i>CPNE1</i>	S:GGAAGTCATCATCCCTCAGC/A: CAAGCCCTGAGTTCTCCAAG	S: GCAGTGAGGGGACTGTAAGC/A: TACCTGATGGCAGCTGTGGAG
<i>CPT1B</i>	S: TGGTCAAGTTGCTGGTCTTG/A: TGACGGGCATCTTCTTCTC	S: CTGGATGGGATCCGTGTG/A: GATTGTTGGTGTTCGTGAG
<i>DST</i>	S: TTGTCGGCATCTGTGATAGG/A: AGTTTCCAACCACTCGCTTG	S: TGAAGAGTTTGATTCCGAACG/A: TCCTGCGGAGTACTGTGATG
<i>EVL</i>	S: TCTCCTCTTTCACCTTGTGG/A: GGATGCCTTCGACTTGGAC	S: GTCCAAGTCGAAGGCATCC/A: GAGCTGGCATTTTCGTGAC
<i>FKBP10</i>	S: TCATTGCAGGTCTCAGATGG/A: GCTCTGCCGTGTCTAATCTTC	S: GCTCCTGCGTCAGTTTC/A: TGGTGAAATCAGGACACTG
<i>FXR1</i>	S: GACCCCTGAAACACTTCTG/A: AACCCCTCTGAAACGGAATC	S: TCTCTTTTCAGGCCCTCTG/A: GAATGGAACGCCTACAGATTG
<i>IK</i>	S: TCACTCCATCTCTCCGTTC/A: TATGCCAAGCTACGCCAAC	S: GATTGGTGAAGGAGTGAGG/A: GAGCTTGGGCTGCACATC
<i>KIAA0476</i>	S:GCCGGTAAGTGAGGTAGGTG/A: TTGTATGAGGGGAAGGAACG	S: GAAGAGCTCCTGATGGAAGC/A: CCTGGACTCTTCCCTTTTCC
<i>KIAA0999</i>	S: TTCCAGGTGAGCTTTGAAGG/A: ACACTCTGCACCTCCCTACG	S: AGGTGAGGAGGCCAGACC/A: CACTGGCCATGACCAACC
<i>LDB1</i>	S: CACAGTCGAGGGACACAAAG/A: GACCCCTGATCCACGCTAC	S:CTGGCTCCAGGTATGTAGGC/A:GAAGATGTCAGTGGGCTGTG
<i>LOC126927</i>	S: TTGGTGAAATCTCCAGAG/A: TTGACTTTGACGACGACTGTG	S: TTGGATCTTTGTGTCCAGGTC/S:CAGACCATGGTGGACACG
<i>MED12</i>	S: GTGGAGGTCTCAGTCCATGTC/A: ACCCACCCTTCTACCAATCC	S: CTGGTGACATAGCCACTG/A: GATTGCAGCAACACACAGG
<i>MTMR12</i>	S: CCACCTAGGATTTGGGCTTC/A: TGAGCTCCAAGACAATTTCG	S: TCTCCTTTGTGGACTCTGTGG/A: TGGAAAAGCAAAACCTGGAC
<i>PDK1</i>	S:TGAATGCGACTCATGTAGAATC/A:TCAGAAACCGACACAATGATG	S: TGAACATTCTGGCTGGTGAC/A: AAGAACATGGTTCAGGCTCTC
<i>PHF21A</i>	S: CCCTTGGGAATTGTTTTCAG/A: GTGCCAGTTACTGATGTGC	S: GAAAGCCCGCTTCCTAC/A: CTCCTCGCTCTTCTTCTCC
<i>RQCD1</i>	S: GCAAGACCATCTTACCCAAG/A: TGCCACATTCATCTCCAG	S: TCACCATGCCTAGAGTGCTG/A: GGTGTCAACACCAAGCAATG
<i>SEC13L1</i>	S: CAGAGACAGCCAGGATGTTG/A: CGTGTGTTCTATTGGACCTG	S: CTGGCTGACCTTCTCTATTC/A: AAAAAGGAGAATGGGGCTTC
<i>SEC24C</i>	S: GTGGCCACATCCACATACTG/A: GTTCCAGCCTCAGACAGGTG	S: AAAAGGAAGGGAGCAAAAGG/A: CAGAAGGGGTGGTGTCTGAG
<i>SFRS3</i>	S: CCATCTAGTCTCTCGGACTGC/A: AACCAAGACGGAATTGGAACG	S: GGCTGCTGCTGTAGTTGTTG/A: TCTCAGCCTTCTCTCCAACC
<i>SHN2</i>	S: CCGCTGGCTTTCTCTAAGG/A: GTTTCATGGGATCATCTCTCG	S: AATGTAAATCTGGGCTTGC/A: AACCTCTGGTCCAGAAAC
<i>SLC20A2</i>	S: ATTGCGACCACTGAGAATCC/A: AGCTGATTGCTTCTCTCTG	S:TCCTGTTGAAAGCAGAAGGTG/A:GATGACGACACAACAGACAGC
<i>SPIRE2</i>	S: TCCCTTCTTCAGACTGCTG/A: GGAAGAGGTGATGGACGTG	S: GCACCAGAACTCAGGTCCAC/A: TGCCCTCGGTTTCTATCTG
<i>SS18</i>	S: CTGCTGCTGTCTGGGTAAAC/A: AGATGCATACCAAGGACAC	S: CTGCTGCTGTCTGGGTAAAC/A: AGATGCATACCAAGGACAC
<i>TCF7L2</i>	S: AGGTAGGGGCTCGTCAGG/A: GCCAAGAGGCAAGATGGAG	S: AGGTAGGGGCTCGTCAGG/A: GCCAAGAGGCAAGATGGAG
<i>UAP1</i>	S: GAAGGCGAGAGCCATTTTC/A: TGATGTCCCTTCATCATTGC	S: TGGTCTCTGACTTCCCATTTG/A: TGATGTCCCTTCATCATTGC
<i>ZNF185</i>	S: TTCTGGGGTACTGGGATCTG/A: TGATTTGAGGGGAAGGATG	S: CGGGGCTAGGCATACTTG/A: GATGGAGGCAGGACCAAGG
<i>PLEKHG4</i>	S: TAAGTCTCCAGGCCCTAC/A: CTCAGAGATCTCGTCCAATG	S: TAAGTCTCCAGGCCCTAC/A: ATCTCCTCCCTAGCTGCTG
<i>RAG1AP1</i>	S: AGGTTTGCAGGAGCCAGTAG/A: TCACCAGCTTATCCGCTTC	S: AGGTTTGCAGGAGCCAGTAG/A: AGTTCTTGTGATTTCAAGGTGCC
<i>ZNF384</i>	S: TACACCTTGGCATGCTTAC/A: AACCTTCAAGTGCCACAAC	S: CTCGATGGAAAGCAAAAGG/A: CTTGGGATCCTGGTATGTGC
<i>MPV17</i>	S: GCTGTAGTTTGGCCAGTTG/A: CGTGTTTCTAGGCTGCTTTC	S: AACAGTTGGGTGATGGCTTC/A: CTGCTTGTCTCTGCTTTC
<i>SFRS16</i>	S: TTCTCTCTTGGGGTGATG/A: GGCACCACAGCAGTAGCC	S: CTCAGGGCGGTACCAATAG/A: ACCACCATGTAAGCCACCTC
<i>CEP192</i>	S: TTCTCATGCTTGTGTGTC/A: TAATGCTGAGGACCCCAATG	S: GGGGGATATATGGTCAGCAG/A: TAAGAGGCACGAGGGAAGC
<i>COL17A1</i>	S: TACAGGAGGGCTCACTACCC/A: ATACGCATGGCGGGTAAC	S: AGTACCCTCGGAAGGAATTTG/A: CTATAGTGCCACGAGCAAAAGG
<i>PRKAR1A</i>	S: TGGGAACGTCTTACGGTAGC/A: AACTCATCCCTCGGTTCTCC	S: ACACTGAGAAAGCGGAAGATG/A: TGCCTGCAGATAGCCCTTAG
<i>PTOV1</i>	S: CTCTCACCTGGGCTCAC/A: CAGCTTTGCAGTGGAGTCAG	S: TCCATCTGTGAGGTGGAGAA/A: AGTATGGCCAGCTGGAAG
<i>ROCK1</i>	S: TGCTGCTGTTAGCATGTTCT/A: TTCCGAAAGACTGATTGTC	S: CTGCTTTGGCACAACACAC/A: TTTCCCCCACTGTACTTGC
<i>RPP30</i>	S: TGGAGGTAGAGACGACTGC/A: GGTCCAAATCTGCAACACC	S: CATGGCAGATGTGATGTATAC/A: CTGGTGACGTGAGTTCAAT
<i>TAOK2</i>	S: GATCCAGAGGACCAAGGATG/A: GCCTCTTGAACAGGATCTTC	S: TAGAACTGCCTGGGTCTTCG/A: TGGGTCCAAAGAGAGCTACC
<i>TRIP6</i>	S: GCAGTGGATCTGGCTCGTAG/A: CTACCACCTGGCTGCTTC	S: GGGGTGTCTGGAATCTAA/A: GGGGGAATCTGCAAGTATC
<i>ERICH1</i>	S: CTGTCGCTGAGATTCTCTG/A: ATCTCATGACCAAGCAAAAG	S: AAAGCTCAAAGGCTCTTCTTC/A: GGGACCCCTCTCTCTGAC
<i>CES1</i>	S: GTGGTTCCGGGCCACAATCA/A: AGACCTTAGTAAGGAATCG	S: CTGGAGCACCCACCTACATG/A: TCTGGTCCCTACTCGAGA

Figura 1. Metodologia de desenho de oligonucleotídeos para eventos de *splicing* localizados no final da sequência. Caixas brancas: exons constitutivos; caixa cinza: exon alternativo; setas: oligonucleotídeos; ORESTES: sequência parcial de RNAm; RefSeq: sequência de referência do gene.

Resultados e Discussão

Análises in silico

Classe A: Sequências com perda e ganho de exon(s) e perda de regiões exônicas ou ganho de regiões intrônicas

Das 197 sequências da classe A, 152 foram re-confirmadas como sequências com eventos de *splicing* ainda não descritos em bancos de dados públicos (Sub-grupo A1).

Vinte e sete sequências foram classificadas em um grupo especial (Sub-grupo A2), por apresentarem múltiplos eventos de *splicing* (mais de um tipo) ou pela dificuldade de análise em função do grande número de isoformas de mRNAs depositados em bancos de dados (como no caso do gene *CREM*). Também foram incluídas no sub-grupo A2 sequências que mostraram alinhamento com a região central de genes que exibem uma quantidade grande de exons (como o gene *TTN* com mais de 313 exons) e cujas ESTs sugerem um número muito elevado de exons alternativos, bem como sequências com suspeita de contaminação genômica ou de *trans-splicing* (partes da sequência alinhadas com genes distintos, mapeado em cromossomos diferentes). Finalmente, 18 eventos de *splicing*, cujos mRNAs correspondentes foram recentemente depositados em bancos de dados públicos, não foram considerados formas novas de *splicing* e, por esse motivo, deixados em um grupo à parte (Sub-grupo A3).

Sub-grupo A1. Das 152 sequências, 46 (30,2%) corresponderam a deleções de exons, 45 (29,6%) a novos exons, 12 (7,9%) a exons truncados e 33 (21,7%) a junções de exons. As 16 sequências restantes corresponderam a eventos combinados, entre eles, 12 (7,9%) casos com deleção de exon / novo exon, dois (1,3%) com novo exon / exon truncado, uma com novo exon / junção de exon (0,7%) e uma com perda de exon / junção de exon (0,7%). Esses eventos podem ser resumidos em 79 inserções (52%) e 73 deleções (48%), sendo 16 deles combinações de *splicing*.

Os tamanhos dos segmentos que participaram do *splicing* variaram de aproximadamente 25pb a 480pb, sendo os mais abundantes na faixa de 80pb a 300pb. Foram detectados também alguns casos com mais de 1Kb.

A regra GT/AG dos sítios doadores/aceptores foi confirmada em 145 (95,4%) eventos enquanto, nos sete casos restantes (4,6%), estiveram presentes os dinucleotídeos AA/AG, AT/AC, AT/AG, CA/AG, CT/AT, TA/GA, seis dos quais em exons truncados. Em 77 casos (50,6%), o *splicing* não modificou a matriz de leitura. Entre os restantes, 17 estavam nas extremidades da sequência, o que impediu a avaliação da matriz.

Dos 152 eventos, 145 ocorreram na região codificadora e apenas 7 na região não traduzível UTR. Os resultados mostraram que a grande maioria ocorreu na região central e na região 3' do RNA. Em relação aos seus efeitos sobre a proteína, foram observados 94 casos (62%) com truncamento do produto em consequência da formação de um códon de terminação. Destes, 58 eventos coincidiram com região de domínio protéico.

A análise realizada mostrou 21 sequências (13,8%) com elementos repetitivos próximos (até 30pb) ou diretamente na região do novo *splicing*. Entre essas sequências repetitivas, foram observadas nove repetições *Alu*, oito *Line*, duas *MIR*, uma *LTR* e uma sequência considerada de baixa complexidade. Também foi observada uma sequência intrônica com todas as características de um minissatélite, cujo consenso aparentemente possui 50pb e é repetido in tandem 27 vezes (Figura 2).

Sequência Consenso:

CTGTGTCCTC TCCAAGGGGC ATTGCTGGGG GGGTGTTACA AGGATAGATG

Figura 2. Sequência consenso de 50pb do candidato a minissatélite, repetida 27 vezes na sequência de referência.

Classe B: Sequências com extensões e deleções da região 5' do exon

Das 203 sequências da classe B, 156 foram confirmadas como apresentando eventos de *splicing* ainda não descritos em bancos de dados públicos. Destas, 123 exibiram eventos individuais de *splicing* (Sub-grupo B1) e 33 continham eventos múltiplos, de mais de um tipo (Sub-grupo B2). Vinte e duas sequências, cujos mRNAs foram recentemente depositados em bancos de dados públicos, não foram consideradas formas novas de *splicing* e, por esse motivo, foram classificados como sub-grupo B3. Finalmente, 25 sequências com suspeita de contaminação genômica foram desconsideradas e reunidas em um Sub-grupo B4.

Sub-grupos B1 e B2. Dos 156 eventos estudados, 121 (77,5%) corresponderam a extensões 5' de exons e dois (1,3%) a deleções (Sub-grupo B1). Entre os eventos combinados (Sub-grupo B2), nove (5,8%) mostraram extensões 5' e 3' de exons (uma delas incluía um novo exon), três (1,9 %) apresentaram junções de exons e extensões 5', cinco (3,2%) combinavam novos exons com extensões 5', dois (1,3%) apresentaram deleções completas de exons (*skipping*) e extensões 5', 12 (7,7%) mostraram deleções nas regiões 5' e 3' de exons (6 incluíam perda de exons) e dois casos (1,3%) apresentaram deleções da extremidade 5' e perda de exons (uma delas mostrou ganho de um novo exon). Tais eventos podem ser resumidos em 141 inserções (88,7%) e 18 deleções (11,3%), sendo três delas (1,9%) eventos combinados de inserção e deleção.

Os tamanhos dos segmentos que participaram do *splicing* variaram de aproximadamente 13pb a 1800pb, sendo os mais abundantes na faixa de 80pb a 500pb. Em 128 (82,1%) do total de 156 eventos, o tamanho não pôde ser deduzido pelo fato da extensão estar em uma das extremidades da sequência. Na verdade, a frequência de extensões que aparecem em exons da ponta da sequência analisada (na maioria das vezes, no seu início) é muito maior que a sua frequência em exons internos. O mesmo é observado nas muitas extensões de ESTs disponíveis na interface gráfica do BLAT (Universidade da Califórnia, Santa Cruz, *Genome Bioinformatics Group*). Uma das explicações seria que essas extensões representam inserção completa do íntron envolvido ou inclusão de sequências de pré-RNA ou de RNAs anti-sense. De qualquer maneira, isto não explica sua posição geralmente no início das sequências e levanta a possibilidade de um artefato decorrente, por exemplo, de erros

de clonagem e/ou sequenciamento das ORESTES. Situações similares foram referidas por De la Grange *et al.*, (2005).

Dos 156 eventos estudados, apenas seis foram localizados nas extremidades não traduzíveis UTRs, sendo a grande maioria presente na região codificadora ou CDS. As análises também mostraram que mais de 90% dos eventos estavam localizados na região central e distal do mRNA.

Como a maioria das extensões estava presente no início da sequência, apenas em 29 sequências foi possível fazer a análise dos sítios doador e receptor. Nesses 29 casos, a regra GT/AG dos sítios doador / aceptor foi confirmada em 21 (72,4%) enquanto, em 8 casos (27,6%), estiveram presentes os dinucleotídeos AA/AG, GG/AG, GC/AG, TT/AG, CG/AT, AT/GG, GC/GT e CC/CC, seis deles em sequências que mantiveram a matriz de leitura. Em 18 casos (62%), o *splicing* não modificou a matriz de leitura. Nas outras 128 sequências, os dinucleotídeos seguintes ao possível início/término da extensão continham bases aleatórias e não obedeciam à regra GT/AG, reforçando a hipótese de artefato ou de que a extensão não estava completa.

Em relação aos efeitos do *splicing* sobre as proteínas, foram observados 13 casos (44,8%) com truncamento do produto decorrente do surgimento de códons de terminação prematuros. Sete deles (24,1%) coincidiram com a região de um domínio protéico.

A análise das isoformas também revelou novos sítios potenciais de fosforilação e glicosilação na proteína. Onze sequências (7%) exibiram elementos repetitivos na região do *splicing* alternativo. Entre essas sequências repetitivas, foram observadas cinco repetições *Alu*, duas *Line*, duas *MIR* e duas sequências consideradas de baixa complexidade. As sequências que apresentaram elementos repetitivos pertenceram unicamente à categoria inserção.

Classe C: Sequências com extensão e deleção da região 3' do exon

Das 240 sequências da classe C, 68 eram referentes a eventos combinados, citados nas classes anteriores. Das 172 sequências remanescentes, 141 foram confirmadas como apresentando eventos de *splicing* ainda não descritos em bancos de dados públicos. Dessas, 101 exibiram eventos individuais de *splicing* (Sub-grupo C1) e 40 continham eventos múltiplos, de mais de um tipo (Sub-grupo C2). Vinte e uma

sequências, cujos mRNAs foram recentemente depositados em bancos de dados públicos, não foram consideradas formas novas de *splicing* e, por esse motivo, foram colocadas no Sub-grupo C3. Finalmente, 10 sequências com suspeita de contaminação genômica foram desconsideradas e reunidas em um Sub-grupo C4.

Entre os 101 eventos do primeiro sub-grupo, 93 (92,1%) corresponderam a extensões 3' de exons e 8 (7,9%) a deleções. No segundo subgrupo, 14 (35%) mostraram extensões 3' e 5' de exons (duas delas incluíam um novo exon), 11 (27,5%) mostraram deleções nas regiões 3' e 5' (sete dessas últimas continham perda de exons), cinco (12,5%) apresentaram junções de exons e extensões 3', oito (20%) combinavam extensões 3' com novos exons e duas sequências (5%) combinavam deleções 3' e perda de exons. Tais eventos podem ser resumidos em 119 inserções (84,4%) e 22 deleções (15,6%).

Os tamanhos de 25 dos 141 segmentos (Subgrupos 1 e 2) envolvidos variaram de aproximadamente 15pb a 800pb, sendo os mais abundantes na faixa de 35pb a 200pb. O tamanho dos 116 (82,2%) restantes não pôde ser deduzido pelo fato da extensão estar em uma das extremidades da sequência. Tais extensões, como referido acima, podem corresponder a artefatos.

Dos 141 eventos estudados, dez apenas foram localizados nas extremidades UTRs e mais de 70% estavam localizados na região central do mRNA.

Das 25 sequências que permitiram análise dos sítios doador e aceptor, a regra GT/AG foi confirmada em 11 (44%) enquanto, em 14 casos (56%), estiveram presentes os dinucleotídeos GC/AG (em três vezes), GG/AG (duas vezes), GC/GG, AT/GG, CG/AG, GA/AG, AT/GA, AG/CA, AA/AG, GT/CT e CC/CC (uma vez). A matriz de leitura não foi alterada em sete das sequências com GT/AG e em quatro daquelas com dinucleotídeos atípicos.

Nas outras 116 sequências que não permitiram análise de um dos sítios doador/aceptor, os dinucleotídeos seguintes ao possível início/término da extensão, após alinhamento com a sequência de referência, mostraram bases aleatórias e não obedeceram à regra GT/AG, reforçando a hipótese de artefato ou de que a extensão era incompleta.

Em relação aos eventuais efeitos do *splicing* sobre as proteínas, foram observados 14 casos (56%) com truncamento do produto decorrente do surgimento de

códons de terminação prematuros. A metade deles coincidiu com a região de um domínio protéico (Figura 3). Como esperado, a análise das isoformas também revelou ganho ou perda de novos sítios potenciais de fosforilação (Figura 4) e glicosilação na proteína. Por outro lado, foi observado um único elemento repetitivo (*Line*), que estava presente em região de *splicing* do gene *HSH2D*.

Das análises realizadas nas classes A, B e C, algumas observações interessantes foram feitas. Em primeiro lugar, apenas a classe A permitiu análises completas, dado que nas outras classes nem sempre a sequência estava completa. Portanto, os resultados da classe A são mais consistentes que os das demais classes, especialmente porque o número de eventos avaliados foi muito maior. De qualquer modo, foi possível concluir que o ganho de exons ou de parte de exons é substancialmente maior que o de perda. A regra GT/AG das extremidades dos íntrons deve atingir valores próximos a 90% dos casos. Aproximadamente 50% dos eventos não modificaram a matriz de leitura, mas a maioria ocorreu na região codificadora, potencialmente levando ao truncamento da proteína, afetando um domínio protéico e removendo ou adicionando novos sítios potenciais de fosforilação e glicosilação. Realmente, é estimado que 75% dos eventos de *splicing* alternativo ocorrem na região traduzida e têm consequências no produto protéico (Kriventseva *et al.*, 2003; Sorek *et al.*, 2004; Stamm *et al.*, 2005).

A presença de elementos repetitivos na região de *splicing* pode introduzir ou interromper sequências *cis* reguladoras. Por exemplo, os elementos *Alu* contêm sítios potenciais de *splicing* e podem evoluir para exons (Lev-Maor *et al.*, 2003). Existem estimativas de que mais de 5% dos exons alternativos humanos são derivados desses elementos (Sorek *et al.*, 2002). Existem também referências da influência de repetições CA intrônicas nas decisões de *splicing* alternativo (Hung *et al.*, 2008).

Figura 3. Representação gráfica da ferramenta ExpASY PROSITE mostrando a perda de um domínio protéico rico em lisina na isoforma alternativa do gene *ZRF1* como consequência de uma deleção de 84pb (28 aminoácidos) na região 3' do exon 9.

Figura 4. Resultados obtidos com a ferramenta NetPhos mostrando os sítios potenciais de fosforilação em região da proteína presente na isoforma constitutiva (A) e ausente na variante alternativa (B) do gene *BRPF1*. A região correspondente à deleção inicia-se no aminoácido 945 da sequência alternativa.

Elementos Reguladores

Os eventos de *splicing* são regulados por elementos *cis* (seqüências no pré-RNA mensageiro) e *trans* (fatores celulares - proteínas e RNAs). Os primeiros participam da identificação dos limites exon-intron e impedem a inclusão de pseudoexons no RNA maduro enquanto os elementos *trans* participam da formação do *spliceosome* e possuem um papel importante na regulação do *splicing* alternativo.

Os elementos *cis* (estimuladores e silenciadores) são encontrados em exons e introns, geralmente próximos aos sítios de *splice*, e atuam nos eventos de *splicing* constitutivos e alternativos (Cartegni *et al.*, 2002). Tanto os estimuladores exônicos (ESEs) como os intrônicos (ISEs) promovem a inclusão de um exon enquanto os silenciadores (ESSs e ISSs) possuem efeito contrário (Black, 2003). Em geral, ambos os tipos, estimuladores e silenciadores, se superpõem, formando um elemento regulador composto, (Caceres & Kornblihtt, 2002) e sua posição em relação a um exon alternativo pode determinar efeito positivo ou negativo (Hui *et al.*, 2005).

Entre os elementos *trans* estão incluídas as proteínas ligadoras de RNA (RBPs), como as SR (ricas em arginina-serina e que geralmente ligam-se a ESEs), e membros da família de proteínas hnRNP (ribonucleoproteínas nucleares heterogêneas), que atuam sobre ESSs e ISSs (Dreyfuss *et al.*, 1993; Tacke & Manley, 1999). Entretanto, existem muitas exceções a essa regra: têm sido descritas SR clássicas com papel repressor e hnRNP que são estimuladoras (revisto por Buratti *et al.*, 2006).

Mutações nos elementos *cis* e alterações nas concentrações de elementos *trans* podem afetar a maquinaria de *splicing* e resultar na síntese de formas novas de RNA ou levar a proporções anormais das isoformas de um tecido. Se os transcritos não usuais não forem degradados, isoformas protéicas aberrantes podem ser sintetizadas e terem, por exemplo, propriedades tumorigênicas (revisto por Venables, 2006; Pajares *et al.*, 2007). Essa regulação é dependente, portanto, não apenas da seqüência do RNA, mas também de um fino balanço entre os fatores *trans* disponíveis na célula, que podem variar de tecido para tecido, durante o desenvolvimento ou na presença de um estímulo externo. A referência da participação de RNAs pequenos nucleolares (snoRNA) na regulação de *splicing* mostra o quão desconhecido e complexo é esse processo (Kishore & Stamm, 2006).

Existem muitas dificuldades na análise do papel dos elementos *cis* e *trans* nos eventos de *splicing*. Em relação aos primeiros, suas seqüências exibem grande diversidade e podem ser altamente degeneradas, o que dificulta a avaliação de seu papel com base apenas na seqüência primária (Ladd & Cooper, 2002). Além disso, a estrutura secundária do RNA também pode influenciar o processo de reconhecimento dos motivos de ligação por fatores de ligação, indicando que a presença de um elemento regulador não é traduzida em uma afinidade de ligação forte (Buratti *et al.*, 2004). Portanto, a validação de predições é ainda um passo fundamental na maioria dos estudos de *splicing*. De qualquer maneira, não existem dúvidas de que abordagens *in silico* acuradas são úteis na caracterização de elementos reguladores (Wang *et al.*, 2004; Han *et al.*, 2005).

Dos 518 exons estudados, a maioria apresentou evidências da presença de elementos reguladores de *splicing*. Esse achado está de acordo com as afirmações da literatura de que a maioria dos exons alternativos é controlada por múltiplos elementos estimuladores e silenciadores (Hui *et al.*, 2005).

Os resultados mostraram 1378 elementos reguladores: 1187 estimuladores (59% do tipo ESE e 27% do tipo PESE) e 192 silenciadores (6% do tipo ESS e 8% do tipo PESS) (Tabela 2). Um exemplo da análise realizada em cada um dos 518 exons está ilustrado na Figura 5 (gene *DST*).

Não foi observado um padrão específico de posição ou presença desses elementos em relação às características do exon alternativo, como perda ou ganho de seqüências, ou do exon vizinho (posição 5' ou 3' em relação ao exon alternativo). Esse último resultado pode ser decorrente da análise manual desenvolvida. Embora a validação experimental das predições seja fundamental, os resultados obtidos abrem caminho para novos estudos *in silico* sobre o papel dos elementos reguladores no *splicing* alternativo.

Tabela 2. Frequência de elementos reguladores em 518 exons analisados de 152 sequências com eventos novos de *splicing* alternativo.

Elementos reguladores	Frequência N (%)	Total
Estimuladores		
ESE	820 (59%)	1187 (86%)
PESE	367 (27%)	
Silenciadores		
ESS	80 (6%)	192 (14%)
PESS	112 (8%)	
Total		1379

Gene DST

>R1_CL308887_CT1(ex90)

1- TTTGGAGACTCCCAGCAACTGCGACTGGTCCGGATCCTGCGGAGTACTGT

51- **GATGGTTCGTGTGGAGGTGGATGGATGGC**ACTTGAATGAGTTCTTAGTGAA
 GATGGTTCGT
 ATGGTTCGTG
 GTTCGTGTTG
 TTCGTGTTGG
 TGTGGAGGT
 GGAGGTGGAT
 GAGGTGGATG
 (Elemento ESS)

101- AAATGATCCTTGCAGGG

>R1_CL308887_CT1 (ex90-b new)

1- TTCAATCATCATGGG**GTAAATGTT**ACGTTCCG**GAATCAA**CTCTCA**ATTACTACTACTCAGCCT** -65
 GTAAATG GAATCA TTACTAC
 TAAATGT AATCAA CTACTAC
 AAAATGTT ATCAA CTACTCA
 AAATGTTA AAATC ACTCAGC
 (Elemento PESE) (Elemento ESE) (Elemento ESE)

61- AGCCTACTATAG

>R1_CL308887_CT1(ex91)

1- CC**AAAGGAAGGACAAACATGGA**CTG**CGTGAGAAGTT**CATTTTAGCAGATGGTGCCAGCC
 AAAGGA TGAGAA
 AAGGAA GAGAAG
 AGGAAG AGAAGT
 GAAGGA GAAGTT
 AAGGAC (Elemento ESE)
 AGGACA
 GACAAA
 CAAACA
 AACATG
 ATGGAA
 TGGAAC (Elemento PESE)
 GAACTG GAAGATCC
 (Elemento ESE) CGAAGATC
 CCGAAGAT
 61- AGGGTATGGCTGCTTTCCGACCCCGAGG**CCGAAGATCC**CGGCCATCATCAGAGGCGCTT
 CGAAGA
 GAAGAT
 AAGATC
 (Elemento ESE)

121- CACCAACAGATCCACTTCTGTGTCCAGTCAGGCTGCGCAGGCGGCTCCCCACAGGTCC

181- CTGCCACCACCACACCCAAG

Figura 5. Exemplo de elementos reguladores posicionados nos exons 90, 90b (novo exon) e 91 do gene *DST*. Em azul, elementos estimuladores e em vermelho elementos silenciadores de *splicing*.

Análises de Expressão Gênica

RT-PCR

Foi analisada, por RT-PCR, a expressão de isoformas constitutivas e alternativas de 43 genes: *ACTN1*, *BAT3*, *CARS*, *CCNL1*, *CEP192*, *CHN2*, *COL17A1*, *CPT1B*, *DST*, *ERICH1*, *FKBP10*, *FXR1*, *GLTSCR2*, *HADHB*, *IK*, *KIAA0476*, *KIAA0999*, *LDB1*, *LOC126927*, *MPV17*, *MTMR12*, *PDK1*, *PHF21A*, *PIK4CA*, *PLEKHG4*, *PRKAR1A*, *PTOV1*, *RAG1AP1*, *ROCK1*, *RPP30*, *RQCD1*, *SEC13L1*, *SFRS16*, *SFRS3*, *SLC20A2*, *SPIRE2*, *TAF4*, *TAOK2*, *TNRC6A*, *TRIP6*, *UAP1*, *ZNF185* e *ZNF384* em amostras procedentes de linhagens, amostras de autópsias, tecido tumoral e margem cirúrgica procedente de pacientes com carcinoma epidermóide de cabeça e pescoço. Alguns exemplos são apresentados nas Figuras 6 e 7. Todos os genes mostraram produtos de amplificação para as isoformas alternativas em todos os tecidos analisados, com exceção dos genes *CARS*, *CCNL1* e *LDB1* para os quais não foram observados transcritos alternativos nas amostras frescas de pacientes com câncer. No entanto, esses produtos estiveram presentes nas linhagens celulares e nos tecidos normais de autópsia.

Seqüenciamento

A Figura 8 mostra os resultados obtidos com a ferramenta BLAT, exibindo os oligonucleotídeos utilizados e os fragmentos sequenciados das isoformas constitutiva e alternativa dos genes *BAT3*, *DST*, *CEP192*, *TRIP6* e *MPV17*, em relação à seqüência de referência do transcrito. As seqüências das isoformas de 21 desses genes foram confirmadas por seqüenciamento (*ACTN1*, *BAT3*, *CEP192*, *CHN2*, *COL17A1*, *CPT1B*, *DST*, *FKBP10*, *FXR1*, *IK*, *LOC126927*, *MPV17*, *PDK1*, *PHF21A*, *PRKAR1A*, *ROCK1*, *RQCD1*, *SFRS16*, *TAOK2*, *TRIP6*, *UAP1*). As Figuras 9 e 10 mostram os cromatogramas obtidos para as isoformas alternativas dos genes *BAT3* e *DST*.

Os dados de validação por RT-PCR e sequenciamento tiveram alto índice de sucesso. Na verdade, todas as isoformas alternativas dos genes de interesse foram confirmadas em muitos tipos celulares. Esses resultados indicam que as ferramentas utilizadas por Reis *et al.* (2005) para análise de ORESTES foram adequadas e que os bancos de dados de seqüências expressas podem fornecer importantes informações sobre os eventos de *splicing* alternativo.

Figura 6. Validação das isoformas procedentes de *splicing* alternativo, correspondentes aos genes *CARS*, *HADHB* e *LDB1* em géis de agarose 1%. Linhagens celulares FaDu e HeLa. Tecidos Normais de autópsia: (1 ou A) cérebro, (2 ou B) testículo, (3) coração, (4) pulmão, (5) estômago, (6) rim, (7) laringe, (8) fígado, (L) marcador de peso molecular.

Figura 7. Expressão das isoformas procedente de *splicing* alternativo, correspondentes aos genes *SFRS3*, *PIK4A1*, *RQCD1* e *TAF4* em géis de agarose 1%. Linhagens FaDu, SiHa e HeLa: 1-5. Tecidos Normais de autópsia: (D) Pulmão, (E) Estômago, (F) Rim, (G) Laringe, (H) Fígado, (L) marcador de peso molecular.

Figura 8. Resultados obtidos com a ferramenta BLAT mostrando os oligonucleotídeos sense (R) e anti-sense (F) utilizados e o fragmento sequenciado da isoforma alternativa dos genes *BAT3*, *DST*, *CEP192*, *TRIP6* e *MPV17*, em relação à sequência de referência do transcrito.

Figura 10. Resultados do seqüenciamento da isoforma alternativa do gene *DST* para os primers sense (C) e antisense (D), visualizados com a ferramenta FinchTV. Em azul está marcada a região homóloga à seqüência do gene.

Transcritos anti-sense

RNAs pequenos anti-sense são geralmente sintetizados a partir de um loco diferente do seu alvo, enquanto a transcrição de RNAs longos anti-sense freqüentemente são codificados no mesmo loco do RNA sense (Vanhee-Brossollet & Vaquero, 1998). No presente trabalho, os dados obtidos sugeriram que não existem transcritos para os genes estudados pertencentes à categoria de RNAs anti-sense longos. Entretanto, não é possível excluir a presença de RNAs intrônicos que atuam sobre o pré-RNA (Vanhee-Brossollet & Vaquero, 1998; Reis *et al.*, 2004).

PCR em tempo real

A análise quantitativa dos produtos alternativos e constitutivos foi realizada para 31 genes (*ACTN1*, *BAT3*, *CEP192*, *CHN2*, *COL17A1*, *CPT1B*, *DST*, *ERICH1*, *FKBP10*, *FXR1*, *IK*, *MPV17*, *MTMR12*, *PDK1*, *PHF21A*, *PLEKHG4*, *PRKAR1A*, *PTOV1*, *RAG1AP1*, *ROCK1*, *RPP30*, *RQCD1*, *SEC13L1*, *SFRS16*, *SLC20A2*, *SPIRE2*, *TAOK2*, *TRIP6*, *UAP1*, *ZNF185*, *ZNF384*). A seleção desses genes levou em conta sua função, o seu envolvimento com tumorigênese e também as características do fragmento envolvido.

Os níveis relativos das isoformas alternativas e constitutivas dos genes selecionados não mostraram um padrão nítido de expressão diferencial. De modo geral, não foi observado um padrão específico de expressão das isoformas nas amostras dos grupos agressivo e não agressivo. Entretanto, os transcritos dos genes *ACTN1*, *BAT3*, *CEP192*, *MPV17*, *PDK1*, *PRKAR1A*, *RAG1AP1* e *TRIP6* mostraram uma aparente expressão diferencial das isoformas quando comparadas entre si ou na comparação dos tecidos tumorais com os normais ou entre os grupos agressivo e menos agressivo. As Figuras 11 e 12 mostram os resultados de PCR em tempo real das isoformas constitutiva e alternativa dos genes *ACTN1* e *MPV17* em amostras de tumores de cabeça e pescoço. No caso do gene *MPV17*, as duas isoformas possuem o mesmo comportamento e estão mais expressas nos tumores agressivos que nos menos agressivos e nas margens cirúrgicas. Como pode ser observado, os transcritos *ACTN1* constitutivo e alternativo são

mais abundantes nos casos agressivos e a forma alternativa é menos expressa que a constitutiva nos dois grupos de tumores.

O gene *ACTN1* está relacionado com a organização do citoesqueleto, com processos de adesão e provavelmente com apoptose. Recentemente, Thorsen *et al.* (2008) descreveram a mesma variante estudada no presente trabalho e sugeriram que ela é tumor-específica para câncer de cólon, bexiga e próstata. Além disso, observaram que seu padrão de *splicing* sofre mudanças dependendo do estágio dos tumores, semelhantemente ao observado para os carcinomas de cabeça e pescoço aqui estudados. O gene *PRKAR1A*, que está envolvido na sinalização dependente de hormônio, possui várias isoformas e uma delas, diferente da analisada no presente trabalho, já foi citada com expressão diferencial em tecidos normais da adrenal (Peverelli *et al.*, 2005). Sobre o gene *MPV17*, codificador de uma proteína da membrana interna da mitocôndria, e sobre os genes *CEP192*, *RAG1AP1*, não existem relatos na literatura sobre sua participação no processo neoplásico. Entre os demais genes que exibiram expressão diferencial de suas isoformas, o *BAT3* já foi associado com sobrevivência maior de células de osteossarcoma (Tsukahara *et al.*, 2009), o *TRIP6* com invasividade (Chastre *et al.*, 2009) e o *PDK1*, um regulador major da atividade quinase, parece conferir vantagem proliferativa e estar associado com motilidade, invasão e metástase (Xie *et al.*, 2006; Pinner and Sahai, 2008).

Dos dados obtidos, podemos concluir que isoformas derivadas de *splicing* alternativo atuam na tumorigênese e são alvos potenciais para tratamento. O presente trabalho identificou e validou variantes derivadas de *splicing* alternativo em algumas dezenas de genes, algumas delas com expressão diferencial em tecidos tumorais e normais, indicando sua importância para o entendimento da biologia dos tumores e para sua utilização como biomarcadores dessa doença ou como alvos terapêuticos.

Figura 11. Gráficos obtidos por PCR em tempo real mostrando expressão diferencial das isoformas constitutiva e alternativa do gene *ACTN1* em amostras de tumores menos agressivos e agressivos. (M) Margem cirúrgica; (T) tecido tumoral e (L) linfonodo. Grupo de amostras menos agressivas (sobrescrito 1); grupo de amostras agressivas (sobrescrito 2). Como controle endógeno foram utilizados os genes *GAPDH*, *ACTB* e *TUBA1C* e como calibrador endógeno foi utilizado o tecido de margem cirúrgica.

Figura 12. Gráficos obtidos por PCR em tempo real mostrando aumento da expressão gênica de ambas as isoformas do gene *MPV17* em amostras de tumores agressivos. (M) Margem cirúrgica; (T) tecido tumoral e (L) linfonodo. Grupo de amostras menos agressivas (sobrescrito 1); grupo de amostras agressivas (sobrescrito 2). Como controle endógeno foram utilizados os genes *GAPDH*, *ACTB* e *TUBA1C* e como calibrador endógeno foi utilizado o tecido de margem cirúrgica.

Abstract

Alternative splicing is a process by which the exons of the primary gene transcript are linked in different ways during RNA processing resulting in distinctive proteins. It is an important mechanism in eukaryotic gene expression that enhances proteome diversity and, therefore, the coding capacity of the genome. Different mechanisms seem affect alternative splicing regulation, including metabolic stresses. In the present study, a detailed informatics analysis of ORESTES sequences from head and neck tissues was performed. This *in silico* analysis revealed that gain of exon sequences is more frequent than exon skipping and GT/AG rule is predominant in splice sites. We observed that alternative splicing usually does not alter the reading frame but may disrupt a protein domain and remove or add new phosphorylation and glycosylation sites. Repetitive and potential regulator elements were frequent in the alternative sequences or in their neighbors. The expression of putative splicing isoforms was investigated in different tissues, including upon stress conditions. We validated approximately 50 new splicing events in normal and tumor cells. Several variants, such as those from *HNRNPK*, *ACTN1*, *BAT3*, *CEP192*, *MPV17*, *PDK1*, *PRKAR1A*, *RAG1AP1* and *TRIP6* genes showed distinctive expression pattern in different cell types, in normal and cancer samples from head and neck carcinoma patients and, in some cases, in different tumor stages. We also validated a new transcript of *RIPK2* gene, which codes a serine/threonine kinase that activates the NF- κ B pathway, and observed a shift in the expression of this variant as a response to temperature stress *in vitro*. It is currently not clear whether alternative splicing is the cause or the consequence of the neoplastic process. Our data add new information to this topic and provide some examples on the importance of RNA processing in gene expression regulation, both in normal and disease conditions.

Resumo

O *splicing* alternativo é um processo pelo qual os exons de um transcrito primário são ligados de diferentes maneiras durante o processamento do RNA, levando à síntese de proteínas distintas. Compreende um importante mecanismo na expressão gênica de eucariotos, responsável pelo aumento da diversidade proteômica e, portanto da capacidade codificante do genoma. Diferentes mecanismos parecem afetar a regulação do *splicing* alternativo, incluindo estresse metabólico. No presente estudo, foi realizada uma análise detalhada de sequências ORESTES de tecidos de cabeça e pescoço. Essa análise revelou que o ganho de sequências exônicas é mais freqüente que sua perda, e que a regra GT/AG é predominante em sítios de *splicing*. Nós observamos que o *splicing* alternativo geralmente não altera a matriz de leitura, mas pode afetar um domínio protéico e remover ou adicionar novos sítios de fosforilação e glicosilação. Elementos reguladores potenciais e elementos repetitivos foram freqüentes nas sequências alternativas e nas suas vizinhas. A expressão de isoformas de *splicing* potenciais foi investigada em diferentes tecidos, incluindo sob condições de estresse. Foram validados cerca de 50 eventos de *splicing* novos em células normais e tumorais. Diversas variantes, tais como aquelas dos genes *HNRNPK*, *ACTN1*, *BAT3*, *CEP192*, *MPV17*, *PDK1*, *PRKAR1A*, *RAG1AP1* e *TRIP6* mostraram padrões de expressão distintos em diferentes tipos celulares, em amostras normais e tumorais de pacientes com carcinoma de cabeça e pescoço e, em alguns casos, em diferentes estágios do tumor. Também foi validado um transcrito novo do gene *RIPK2*, responsável por codificar uma quinase de serine/treonine que ativa a via de sinalização NF- κ B, e foi observada uma mudança na expressão dessa variante em resposta ao estresse térmico *in vitro*. Ainda não está claramente definido se o *splicing* alternativo é causa ou consequência do processo neoplásico. Nossos dados adicionam informações novas a esse tópico e fornecem alguns exemplos que evidenciam a importância do processamento do RNA na regulação da expressão gênica, tanto em condições normais como de doença.

Conclusões

O presente trabalho teve as seguintes conclusões:

1. As ferramentas utilizadas por Reis et al. (2005) para análise de ORESTES foram adequadas e eficientes para detecção de novos eventos de *splicing* alternativo.
2. O ganho de exons ou de parte de exons é substancialmente maior que o de perda em eventos de *splicing* alternativo. As sequências GT/AG das extremidades dos íntrons são os sítios de *splice* mais freqüentes. Uma grande parte dos eventos de *splicing* alternativo não modifica a matriz de leitura, embora a maioria ocorra na região codificadora do transcrito e potencialmente pode levar ao truncamento da proteína, afetar um domínio protéico e remover ou adicionar novos sítios potenciais de fosforilação e glicosilação.
3. Elementos reguladores potenciais e elementos repetitivos são freqüentes nas sequências alternativas e nas suas vizinhas.
4. Elementos reguladores potenciais e elementos repetitivos são freqüentes nas sequências alternativas e nas suas vizinhas.
5. As isoformas derivadas de *splicing* alternativo dos genes *HNRPK*, *ACTN1*, *BAT3*, *CEP192*, *MPV17*, *PDK1*, *PRKAR1A*, *RAG1AP1* e *TRIP6* possuem expressão diferencial quando comparadas entre si, na comparação dos tecidos tumorais com os normais ou entre os grupos de tumores agressivos e menos agressivos.
6. As variantes do gene *RIPK2*, derivadas de *splicing* alternativo, possuem expressão diferencial *in vitro* em resposta à temperatura.

Referencias Bibliográficas

Aebi, M., H. Hornig, et al. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell*, v.47, n.4, Nov 21, p.555-65. 1986.

Ambrosch, P., M. Kron, et al. Clinical staging of oropharyngeal carcinoma: a critical evaluation of a new stage grouping proposal. *Cancer*, v.82, n.9, May 1, p.1613-20. 1998.

Andreadis, A. Misregulation of tau alternative splicing in neurodegeneration and dementia. *Prog Mol Subcell Biol*, v.44, p.89-107. 2006.

Ben-Dov, C., B. Hartmann, et al. Genome-wide analysis of alternative pre-mRNA splicing. *J Biol Chem*, v.283, n.3, Jan 18, p.1229-33. 2008.

Black, D. L. Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell*, v.103, n.3, Oct 27, p.367-70. 2000.

Black DL Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, v.72, p.291-336. 2003.

Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell*, v.126, n.1, Jul 14, p.37-47. 2006.

Bonnal, S., C. Martinez, et al. RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol Cell*, v.32, n.1, Oct 10, p.81-95. 2008.

Boon, K. L., R. J. Grainger, et al. prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nat Struct Mol Biol*, v.14, n.11, Nov, p.1077-83. 2007.

Bourdon, J. C. p53 and its isoforms in cancer. *Br J Cancer*, v.97, n.3, Aug 6, p.277-82. 2007.

Buratti. hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res*, v.32, n.14, p.4224-36. 2004.

Buratti, E., M. Baralle, et al. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res*, v.34, n.12, p.3494-510. 2006.

Burghes, A. H. e C. E. Beattie. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nat Rev Neurosci*, Jul 8. 2009.

Caceres, J. F. e A. R. Kornblihtt. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet*, v.18, n.4, Apr, p.186-93. 2002.

Cartegni, L., S. L. Chew, et al. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*, v.3, n.4, Apr, p.285-98. 2002.

Chastre, E., M. Abdessamad, et al. TRIP6, a novel molecular partner of the MAGI-1 scaffolding molecule, promotes invasiveness. *Faseb J*, v.23, n.3, Mar, p.916-28. 2009.

Consortium, C. E. S. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, v.282, n.5396, Dec 11, p.2012-8. 1998.

Cooper, T. A. e W. Mattox. The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet*, v.61, n.2, Aug, p.259-66. 1997.

Cooper, T. A., L. Wan, et al. RNA and disease. *Cell*, v.136, n.4, Feb 20, p.777-93. 2009.

Cork, D. M., T. W. Lennard, et al. Alternative splicing and the progesterone receptor in breast cancer. *Breast Cancer Res*, v.10, n.3, p.207. 2008.

De La Grange, P., M. Dutertre, et al. FAST DB: a website resource for the study of the expression regulation of human gene products. *Nucleic Acids Res*, v.33, n.13, p.4276-84. 2005.

Dreyfuss, G., M. J. Matunis, et al. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem*, v.62, p.289-321. 1993.

Gallo, J. M., W. Noble, et al. RNA and protein-dependent mechanisms in tauopathies: consequences for therapeutic strategies. *Cell Mol Life Sci*, v.64, n.13, Jul, p.1701-14. 2007.

Glisovic, T., J. L. Bachorik, et al. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*, v.582, n.14, Jun 18, p.1977-86. 2008.

Goldstrohm, A. C., A. L. Greenleaf, et al. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene*, v.277, n.1-2, Oct 17, p.31-47. 2001.

Gozani, O., R. Feld, et al. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev*, v.10, n.2, Jan 15, p.233-43. 1996.

Graveley, B. R. Sorting out the complexity of SR protein functions. *Rna*, v.6, n.9, Sep, p.1197-211. 2000.

Han, S. e R. E. Viola. Splicing of unnatural amino acids into proteins: a peptide model study. *Protein Pept Lett*, v.11, n.2, Apr, p.107-14. 2004.

Han, S., Lee, BC., Yu, ST., Jeong, CS., Lee, S., Kim, D. Fold recognition by combining profile-profile alignment and support vector machine. *Bioinformatics*. 1;21(11):2667-73. 2005

Hui, J., L. H. Hung, et al. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *Embo J*, v.24, n.11, Jun 1, p.1988-98. 2005.

Hung, L. H., M. Heiner, et al. Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *Rna*, v.14, n.2, Feb, p.284-96. 2008.

Irimia, M., D. Penny, et al. Coevolution of genomic intron number and splice sites. *Trends Genet*, v.23, n.7, Jul, p.321-5. 2007.

Irimia, M. e S. W. Roy. Evolutionary convergence on highly-conserved 3' intron structures in intron-poor eukaryotes and insights into the ancestral eukaryotic genome. *PLoS Genet*, v.4, n.8, p.e1000148. 2008.

Irimia, M., J. L. Rukov, et al. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evol Biol*, v.7, p.188. 2007.

Irimia M, Rukov JL, Roy SW, Vinther J, Garcia-Fernandez J. Quantitative regulation of alternative splicing in evolution and development. *Bioessays*, v.31, n.1, Jan, p.40-50. 2009.

Kampa, D., J. Cheng, et al. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res*, v.14, n.3, Mar, p.331-42. 2004.

Kim, E., A. Magen, et al. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res*, v.35, n.1, p.125-31. 2007.

Kishore, S. e S. Stamm. Regulation of alternative splicing by snoRNAs. *Cold Spring Harb Symp Quant Biol*, v.71, p.329-34. 2006.

Konarska, M. M., J. Vilardell, et al. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell*, v.21, n.4, Feb 17, p.543-53. 2006.

Kramer, A. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem*, v.65, p.367-409. 1996.

Kriventseva, E. V., I. Koch, et al. Increase of functional diversity by alternative splicing. *Trends Genet*, v.19, n.3, Mar, p.124-8. 2003.

Ladd, A. N. e T. A. Cooper. Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol*, v.3, n.11, Oct 23, p.reviews0008. 2002.

Lallena, M. J., K. J. Chalmers, et al. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell*, v.109, n.3, May 3, p.285-96. 2002.

Lander, E. S., L. M. Linton, et al. Initial sequencing and analysis of the human genome. *Nature*, v.409, n.6822, Feb 15, p.860-921. 2001.

Lewin B, *Genes VII*. Oxford University Press. 2000.

Lev-Maor, G., R. Sorek, et al. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science*, v.300, n.5623, May 23, p.1288-91. 2003.

Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E. *Molecular Cell Biology* New York: W. H. Freeman & Co.; c1999

Mardon, H. J., G. Sebastio, et al. A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic Acids Res*, v.15, n.19, Oct 12, p.7725-33. 1987.

Nelson DL. & Cox MM. *Lehninger, Principles of Biochemistry*. Third Edition. Worth Publishers. 2000.

Pajares, M. J., T. Ezponda, et al. Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol*, v.8, n.4, Apr, p.349-57. 2007.

Pan, Q., O. Shai, et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, v.40, n.12, Dec, p.1413-5. 2008.

Patel, S. B. e M. Bellini. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, v.36, n.20, Nov, p.6482-93. 2008.

Peverelli, E., G. Mantovani, et al. Expression of the two alternatively spliced PRKAR1A RNAs in human endocrine glands. *Mol Cell Endocrinol*, v.238, n.1-2, Jun 30, p.51-5. 2005.

Pinner, S. e E. Sahai. PDK1 regulates cancer cell motility by antagonising inhibition of ROCK1 by RhoE. *Nat Cell Biol*, v.10, n.2, Feb, p.127-37. 2008.

Reis, E. M., H. I. Nakaya, et al. Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene*, v.23, n.39, Aug 26, p.6684-92. 2004.

Reis, E. M., E. P. Ojopi, et al. Large-scale transcriptome analyses reveal new genetic marker candidates of head, neck, and thyroid cancer. *Cancer Res*, v.65, n.5, Mar 1, p.1693-9. 2005.

Romero, P. R., S. Zaidi, et al. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A*, v.103, n.22, May 30, p.8390-5. 2006.

Rowen, L., J. Young, et al. Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics*, v.79, n.4, Apr, p.587-97. 2002.

Roy, S. W. e D. Penny. Intron length distributions and gene prediction. *Nucleic Acids Res*, v.35, n.14, p.4737-42. 2007.

Sakabe, N. J. e S. J. De Souza. Sequence features responsible for intron retention in human. *BMC Genomics*, v.8, p.59. 2007.

Sambrook J. & Russell D.W. *Molecular Cloning, a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. Third Edition, 2001

Sharma, S., L. A. Kohlstaedt, et al. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol*, v.15, n.2, Feb, p.183-91. 2008.

Singh, R. e J. Valcarcel. Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol*, v.12, n.8, Aug, p.645-53. 2005.

Smith, C. W. e J. Valcarcel. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, v.25, n.8, Aug, p.381-8. 2000.

Sorek, R., G. Ast, et al. Alu-containing exons are alternatively spliced. *Genome Res*, v.12, n.7, Jul, p.1060-7. 2002.

Sorek, R., R. Shamir, et al. How prevalent is functional alternative splicing in the human genome? *Trends Genet*, v.20, n.2, Feb, p.68-71. 2004.

Srebrow, A., M. Blaustein, et al. Regulation of fibronectin alternative splicing by a basement membrane-like extracellular matrix. *FEBS Lett*, v.514, n.2-3, Mar 13, p.285-9. 2002.

Stamm, S., S. Ben-Ari, et al. Function of alternative splicing. *Gene*, v.344, Jan 3, p.1-20. 2005.

Tacke, R. e J. L. Manley. Functions of SR and Tra2 proteins in pre-mRNA splicing regulation. *Proc Soc Exp Biol Med*, v.220, n.2, Feb, p.59-63. 1999.

Tazi, J., N. Bakkour, et al. Alternative splicing and disease. *Biochim Biophys Acta*, v.1792, n.1, Jan, p.14-26. 2009.

The *C. elegans* Sequencing Consortium** (See www.genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/ for a list of authors). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. 11;282(5396):2012-8, 1998; 1;283(5398):35, 1999; 26;283(5410):2103, 1999 and 3;285(5433):1493, 1999.

Thorsen, K., K. D. Sorensen, et al. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol Cell Proteomics*, v.7, n.7, Jul, p.1214-24. 2008.

Tsukahara, T., S. Kimura, et al. Scythe/BAT3 regulates apoptotic cell death induced by papillomavirus binding factor in human osteosarcoma. *Cancer Sci*, v.100, n.1, Jan, p.47-53. 2009.

Valcarcel, J., R. K. Gaur, et al. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science*, v.273, n.5282, Sep 20, p.1706-9. 1996.

Van Alphen, R. J., E. A. Wiemer, et al. The spliceosome as target for anticancer treatment. *Br J Cancer*, v.100, n.2, Jan 27, p.228-32. 2009.

Vanhee-Brossollet, C. e C. Vaquero. Do natural antisense transcripts make sense in eukaryotes? *Gene*, v.211, n.1, Apr 28, p.1-9. 1998.

Varani, G. e K. Nagai. RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct*, v.27, p.407-45. 1998.

Venables, J. P. Unbalanced alternative splicing and its significance in cancer. *Bioessays*, v.28, n.4, Apr, p.378-86. 2006.

Venter, J. C., M. D. Adams, et al. The sequence of the human genome. *Science*, v.291, n.5507, Feb 16, p.1304-51. 2001.

Villate, O., A. Rastrojo, et al. Differential splicing, disease and drug targets. *Infect Disord Drug Targets*, v.8, n.4, Dec, p.241-51. 2008.

Vithana, E. N., L. Abu-Safieh, et al. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell*, v.8, n.2, Aug, p.375-81. 2001.

Wahl, M. C., C. L. Will, et al. The spliceosome: design principles of a dynamic RNP machine. *Cell*, v.136, n.4, Feb 20, p.701-18. 2009.

Wang, E. T., R. Sandberg, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*, v.456, n.7221, Nov 27, p.470-6. 2008.

Wang, Z. e C. B. Burge. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna*, v.14, n.5, May, p.802-13. 2008.

Wang, Z., M. E. Rolish, et al. Systematic identification and analysis of exonic splicing silencers. *Cell*, v.119, n.6, Dec 17, p.831-45. 2004.

Wilkie, S. E., V. Vaclavik, et al. Disease mechanism for retinitis pigmentosa (RP11) caused by missense mutations in the splicing factor gene PRPF31. *Mol Vis*, v.14, p.683-90. 2008.

Will, C. L., C. Schneider, et al. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *Embo J*, v.20, n.16, Aug 15, p.4536-46. 2001.

Wirth, B., L. Brichta, et al. Spinal muscular atrophy and therapeutic prospects. *Prog Mol Subcell Biol*, v.44, p.109-32. 2006.

Xie, Z., H. Yuan, et al. 3-phosphoinositide-dependent protein kinase-1 (PDK1) promotes invasion and activation of matrix metalloproteinases. *BMC Cancer*, v.6, p.77. 2006.

Zhang, Z., F. Lotti, et al. SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell*, v.133, n.4, May 16, p.585-600. 2008.