

**ANÁLISE CORRELACIONAL ENTRE A
EXPRESSÃO DOS FATORES DE *SPLICING* E A
OCORRÊNCIA DE *SPLICING* ALTERNATIVO
EM TECIDOS HUMANOS E DE CAMUNDONGOS**

JULIO CÉSAR NUNES

**Dissertação apresentada à Fundação Antônio
Prudente para a obtenção do título de Mestre em
Ciências**

Área de Concentração: Oncologia

Orientador: Dr. Sandro José de Souza

São Paulo

2008

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Nunes, Julio César

Análise correlacional entre a expressão dos fatores de *splicing* e a ocorrência de *splicing* alternativo em tecidos humanos e de camundongos /

Julio César Nunes – São Paulo, 2008.

79p.

Dissertação (Mestrado) - Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Sandro José Souza

Descritores: 1. SPLICING ALTERNATIVO 2. BIOLOGIA MOLECULAR COMPUTACIONAL 3. CÂNCER 4. GENOMICA.

AGRADECIMENTOS

Agradeço à FAPESP e CAPES pela bolsa de Mestrado.

Ao Sandro José de Souza agradeço toda orientação e conhecimento oferecido.

Meus especiais agradecimentos ao Pedro Alexandre Favoretto Galante que dedicou atenção a minha formação no processo de Pós-Graduação na Fundação Antônio Prudente, bem como pela sua oficiosa co-orientação ao projeto de pesquisa.

À grande família e amigos pela dedicação e incentivo a minha formação acadêmica.

À Fundação Antônio Prudente, Hospital do Câncer e Instituto Ludwig de Pesquisa sobre o Câncer dedico os meus nobres agradecimentos finais.

RESUMO

Nunes JC. **Análise correlacional entre a expressão dos fatores de *splicing* e a ocorrência de *splicing* alternativo em tecidos humanos e de camundongos.** São Paulo; 2007. [Dissertação de Mestrado - Fundação Antônio Prudente]

Splicing alternativo desempenha uma significativa função no aumento da complexidade genômica, produzindo um extenso número de mRNA e isoformas protéicas. *Splicing* alternativos em genes humanos são estimados a ocorrerem na frequência de 40% a 60%, tornando assim este evento mais propriamente uma regra do que exceção. Recentes abordagens experimentais e *in silico* indicam que amostras derivadas de tumores frequentemente apresentam isoformas diferentes de *splicing*, o que sugere que padrões alternativos de *splicing* estão amplamente presentes em neoplasias. Interações entre fatores de *splicing* e elementos auxiliares presentes nas moléculas de RNA (elementos em *cis*) constituem um modo de controle do *splicing* alternativo. Este estudo incorpora à investigação *in silico* o objetivo geral de efetuar-se uma análise correlacional entre os perfis de expressão gênica dos fatores de *splicing* e a presença de eventos de *splicing* alternativos em ambos humanos e camundongos. Nossos objetivos específicos foram compostos em quatro partes: primeiro, reconhecer um conjunto favorável de fatores de *splicing* humanos; segundo, selecionar os fatores de *splicing* ortólogos em camundongos; terceiro, identificar os eventos de *splicing* alternativo em ambas as espécies; quarto, analisar as especificidades teciduais normais e neoplásicas em humanos. Os resultados proporcionaram respostas conclusivas a uma análise compreensiva ao escopo dos fatores de *splicing* em sua totalidade. Apresenta-se como resultados um conjunto final de 124 fatores de *splicing* ortólogos em humanos e camundongos; coexpressão diferencial dos fatores de *splicing* snRNP, SR, hnRNP e Sm, bem como incidente ocorrência de eventos de *splicing* alternativos, preferencialmente em sistema nervoso e tecidos sexo específicos; conjunto de fatores de *splicing* com expressão gênica superior em bibliotecas tumorais de Massively Parallel Signature Sequencing – MPSS, e promissores candidatos a investigações experimentais específicas, que corroborem em outros métodos os seus envolvimento na tumorigênese.

SUMMARY

Nunes JC. [**Correlation analysis of splicing factor expression and the occurrence of alternative splicing in human and mice tissues**]. São Paulo; 2007. [Dissertação de Mestrado - Fundação Antônio Prudente]

Alternative splicing plays a significant role in increasing the level of genomic complexity, thereby resulting in a large number of mRNA and protein isoforms. Alternative splicing in human genes are estimated to occur at a frequency of 40% to 60% thus making this event a rule rather than an exception. Recent experimental and in silico approaches have shown that samples from tumor often present different splicing isoforms, which suggests that alternative patterns of splicing are widely present in neoplasies. Interactions between splicing factors and auxiliary elements in RNA molecules (elements in *cis*) constitute a way of controlling alternative splicing. This study incorporates the research in silico to its general objective of performing a correlation analysis between profiles of splicing factor gene expression and the presence of alternative splicing events in both human and mice. Our specific objectives were fourfold: first, to recognize a favorable set of human splicing factors; second, select splicing factors of orthologs in mice; third, identify alternative splicing events in both species; and fourth, to analyze the specificities of normal and neoplastic human tissues. The results provided conclusive responses to an encompassing analysis of the range of splicing factors as a whole. Presented as results are a final set of 124 ortholog factors in human and mice; a differential co-expression of snRNP, SR, hnRNP and Sm splicing factors, as well incident occurrence of alternative splicing events, preferably in the nervous system and gender-specific tissues; a set of splicing factors with higher gene expression in Massively Parallel Signature Sequencing - MPSS tumor libraries; as well as promising candidates to specific experimental investigation, which may corroborate its involvement in tumor genesis through other methods.

LISTA DE FIGURAS

Figura 1	Os sinais de <i>splicing</i>	3
Figura 2	Elementos de seqüências indicando íntrons	3
Figura 3	Modos alternativos de <i>splicing</i>	4
Figura 4	Elementos de seqüências e fatores de <i>splicing</i>	7
Figura 5	Funções das proteínas SR na montagem do spliceossomo	10
Figura 6	Modo anormal de <i>splicing</i> de mRNA originando isoforma protéica com propriedades oncogênicas	13
Figura 7	Formação e rearranjo do spliceossomo durante a reação de <i>splicing</i>	16
Figura 8	Esquema geral da abordagem aplicada à pesquisa	19
Figura 9	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> snRNP, em bibliotecas de MPSS de humanos	38
Figura 10	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> SR, em bibliotecas de MPSS de humanos	38
Figura 11	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> hnRNP, em bibliotecas de MPSS de humanos	38
Figura 12	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> Sm, em bibliotecas de MPSS de humanos	39
Figura 13	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> snRNP, em bibliotecas de MPSS de camundongos (macho)	40
Figura 14	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> SR, em bibliotecas de MPSS de camundongos (macho)	40
Figura 15	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> hnRNP, em bibliotecas de MPSS de camundongos (macho)	41
Figura 16	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> Sm, em bibliotecas de MPSS de camundongos (macho)	42

Figura 17	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> snRNP, em bibliotecas de MPSS de camundongos (fêmea)	43
Figura 18	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> SR, em bibliotecas de MPSS de camundongos (fêmea)	43
Figura 19	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> hnRNP, em bibliotecas de MPSS de camundongos (fêmea)	44
Figura 20	Gráfico da distribuição dos valores de expressão gênica dos fatores de <i>splicing</i> Sm, em bibliotecas de MPSS de camundongos (fêmea)	44

LISTA DE TABELAS

Tabela 1	Os 15 maiores números de eventos de <i>splicing</i> alternativos em humanos	46
Tabela 2	Os 15 maiores números de eventos de <i>splicing</i> alternativos em camundongos	46
Tabela 3	Números de eventos de <i>splicing</i> alternativos encontrado em humanos, com presença indicativa dos 41 fatores de <i>splicing</i> constitutivos	48
Tabela 4	Números de eventos de <i>splicing</i> alternativos encontrado em camundongos, contendo a presença indicativa dos 31 fatores de <i>splicing</i> ortólogos constitutivos	49
Tabela 5	Os genes ortólogos resultantes, comparados e pareados quanto aos seus índices de presença em distintos eventos de <i>splicing</i> alternativos	50
Tabela 6	Listagem dos genes presentes preferencialmente em transcritos variantes tumorais (por um fator de dois ou mais).	51

LISTA DE QUADROS

Quadro 1	UniGene <i>clusters</i> dos fatores de <i>splicing</i> constitutivos de humanos	31
Quadro 2	UniGene <i>clusters</i> dos fatores de <i>splicing</i> constitutivos de camundongos	32
Quadro 3	Bibliotecas e suas respectivas origens sexo específicas em humanos	32
Quadro 4	Bibliotecas e suas distintas origens sexo específicas em camundongos	33
Quadro 5	Valores diferenciais de coexpressão gênica dos fatores de <i>splicing</i> constitutivos entre todas as bibliotecas indistintamente	35
Quadro 6	Valores diferenciais de coexpressão gênica dos fatores de <i>splicing</i> constitutivos entre as bibliotecas de determinado grupo	36
Quadro 7	Grupos teciduais	36
Quadro 8	Conjunto de bibliotecas humanas tumorais de MPSS com seus respectivos correspondentes tipos normais	52
Quadro 8.1	Conjunto de bibliotecas humanas modificadas de MPSS com seus respectivos correspondentes tipos normais	52
Quadro 9	Números de fatores de <i>splicing</i> humanos com presença indicativa em padrões de eventos de <i>splicing</i> alternativos	54
Quadro 10	Fatores de <i>splicing</i> humanos encontrados apenas em transcritos variantes tumorais	55
Quadro 11	Fatores de <i>splicing</i> citados em referências científicas correlatas pelo seu envolvimento na tumorigênese	56
Quadro 12	Fatores de <i>splicing</i> humanos encontrados em transcritos variantes normais e tumorais, e com expressão gênica superior em bibliotecas tumorais de MPSS	58
Quadro 13	Fatores de <i>splicing</i> humanos ausentes nos transcritos variantes tumorais e com expressão gênica superior em bibliotecas tumorais de MPSS	59

LISTA DE ABREVIATURAS

AS	<i>Alternative splicing</i>
ASS	Sítios doadores e/ou aceptores alternativos de <i>splicing</i>
ATP	<i>Adenosine 5'-triphosphate</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
cDNA	<i>Complementary DNA</i>
COMP BIO	Laboratório de Biologia Computacional
domínio RS	Domínio rico em arginina e serina
DNA	<i>Deoxyribonucleic acid</i>
ES	Uso alternativo do éxon
ESE	<i>Exonic splicing enhancer</i>
ESS	<i>Exonic splicing silencer</i>
EST	<i>Expressed sequence tag</i>
hnRNP	<i>Heterogenous nuclear ribonucleoprotein</i>
HTC	<i>High-throughput cDNAs</i>
ILPC	Instituto Ludwig de Pesquisa sobre o Câncer
ISE	<i>Intronic splicing enhancer</i>
ISS	<i>Intronic splicing silencer</i>
IR	Retenção de íntron
LIBID	<i>Library identification</i>
MGC	<i>Mammalian Gene Collection</i>
MPSS	<i>Massively Parallel Signature Sequencing</i>
mRNA	<i>Messenger RNA</i>
NCBI	<i>National Center for Biotechnology Information</i>
NMD	<i>Nonsense-mediated decay</i>
nt	Nucleotídeo
r	Coeficiente de correlação de Pearson
PTB	<i>Polypyrimidine tract binding protein</i>
RRM	<i>RNA recognition motif</i>
RT-PCR	<i>Reverse transcription followed by polymerase chain reaction</i>
RNA	<i>Ribonucleic acid</i>
SAF	<i>Scaffold attachment factor (A or B)</i>
SAGE	<i>Serial Analysis of Gene Expression</i>

SAP155	<i>Spliceosome-associated protein 155</i>
SF3b155	<i>Splicing factor 3B subunit 1/Spliceosome-associated protein 155</i>
SFPQ	<i>Splicing factor proline/glutamine-rich</i>
SFRS14	<i>Splicing factor, arginine/serine-rich 14</i>
SC35	<i>Splicing component, 35 kDa; splicing factor, arginine/serine-rich</i>
snRNP	<i>Small nuclear ribonucleoprotein particle</i>
SR	<i>Serine-arginine-rich protein</i>
SRm 160/300	<i>SR-related nuclear matrix proteins of 160 and 300 kDa</i>
Tra2	<i>Transformer 2</i>
UCSC	<i>University of California, Santa Cruz</i>
U1 70K	<i>U1 snRNP 70 kDa protein</i>
U2AF	<i>U2 snRNP auxiliary factor (35 or 65 kDa)</i>
UTR	<i>Untranslated region</i>

ÍNDICE

1	INTRODUÇÃO	1
1.1	O <i>splicing</i> constitutivo e a inerente dinâmica de base	1
1.1.1	Modos alternativos de <i>splicing</i>	4
1.1.2	O acurado e complexo spliceossomo	5
1.1.3	O mecanismo de ação dos fatores de <i>splicing</i>	6
1.1.4	Grupo de fatores de <i>splicing</i> SR	7
1.1.5	A função do grupo SR no <i>splicing</i> constitutivo e alternativo	8
1.1.6	A família de proteínas hnRNP	11
1.2	<i>Splicing</i> alternativo e doenças humanas	12
1.2.1	<i>Splicing</i> alternativo e câncer	12
1.3	O mecanismo de <i>splicing</i>	14
1.4	<i>Massively Parallel Signature Sequencing</i> (MPSS)	17
2	ABORDAGEM GERAL APLICADA À PESQUISA	18
3	MATERIAIS E MÉTODOS	20
3.1	Seleção curada dos fatores de <i>splicing</i>	20
3.2	Obtenção das seqüências de cDNA	20
3.3	Extração das <i>tags</i> virtuais	20
3.4	Seleção dos genes dos fatores de <i>splicing</i> ortólogos de camundongos	21
3.5	Submissão das <i>tags</i> virtuais em bancos de dados de MPSS	22
3.6	Análise na correlação de expressão gênica dos fatores de <i>splicing</i> constitutivos entre as bibliotecas de MPSS	22
3.7	Protocolo de busca aos eventos de <i>splicing</i> alternativos	24
3.8	Investigação dos níveis de <i>splicing</i> alternativos	25
3.9	Análise comparativa entre os valores de expressão gênica dos fatores de <i>splicing</i>	26
3.10	Correlação na expressão gênica dos fatores de <i>splicing</i> presentes nos eventos de <i>splicing</i> alternativos	27

3.11	Simulação <i>in silico</i> aos eventos de <i>splicing</i> alternativos de humanos	28
4	RESULTADOS	29
4.1	Padrões de expressão gênica dos fatores de <i>splicing</i> constitutivos	29
4.2	Eventos de <i>splicing</i> alternativos em humanos e camundongos	45
4.3	Fatores de <i>splicing</i> constitutivos presentes nos eventos de <i>splicing</i> alternativos	47
4.4	Expressão gênica diferencial de todos os fatores de <i>splicing</i> humanos presentes em transcritos variantes normais e tumorais	52
5	DISCUSSÃO	60
5.1	Padrão diferencial de coexpressão gênica dos fatores de <i>splicing</i> nos tecidos do sistema nervoso e sexo específico	60
5.2	A regulação dos eventos de <i>splicing</i> alternativos sob os fatores de <i>splicing</i>	62
5.3	Os fatores de <i>splicing</i> presentes preferencialmente nos transcritos variantes tumorais apresentam valores de expressão gênica comparativamente superiores aos normais	63
6	CONCLUSÃO	66
6.1	Conjunto de fatores de <i>splicing</i> ortólogos em humanos e camundongos	66
6.2	Padrões de coexpressão diferencial dos fatores de <i>splicing</i> constitutivos	67
6.3	Os transcritos variantes e a expressão gênica dos fatores de <i>splicing</i> em bibliotecas de MPSS normais <i>versus</i> tumorais de humanos	67
7	REFERÊNCIAS BIBLIOGRÁFICAS	69

ANEXOS

- Anexo 1** Bibliotecas de MPSS de humanos
- Anexo 2** Bibliotecas de MPSS de camundongos
- Anexo 3** Anotações gênicas dos fatores de *splicing*
- Anexo 4** Conjunto de 211 *clusters* gênicos ortólogos curados: fatores de *splicing* e proteínas envolvidas no processo de *splicing* alternativo
- Anexo 5** Conjunto final de 124 fatores de *splicing* ortólogos
- Anexo 6** 88 fatores de *splicing* com expressões gênicas de MPSS com valores superiores em bibliotecas tumorais *versus* normais (mais do que 3 vezes)
- Anexo 7** 72 fatores de *splicing* com expressões gênicas de MPSS com valores superiores em bibliotecas tumorais *versus* normais (mais do que 5 vezes)
- Anexo 8** Fatores de *splicing* humanos com presença indicativa em padrões de eventos de *splicing* alternativos

1 INTRODUÇÃO

Apresentamos o presente estudo que alcançou o objetivo geral de analisar a correlação entre a expressão dos fatores de *splicing* e a ocorrência dos eventos de *splicing* alternativos em humanos e camundongos, que resultaram em respostas aos perfis de expressão gênica dos fatores de *splicing* investigados. Para tanto foram alcançados os seguintes objetivos específicos:

- Identificação dos genes candidatos de humanos e camundongos que representem os referidos fatores de *splicing*;
- Análise da expressão dos fatores de *splicing* através dos dados de MPSS;
- Identificação dos eventos de *splicing* alternativos em humanos e camundongos;
- Estudo comparativo entre a expressão dos fatores de *splicing* e a ocorrência de *splicing* alternativos, em tecidos humanos tumorais e normais;
- Apresentação de conclusões que favoreceram o conhecimento de uma proeminente e sugestiva análise funcional aos fatores de *splicing* e modos alternativos de *splicing*.

1.1 O *SPLICING* CONSTITUTIVO E A INERENTE DINÂMICA DE BASE

Com a descoberta da estrutura do DNA por WATSON e CRICK (1953), os genes foram postulados como seqüências contíguas de bases nitrogenadas, pelos

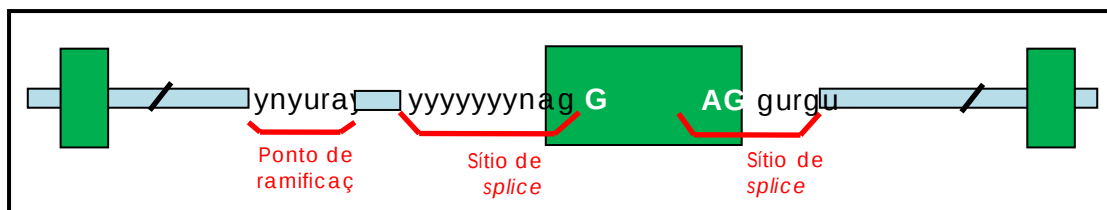
quais as informações eram transferidas para a síntese protéica (SHARP 2005). A caracterização da biologia molecular manteve-se assim atrelada ao preceito básico de um código genético descrevendo a relação contínua e única entre a sequência de DNA e a sequência de proteína. O processo representado permanece fidedigno em organismos procariotos, mas com contraposição em eucariotos, os quais apresentam discrepâncias comparativas entre as dimensões das sequências de DNA e mRNA.

Os autores CHOW et al. (1977) e BERGET et al. (1977) evidenciaram sequências adicionais inclusas no pré-mRNA nuclear, quando comparado com o mRNA citoplasmático, sendo estas removidas no *splicing*. Tais sequências intervenientes, denominadas íntrons, são excluídas quando o transcrito primário é processado no núcleo, originando o mRNA maduro que apresenta apenas as sequências exônicas.

O processo de *splicing* do pré-mRNA ocorre após a clivagem no sítio de poli(A) e poliadenilação da extremidade 3' do transcrito primário, quando em unidades transcricionais curtas, e inicia-se previamente sobre o pré-mRNA nascente antes que a transcrição esteja completa, em se tratando de unidades transcricionais longas (SHATKIN e MANLEY 2000).

Um típico gene humano contém em média 8.8 éxons. Os éxons possuem em média 145 nucleotídeos (nt), sendo que os íntrons possuem uma dimensão de grandeza superior, com um número 10 vezes maior, ou mais (LANDER et al. 2001). Os limites dos éxons são definidos pelas curtas e degeneradas sequências de sítios de *splice* existentes nas bordas íntron/éxon (sítio de *splice* 5', sítio de *splice* 3', e o ponto de ramificação) (Figura 1).

A maior classe de íntrons humanos (> 99%) contém o clássico sinal de *splicing* GT-AG mostrados na Figura 2.



Legenda: (n = G, A, U ou C; y = pirimidina; r = purina). Os íntrons são indicados pelas linhas azuis e estreitas, e os éxons pelas caixas verdes. Mostram-se apenas os sítios ao redor do éxon central.

Fonte: Adaptado de FAUSTINO e COOPER (2003)

Figura 1 - Os sinais de *splicing*.

Elementos	Sequências consensuais
Sítio de <i>splice</i> 5' (doador)	YRG / <u>G</u> URAGU
Sítio de <i>splice</i> 3' (aceitador) precedido por um trato de polipirimidina	Y ₁₂ NY <u>A</u> G
Ponto de ramificação localizado 18-200 nucleotídeos a montante do sítio de <i>splice</i> 3'	YNYUR <u>A</u> Y

Legenda: (Y = pirimidina; R = purina; N = qualquer nucleotídeo). A barra indica a borda éxon-íntron. Os nucleotídeos invariantes estão sublinhados.

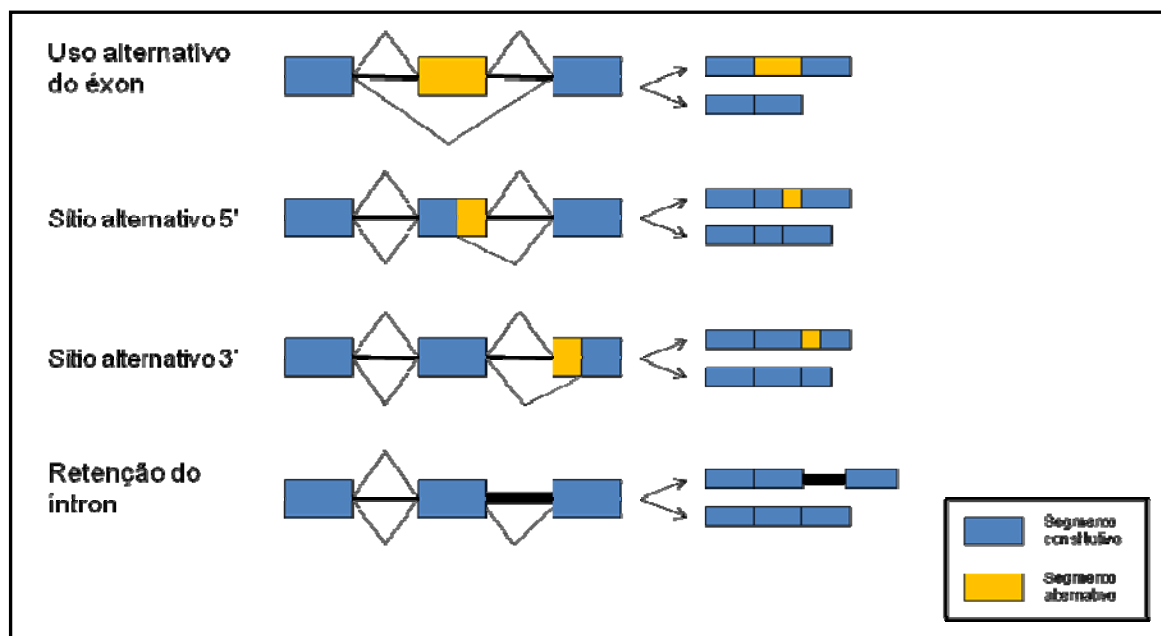
Figura 2 - Elementos de sequências indicando íntrons.

Íntrons com as bordas GT-AG são considerados íntrons do tipo U2. Uma nova classe de íntrons fora encontrada tendo em vista os seus sítios de *splice* incomuns (HALL e PAGETT 1994). Estes íntrons contêm AT e AC nos sítios de *splice* 5' e 3' respectivamente. Este tipo de íntron fora denominado de íntron U12. O tipo de íntron U12 está presente no núcleo dos vertebrados, insetos, e plantas (WU et al. 1996). As análises efetuadas nos pares de junção do *splice* dos genes anotados de mamíferos mostraram que 98.71% são conformados pelo canônico GT-AC, 0.56%

ao não-canônico GC-AG e 0.73% a outras bordas de *splice* não-canônicas (BURSET et al. 2001).

1.1.1 Modos alternativos de *splicing*

Uma via divergente, resultante de uma alteração dos limites das regiões estabelecidas a sofrerem *splicing*, constitui um modo alternativo de *splicing*, com possíveis excisões ou adições de seqüências nucleotídicas sendo incorporadas prontamente ao mRNA maduro (BLACK 1995). Os eventos de *splicing* alternativos podem ser classificados em três padrões básicos de *splicing*: uso alternativo de éxon; sítios alternativos 5' e 3'; retenção de íntron (Figura 3).



Legenda: As caixas azuis indicam os segmentos constitutivos e as caixas amarelas indicam os segmentos alternativos incorporados.

Figura 3 - Modos alternativos de *splicing*.

O uso alternativo do éxon compreende o modo mais incidente, embora os complexos padrões de *splicing* evidenciados nos sítios alternativos 5'e 3' sejam também freqüentemente observados. Estima-se que 75% de todos os padrões de *splicing* alternativos alterem a seqüência codificante (OKAZAKI et al. 2002).

1.1.2 O acurado e complexo spliceossomo

O spliceossomo é um complexo maquinário macromolecular formado por uma associação de proteínas e cinco pequenos RNA nucleares, as snRNA, formando as subcomplexas e pequenas partículas ribonucleoprotéicas nucleares, as snRNP U1, U2, U4, U5 e U6 (LERNER et al. 1981). As possíveis interações intrônicas e exônicas com as snRNP, e destas entre si, são altamente dinâmicas, modificando-se progressivamente através do processo de *splicing* (BROW 2002). Investigações em espectrometrias de massas, sobre a complexidade protéica envolta ao spliceossomo e *splicing* alternativo, identificam neste complexo cerca de 300 proteínas intimamente relacionadas (JURICA e MOORE 2003). O reconhecimento de ambos os sítios de *splice* ocorre durante o processo de montagem do spliceossomo devido a específicas interações com diferentes componentes de montagem. Incluem-se a estas, interações com os elementos regulatórios *cis*-atuantes (elementos em *cis*) localizados no pré-mRNA, necessárias para o reconhecimento do éxon, e para o reconhecimento de um conjunto de proteínas denominadas fatores *trans*-atuantes controladoras do *splicing* alternativo.

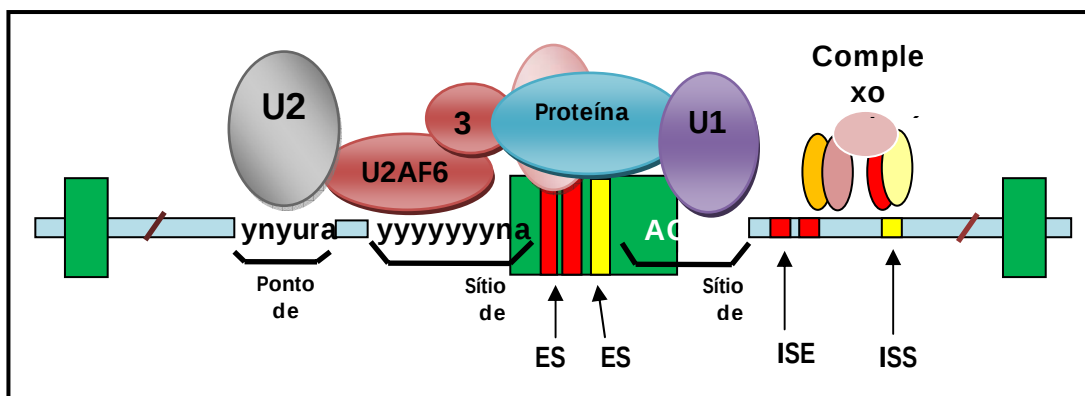
Os fatores *trans*-atuantes contêm domínios ligantes de mRNA e várias proteínas, sendo que tais interações domínio protéicas permitem as interações de membros individuais destas famílias de proteínas. Como resultado, temos a formação

de uma complexa rede protéica sobre o pré-mRNA, acerca dos éxons e íntrons, auxiliando nos seus reconhecimentos. As interações individuais entre os elementos em *cis* e em *trans* envolvidos na seleção dos sítios de *splice* são consideradas fracas. Somente através da ação de diversas outras ligações, interpondo-se a um íntron e éxon, é que o reconhecimento será efetivado. A relativa concentração dos fatores *trans*-atuantes é variada entre os tipos celulares e teciduais bem como durante seu desenvolvimento. Padrões na seleção dos sítios de *splice* mudam dependendo dos níveis de concentração local dos fatores de *splicing* e/ou de específicos genes reguladores.

Devido a este conjunto de controle, um amplo número de éxons alternativos pode ser regulado por um delimitado número de proteínas reguladoras. Isto explica a importância da modulação da seleção dos sítios de *splice*, dependendo do estágio de desenvolvimento, diferenciação tecidual, ou mudanças metabólicas das células (BLACK 1995).

1.1.3 O mecanismo de ação dos fatores de *splicing*

A modulação das reações de *splicing* é concebida pelos fatores de *splicing trans*-atuantes quando estes reconhecem um arranjo positivo (*splicing enhancer*) e/ou negativo (*splicing silencer*) dos elementos de seqüências *cis*-atuantes. Estes elementos podem ser exônicos: (ESE) *exonic splicing enhancer*, (ESS) *exonic splicing silencer*; ou intrônicos: (ISE) *intronic splicing enhancer*, (ISS) *intronic splicing silencer* (Figura 4).



Legenda: Os elementos de seqüências *exonic splicing enhancer* e *intronic splicing enhancer* são ilustrados em vermelho, e os elementos de seqüências *exonic splicing silencer* e *intronic splicing silencer*, em amarelo. Os fatores de *splicing* são figurados em formas de elipse. Os elementos intrônicos também servem para modular o uso alternativo de éxon pela ligação de um complexo regulatório.

Fonte: Adaptado de FAUSTINO e COOPER (2003).

Figura 4 - Elementos de seqüências e fatores de *splicing*.

Estes elementos regulatórios são comumente requeridos para os eficientes *splicing* constitutivos e alternativos (LADD e COOPER 2002). Os *enhancers cis*-atuantes podem auxiliar no recrutamento dos fatores de *splicing*, nos casos de quando as distâncias entre os sítios de *splice* são desfavoráveis, ou quando os sítios de *splice* são considerados fracos (BLACK 1995).

As proteínas ligantes específicas às seqüências *enhancers* ou *silencers*, e moduladoras da seleção do sítio de *splice* alternativo, podem ser subdivididas em dois grupos: membros da família de proteínas SR (TACKE e MANLEY 1996) e hnRNP (WEIGHARDT 1996).

1.1.4 Grupo de fatores de *splicing* SR

As proteínas do grupo SR são fatores de *splicing* essenciais (TACKE e MANLEY 1996). Eles pertencem a uma família de proteínas altamente conservada em metazoários, estas proteínas são requeridas no *splicing* constitutivo e também

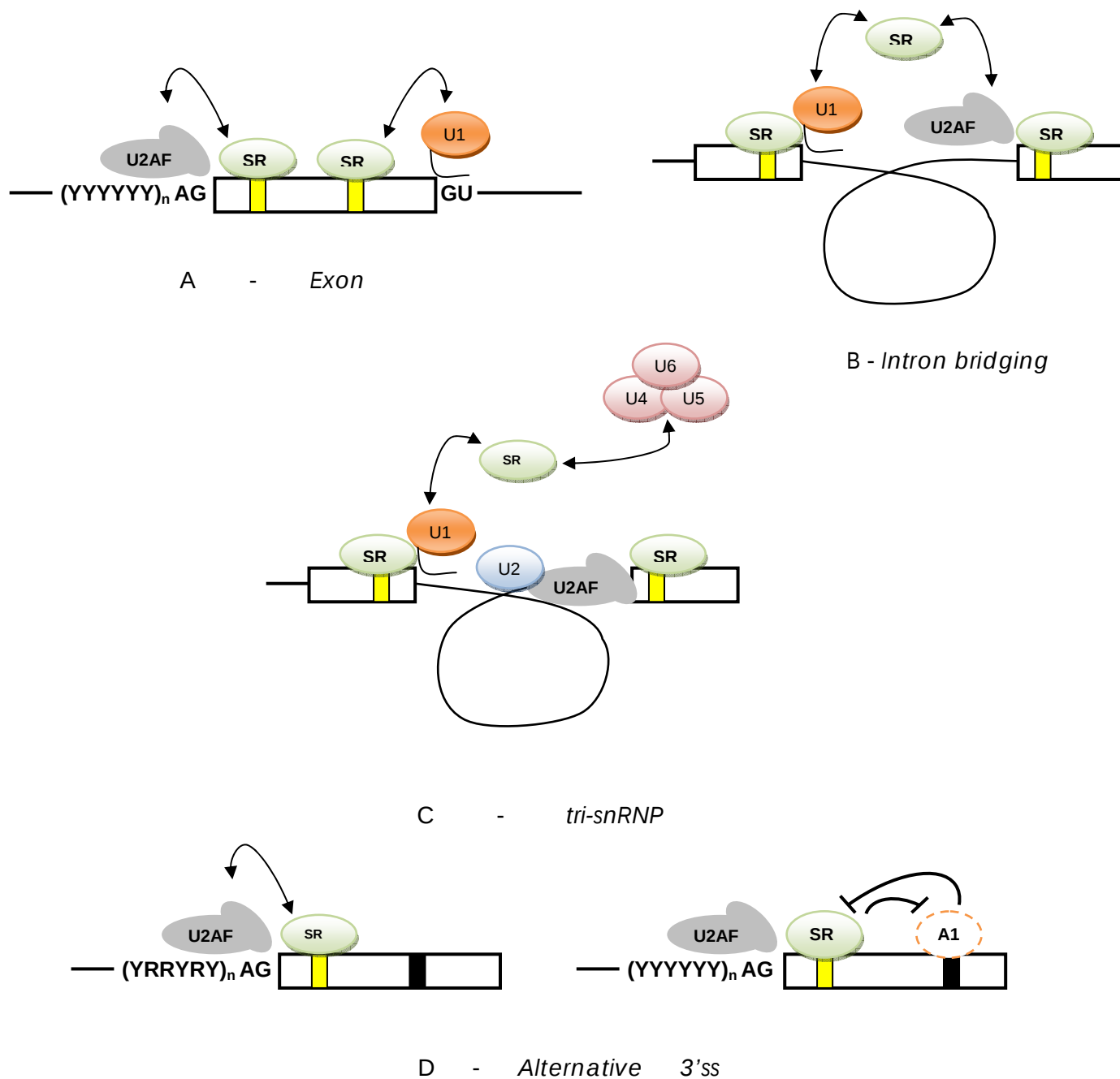
para a regulação da seleção alternativa do sítio de *splice* (GRAVELEY 2000). As proteínas possuem uma estrutura modular, consistindo de uma ou duas cópias de domínios RNA-ligantes N-terminal, com funções RNA ligantes e um domínio RS. O domínio RS destas proteínas é rico em resíduos alternantes de serina e arginina. Isto auxilia na mediação da proteína: interações protéicas do spliceossomo. O domínio RS contém múltiplos sítios de fosforilação de serina. A fosforilação de serina é importante na regulação das atividades e localização de proteínas SR (SANFORD et al. 2003). As proteínas *SR-related* (SRrp) pertencem a outra classe de proteínas contendo domínios RS. As maiorias destas proteínas contêm RRM. Dentre estas se incluem as proteínas U1-70K, ambas as subunidades de U2AF, SRm 160/300, bem como os reguladores de *splicing* alternativos dos genes Tra e Tra2, nos quais estão envolvidos na seleção do sítio de *splice*.

1.1.5 A função do grupo SR no *splicing* constitutivo e alternativo

As proteínas SR e *SR-related* auxiliam na seleção do sítio de *splice* e montagem do spliceossomo pela suas interações com outros fatores de *splicing*, via seu domínio RS. Estas proteínas são componentes recrutadores do aparato central do *splicing* promotores do pareamento no sítio de *splice* (TACKE e MANLEY 1999). Funcionam como conectores entre o pré-mRNA e o maquinário de base do *splicing*. Em adição ao processamento do pré-mRNA, estas específicas sequências de proteínas RNA ligantes possuem um significativo papel no transporte, estabilidade e tradução do mRNA. A fosforilação do domínio RS regula as atividades e localização das proteínas SR (SANFORD et al. 2003). Uma das funções das famílias de proteínas SR e *SR-related* é a de ativar sítios de *splice* subotimos e adjacentes

(BLENCOWE et al. 2006) no *splicing* alternativo. Propõe-se que a função das proteínas SR é a de estimular o reconhecimento de fracos sítios de *splice* 3' a montante, pelo recrutamento de U2AF, ou para facilitar o ligamento do snRNP U1 ao sítio de *splice* 5' (BLACK 2003). Certas proteínas SR possuem efeitos antagonistas no *splicing* alternativo, no qual fora evidenciado na regulação da β -tropomiosina pela ação oposta de SF2/ASF E SC35 (GALLEGO et al. 1997). As proteínas SR e SR-*related* interagem com múltiplos elementos *cis*-atuantes localizados dentro de seqüências intrônicas e exônicas. A seleção dos sítios de *splice*, portanto depende da interação de proteínas SR com estes elementos, e subsequente participação em múltiplas fases de montagem do spliceossomo. Os efeitos dos elementos *cis*-atuantes são dependentes da posição. Sítios das proteínas SR-ligantes contendo ESE possuem um efeito positivo na seleção do sítio de *splice*. As interações entre as proteínas SR e ESE favorecem o recrutamento e estabilização do snRNP U1 e U2AF aos sítios de *splice* 5' e 3'.

Este processo é definido como *exon definition* (BOUKIS et al. 2004) (Figura 5).



Legenda: (A) U2AF é ilustrado em cinza e está a montante do sítio de *splice* 3', e o U1 snRNP, o qual é ilustrado em laranja e situado a jusante do sítio de *splice* 5'. A ligação ao RNA é facilitada pelas proteínas SR acerca do ESE (faixas em amarelo). O trato de polipirimidina (YYYYYY)_n é uma parte do sítio de *splice* 3'. (B) Os sítios de *splice* 5' e 3' podem estar justapostos precocemente na reação de *splicing* pelas interações *intron bridging* entre as proteínas SR e o domínio RS contendo subunidades de snRNP U1 e U2AF. (C) As proteínas SR recrutam o U4/U6.U5 tri-snRNP para o spliceossomo. (D) As proteínas SR acerca do ESE promovem a seleção alternativa do sítio de *splice* 3' pelo recrutamento do U2AF ao sítio de *splice* 3'. Alternativamente, *exonic splicing silencers*, ilustrados como faixas pretas, podem recrutar proteínas repressoras de *splicing* tais como hnRNP A1 e bloquear a seleção do sítio de *splice* 3' pelo U2AF.

Fonte: Adaptado de SANFORD et al. (2005).

Figura 5 - Funções das proteínas SR na montagem do spliceossomo.

O reconhecimento do éxon e seleção do sítio de *splice* são realizados por uma coordenada ação de ambos os controles regulatórios positivos e negativos, providenciados pelas proteínas SR e SR-like e proteínas hnRNP, respectivamente. Os fatores que freqüentemente se opõem a ação da família de proteínas SR são as ribonucleoproteínas heterogêneas (hnRNP).

1.1.6 A família de proteínas hnRNP

As hnRNP foram primeiramente descritas como um grupo de proteínas nucleares RNA-ligantes. As hnRNP pertencem a uma abundante família de proteínas que se associam com pré-mRNA heterogêneos durante a transcrição e mantêm-se associados com mRNA após o *splicing* (NAKIELNY et al. 1997).

Estas proteínas são evidenciadas pelos seus envolvimento na biogênese e transporte nucleocitoplasmático do mRNA (DREYFUSS et al. 1993). A proteína hnRNP A1 fora amplamente estudada e preconizada pela sua ação antagonista às proteínas SR que promovem o uso do sítio de *splice* 5' em pré-mRNA de β -globina (MAYEDA e KRAINER 1992).

Outra proteína providenciando um exemplo de controle regulatório negativo de escolha ao sítio de *splice* é o *polypyrimidine tract binding protein* (PTB). PTB fora descoberto como uma proteína que se liga ao trato de polipirimidina U-rich de vários íntrons (GARCIA-BLANCO 1989). PTB media o silenciamento de éxons pela ligação a um amplo número de *splicing silencers* intrônicos dos pré-mRNA (GARCIA-BLANCO 2004). Isto sugere que tais proteínas agem como repressores globais de éxons regulados.

1.2 ***SPLICING* ALTERNATIVO E DOENÇAS HUMANAS**

Os eventos de *splicing* alternativos possuem uma importante função na expressão gênica e diversidade proteômica em humanos, possuindo alta relevância de investigação em possíveis relações com doenças e terapias. Formas aberrantes de *splicing* são reconhecidas como origem de um diverso número de doenças (FAUSTINO e COOPER 2003).

A sujeição incidente à imprecisão genética originada dos modos distintos de *splicing* torna-se evidente quando cerca de 15% de todas as doenças genéticas humanas surgem de mutações em seqüências consensuais de sítios de *splicing* (KRAWCZAK et al. 1992).

Nonsense-mediated mRNA decay (NMD) é considerado um mecanismo de controle no qual remove formas inapropriadas de mRNA oriundas de modos aberrantes de *splicing*. Em certas condições patológicas formas aberrantes de mRNA podem ser geradas e manterem-se despercebidas pelo NMD, e assim sendo serem traduzidas em proteínas. Tais proteínas podem resultar na causa primária de diversas doenças, incluindo-se fibrose cística, retinite pigmentosa, atrofia muscular espinhal, neurofibromatose tipo 1, hemofilia A, β -talassemia, e várias tipos de neoplasias.

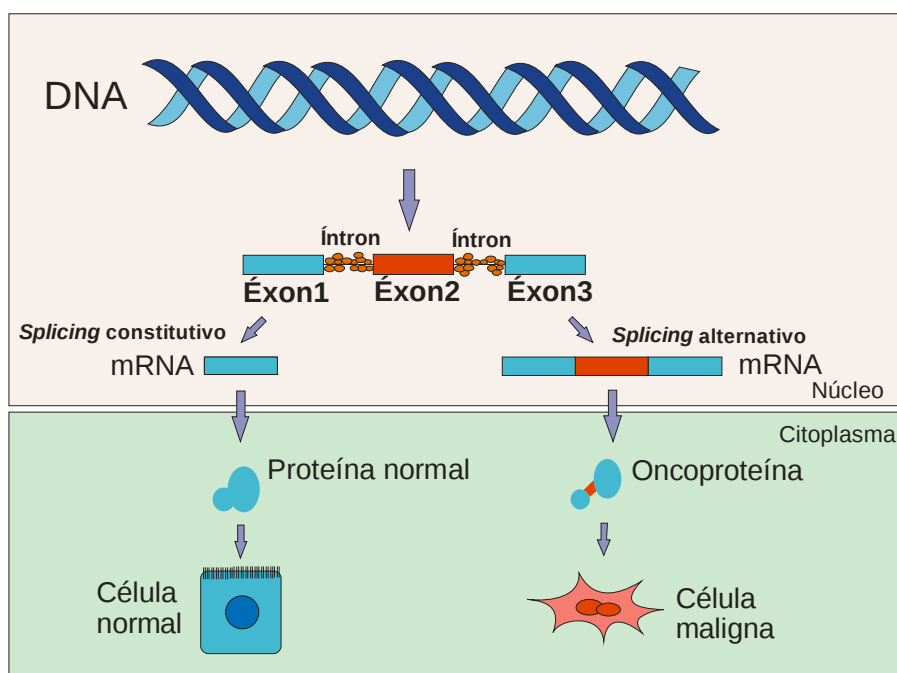
1.2.1 ***Splicing* alternativo e câncer**

Diversos experimentos e estudos *in silico* têm sido realizados com o principal objetivo de identificar variantes de *splicing* câncer-associado (Figura 6), resultando em um amplo número de genes candidatos (ROY et al. 2005). Abordagens experimentais utilizando-se de técnicas de RT-PCR evidenciam modos aberrantes de

splicing relacionados com a ativação dos oncogenes e inibição dos supressores de tumores em uma variedade de tipos de câncer em humanos (VENABLES 2006).

Alterações na concentração, localização, composição, ou atividade de fatores regulatórios *trans*-atuantes, tais como as proteínas hnRNP e SR, podem resultar em mudanças do processo de *splicing* (KARNI et al. 2007).

Os eventos de *splicing* alternativos sob os genes tumorais podem ser altamente relevantes em todos os principais aspectos da biologia celular do câncer, incluindo controle do ciclo celular, vias de transdução de sinal, apoptose, angiogênese, invasão e motilidade, e metástase (BLENCOWE 2006).



Fonte: Adaptado de PAJARES et al. (2007).

Figura 6 - Modo anormal de *splicing* de mRNA originando isoforma protéica com propriedades oncogênicas.

1.3 O MECANISMO DE *SPLICING*

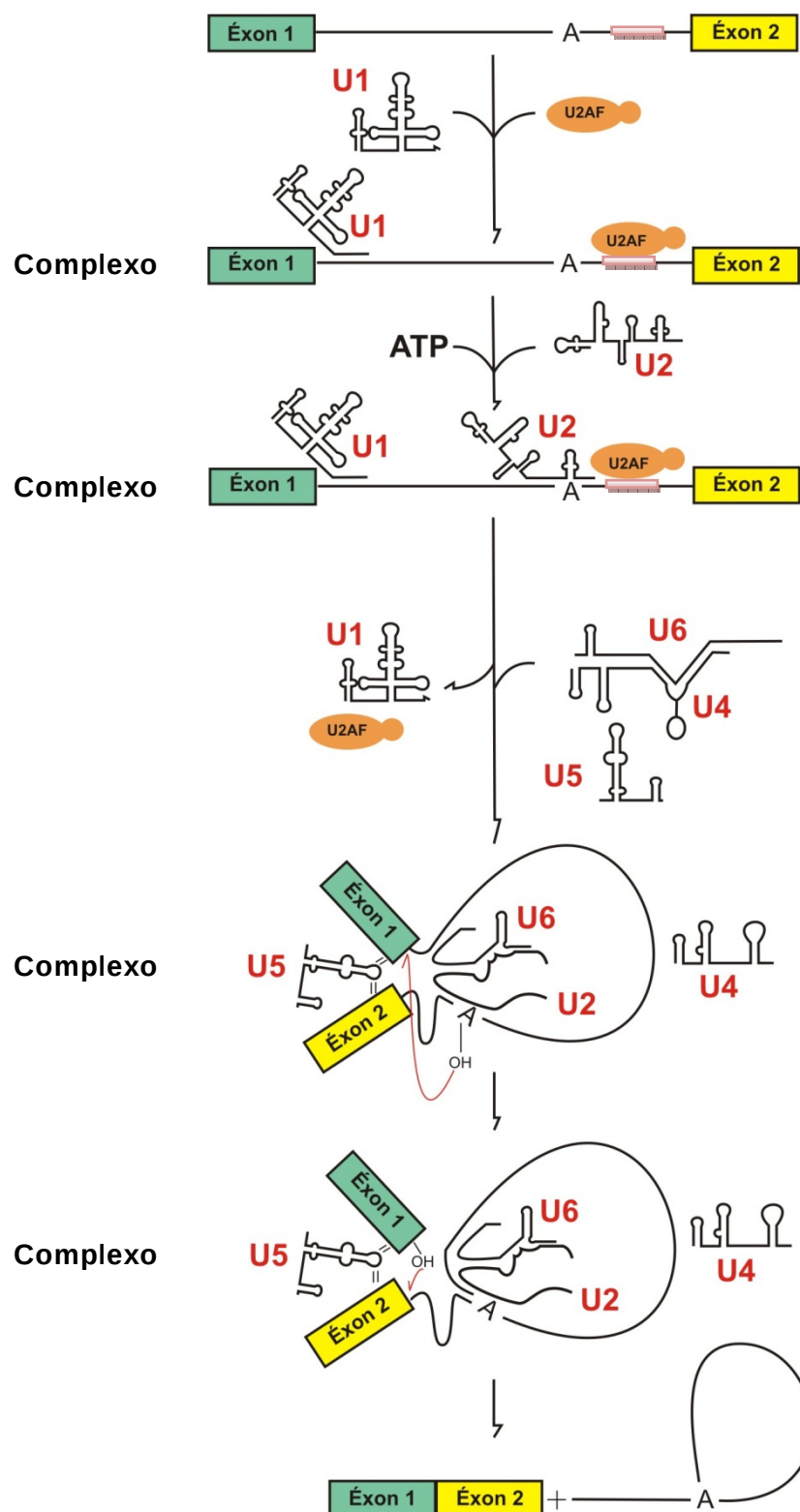
O spliceossomo é um dinâmico complexo molecular. Este amplo maquinário é composto de 5 pequenas partículas ribonucleoprotéicas, snRNP: U1, U2, U4 e U6 (na classe maior de spliceossomo - tipo U2) e outros 5 snRNP: U11, U12, U4 e U6 (na menor classe de spliceossomo - tipo U12), e aproximadamente 50-100 fatores de *splicing* considerados não-snRNP (KRAMER et al. 1996). Cada snRNP é composto de uma pequeno RNA nuclear (snRNA) e múltiplas proteínas. Os sítios de ligação e funções destas partículas são muito específicos. Por exemplo, o snRNP U1 liga-se ao sítio de *splice* 5', enquanto que snRNP U2 liga-se ao sítio de ramificação via RNA: interações entre snRNA e o pré-mRNA (Figura 7).

O processo de *splicing* inicia-se com a formação do Complexo E. A montagem do complexo E envolvem os reconhecimentos do sítio de *splice* 5', do trato de polipirimidina e sítio de *splice* 3', pelos snRNP U1 e fator de *splicing* heterodimérico U2AF (fator auxiliar snRNP U2), consistindo-se este pelo fator auxiliar U2-65 (U2AF65) e fator auxiliar U2-35 (U2AF35). O ponto de ramificação é reconhecido pelo fator de *splicing* 1 (SF1). Diversos fatores de *splicing* não-snRNP tais como proteínas serina/arginina (SR) e SR-related também se associam com o pré-mRNA nesta etapa. O tri-snRNP U4/U6.U5 podem associar-se com o primeiro éxon perto do sítio de *splice* 5' no Complexo E. Esta associação é dependente de ATP. O pareamento de base ATP-dependente do snRNP U2 com o ponto de ramificação forma o Complexo A. O Complexo B é formado pelo recrutamento do tri-snRNP U4/U6.U5 ao pré-spliceossomo. O duplo U6/U4 é rompido e uma nova associação entre U6 e o sítio de *splice* 5' é formado, resultando no deslocamento do

snRNP U1. O sítio de *splice* 5' é trazido para perto do ponto de ramificação e sítio de *splice* 3', através do pareamento de base do snRNP U6/U2 e da interação do snRNP U5 com ambos os éxons próximos aos sítios de *splice*.

Neste momento do processo, o snRNP U4 deixa o complexo e ocorre a primeira fase catalítica do *splicing*, formando um laço intrônico. Finalmente, com o pareamento de bases do snRNP U5 com ambos os limites 5' e 3' dos éxons, são posicionadas as terminações dos dois éxons para a subsequente segunda etapa catalítica do *splicing*.

Após a segunda etapa ter sido completada, os exóns estão ligados, o laço intrônico é liberado e os componentes do spliceossomo dissociam-se para serem reutilizados em outros eventos de *splicing*. A Figura 7 representa esquematicamente a montagem do spliceossomo, e o processo resultante na excisão dos íntrons do pré-mRNA.



Fonte: Adaptada de PATEL E STEITZ (2003).

Figura 7 - Formação e rearranjo do spliceossomo durante a reação de *splicing*.

1.4 MASSIVELY PARALLEL SIGNATURE SEQUENCING (MPSS)

Tecnologias de medição de expressão gênica em larga escala são desenvolvidas para serem capazes de mensurar a abundância de muitos transcritos de mRNA de uma amostra. Nestas estão inclusas a tecnologias de *microarray* (SCHENA et al. 1995), *Serial Analysis of Gene Expression* - SAGE (VELCULESCU et al. 1995), e mais recentemente *Massively Parallel Signature Sequencing* - MPSS. Comparado com a tecnologia de SAGE, MPSS é mais sensível e pode ser usado para mensurar com confiança transcritos extremamente raros (BRENNER et al. 2000; CHEN e RATTRAY 2006).

Há dois métodos básicos de MPSS: o método *Classical* e *Signature*. A diferença entre estes métodos é que para o método *Classical*, todo o fragmento do sítio *DpnII* (GATC) ao início do segmento de *poly(A)* é clonado e submetido ao sequenciamento. No método *Signature*, durante a clonagem, uma enzima de restrição corta a 21 ou 22 pb a jusante do sítio de *DpnII*. O método *Signature* é capaz de remover viéses durante o processamento de clonagem (MEYERS et al. 2004).

2 ABORDAGEM GERAL APLICADA À PESQUISA

A primeira parte desta pesquisa fora focada na seleção curada e anotação gênica dos fatores de *splicing* humanos e das proteínas *splicing*-associadas indicadas em BARBOSA-MORAIS et al. (2006), seguido da busca aos respectivos genes ortólogos em camundongos, resultando no conjunto final de genes ortólogos curados envolvidos no processo de *splicing* em ambas as espécies. Posteriormente às análises restritivas efetuadas sobre as seqüências dos respectivos cDNA, foram então extraídas as suas *tags* virtuais. No próximo passo, utilizamos as *tags* virtuais e as *tags* experimentais para avaliar a expressão dos genes de interesse.

A segunda parte da pesquisa concentrou-se nos grupos protéicos de fatores de *splicing*: snRNP, SR, hnRNP e Sm. Inicialmente, pelo computo dos coeficientes de correlação (r) de Pearson (produto-momento) existentes entre os valores de expressão gênica destes grupos de fatores. Posteriormente, pela inferência estatística e funcional ao padrão de coexpressão gênica destes grupos, entre os pares de bibliotecas de MPSS, encontrados nas distintas espécies. Finalmente, pela investigação comparativa quanto aos eventos de *splicing* alternativos presentes nos grupos de seus fatores de *splicing*, e ocorrendo em tecidos normais e tumorais.

Na terceira parte da pesquisa finalizou-se a obtenção e análise dos distintos valores de expressão gênica (bibliotecas normais e tumorais de MPSS) de todos os fatores de *splicing* curados de humanos, em correlação aos respectivos eventos de *splicing* alternativos aos quais se encontram presentes.

A Figura 8 representa um esquema geral da abordagem geral desta pesquisa, sendo possível visualizar as principais etapas.

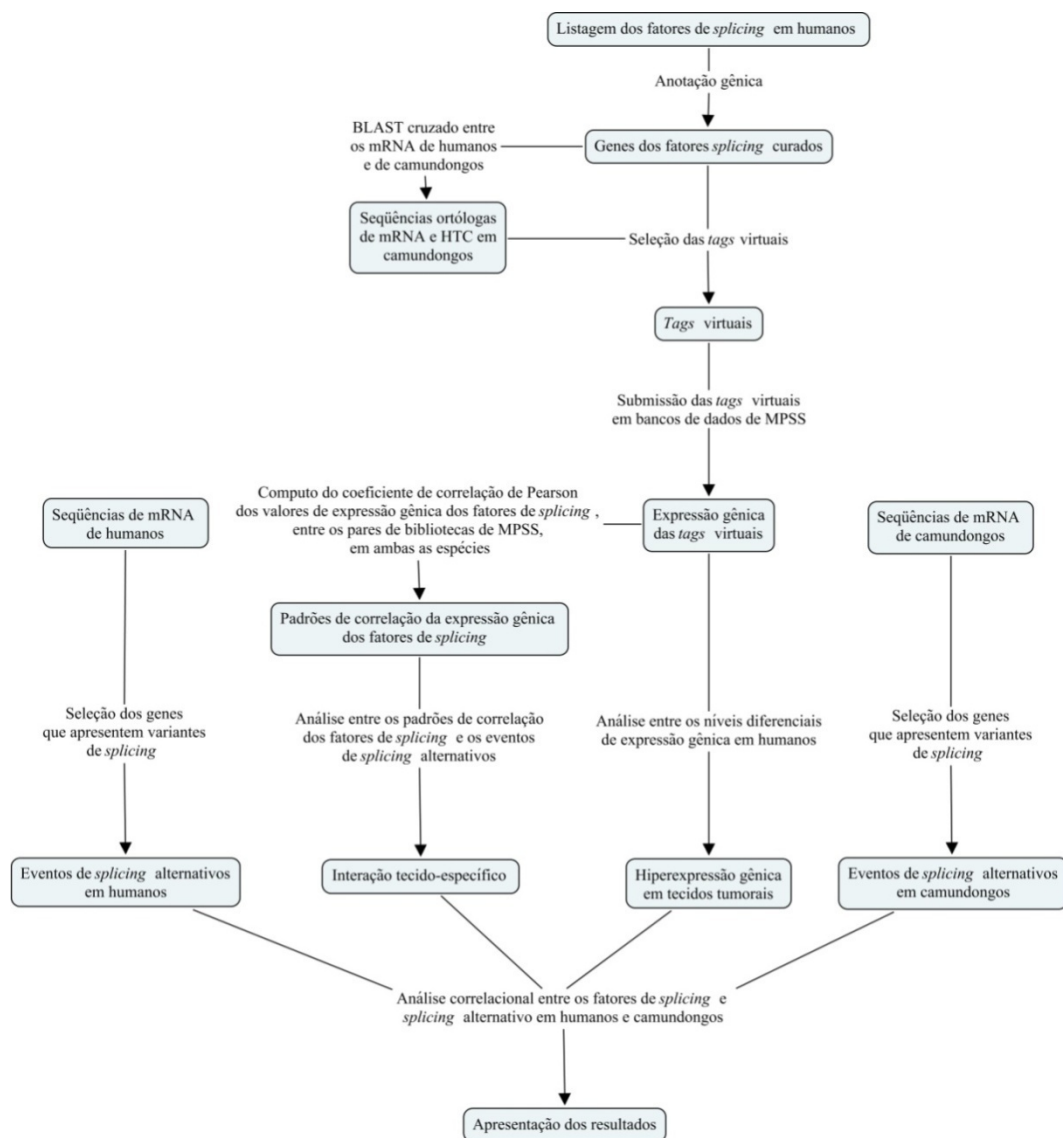


Figura 8 - Esquema geral da abordagem aplicada à pesquisa.

3 MATERIAIS E MÉTODOS

3.1 SELEÇÃO CURADA DOS FATORES DE *SPLICING*

Foram selecionadas as proteínas pelas quais há evidência experimental de seu envolvimento em modos de *splicing*, tendo-se como referência de escolha o conjunto total de 254 proteínas, disponível em BARBOSA-MORAIS et al. (2006). Foram coletadas e analisadas as suas respectivas anotações gênicas, através de pesquisas efetuadas com os seus referidos *clusters* gênicos encontrados no UniGene (BOGUSKI e SHULER 1995).

3.2 OBTENÇÃO DAS SEQUÊNCIAS DE cDNA

Todas as EST públicas de humanos e camundongos foram obtidas do dbEST (BOGUSKI et al. 1993). Os cDNA *full length* e *full inserts*, ambos denominados de mRNA no presente projeto, foram obtidos do RefSeq (PRUITT e MAGLOTT 2001) do MGC (STRAUSBERG et al. 2002), e do UniGene (BOGUSKI e SCHULER 1995) – *Homo sapiens*: versão 190; *Mus musculus*: versão 152.

3.3 EXTRAÇÃO DAS TAGS VIRTUAIS

As *tags* virtuais foram extraídas das sequências de cDNA previamente selecionadas (contendo os UniGene *clusters* dos fatores de *splicing* curados).

Limitou-se a seleção dos segmentos de 13 nucleotídeos (em humanos) e 16 nucleotídeos (em camundongos) adjacentes ao sítio mais a jusante da enzima de restrição *DpnII* (GATC). Reiteramos que a confiabilidade de uma *tag* virtual é dada pela qualidade da região 3' da qual fora derivada. Quando há evidências de que a região 3' está completa, a *tag* virtual nos é considerada confiável. Como descrito em BOON et al. (2002), fora utilizado a presença do sinal de poliadenilação (AAUAAA ou AUUAAA, sinais canônicos) e/ou presença da cauda de poli(A) (definida como conjunto mínimo de 8 adeninas no final 3' do cDNA) como indicativos de total seqüenciamento da região 3' para cada cDNA. O conjunto de *tags* virtuais providenciou uma associação única entre uma seqüência específica de um transcrito dos fatores de *splicing*, e o seu restrito segmento terminal. Foram descartados aqueles casos em que os alinhamentos divergentes nas 3' UTR não compreenderem poliadenilação alternativa e obedecerem ao padrão mínimo de 8 nt de adeninas em uma seqüência de 10 nt, nas 30 últimas posições de seqüências genômicas, então considerados casos de artefatos de seqüenciamento, denominados aqui de *internal priming*.

3.4 SELEÇÃO DOS GENES DOS FATORES DE *SPLICING* ORTÓLOGOS DE CAMUNDONGOS

A pesquisa de possíveis ortólogos em camundongos fora efetuada através do alinhamento local cruzado (humanos *versus* camundongos, e camundongos *versus* humanos), parâmetro *default* e a seleção dos melhores *Hits*, no processamento do programa *Basic Local Alignment Search Tool* - BLAST (ALTSCHUL 1990). A

estratégia de alinhamento cruzado é adotada em larga escala por diversos outros projetos, na busca de genes ortólogos, citando-se o projeto HomoloGene do *National Center for Biotechnology Information* (NCBI) em 2006a.

3.5 SUBMISSÃO DAS TAGS VIRTUAIS EM BANCOS DE DADOS DE MPSS

Os valores de expressão gênica dos candidatos foram obtidos pela busca de identidade das *tags* virtuais previamente coletadas, com todas as *tags* de bibliotecas de MPSS de ambas as espécies, em banco de dados local. Constam 51 e 81 bibliotecas de MPSS de humanos (tecidos normais e tumorais) (Anexo 1) e de camundongos (tecidos normais) (Anexo 2) respectivamente. Foram excluídos os genes que apresentam valores de expressão gênica nulo em mais de 2/3 das bibliotecas de MPSS.

3.6 ANÁLISE NA CORRELAÇÃO DE EXPRESSÃO GÊNICA DOS FATORES DE *SPLICING* CONSTITUTIVOS ENTRE AS BIBLIOTECAS DE MPSS

A variação na coexpressão gênica dos fatores de *splicing* constitutivos (snRNP, SR, hnRNP e Sm) existente entre as bibliotecas de MPSS de ambas as espécies, fora analisada pelo computo do coeficiente de correlação (r) de Pearson (produto-momento). Buscou-se uma possível semelhança funcional aos padrões de

coexpressão dos fatores de *splicing* constitutivos dispostos nas bibliotecas de MPSS das espécies.

A ausência de correlação entre as bibliotecas é apontada pelo coeficiente de correlação negativo ($r \leq 0$). Os valores de r ($0 < r < 0.5$) compreendidos na escala de correlação positiva, indicam um baixo grau de concordância na expressão relativa dos mRNA dos conjunto de fatores de *splicing*.

Os valores r compreendidos entre os limites maior ou igual a 0.5 e menor que 0.75 correspondem a uma moderada concordância. Os outros valores superiores de r ($r \geq 0.75$) indicam um alto índice de correlação.

Nestes critérios abordados, os valores superiores de r ($r \geq 0.5$) são índices adequados a conclusão de uma favorável correlação positiva existente na coexpressão gênica dos fatores de *splicing* entre as bibliotecas.

Como exposto, enfatizou-se que o critério para se definir um limiar inferior a existência de correlação teve como base a inferência estatística aos valores de r , como abordado por outros relevantes estudos de referência (ZHOU et al. 2003). A biblioteca que obtiver o maior montante (acima ou igual a 50%) dos valores de r ($r < 0.5$) (comparativamente às outras bibliotecas) será considerada um *outlier*, ou seja, possuem um diferencial negativo na coexpressão gênica dos fatores de *splicing*.

Tal constatação possui favorável parecer a uma subsequente investigação correlacional, quanto à distinta coexpressão dos tecidos considerados *outliers*, frente aos eventos de *splicing* alternativos encontrados nestes (YEO et al. 2004).

O padrão comparativo de coexpressões tornou-se uma favorável alternativa de investigação correlacional com os outros padrões diferenciais de *splicing* alternativos encontrados.

Foram analisadas as correlações existentes nas bibliotecas semelhantes funcionalmente (pertencentes ao mesmo grupo tecidual), e também analisadas estas com as outras sem semelhanças funcionais (pertencentes a outros distintos grupos teciduais), obtivemos assim dois resultados de correlação de Pearson.

3.7 PROTOCOLO DE BUSCA AOS EVENTOS DE *SPLICING* ALTERNATIVOS

Para a investigação de ocorrência de eventos de *splicing* alternativos nas espécies, foram utilizados os respectivos totais de mRNA, previamente selecionados. Esta busca é baseada na comparação pareada entre os alinhamentos nos genomas dos cDNA dos referentes *clusters* alocados em bancos de dados local. O protocolo utilizado é referido por MIRONOV et al. (1999); MODREK et al. (2001) e GALANTE et al. (2004). O protocolo adaptado possui em seu processo os seguintes momentos: Mapeamento dos cDNA no genoma e clusterização.

Todos os cDNA disponíveis no dbEST, e as seqüências de mRNA do UniGene versão 153, e em camundongos do UniGene versão 15, são alinhados ao conjunto de seqüências do genoma de humanos (Hs 17), e do genoma de camundongos (Mm 7), pelo uso do pp-Blast (OSÓRIO et al. 2003), e uma implementação do MegaBlast (ZHANG et al. 2002) para o modo *cluster* paralelo. As análises dos altos níveis de *scores* foram associadas com as altas identidades sobre todos os alinhamentos. Somente as seqüências com os mais altos *scores* foram utilizadas na presente análise.

A clusterização das seqüências de cDNA fora baseada nas suas coordenadas genômicas como descrito por SAKABE et al. (2003) e GALANTE et al. (2004). As seqüências foram criterizadas como integrantes de mesmo *cluster* quando estas partilharam de mesma estrutura gênica. Após a construção do banco de dados, seguimos para a análise do *splicing* alternativo.

São classificados como modos alternativos de *splicing* os genes apresentando:

- I. Uso alternativo de éxon: ocorre quando o gene apresenta alguns transcritos contendo um determinado éxon, em detrimento a outros com ausência deste éxon.
- II. Sítio alternativo 3': ocorre quando a parte 5' do éxon apresenta variações, ou seja, há momentos em que o éxon possui um segmento da seqüência que no transcrito variante é parte do íntron.
- III. Sítio alternativo 5': modo similar e distinto ao anterior com a parte 3' do éxon possuindo o ganho de um segmento da seqüência que no transcrito variante é parte do íntron.
- IV. Retenção de íntron: Ocorre quando um transcrito maduro do gene apresenta uma seqüência que é intrônica em outro transcrito. No modo variante o transcrito apresenta um novo éxon, contendo o íntron e os dois éxons adjacentes a este.

3.8 INVESTIGAÇÃO DOS NÍVEIS DE *SPLICING* ALTERNATIVOS

Para a identificação dos tecidos de onde as EST foram geradas, utilizamos os resultados apresentados pelo projeto eVOC (KELSO et al. 2003). No presente estudo

as bibliotecas de expressão gênica do projeto eVOC foram selecionadas e agrupadas por semelhança funcional (Ontologia). Como resultado obtivemos grupos específicos de tipos teciduais que foram utilizados como quadro representativo *chave* para a soma dos valores de eventos de *splicing* alternativos, presentes nos distintos grupos de bibliotecas. Tendo em vista o exposto, todas as identidades das bibliotecas com tais semelhanças foram agrupadas em um tipo tecidual único que abranja funcionalmente o grupo.

Em camundongos não há uma anotação como o eVOC. No estudo em camundongos, adotou-se uma estratégia similar, de através das 1.220 bibliotecas identificadas no arquivo *Mm.lib.info* – NCBI (2006b) efetuar-se uma análise manual curada, restringindo-se e agrupando-se os tipos teciduais por origem e função inerentes aos mesmos. Nesta espécie buscou-se quantificar os eventos de *splicing* alternativos num tecido, pela busca de identidade entre as bibliotecas apresentadas nos respectivos eventos com as bibliotecas presentes em um tipo tecidual *chave*. A classificação efetuada está disponível em COMPBIO LUDWIG (2007).

3.9 ANÁLISE COMPARATIVA ENTRE OS VALORES DE EXPRESSÃO GÊNICA DOS FATORES DE *SPLICING*

Inicialmente os valores de expressão gênica dos fatores de *splicing* constitutivos foram analisados em ambas as espécies. Selecionamos os três tipos teciduais nos quais se encontram os maiores valores de expressão gênica de cada grupamento de fatores de *splicing* constitutivos. Investigamos os possíveis tipos

teciduais nos quais os grupamentos de fatores de *splicing* constitutivos apresentam os maiores valores de expressão gênica.

Foram confrontados os distintos valores de expressão gênica de todos os fatores de *splicing* de humanos, encontrados em bibliotecas normais e tumorais de MPSS. Para toda comparação efetuada entre os valores de expressão dos fatores de *splicing*, considera-se expressão gênica superior quando ocorrer valores de expressão em tumores maiores do que o dobro dos valores de expressão obtidos em bibliotecas normais.

3.10 CORRELAÇÃO NA EXPRESSÃO GÊNICA DOS FATORES DE *SPLICING* PRESENTES NOS EVENTOS DE *SPLICING* ALTERNATIVOS

Foram selecionados todos os fatores de *splicing* com valores de expressões gênicas superiores (mais do que o dobro) em tecidos tumorais de humanos, e/ou com padrão de coexpressão diferencial tecido-preferencial. Estes fatores de *splicing* foram investigados quanto ao seu envolvimento em padrões de eventos de *splicing* alternativos, ou seja, se são encontrados em transcritos variantes de *splicing*. Foram listados os eventos de *splicing* alternativos sob presença indicativa de todos os fatores de *splicing*.

Foram comparados e analisados os valores de expressão gênica de todos os fatores de *splicing* sob presença indicativa em eventos de *splicing* alternativos, em tecidos humanos tumorais *versus* normais, e confrontados estes resultados finais *in*

silico, com outros distintos dados sob validação experimental, encontrados em referências científicas correlatas.

3.11 SIMULAÇÃO *IN SILICO* AOS EVENTOS DE *SPLICING* ALTERNATIVOS DE HUMANOS

Para o controle nas diferenças numéricas de EST dentre as bibliotecas de ambas as espécies, foram considerados os genes que tenham um número favorável de 20 seqüências de EST alinhadas, e presentes nos dados de *splicing* alternativos.

Foram gerados 1000 conjuntos de 5 EST randomicamente, presentes nas bibliotecas tumorais (Glândula adrenal, Glândula mamária, Fígado, Sistema nervoso, Glândula pituitária, Próstata, Estômago, Timo, Bexiga, Pulmão, Placenta e Útero) e normais (Glândula adrenal, Glândula mamária, Fígado, Sistema nervoso, Glândula pituitária, Próstata, Estômago, Timo, Bexiga, Intestino, Pulmão, Placenta e Útero). Tais conjuntos de EST foram investigados quanto à ocorrência de eventos de *splicing* alternativos sob presença dos fatores de *splicing* constitutivos (snRNP, SR, hnRNP e Sm).

4 RESULTADOS

4.1 PADRÕES DE EXPRESSÃO GÊNICA DOS FATORES DE *SPLICING* CONSTITUTIVOS

Estimativas atuais sobre o número de proteínas envolvidas aos eventos de *splicing* alternativos ou intimamente relacionadas ao spliceossomo são consensuais sobre o montante aproximado de 300 unidades protéicas (JURICA e MOORE 2003), como abordado em referencial teórico. Estudos relevantes em análise proteômica do spliceossomo de humanos, como os resultados apresentados por ZHOU et al. (2003), e mais recentemente BARBOSA-MORAIS et al. (2006), foram o suporte referencial ao conjunto investigado de proteínas intimamente relacionadas aos eventos de *splicing* alternativos. Com o propósito alcançado de curar-se manualmente 254 proteínas, pelos seus respectivos números de acesso SwissProt (2006), fora possível coletar e estabelecer 253 anotações gênicas ao referido conjunto, como consta em Anexo 3.

Todas as EST públicas de humanos e camundongos foram obtidas do dbEST (BOGUSKI et al. 1993), através do acesso ao UniGene: *Homo sapiens* versão 190 (6.877.831 seqüências em *clusters*; 154.128 mRNA; 48.452 HTC); e *Mus musculus* versão 152 (4.444.598 seqüências em *clusters*; 75.913 mRNA; 129.571 HTC). Com o conjunto total de seqüências (mRNA e HTC) e o acesso aos formatos FASTAS, restringiu-se a investigação ao conjunto disponível de 3.567 seqüências, com referências gênicas aos 253 *clusters* anotados. Fora utilizado a presença do sinal de

poliadenilação e ou presença da cauda de poli(A) como indicativos de região 3' completa para cada cDNA, resultando em 1.311 seqüências com cauda de poli(A), 1.837 seqüências com sinal de poliadenilação, e 982 seqüências com cauda de poli(A) e concomitante sinal de poliadenilação.

A subsequente análise restritiva às ocorrências de poliadenilação alternativa e *internal priming*, foram suportadas pela seguinte ferramenta de análise: *Web Genome Gateway* alocada na *UCSC Genome Bioinformatics* [2006]. Selecionando-se aqueles casos em que ocorram alinhamentos plenos das seqüências dispostas por *clusters*, e os distintos casos de poliadenilações alternativas. Tal análise restritiva resultou em 238 *clusters* de humanos, representados todos pelas suas respectivas terminações nucleotídicas, as *tags* virtuais.

A pesquisa de possíveis ortólogos em camundongos, a partir de 238 *clusters* em humanos, fora efetuada através do alinhamento local cruzado (humanos *versus* camundongos, e camundongos *versus* humanos) no parâmetro *default*, com o processamento do programa BLAST, gerando um conjunto de 218 *clusters* ortólogos em camundongos.

O confronto dos conjuntos de *clusters* presentes em ambas as espécies, resultou num número de 38 *clusters* sem ortólogos correspondentes que obedeçam aos critérios restritivos adotados. Com tal análise restritiva aos ortólogos, obteve-se um total correspondente de 211 *clusters* gênicos (Anexo 4). As *tags* virtuais foram extraídas das seqüências de cDNA dos 211 *clusters* ortólogos de camundongos.

A seleção das frequências normalizadas de expressão gênica de MPSS dos 211 *clusters* fora obtida pela comparação das *tags* virtuais com todas as *tags* de

bibliotecas de MPSS de humanos (51 bibliotecas), e camundongos (81 bibliotecas), disponibilizadas em banco de dados local.

Efetua-se o cálculo dos coeficientes r entre os valores de expressão gênica dos fatores de *splicing* snRNP (13 genes), hnRNP (18 genes), SR (7 genes) e Sm (6 genes) de humanos (Quadro 1), dispostos entre as bibliotecas de MPSS tipo sexo específico (17 bibliotecas presentes em machos e *pool*; 17 bibliotecas presentes em fêmea e *pool*) (Quadro 3).

Quadro 1 - UniGene *clusters* dos fatores de *splicing* constitutivos de humanos.

snRNP	SR	hnRNP	Sm
Hs.181368	Hs.533122	Hs.508848	Hs.516076
Hs.151787	Hs.6891	Hs.546271	Hs.111632
Hs.246112	Hs.369624	Hs.522257	Hs.515255
Hs.469173	Hs.405144	Hs.465808	Hs.564847
Hs.182255	Hs.68714	Hs.589594	Hs.425311
Hs.11776	Hs.584801	Hs.380118	Hs.424908
Hs.374973	Hs.479693	Hs.2853	
Hs.280378		Hs.571177	
Hs.406423		Hs.487774	
Hs.177861		Hs.501309	
Hs.1063		Hs.172550	
Hs.466775		Hs.166463	
Hs.528763		Hs.808	
		Hs.546261	
		Hs.432485	
		Hs.96996	
		Hs.573762	
		Hs.516539	

Quadro 2 - UniGene *clusters* dos fatores de *splicing* constitutivos de camundongos.

snRNP	SR	hnRNP	Sm
Mm.3757	Mm.2478	Mm.390303	Mm.276802
Mm.299312	Mm.5222	Mm.9043	Mm.379101
Mm.34562	Mm.287826	Mm.390606	Mm.28694
Mm.386890	Mm.223946	Mm.379375	Mm.45683
Mm.308514	Mm.43331	Mm.331640	Mm.45151
Mm.102627		Mm.21740	Mm.248188
Mm.821		Mm.317706	Mm.165735
Mm.873		Mm.286394	
Mm.216386		Mm.286408	
Mm.215860		Mm.332474	
Mm.156914		Mm.254223	
Mm.196532			
Mm.281900			

Quadro 3 - Bibliotecas e suas respectivas origens sexo específicas em humanos.

Bibliotecas	Origem
Hipotálamo	<i>Pool</i>
Núcleo caudato	<i>Pool</i>
Amígdala	<i>Pool</i>
Tálamo	<i>Pool</i>
Medula espinhal	<i>Pool</i>
Cerebelo	<i>Pool</i>
Gl. adrenal	<i>Pool</i>
Gl. pituitária	<i>Pool</i>
Baço	<i>Pool</i>
Timo	<i>Pool</i>
Testículo	Macho
Próstata	Macho
Fígado	Sem especificação
Bexiga	<i>Pool</i>
Pulmão	Macho
Intestino delgado	<i>Pool</i>
Estômago	<i>Pool</i>
Útero	Fêmea
Gl. mamária	Fêmea
Placenta	Fêmea

Legenda: Pool incluem Macho e Fêmea.

Efetuuou-se concomitantemente o cálculo dos coeficientes r entre os valores de expressão gênica dos fatores de *splicing* snRNP (13 genes), hnRNP (11 genes) , SR (5 genes) e Sm (7 genes) de camundongos (Quadro 2), dispostos entre as

bibliotecas de MPSS tipo sexo específico (16 bibliotecas de tecidos normais de origem exclusiva em macho; 18 bibliotecas de tecidos normais de origem exclusiva em fêmea) (Quadro 4). Analisou-se a expressão gênica dos fatores de *splicing*, através de uma busca por fraca concordância, dentre os coeficientes r concordantes.

Correlacionou-se os pares teciduais normais tipos sexo específico (17 bibliotecas presentes em macho; 17 bibliotecas presentes em fêmea) contendo os valores de expressão gênica dos fatores de *splicing*: snRNP (13 genes), hnRNP (18 genes), SR (7 genes) e Sm (6 genes), com as 7 bibliotecas tumorais (*Breast Cancer, ER- cell line; Breast Cancer, ER+ cell line; HB4A modified C5.2, ErbB2 expr; Breast Cancer; Lung Cancer Cells; Melanoma; Melanoma Biopsies*), em humanos.

Obtivemos como resultado, um padrão de coexpressão gênica diferencial dos fatores de *splicing* constitutivos, entre as duas entidades de bibliotecas (normais e tumorais). As bibliotecas de MPSS de origens tumorais comportaram-se como *outliers* em relação às bibliotecas de MPSS de origens normais, em humanos.

Quadro 4 - Bibliotecas e suas distintas origens sexo específicas em camundongos.

Bibliotecas em Fêmeas	Bibliotecas em Machos
Tálamo	Tálamo
Medula espinhal: inteira	Medula espinhal: inteira
Hipotálamo	Hipotálamo
Caudato, Putamen, Acumbens	Caudato, Putamen, Acumbens
Cerebelo	Cerebelo
Amígdala	Amígdala
Gl. mamária	Próstata
Útero	Testículo
Gl. pituitária	Gl. pituitária
Gl. adrenal	Gl. adrenal
Timo	Timo
Baço	Baço
Placenta-E18	Intestino delgado
Estômago	Pulmão
Intestino delgado	Bexiga
Pulmão	Fígado: lobo direito
Bexiga	
Fígado: lobo direito	

Examinamos a correlação na expressão gênica dos fatores de *splicing* constitutivos entre todos os pares de tipos teciduais presentes mutuamente em ambas as espécies, e obtivemos os seguintes resultados representados como seguem abaixo nos Quadros 5 e 6. A inferência estatística aplicada nestes baseia-se nos padrões de distribuição dos *outliers* (relativos grupos diferenciais de correlação). Os *outliers* apresentam coeficientes inferiores ($r < 0,5$) entre os níveis de expressão dos fatores de *splicing*, entre os tipos teciduais (presentes acima ou igual a 50% nos valores de r do tipo tecidual). As investigações dos *outliers* obedeceram dois momentos: comparar todos os tecidos (Quadro 5); comparar os tecidos de determinado grupo (Quadro 6). O Quadro 7 apresenta e discrimina os quatro grupos: sistema nervoso; glandulares; sexo específico e outros órgãos.

As diferentes perspectivas de análise, efetuadas nos dois momentos comparativos (Quadro 5 e 6), quanto ao padrão diferencial de coexpressão gênica entre os fatores de *splicing*, resultaram numa presença preferencial de *outliers* nos grupos sistema nervoso e sexo específico.

No Quadro 5 os valores diferenciais de coexpressão gênica dos fatores de *splicing* constitutivos foram comparados de acordo com a distribuição dos seus coeficientes de correlação entre todas as bibliotecas indistintamente. No primeiro caso apresentado todas as bibliotecas foram comparadas sem restringir-se o grupo tecidual ao qual pertencem. Como especificado, no primeiro modo comparativo adotado, verificamos que preferencialmente as bibliotecas pertencentes aos grupos teciduais sistema nervoso e sexo específico foram *outliers*. As distintas análises em macho e fêmea foram efetuadas entre as bibliotecas dos tipos: sexo específico correlato e *pool*, ou seja foram agrupadas bibliotecas de mesma origem sexual com

bibliotecas *pool*, e comparados seus valores de correlação. No Quadro 5 as correlações de Pearson entre os valores de expressão gênica dos fatores de *splicing* snRNP, indicam a biblioteca tipo tecidual medula espinhal como *outlier*, único então verificado em ambas as espécies.

Quadro 5 - Valores diferenciais de coexpressão gênica dos fatores de *splicing* constitutivos entre todas as bibliotecas indistintamente.

		snRNP	SR	hnRNP	Sm
Humanos	Macho	Estômago	-	-	Cerebelo Próstata
	Fêmea	Medula espinhal Tálamo Estômago	-	-	Cerebelo Gl. mamária
Camundongos	Macho	Medula espinhal Testículo	Hipotálamo	Testículo Baço Bexiga Gl. pituitária	Medula espinhal Caudato Amígdala Timo
	Fêmea	Placenta Fígado	-	Placenta Bexiga Útero Cerebelo	Amígdala
Humanos e Camundongos	Macho	Medula espinhal	-	-	-
	Fêmea	-	-	-	-

Legenda: A biblioteca que obtiver o maior montante de valores de $r < 0,5$ (comparativamente às outras bibliotecas) será considerada um *outlier*. As distintas análises em macho e fêmea foram efetuadas entre as bibliotecas tipos: sexo específico correlato e *poll*.

No Quadro 6 os valores diferenciais de coexpressão gênica dos fatores de *splicing* constitutivos foram comparados de acordo com a distribuição dos seus coeficientes de correlação entre as bibliotecas pertencentes ao mesmo grupo tecidual. Como visto compara-se apenas as bibliotecas de mesmo grupo tecidual. No segundo modo comparativo adotado verificamos que preferencialmente os grupos teciduais sistema nervoso e sexo específico possuem bibliotecas *outliers*. As distintas análises

em macho e fêmea foram efetuadas entre as bibliotecas dos tipos: sexo específico correlato e *pool*, ou seja foram agrupadas bibliotecas de mesma origem sexual com bibliotecas *pool*, e comparados seus valores de correlação. No Quadro 6 as correlações de Pearson entre os valores de expressão gênica dos fatores de *splicing* snRNP e Sm indicam as bibliotecas tipos teciduais medula espinhal e cerebelo como *outliers*, únicas então verificadas em ambas as espécies.

Quadro 6 - Valores diferenciais de coexpressão gênica dos fatores de *splicing* constitutivos entre as bibliotecas de determinado grupo.

		snRNP	SR	hnRNP	Sm
Humanos	Macho	Medula espinhal Fígado Estômago	Hipotálamo	Medula espinhal Testículo Gl. adrenal Bexiga Gl. pituitária	Medula espinhal Cerebelo Timo Amígdala Caudato
	Fêmea	Medula espinhal	-	-	Cerebelo Gl. mamária Placenta
Camundongos	Macho	Medula espinhal Testículo Gl. adrenal	-	-	Cerebelo Próstata
	Fêmea	Hipotálamo Placenta Fígado	-	Placenta Bexiga	Útero Baço
Humanos e Camundongos	Macho	Medula espinhal	-	-	Cerebelo
	Fêmea	-	-	-	-

Legenda: A biblioteca que obtiver o maior montante de valores de $r < 0,5$ (comparativamente às outras bibliotecas do mesmo grupo) será considerada um *outlier*. As distintas análises em macho e fêmea foram efetuadas entre as bibliotecas tipos: sexo específico correlato e *pool*.

Quadro 7 - Grupos teciduais.

Sistema nervoso	Glandulares	Outros órgãos	Sexo específico
Núcleo caudato	Gl. adrenal	Fígado	Útero
Amígdala	Gl. pituitária	Bexiga	Gl. mamária
Tálamo	Baço	Pulmão	Placenta
Medula espinhal	Timo	Intestino delgado	Testículo
Cerebelo		Estômago	Próstata

As Figuras 9-20 ilustram os valores de expressões gênicas dos grupamentos de fatores de *splicing* constitutivos, encontrados nas respectivas bibliotecas de MPSS de humanos e camundongos.

Foram selecionados os três primeiros maiores valores de expressões gênicas de cada grupamento de fatores de *splicing* verificados nas respectivas bibliotecas de MPSS de ambas as espécies.

Na Figura 9 verificou-se que os três primeiros valores de expressões gênicas do grupamento de fatores de *splicing* snRNP (humanos) foram os dos UniGene *clusters*: Hs.246112 (Timo), Hs.11776 (Timo) e Hs.181368 (Glândula adrenal).

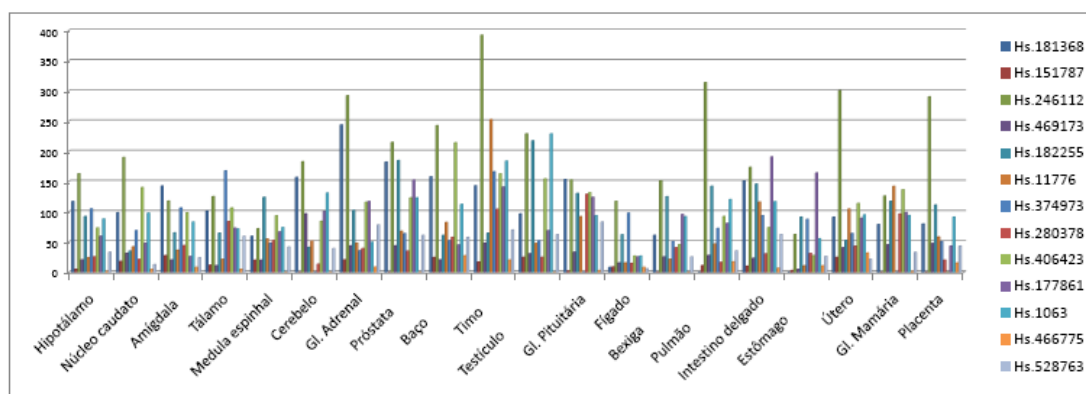


Figura 9 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* snRNP, em bibliotecas de MPSS de humanos.

Na Figura 10 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* SR (humanos) foram os dos UniGene *clusters*: Hs.369624 (Testículo), Hs.584801 (Timo) e Hs.68714 (Glândula pituitária).

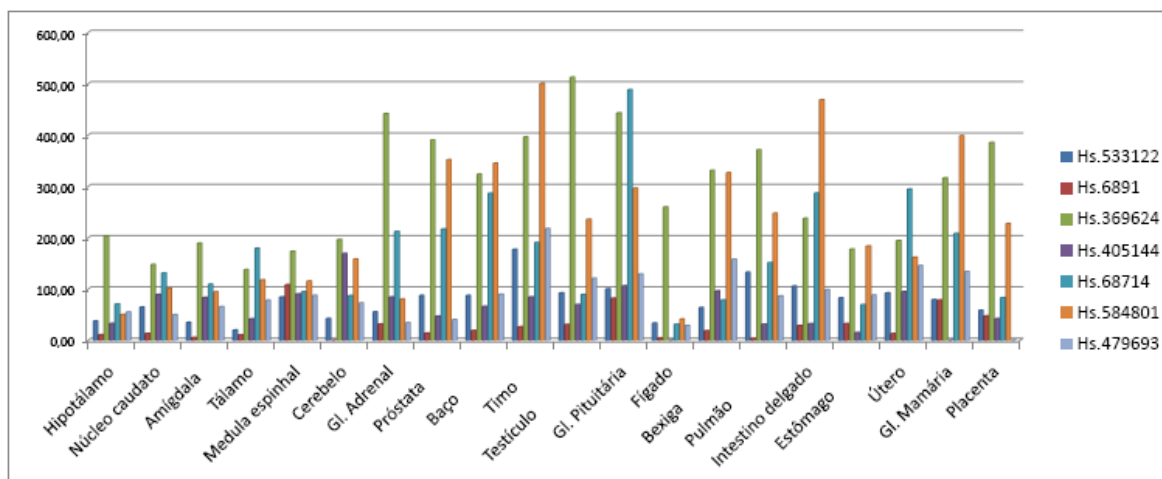


Figura 10 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* SR, em bibliotecas de MPSS de humanos.

Na Figura 11 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* hnRNP (humanos) foram os dos UniGene *clusters*: Hs.2853 (Glândula adrenal), Hs.571177 (Testículo) e Hs.487774 (Glândula mamária).

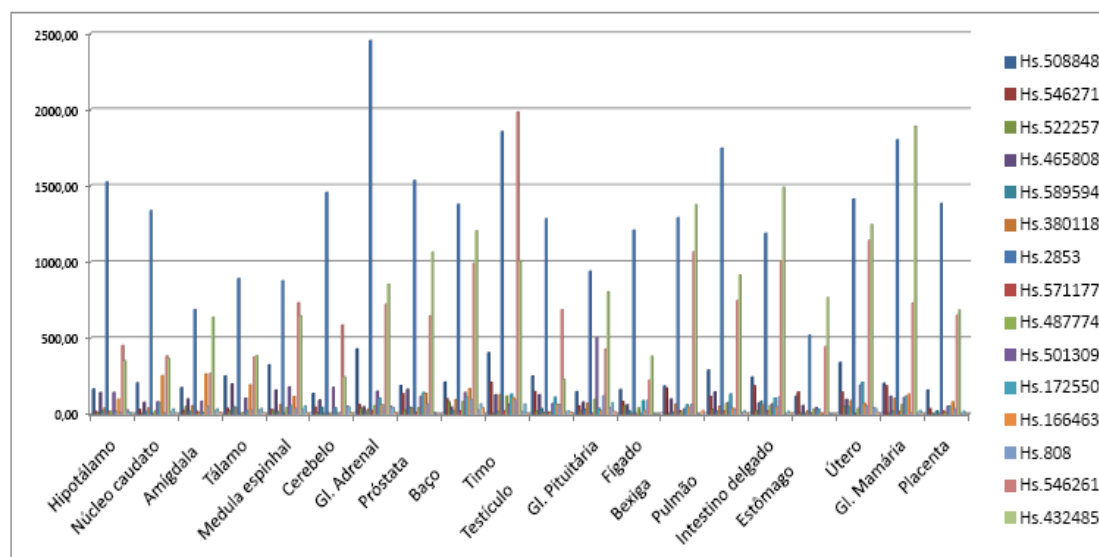


Figura 11 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* hnRNP, em bibliotecas de MPSS de humanos.

Na Figura 12 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* Sm (humanos) foram os dos UniGene *clusters*: Hs.564847 (Hipotálamo), Hs.516076 (Testículo) e Hs.455311 (Intestino Delgado).

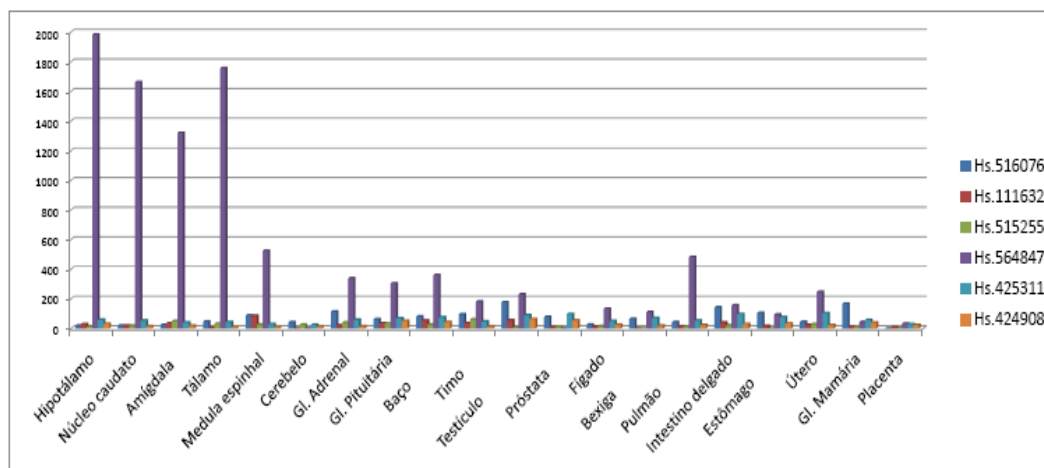


Figura 12 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* Sm, em bibliotecas de MPSS de humanos.

Verificou-se nas Figuras 9-12 que os tecidos timo, testículo, glândula mamária e sistema nervoso, possuem os três primeiros valores de expressão gênica dos grupamentos de fatores de *splicing* snRNP, SR, hnRNP e Sm. Correlacionou-se tal fato verificado com o suposto de que tais tecidos são preferenciais aos acometimentos de eventos de *splicing* alternativos, e podemos assim sugerir uma funcional e importante regulação destes fatores de *splicing* nos respectivos tecidos.

Na Figura 13 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* snRNP (camundongos machos) foram os dos UniGene *clusters*: Mm.3757 (Hipotálamo), Mm.34562 (Timo) e Mm.308514 (Testículo).

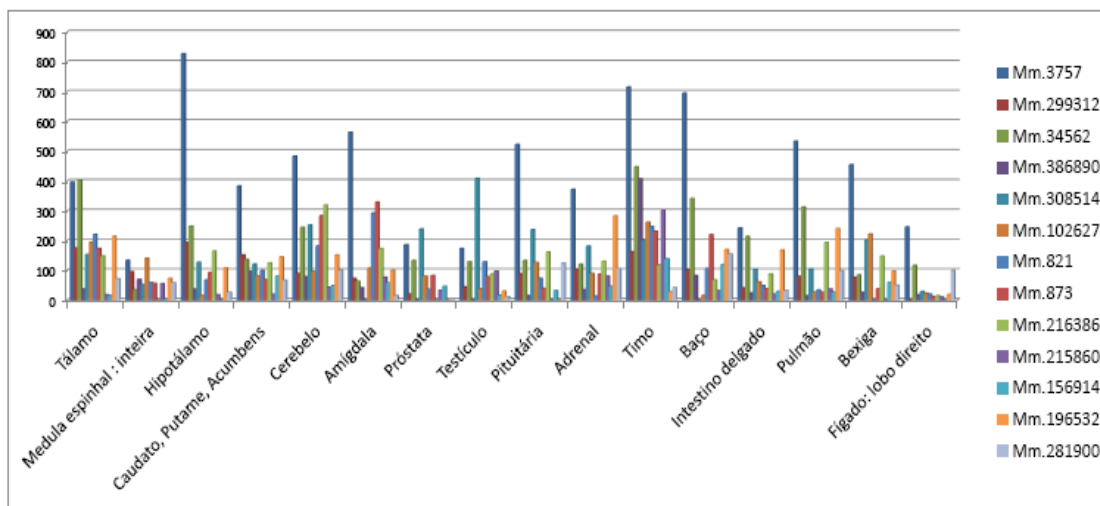


Figura 13 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* snRNP, em bibliotecas de MPSS de camundongos (macho).

Na Figura 14 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* SR (camundongos machos) foram os dos UniGene *clusters*: Mm.43331 (Cerebelo), Mm.5222 (Hipotálamo) e Mm.287826 (Timo).

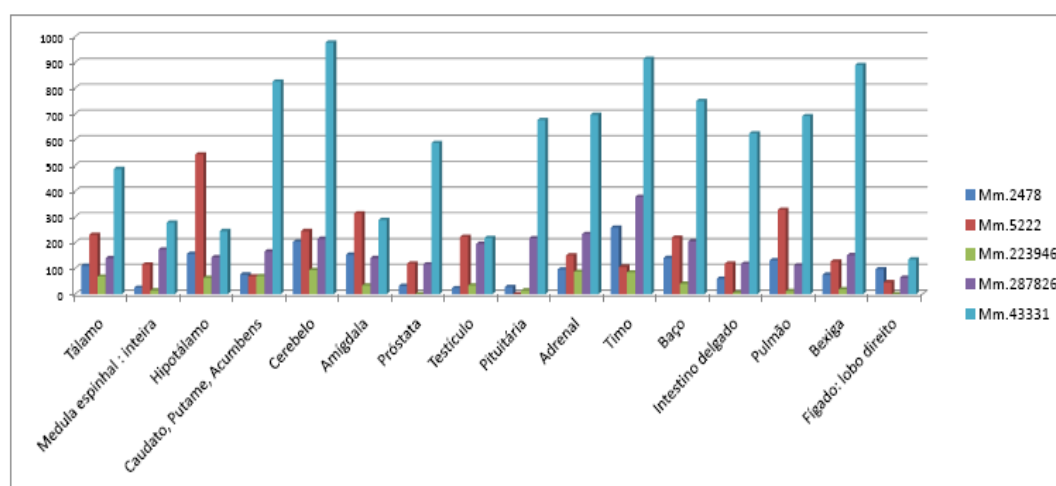


Figura 14 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* SR, em bibliotecas de MPSS de camundongos (macho).

Na Figura 15 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* hnRNP (camundongos machos) foram os dos UniGene *clusters*: Mm.9043 (Timo), Mm.331640 (Baço) e Mm.286408 (Timo).

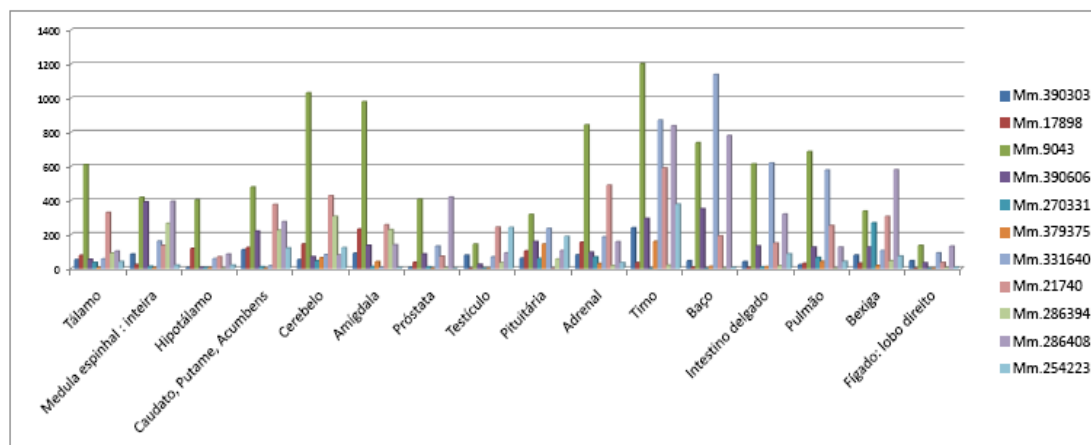


Figura 15 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* hnRNP, em bibliotecas de MPSS de camundongos (macho).

Na Figura 16 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* Sm (camundongos machos) foram os dos UniGene *clusters*: Mm.248188 (Testículo), Mm.276802 (Timo) e Mm.45683 (Testículo).

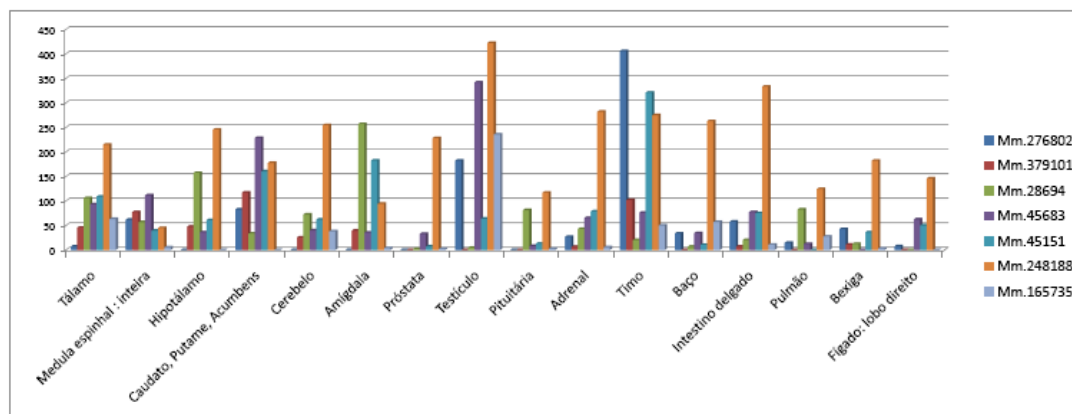


Figura 16 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* Sm, em bibliotecas de MPSS de camundongos (macho).

Na Figura 17 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* snRNP (camundongos fêmeas) foram os dos UniGene *clusters*: Mm.34562 (Pulmão), Mm.3757 (Glândula adrenal) e Mm.216386 (Timo).

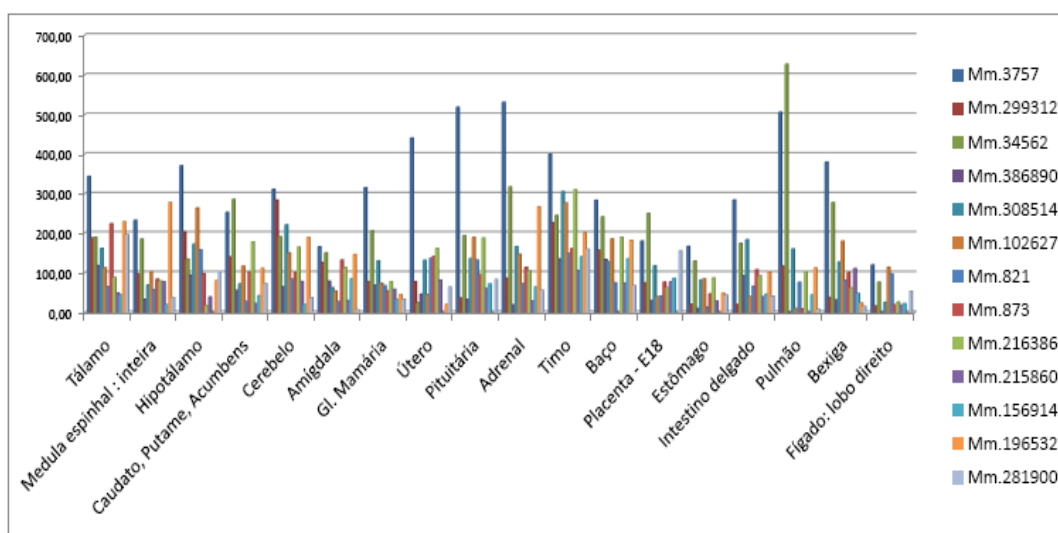


Figura 17 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* snRNP, em bibliotecas de MPSS de camundongos (fêmea).

Na Figura 18 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* SR (camundongos fêmeas) foram os

dos UniGene *clusters*: Mm.43331 (Glândula pituitária), Mm.287826 (Útero) e Mm.5222 (Cerebelo).

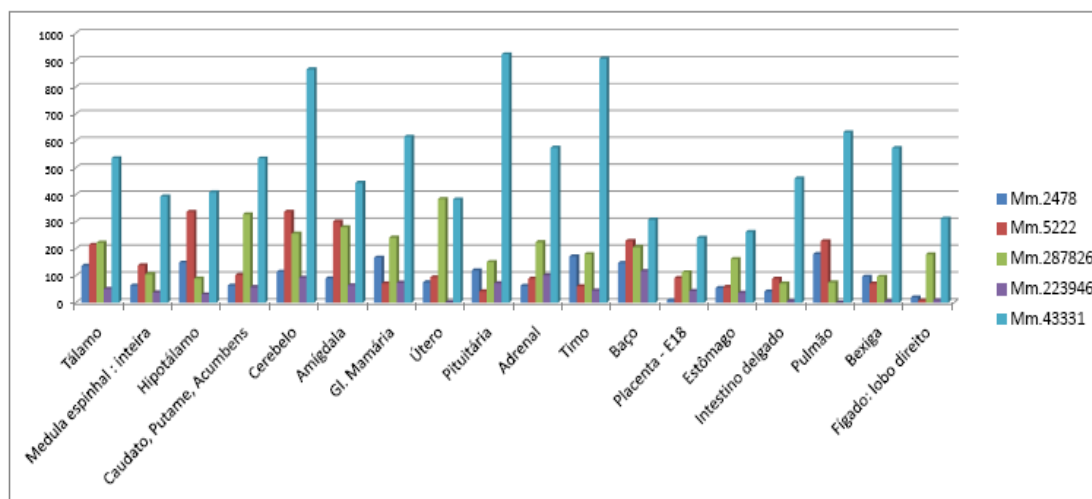


Figura 18 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* SR, em bibliotecas de MPSS de camundongos (fêmea).

Na Figura 19 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* hnRNP (camundongos fêmeas) foram os dos UniGene *clusters*: Mm.286408 (Baço), Mm.9043 (Timo) e Mm.317706 (Timo).

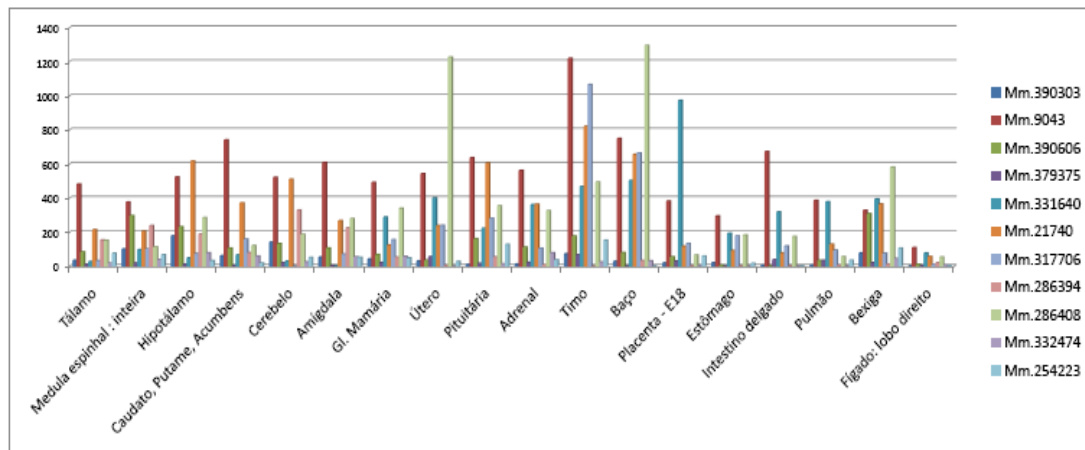


Figura 19 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* hnRNP, em bibliotecas de MPSS de camundongos (fêmea).

Na Figura 20 verificou-se que os três primeiros maiores valores de expressões gênicas do grupamento de fatores de *splicing* Sm (camundongos fêmeas) foram os dos UniGene *clusters*: Mm.248188 (Intestino Delgado), Mm.45151 (Amígdala) e Mm.45683 (Baço).

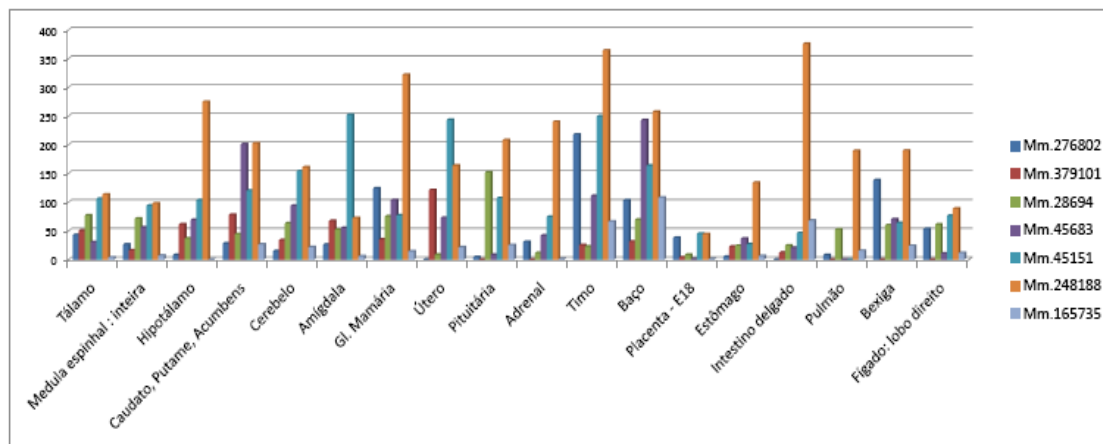


Figura 20 - Gráfico da distribuição dos valores de expressão gênica dos fatores de *splicing* Sm, em bibliotecas de MPSS de camundongos (fêmea).

Verificou-se nas Figuras 13-20 que os tecidos timo, baço, testículo, e sistema nervoso, possuem os três primeiros valores de expressão gênica dos grupamentos de fatores de *splicing* snRNP, SR, hnRNP e Sm.

Visto que os tecidos do sistema nervoso e sexo específicos possuem altos valores de expressão gênica em humanos e camundongos, e com o suposto de que tais tecidos são preferenciais aos acometimentos de eventos de *splicing* alternativos, podemos retomar a sugestão de uma funcional e importante regulação dos fatores de *splicing* nos respectivos tecidos.

4.2 EVENTOS DE *SPLICING* ALTERNATIVOS EM HUMANOS E CAMUNDONGOS

Foram investigadas as ocorrências de eventos de *splicing* alternativos nas distintas espécies. Os critérios conceituais adotados para os padrões de modos alternativos de *splicing* foram os genes que apresentem: Uso alternativo de éxon; Sítios doadores e ou aceptores alternativos; Retenção de íntron.

Foram investigados os tecidos com os maiores números de genes contendo eventos de *splicing* alternativos ocorrendo em todas as bibliotecas de ambas as espécies. Apresenta-se nas Tabelas 1 e 2 os resultados dos 15 tecidos com os maiores níveis de eventos de *splicing* alternativos teciduais, ordenados pelos números de genes indicados nos eventos.

Tabela 1 - Os 15 tecidos com os maiores números de genes contendo eventos de *splicing* alternativos em humanos.

Número de genes contendo eventos de <i>splicing</i> alternativos nas respectivas bibliotecas de humanos					
Uso alternativo de éxon		Sítios doadores e/ou aceptores alternativos		Retenção de íntron	
4410	Sistema nervoso	8598	Sistema nervoso	2729	Sistema nervoso
2112	Testículo	3408	Fígado	1544	Fígado
1589	Placenta	3360	Pulmão	1039	Pulmão
1534	Pulmão	3243	Placenta	1039	Placenta
1233	Olho	3135	Testículo	833	Testículo
1126	Pele	2862	Pele	737	Pele
1115	Útero	2362	Útero	702	Olho
1096	Fígado	2248	Olho	619	Útero
1023	Rim	2115	Rim	521	Rim
668	Cólon	1686	Cólon	448	Pâncreas
599	Pâncreas	1440	Pâncreas	441	Gl. adrenal
586	Próstata	1326	Próstata	370	Cólon
543	Estômago	1084	Estômago	338	Estômago
487	Gl. mamária	1083	Gl. mamária	324	Ovário
468	Músculo estriado	1065	Ovário	310	Próstata

Tabela 2 - Os 15 tecidos com os maiores números de genes contendo eventos de *splicing* alternativos em camundongos.

Número de genes contendo eventos de <i>splicing</i> alternativos nas respectivas bibliotecas de camundongos					
Uso alternativo de éxon		Sítios doadores e/ou aceptores alternativos		Retenção de íntron	
5507	Sistema nervoso	8344	Gl. mamária	2256	Sistema nervoso
4456	Gl. mamária	8207	Sistema nervoso	1940	Gl. mamária
3105	Rim	3806	Rim	551	Olho
1883	Fígado	3032	Fígado	483	Rim
1576	Testículo	2169	Testículo	406	Fígado
1282	Olho	1712	Olho	392	Testículo
867	Baço	1094	Timo	292	Pulmão
766	Timo	1083	Cabeça	270	Timo
739	Cabeça	1013	Pulmão	249	Cabeça
594	Próstata	1003	Pâncreas	208	Pâncreas
546	Pulmão	975	Baço	201	Baço
436	Placenta	831	Cólon	200	Próstata
356	Pâncreas	775	Próstata	174	Cólon
294	Coração	629	Coração	140	Placenta

4.3 FATORES DE *SPLICING* CONSTITUTIVOS PRESENTES NOS EVENTOS DE *SPLICING* ALTERNATIVOS

Com a investigação dos 44 fatores de *splicing* dos grupamentos constitutivos de humanos (Bibliotecas tumorais e normais) e dos 36 fatores de *splicing* dos mesmos grupamentos em camundongos (Bibliotecas normais), presentes nos eventos de *splicing* alternativos de ambas as espécies, obtivemos como resultado: 41 genes dos fatores de *splicing* de humanos contidos em 4.241 eventos de *splicing* alternativos (Bibliotecas normais e tumorais) (Tabela 3) e 31 genes dos fatores *splicing* de camundongos (Tabela 4) contidos em 850 eventos de *splicing* alternativos. No intuito de ser investigado uma possível correlação entre os 41 fatores de *splicing* de humanos (Tabela 3) e os 31 fatores de *splicing* de camundongos (Tabela 4), aos distintos eventos de *splicing* alternativos, efetuou-se um pareamento associativo destes fatores de *splicing* ortólogos (Tabela 5).

Tabela 3 - Números de eventos de *splicing* alternativos encontrado em humanos, com presença indicativa dos 41 fatores de *splicing* constitutivos.

Genes ortólogos	Humanos		Fatores de <i>splicing</i>	Camundongos
	AS em bibliotecas normais	AS em bibliotecas tumorais		Genes ortólogos
Hs.1063	2	1	snRNP	Mm.308514
Hs.11776	2	5	snRNP	Mm.279872
Hs.151787	39	110	snRNP	Mm.873
Hs.166463	103	104	hnRNP	Mm.2115
Hs.172550	26	80	hnRNP	Mm.265610
Hs.177861	1	0	snRNP	Mm.102627
Hs.181368	24	29	snRNP	Mm.3757
Hs.182255	43	40	snRNP	Mm.299312
Hs.246112	23	33	snRNP	Mm.215860
Hs.280378	28	64	snRNP	Mm.1323
Hs.369624	6	3	SR	Mm.287826
Hs.374973	7	22	snRNP	Mm.30660
Hs.380118	51	221	hnRNP	Mm.28275
Hs.405144	15	9	SR	Mm.358634
Hs.406423	10	56	snRNP	Mm.196532
Hs.424908	28	26	Sm	Mm.25642
Hs.425311	17	27	Sm	Mm.30198
Hs.432485	1	1	hnRNP	Mm.286408
Hs.465808	12	143	hnRNP	Mm.311439
Hs.466775	2	3	snRNP	Mm.386890
Hs.469173	31	45	snRNP	Mm.281900
Hs.479693	49	51	SR	Mm.223946
Hs.487774	83	136	hnRNP	Mm.155896
Hs.501309	92	144	hnRNP	Mm.17898
Hs.508848	124	363	hnRNP	Mm.274690
Hs.515255	3	2	Sm	Mm.248188
Hs.516076	1	1	Sm	Mm.276802
Hs.516539	14	39	hnRNP	Mm.379375
Hs.522257	135	237	hnRNP	Mm.142872
Hs.528763	26	42	snRNP	Mm.821
Hs.533122	50	85	SR	Mm.210352
Hs.546261	65	155	hnRNP	Mm.299367
Hs.546271	77	184	hnRNP	Mm.236513
Hs.564847	122	73	Sm	Mm.274995
Hs.571177	14	38	hnRNP	Mm.260545
Hs.584801	92	115	SR	Mm.21841
Hs.589594	2	3	hnRNP	Mm.9043
Hs.68714	19	24	SR	Mm.45645
Hs.6891	7	27	SR	Mm.24042
Hs.808	14	39	hnRNP	Mm.317706
Hs.96996	1	0	hnRNP	Mm.390606

Legenda: Listagem dos seus respectivos ortólogos em camundongos; *Splicing* alternativo (AS).

Tabela 4 - Números de eventos de *splicing* alternativos encontrado em camundongos, contendo a presença indicativa dos 31 fatores de *splicing* ortólogos constitutivos.

Camundongos			Humanos
Genes ortólogos	AS em bibliotecas normais	Fatores de <i>splicing</i>	Genes ortólogos
Mm.43331	141	SR	Hs.166975
Mm.21740	86	hnRNP	Hs.202166
Mm.196532	72	snRNP	Hs.406423
Mm.165735	46	Sm	Hs.103106
Mm.223946	46	SR	Hs.479693
Mm.45683	46	Sm	Hs.565094
Mm.248188	43	Sm	Hs.515255
Mm.286394	41	hnRNP	Hs.20930
Mm.873	41	snRNP	Hs.151787
Mm.386890	39	snRNP	Hs.466775
Mm.216386	33	snRNP	Hs.467097
Mm.215860	31	snRNP	Hs.246112
Mm.276802	28	Sm	Hs.516076
Mm.821	28	snRNP	Hs.528763
Mm.254223	25	hnRNP	Hs.155218
Mm.379375	23	hnRNP	Hs.516539
Mm.5222	19	SR	Hs.433343
Mm.3757	12	snRNP	Hs.181368
Mm.28694	11	Sm	Hs.190520
Mm.9043	11	hnRNP	Hs.589594
Mm.317706	6	hnRNP	Hs.808
Mm.281900	4	snRNP	Hs.469173
Mm.299312	4	snRNP	Hs.182255
Mm.287826	3	SR	Hs.369624
Mm.102627	2	snRNP	Hs.177861
Mm.2478	2	SR	Hs.469970
Mm.286408	2	hnRNP	Hs.432485
Mm.34562	2	snRNP	Hs.502883
Mm.156914	1	snRNP	Hs.406277
Mm.308514	1	snRNP	Hs.1063
Mm.379101	1	Sm	Hs.512610

Legenda: Listagem dos seus respectivos ortólogos em humanos.

Através desta associação cruzada obteve-se uma correlação entre um conjunto de 18 fatores de *splicing* ortólogos (Tabela 5), presentes comumente em distintos eventos de *splicing* alternativos em ambas as espécies.

Tabela 5 - Os genes ortólogos resultantes, comparados e pareados quanto aos seus índices de presença em distintos eventos de *splicing* alternativos.

Humanos			Camundongos		
Genes ortólogos	AS em bibliotecas normais	AS em bibliotecas tumorais	Genes ortólogos	AS em bibliotecas normais	Fatores de <i>splicing</i>
Hs.522257	135	237	Mm.142872	-	hnRNP
Hs.508848	124	363	Mm.274690	-	hnRNP
Hs.564847	122	73	Mm.274995	-	Sm
Hs.166463	103	104	Mm.2115	-	hnRNP
Hs.501309	92	144	Mm.17898	-	hnRNP
Hs.584801	92	115	Mm.21841	-	SR
Hs.487774	83	136	Mm.155896	-	hnRNP
Hs.546271	77	184	Mm.236513	-	hnRNP
Hs.546261	65	155	Mm.299367	-	hnRNP
Hs.380118	51	221	Mm.28275	-	hnRNP
Hs.533122	50	85	Mm.210352	-	SR
Hs.479693	49	51	Mm.223946	46	SR
Hs.182255	43	40	Mm.299312	4	snRNP
Hs.151787	39	110	Mm.873	41	snRNP
Hs.469173	31	45	Mm.281900	4	snRNP
Hs.280378	28	64	Mm.1323	-	snRNP
Hs.424908	28	26	Mm.25642	-	Sm
Hs.172550	26	80	Mm.265610	-	hnRNP
Hs.528763	26	42	Mm.821	28	snRNP
Hs.181368	24	29	Mm.3757	12	snRNP
Hs.246112	23	33	Mm.215860	31	snRNP
Hs.68714	19	24	Mm.45645	-	SR
Hs.425311	17	27	Mm.30198	-	Sm
Hs.405144	15	9	Mm.358634	-	SR
Hs.516539	14	39	Mm.379375	23	hnRNP
Hs.808	14	39	Mm.317706	6	hnRNP
Hs.571177	14	38	Mm.260545	-	hnRNP
Hs.465808	12	143	Mm.311439	-	hnRNP
Hs.406423	10	56	Mm.196532	72	snRNP
Hs.6891	7	27	Mm.24042	-	SR
Hs.374973	7	22	Mm.30660	-	snRNP
Hs.369624	6	3	Mm.287826	3	SR
Hs.515255	3	2	Mm.248188	43	Sm
Hs.11776	2	5	Mm.279872	-	snRNP
Hs.589594	2	3	Mm.9043	11	hnRNP
Hs.466775	2	3	Mm.386890	39	snRNP
Hs.1063	2	1	Mm.308514	1	snRNP
Hs.432485	1	1	Mm.286408	2	hnRNP
Hs.516076	1	1	Mm.276802	28	Sm
Hs.96996	1	0	Mm.390606	-	hnRNP
Hs.177861	1	0	Mm.102627	2	snRNP

Tabela 6 - Listagem dos genes presentes preferencialmente em transcritos variantes tumorais (valores maiores que o dobro).

Genes	Fatores de <i>Splicing</i>	Eventos de AS em Bibliotecas Normais	Eventos de AS em Bibliotecas Tumorais
Hs.172550	hnRNP	26	80
Hs.380118	hnRNP	51	221
Hs.465808	hnRNP	12	143
Hs.487774	hnRNP	83	136
Hs.508848	hnRNP	124	363
Hs.516539	hnRNP	14	39
Hs.522257	hnRNP	135	237
Hs.546261	hnRNP	65	155
Hs.546271	hnRNP	77	184
Hs.571177	hnRNP	14	38
Hs.808	hnRNP	14	39
Hs.151787	snRNP	39	110
Hs.280378	snRNP	28	64
Hs.374973	snRNP	7	22
Hs.406423	snRNP	10	56
Hs.68714	SR	19	24
Hs.6891	SR	7	27

Frente às investigações comparativas entre os números de eventos de *splicing* alternativos ocorrendo em humanos, nos 41 genes dos fatores de *splicing*, encontrados nas bibliotecas de tecidos normais e tumorais, obteve-se como resultado os seguintes genes presentes principalmente em transcritos variantes tumorais (valores maiores do que o dobro) (Tabela 6). Para o controle nas diferenças numéricas de EST dentre as bibliotecas tumorais e normais, efetuou-se uma simulação dos eventos de *splicing* alternativos *in silico*, e foram então gerados randomicamente níveis de eventos de *splicing* alternativos, presentes nas bibliotecas tumorais e normais de humanos. Tais conjuntos de EST foram investigados quanto à ocorrência de eventos de *splicing* alternativos sob presença dos fatores de *splicing* constitutivos. Evidenciou-se os mesmos padrões de distribuição dos eventos de *splicing* alternativos previamente obtidos.

4.4 EXPRESSÃO GÊNICA DIFERENCIAL DE TODOS OS FATORES DE *SPLICING* HUMANOS PRESENTES EM TRANSCRITOS VARIANTES NORMAIS E TUMORAIS

Dentre o conjunto de 211 genes humanos curados (Anexo 4), com pleno envolvimento em eventos de *splicing* alternativos, foram selecionados todos aqueles descritos como coadjuvantes do spliceossomo, e então confrontados os seus níveis de expressão gênica de MPSS (de mesmo método) com tecidos e/ou linhagens celulares normais e tumorais (Quadro 8), bem como os seus valores diferenciais de expressão gênica em bibliotecas de MPSS (de mesmo método) modificadas *versus* bibliotecas normais (Quadro 8.1).

Quadro 8 - Conjunto de bibliotecas humanas tumorais de MPSS com seus respectivos correspondentes tipos normais.

Bibliotecas de MPSS		
Tumorais	Normais	Método
Câncer de Mama (ER -) <i>cell line</i>	Gl. mamária	<i>Classical</i>
Câncer de Mama (ER +) <i>cell line</i>	Gl. mamária	<i>Classical</i>
Câncer de Mama	HB4A <i>normal sample</i>	<i>Signature</i>
HB4A modificado C5.2; ErbB2 <i>expr</i>	HB4A <i>normal sample</i>	<i>Signature</i>
Melanoma (1)	Células epiteliais	<i>Signature</i>
Melanoma <i>Biopsies</i> (2)	Células epiteliais	<i>Signature</i>
Melanoma (1)	Melanócito	<i>Signature</i>
Melanoma <i>Biopsies</i> (2)	Melanócito	<i>Signature</i>

Quadro 8.1 - Conjunto de bibliotecas humanas modificadas de MPSS com seus respectivos correspondentes tipos normais.

Bibliotecas de MPSS		
Modificada	Normal	Método
<i>Colon1; p53 -/- normal</i>	<i>Colon3; p53 +/+ normal</i>	<i>Signature</i>
<i>Colon2; p53 -/- anaerobic</i>	<i>Colon3; p53 +/+ normal</i>	<i>Signature</i>
<i>Colon4; p53 +/+ anaerobic</i>	<i>Colon3; p53 +/+ normal</i>	<i>Signature</i>

O processo de busca por um número total de fatores de *splicing* resultou em 124 fatores de *splicing* (Anexo 5), contendo 104 fatores de *splicing* humanos com expressão gênica superior (mais do que o dobro) em bibliotecas tumorais de MPSS (de mesmo método). Os 124 fatores de *splicing* selecionados foram investigados quanto a sua presença em transcritos variantes normais e/ou tumorais dos tecidos: Glândula mamária, Pulmão, Fígado, Próstata, Sistema nervoso, Pele e Cólon (Quadro 9). Os respectivos valores encontrados no Quadro 9 são detalhados no Anexo 8.

Faz-se necessário reiterar que o critério de escolha de expressão gênica superior ser o valor maior do que o dobro, não é um parâmetro considerado como muito restrigente, entretanto o escolhemos porque analisamos manualmente na literatura científica correlata cada um dos 104 fatores de *splicing*. Efetuamos o processo com a busca de fatores de *splicing* com expressão gênica superior (mais do que 3 vezes) resultando em 88 fatores de *splicing* (Anexo 6), e com expressão superior (mais do que 5 vezes) resultando em 72 fatores de *splicing* (Anexo 7).

Quadro 9 – Números de fatores de *splicing* humanos com presença indicativa em padrões de eventos de *splicing* alternativos.

	Conjunto de <i>clusters</i> gênicos dos fatores de <i>splicing</i> presentes em eventos de <i>splicing</i> alternativos					
	Bibliotecas Normais			Bibliotecas Tumorais		
Glândula mamária	IR	ES	ASS	IR	ES	ASS
			2	7	3	10
Pulmão	IR	ES	ASS	IR	ES	ASS
	14	5	19	7	7	11
Fígado	IR	ES	ASS	IR	ES	ASS
	5	3	13	13	3	20
Próstata	IR	ES	ASS	IR	ES	ASS
	1	5	9	5	4	9
Sistema nervoso	IR	ES	ASS	IR	ES	ASS
	8	14	38	11	14	23
Pele	IR	ES	ASS	IR	ES	ASS
	3	6	11	8	13	23
Cólon	IR	ES	ASS	IR	ES	ASS
		1	2	5	8	21

IR – Retenção de íntron; ES – Uso alternativo do Éxon; ASS – Sítios doadores e/ou aceptores alternativos de *splicing*.

Através da discriminação dos fatores de *splicing* presentes nos padrões de eventos de *splicing* alternativos, em tipos teciduais normais e tumorais (Quadro 9) (Anexo 8), foram evidenciados altos números de transcritos variantes contendo sítios doadores e/ou aceptores alternativos de *splicing*.

Quadro 10 - Fatores de *splicing* humanos encontrados apenas em transcritos variantes tumorais.

Fatores de <i>splicing</i>
Hs.151787 - SNRP116
Hs.155218 - HNRPUL1
Hs.166463 - HNRNPU
Hs.18192 - SRRM1
Hs.20225 - TFIP11
Hs.311609 - DDX39
Hs.356549 - SNRPD3
Hs.374973 - PRPF4
Hs.421576 - MAGOH
Hs.425311 - LSM1
Hs.444520 - DHX35
Hs.465498 - U5-15KD
Hs.469173 - SAD1
Hs.469970 - SFRS4
Hs.485060 - DHX16
Hs.502829 - SF1
Hs.502883 - SART1
Hs.512610 - LSM7
Hs.512661 - ISY1
Hs.516539 - HNRPA3
Hs.519347 - SFRS12
Hs.533245 - DDX46
Hs.808 - HNRPF

O Quadro 10 lista um subconjunto de 23 fatores de *splicing* encontrados apenas em transcritos variantes tumorais (Quadro 9). Dentre os 23 fatores de *splicing* listados no Quadro 10, estão com expressão gênica superior em bibliotecas tumorais de MPSS os 19 genes: SNRP116, HNRPUL1, HNRNPU, SRRM1, TFIP11, DDX39, SNRPD3, PRPF4, MAGOH, SAD1, LSM1, SFRS4, DHX16, SF1, SART1, HNRPA3, SFRS12, DDX46, HNRPF.

Os genes SNRN116, HNRPU, HNRPUL1, SRRM1, SNRPD3, LSM1, SFRS4, SART1, HNRPA3, SFRS12 e HNRPF (Quadro 10), são citados em referências científicas correlatas pelos seus envolvimento na tumorigênese (Quadro 11).

Estão presentes nos transcritos variantes tumorais de Pele, Glândula mamária e Cólon, e com expressão gênica superior em respectivas bibliotecas tumorais de MPSS, os seguintes genes: HNRPU, HNRPA3, SNRP116, HNRPUL1, DDX39 (Melanoma); HNRPA3, SART1, HNRPF (Câncer de Mama); SNRP116 (Cólon) (Quadro 10).

Quadro 11 - Fatores de *splicing* citados em referências científicas correlatas pelo seu envolvimento na tumorigênese.

Fatores de <i>splicing</i> envolvidos em processos tumorigênicos		
Fatores de <i>splicing</i> encontrados apenas em transcritos variantes tumorais.	Hs.808 - HNRPF	BALASUBRAMANI et al. (2006)
	Hs.151787 - SNRP116	KUBO et al. (2002); KIM et al. (2003)
	Hs.155218 - HNRPUL1	BARRAL et al. (2005)
	Hs.166463 - HNRNPU	YUGAMI et al. (2007); KATERINAKI et al. (2003); SPRAGGON et al. (2007).
	Hs.18192 - SRRM1	CHENG e SHARP (2006); PATRAWALA et al. (2006)
	Hs.356549 - SNRPD3	SCHENKEL et al. (2002); MATHUR e SAMUELS (2007)
	Hs.425311 - LSM1	SCHWEINFEST et al. (1997); FRASER et al. (2005)
	Hs.469970 - SFRS4	WATERMANN et al. (2006)
	Hs.502883 - SART1	FAUSTINO e COOPER (2003); VILLA et al. (2002)
	Hs.516539 - HNRPA3	HE et al. (2005)
Fatores de <i>splicing</i> encontrados em transcritos variantes normais e tumorais.	Hs.519347 - SFRS12	PETTIGREW et al. (2005)
	Hs.202166 - HNRPH1	MARKOVTSOV et al. (2000); GRABOWSKI (2004)
	Hs.355934 - SFPQ	MATHUR e SAMUELS (2007)
	Hs.480073 - HNRPD	AUDIC e HARTLEY (2004)
	Hs.487774 - HNRNPA2B1	ZECH et al. (2006)
	Hs.498548 - SPF45	SAMPATH et al. (2003)
	Hs.516076 - SNRPG	CONTE et al. (2002)
	Hs.528007 - U2AF2	MAEDA et al. (1999)
	Hs.546261 - HNRNPA1	KARNI et al. (2007)
	Hs.546271 - HNRPE2	ROYCHOUDHURY et al. (2007)
	Hs.570079 - DHX38	WEI-DONG et al. (2004)
	Hs.589594 - HNRNPL	ITO et al. (2001)
	Hs.6891 - SFRS6	KARNI et al. (2007)
	Hs.68714 - SF2/ASF	KARNI et al. (2007)
	Hs.432485 - RPL36A	KIM et al. (2004)
Fatores de <i>splicing</i> ausentes nos transcritos variantes tumorais.	Hs.9822 - XAB2	YUGAMI et al. (2007); WAN et al. (2004)
	Hs.443861 - SRPK1	HAYES et al. (2007)
	Hs.1063 - SNRPC	SHETTY (2005)

Oito genes do Quadro 10 estão com expressão gênica superior em bibliotecas tumorais e sob eventos de *splicing* alternativos apenas em tecidos tumorais (Quadro 9), mas sem menção na literatura científica como estando envolvidos em processos tumorigênicos: TFIP11, DDX39, PRPF4, MAGOH, SAD1, DHX16, SF1 e DDX46.

O Quadro 12 lista um subconjunto de 47 fatores de *splicing* encontrados nos transcritos variantes normais e tumorais do Quadro 9, e com expressão gênica superior em bibliotecas tumorais de MPSS.

Dentre os fatores de *splicing* listados no Quadro 12, os 15 genes HNRPH1, SFPQ, RPL36A, HNRPD, HNRNPA2B1, SPF45, SNRPG, U2AF2, HNRNPA1, HNRPE2, DHX38, HNRNPL, SF2/ASF, SRp55 e XAB2 são citados em referências científicas correlatas pelos seus envolvimento na tumorigênese.

No Quadro 12, há um subconjunto de 32 genes com expressão gênica superior em bibliotecas tumorais e sob presença de eventos de *splicing* alternativos em tecidos tumorais e normais (Quadro 9), mas sem menção na literatura científica como estando envolvidos em processos tumorigênicos: LSM2, DDX23, RALY, PTBP1, PRPF8, ASCC3L1, PLRG1, BAT1, DDX5, RBM39, PRPF6, HSPC225, SFRS3, SF3B2, CLK1, HNRPM, SNRPA, SFRS16, SFRS11, DDX41, PRPF19, HNRNPC, SF3B3, PRPF31, SF3B4, PUF60, HNRPDL, DDX17, SFRS10, RY1, SNW1 e SFRS2.

No Quadro 13 lista-se um subconjunto de 15 fatores de *splicing* ausentes em transcritos variantes tumorais (Quadro 9) e com expressão gênica superior em bibliotecas tumorais de MPSS. Dentre os fatores de *splicing* listados no Quadro 13, os genes SRPK1 e SNRPC são citados em referências científicas correlatas pelo seu envolvimento na tumorigênese.

Quadro 12 - Fatores de *splicing* humanos encontrados em transcritos variantes normais e tumorais, e com expressão gênica superior em bibliotecas tumorais de MPSS.

Fatores de <i>splicing</i>
Hs.103106 - LSM2
Hs.130098 - DDX23
Hs.136947 - RALY
Hs.172550 - PTBP1
Hs.181368 - PRPF8
Hs.202166 - HNRPH1
Hs.246112 - ASCC3L1
Hs.249996 - PLRG1
Hs.254042 - BAT1
Hs.279806 - DDX5
Hs.282901 - RBM39
Hs.31334 - PRPF6
Hs.33104 - HSPC225
Hs.355934 - SFPQ
Hs.405144 - SFRS3
Hs.406423 - SF3B2
Hs.432485 - RPL36A
Hs.433732 - CLK1
Hs.465808 - HNRPM
Hs.466775 - SNRPA
Hs.466917 - SFRS16
Hs.479693 - SFRS11
Hs.480073 - HNRPD
Hs.484288 - DDX41
Hs.487774 - HNRNPA2B1
Hs.498548 - SPF45
Hs.502705 - PRPF19
Hs.508848 - HNRNPC
Hs.514435 - SF3B3
Hs.515598 - PRPF31
Hs.516076 - SNRPG
Hs.516160 - SF3B4
Hs.521924 - PUF60
Hs.527105 - HNRPDL
Hs.528007 - U2AF2
Hs.528305 - DDX17
Hs.533122 - SFRS10
Hs.546261 - HNRNPA1
Hs.546271 - HNRPE2
Hs.54649 - RY1
Hs.546550 - SNW1
Hs.570079 - DHX38
Hs.584801 - SFRS2
Hs.589594 - HNRNPL
Hs.68714 - SF2/ASF
Hs.6891 - SRp55
Hs.9822 - XAB2

Quadro 13 - Fatores de *splicing* humanos ausentes nos transcritos variantes tumorais e com expressão gênica superior em bibliotecas tumorais de MPSS.

Fatores de <i>splicing</i>
Hs.1063 - SNRPC
Hs.11776 - PRPF3
Hs.177861 - SF3B14
Hs.182255 - SNU13
Hs.20013 - SYF2
Hs.274531 - SKIV2L2
Hs.27693 - PPIL1
Hs.280378 - SNRPB2
Hs.365116 - U2AF1
Hs.443861 - SRPK1
Hs.485471 - CDC5L
Hs.515255 - LSM4
Hs.528763 - SNRPA1
Hs.571177 - HNRPQ1
Hs.73986 - CLK2

5 DISCUSSÃO

5.1 PADRÃO DIFERENCIAL DE COEXPRESSÃO GÊNICA DOS FATORES DE *SPLICING* NOS TECIDOS DO SISTEMA NERVOSO E SEXO ESPECÍFICO

O *splicing* alternativo é um processo que gera diferentes mRNA. Estes geralmente codificam diversos produtos protéicos a partir de um gene. Isto então aumenta drasticamente a capacidade codificante dos genes. O *splicing* alternativo é usualmente regulado de modo tecido-preferencial ou estágio-específico durante o processo de desenvolvimento biológico. Múltiplas interações são requeridas e importantes para o estabelecimento do complexo que consigna o pré-mRNA para o *splicing*. Tal complexo controle é finamente regulado pelo spliceossomo.

A primeira e segunda parte deste estudo inclui a caracterização dos grupamentos de todos os fatores de *splicing* e a sua relação tecido-preferencial, focando-se nos grupos de fatores de *splicing*: snRNP, hnRNP, SR e Sm. A seleção destes grupamentos deve-se ao seu papel constitutivo aos eventos de *splicing*, sendo estes requeridos em algum momento do processo. Estes grupamentos foram então referenciais nas inferências intervenientes sobre os prováveis padrões de coexpressão encontrados entre os fatores de *splicing* ditos constitutivos. Nesta etapa o principal objetivo específico fora o de investigar apenas a expressão gênica dos fatores de *splicing* nas bibliotecas de MPSS de origem tecidual e celular normais.

O primeiro momento caracteriza-se por ser altamente restritivo quanto a análise à origem das proteínas referenciadas em BARBOSA-MORAIS et al. (2006). Permaneceram sendo investigadas apenas as proteínas com as devidas anotações gênicas que as caracterizem como fatores de *splicing*. A busca pelas *tags* virtuais e poliadenilação alternativa foram outros fatores restritivos limitantes quanto à qualidade das seqüências de mRNA e HTC geradas. O grupo final de fatores de *splicing* de ambas as espécies são referências confiáveis a qualquer estudo correlato. Não há precedentes de referências científicas que apresentem um conjunto de fatores de *splicing* de camundongos, submetidos a tais critérios de restrição.

Os ensaios comparativos dos genomas de *Homo sapiens* e *Mus musculus*, citados por WATERSTON et al. (2002), inferem que o genoma deste é aproximadamente 14% menor, e ambos apresentam cerca de 90% de correspondentes regiões com sintenia, refletindo segmentos nos quais contêm aproximadamente 40% dos nucleotídeos do genoma de *Homo sapiens* sendo alinhados com o genoma de *Mus musculus*, representando os possíveis ortólogos remanescentes de um ancestral comum. A proporção dos genes de *Mus musculus* com um único e identificado ortólogo no genoma de *Homo sapiens* é de aproximadamente 80%. A relação de genes de *Mus musculus* sem qualquer homólogo detectado em genoma de *Homo sapiens*, e vice-verso, é de aproximadamente 1%. Indica-se como um caráter qualitativo a sua alta homologia, e assim relevante aos estudos comparativos efetuados nas respectivas espécies.

As análises comparativas de busca a um padrão diferencial de coexpressão gênica entre os fatores de *splicing* resultam numa prevalente presença de *outliers* nos tecidos do sistema nervoso e sexo específico. A importância da escolha dos *outliers*

deve-se ao seu distinto padrão de coexpressão como abordado em outras similares abordagens de análise em estudos de expressão gênica (YEO et al. 2004). Fazemos aqui necessária menção às revisões científicas sobre o *splicing* alternativo (ULE et al. 2003; BLACK 2003), nas quais citam preferencialmente o sistema nervoso como sendo o grupo amplamente envolvido nos maiores índices de eventos de *splicing* alternativos. O encontro de *outliers* no grupo sexo específico indica uma eventual correlação com o suposto de que as etapas de determinação e diferenciação das gônadas, em testículos ou em ovários, e a diferenciação dos genitais externos masculinos ou femininos, envolvam uma expressão específica de uma cascata de genes. Os genes dos fatores de *splicing* de bibliotecas do grupo sistema nervoso e sexo específico, com seus respectivos padrões de coexpressão, são abordados neste estudo como favoráveis à interpretação de uma funcional correlação entre a expressão diferencial destes e os seus altos níveis de incidência dos eventos de *splicing* alternativos, em ambas as espécies.

Uma oportuna análise sobre os diferenciais níveis de coexpressão gênica encontrados entre os fatores de *splicing* ditos constitutivos, entre tecidos normais e tumorais, resultaram em esperadas correlações negativas, sugere-se então a existência de distintas preferências tipos tecidual normal e tumoral.

5.2 A REGULAÇÃO DOS EVENTOS DE *SPLICING* ALTERNATIVOS SOB OS FATORES DE *SPLICING*

Estudos correlatos sobre importância dos eventos de *splicing* atuantes nos fatores de *splicing* sugerem que os modos alternativos de *splicing* possam manter

uma homeostase dos fatores de *splicing*, bem como permitir a sua expressão gênica tecido-preferencial. Tais observações associam os modos alternativos de *splicing* a um processo regulatório que afetaria a expressão gênica de fatores de *splicing* (LAREAU et al. 2007).

Investigando-se a presença dos fatores de *splicing* ditos constitutivos em transcritos variantes normais e tumorais de humanos, verificou-se que a maioria (93%) destes fatores de *splicing* de humanos, e 86% dos fatores ortólogos em camundongos, sofrem *splicing* alternativo.

Através de uma análise comparativa entre a presença dos fatores de *splicing* em transcritos variantes normais de humanos e camundongos obteve-se um conjunto de 18 fatores de *splicing* encontrados em ambas as espécies. Do montante total de fatores de *splicing* ortólogos preservados em transcritos variantes normais, 10 são do grupamento snRNP, 2 do grupamento SR, 4 do grupamento hnRNP e 2 do grupamento Sm. O número final de genes correlacionados e preservados nas distintas espécies é relevante nas possíveis inferências às suas funcionalidades regulatórias tecido-preferencial.

5.3 OS FATORES DE *SPLICING* PRESENTES PREFERENCIALMENTE NOS TRANSCRITOS VARIANTES TUMORAIS APRESENTAM VALORES DE EXPRESSÃO GÊNICA COMPARATIVAMENTE SUPERIORES AOS NORMAIS

Embora os modos aberrantes de mRNA sejam degradados pelo mecanismo *nonsense-mediated mRNA decay* (CARTEGNI et al. 2002), algumas alterações

disfuncionais do processo de *splicing* em células cancerígenas resultam na produção de novas isoformas de mRNA que acarretam na aquisição de outras propriedades carcinogênicas. Diversas revisões fornecem listas de genes com alterações nos modos de *splicing* relacionados ao câncer (KALNINA et al. 2005; SCHWERK e SCHULZE-OSTHOFF 2005; VENABLES 2006). Alterações na concentração, localização, composição ou atividade dos fatores regulatório *trans*-atuantes, tais como os grupamentos protéicos de hnRNP e SR, podem resultar em modificações no processo de *splicing* (KARNI et al. 2007), somando-se ao exposto que, outras análises computacionais em larga escala aos dados de EST trouxeram como conclusão que alguns fatores de *splicing* são *over*-expressos em células de câncer *versus* normal (KIRSCHBAUM-SLAGER et al. 2004).

Nos resultados apresentados constam outros estudos científicos reportando alguns fatores de *splicing* candidatos, aqui obtidos, com amplo envolvimento na tumorigênese (Quadro 11). Fora possível corroborar *in silico* os dados de validação experimental citados em KARNI et al. (2007), com os proto-oncogenes de consenso (Tabela 9): SF2/ASF (Hs.68714), HNRNPA1 (Hs.546261) e SFRS6 (Hs.6891). O fator de *splicing* ASF/SF2 atua de modo antagônico ao HNRNPA1 na regulação do *splicing* alternativo: as concentrações aumentadas de ASF/SF2 selecionam sítios de *splice* intron-proximal; concentrações aumentadas de HNRNPA1 promovem a seleção dos sítios de *splice* intron-distal (KARNI et al. 2007).

Aqueles fatores de *splicing* candidatos que não foram encontrados em estudos científicos correlatos mantêm-se numa categoria de subgrupo favorável a uma validação experimental que corrobore sua possível expressão gênica diferencial em tecidos tumorais e envolvimento na carcinogênese. Os fatores de *splicing* candidatos,

somam 8 genes com expressão gênica superior em bibliotecas tumorais e presentes apenas em transcritos variantes tumorais (Quadro 9): TFIP11, DDX39, PRPF4, MAGOH, SAD1, DHX16, SF1 e DDX46.

Visto que a concentração aumentada de determinados fatores de *splicing*, especificamente os grupos SR e hnRNP, possam resultar em variantes de *splicing* câncer-associado, podemos acrescentar um subgrupo de 32 novos genes candidatos a uma validação experimental. Tal subgrupo gênico está presente em ambos transcritos variantes tumorais e normais, mas com expressão gênica superior em bibliotecas tumorais de MPSS: LSM2, DDX23, RALY, PTBP1, PRPF8, ASCC3L1, PLRG1, BAT1, DDX5, RBM39, PRPF6, HSPC225, SFRS3, SF3B2, CLK1, HNRPM, SNRPA, SFRS16, SFRS11, DDX41, PRPF19, HNRNPC, SF3B3, PRPF31, SF3B4, PUF60, HNRPDL, DDX17, SFRS10, RY1, SNW1 e SFRS2.

6 CONCLUSÃO

Esta dissertação resulta de um trabalho de investigação em genômica comparativa com a busca inicial e abrangente sobre padrões de expressão dos fatores de *splicing* constitutivos em humanos e camundongos, correlacionados à ocorrência de eventos de *splicing* alternativos.

As funcionalidades dos fatores de *splicing* e dos eventos de *splicing* alternativos foram confrontadas e investigadas em tecidos tumorais *versus* normais de humanos.

Em virtude dos fatos mencionados, as conclusões obtidas neste estudo *in silico* são listadas nos seguintes tópicos:

6.1 CONJUNTO DE FATORES DE *SPLICING* ORTÓLOGOS EM HUMANOS E CAMUNDONGOS

Através da adoção de um método restritivo de investigação *in silico*, sugere-se a presença de um conjunto final de 124 fatores de *splicing* ortólogos em humanos e camundongos.

6.2 PADRÕES DE COEXPRESSÃO DIFERENCIAL DOS FATORES DE *SPLICING* CONSTITUTIVOS

Os genes dos fatores de *splicing* constitutivos expressos nos grupos de bibliotecas de MPSS, sistema nervoso e sexo específico, apresentam padrões de coexpressão gênica negativos, em relação aos demais grupos de bibliotecas. Tais *outliers* são abordados neste estudo como favoráveis à interpretação de uma correlação funcional existente entre a coexpressão diferencial dos fatores de *splicing* constitutivos e uma incidente ocorrência de evento de *splicing* alternativo presente nestes, em ambas as espécies.

6.3 OS TRANSCRITOS VARIANTES E A EXPRESSÃO GÊNICA DOS FATORES DE *SPLICING* EM BIBLIOTECAS DE MPSS NORMAIS VERSUS TUMORAIS DE HUMANOS

Pela observação dos aspectos analisados, concluímos que dentre um conjunto final de 124 fatores de *splicing* humanos:

- 104 fatores de *splicing* possuem expressão gênica superior em bibliotecas tumorais de MPSS;
- 23 fatores de *splicing* encontrados apenas em transcritos variantes tumorais e 19 destes genes se encontram com expressão gênica superior em bibliotecas tumorais de MPSS;

- 47 fatores de *splicing* encontrados em transcritos variantes normais e tumorais, e com expressão gênica superior em bibliotecas tumorais de MPSS;
- 15 fatores de *splicing* ausentes em transcritos variantes tumorais e com expressão gênica superior em bibliotecas tumorais de MPSS.
- Há 28 fatores de *splicing* sendo referenciados por outros autores, pelo seu envolvimento no processo de tumorigênese;
- 40 fatores de *splicing* (sem referência científica correlata) são promissores candidatos a outras investigações experimentais específicas, que corroborem em outros métodos os seus envolvimento na tumorigênese: TFIP11, DDX39, PRPF4, MAGOH, SAD1, DHX16, SF1, DDX46, LSM2, DDX23, RALY, PTBP1, PRPF8, ASCC3L1, PLRG1, BAT1, DDX5, RBM39, PRPF6, HSPC225, SFRS3, SF3B2, CLK1, HNRPM, SNRPA, SFRS16, SFRS11, DDX41, PRPF19, HNRNPC, SF3B3, PRPF31, SF3B4, PUF60, HNRPDL, DDX17, SFRS10, RY1, SNW1 e SFRS2.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Altschul SF. Basic local alignment search tool. **J Mol Biol** 1990; 215:403-10.

Audic Y, Hartley RS. Post-transcriptional regulation in cancer. **Biol Cell** 2004; 96:479-98.

Balasubramani M, Day BW, Schoen RE, Getzenberg RH. Altered expression and localization of creatine kinase B, heterogeneous nuclear ribonucleoprotein F, and high mobility group box 1 protein in the nuclear matrix associated with colon cancer. **Cancer Res** 2006; 66:763-9

Barbosa-Morais NL, Carmo-Fonseca M, Aparício S. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. **Genome Res** 2006; 16:66-77.

Barral PM, Rusch A, Turnell AS et al. The interaction of the hnRNP family member E1B-AP5 with p53. **FEBS Lett** 2005; 579:2752-8.

Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. **Proc Natl Acad Sci USA** 1977; 74:3171-5.

Black DL. Finding splice sites within a wilderness of RNA. **RNA** 1995; 1:763-71.

Black DL. Mechanisms of alternative pre-messenger RNA Splicing. **Annu Rev Biochem** 2003; 72:27-48.

Blencowe BJ. Alternative splicing: New insights from global analyses. **Cell** 2006, 126:37-47.

Boguski DE, Shuler MS. Establishing a human transcript map. **Nature Genet** 1995; 10: 369-71.

Boguski MS, Lowe TM, Tolstoshev CM. dbEST-database for 'expressed sequence tags'. **Nature Genet** 1993; 4:332-3.

Boon K, Osorio EC, Greenhut SF, et al. An anatomy of normal and malignant gene expression. **Proc Natl Acad Sci U S A** 2002; 99:11287-92.

Boukris LA, Liu N, Furuyama S, Bruzik JP. Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. **J Biol Chem** 2004; 279:29647-53.

Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. **Nat Biotechnol** 2000; 18:630-4.

Brow DA. Allosteric cascade of spliceosome activation. **Annu Rev Genet** 2002; 36:333-60.

Burset M, Seledtsov IA, Solovyev VV. SpliceDB: database of canonical and non-canonical mammalian splice sites. **Nucleic Acids Res** 2001; 29:255-9.

Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. **Nat Rev Genet** 2002; 3:285-98.

Chen J, Rattray M. Analysis of tag-position bias in MPSS technology. **BMC Genomics**. 2006 Apr 7;7:77.

Cheng C, Sharp PA. Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. **Mol Cell Biol** 2006; 26:362-70.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. **Cell** 1977; 12:1-8.

[COMPBio Ludwig] Laboratório de Biologia Computacional, Instituto Ludwig de Pesquisa sobre o Câncer. **JNUNES**. Disponível em: URL<http://www.compbio.ludwig.org.br/~jnunes/class_libs_mouse> [2007 Dec 01]

Conte N, Charafe-Jauffret E, Delaval B, et al. Carcinogenesis and translational controls: TACC1 is down-regulated in human cancers and associates with mRNA regulators. **Oncogene** 2002; 21:5619-30.

Doerge RW. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Edited by Gentleman R, Carey V, Huber W, Irizarry R, and Dudoit S. **Biometrics** 2006; 62:1270–1.

Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. **Annu Rev Biochem** 1993; 62:289-321.

Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. **Genes Dev** 2003; 17:419-37.

Fraser MM. CaSm-mediated cellular transformation is associated with altered gene expression and messenger RNA stability. **Cancer Res** 2005; 65:6228-36.

Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ. Detection and evaluation of intron retention events in the human transcriptome. **RNA** 2004; 10:757-65.

Gallego ME, Gattoni R, Stevenin J, Marie J, Expert-Bezancon A. The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancerdependent splicing of the beta-tropomyosin alternative exon 6A. **EMBO J** 1997; 16:1772-84.

Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. **Nat Biotechnol** 2004; 22, 535-46.

Garcia-Blanco MA, Jamison SF, Sharp PA. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. **Genes Dev** 1989; 3:1874-86.

Grabowski PJ. A molecular code for splicing silencing: configurations of guanosine-rich motifs. **Biochem Soc Trans** 2004; 32:924-7.

Graveley BR. Sorting out the complexity of SR protein functions. **RNA** 2000; 6:1197-211.

Hayes GM, Carrigan PE, Miller LJ. Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas. **Cancer Res** 2007; 67:2072-80.

He Y, Brown MA, Rothnagel JA, Saunders NA, Smith R. Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation. **J Cell Sci** 2005; 118:3173-83.

Ito M, Shichijo S, Tsuda N et al. Molecular basis of T cell-mediated recognition of pancreatic cancer cells. **Cancer Res** 2001; 61:2038-46.

Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. **Mol Cell** 2003; 12:5-14.

Kalnina Z, Zayakin P, Silina K, Line A. Alterations of pre-mRNA splicing in cancer. **Genes Chromosomes Cancer** 2005; 42:342-57.

Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. **Nat Struct Mol Biol** 2007; 14:185-93.

Katerinaki E, Evans GS, Lorigan PC, MacNeil S. TNF-alpha increases human melanoma cell invasion and migration in vitro: the role of proteolytic enzymes. **Br J Cancer** 2003; 89:1123-9.

Kelso J. eVOC: a controlled vocabulary for unifying gene expression data. **Genome Res** 2003; 13(6A):1222-30.

Kim B, Bang S, Lee S, et al. Expression profiling and subtype-specific expression of stomach cancer. **Cancer Res** 2003; 63:8248-55.

Kim JH, You KR, Kim IH, Cho BH, Kim CY, Kim DG. Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma. **Hepatology** 2004; 39:129-38.

Kirschbaum-Slager N, Lopes GM, Galante PA, Riggins GJ, de Souza SJ. Splicing factors are differentially expressed in tumors. **Genet Mol Res** 2004; 3:512-20.

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

Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. **Hum Genet** 1992; 90:41-54.

Kubo M, Ihn H, Kuwana M et al. Anti-U5 snRNP antibody as a possible serological marker for scleroderma-polymyositis overlap. **Rheumatology (Oxford)** 2002; 41:531-4.

Ladd AN, Cooper TA. Finding signals that regulate alternative splicing in the post-genomic era. **Genome Biol** 2002; 3:8.

Lander ES, Linton LM, Birren B et al. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. **Nature** 2001; 409: 860-921.

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. **Nature** 2007; 446:926-9.

Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. Are snRNPs involved in splicing? **Nature** 1981; 283:220-4.

Maeda T, Hiranuma H, Jikko A. Differential expression of the splicing regulatory factor genes during two-step chemical transformation in a BALB/3T3-derived cell line, MT-5. **Carcinogenesis** 1999; 20:2341-4.

Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, Black DL. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. **Mol Cell Biol** 2000; 20:7463-79.

Mathur M, Samuels HH. Role of PSF-TFE3 oncoprotein in the development of papillary renal cell carcinomas. **Oncogene** 2007; 26:277-83.

Mayeda A, Krainer AR. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. **Cell** 1992; 68:365-75.

Meyers BC, Tej SS, Vu TH, Haudenschild CD, et al. The Use of MPSS for Whole-Genome Transcriptional Analysis in Arabidopsis. **Genome Res** 2004; 14:1641-53.

Mironov AA, Koonin EV, Roytberg MA, Gelfand MS. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. **Nucleic Acids Res** 1999; 27:2981-9.

Modrek B, Resch A, Grasso C, Lee C. Genome-wide detection of alternative splicing in expressed sequences of human genes. **Nucleic Acids Res** 2001; 29:2850-9.

[NCBI] National Center for Biotechnology Information. **HomoloGene**. Available from: <URL:<http://www.ncbi.nlm.nih.gov/HomoloGene>> [2006a abr 16]

[NCBI] National Center for Biotechnology Information. **UniGene**. Available from: <URL:<http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>> [2006b abr 16]

Nakielnny S, Fischer U, Michael W M, Dreyfuss G. **Annu Rev Neurosci** 1997; 20:269-301.

Okazaki Y, Furuno M, Kasukawa T, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. **Nature** 2002; 420:563-73.

Osório EC, de Souza JE, Zaiats AC, de Oliveira PS, de Souza SJ. pp-Blast: a "pseudo-parallel" Blast. **Braz J Med Biol Res** 2003; 36:463-4.

Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM. Alternative splicing: an emerging topic in molecular and clinical oncology. **Lancet Oncol** 2007; 8:349-57.

Patel AA, Steitz JA. Splicing double: insights from the second spliceosome. **Nat Rev Mol Cell Biol** 2003; 4:960-70.

Patrawala L. Highly purified CD44⁺ prostate cancer cells from xenograft human tumors are enriched in tumorigenic and metastatic progenitor cells. **Oncogene** 2006; 25:1696-708.

Pettigrew C, Wayte N, Lovelock PK, et al. Evolutionary conservation analysis increases the colocalization of predicted exonic splicing enhancers in the BRCA1 gene with missense sequence changes and in-frame deletions, but not polymorphisms. **Breast Cancer Res** 2005; 7:R929-39.

Pruitt KD, Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. **Nucleic Acids Res** 2001; 29:137-40.

Roy M, Xu Q, Lee C. Evidence that public database records for many cancer associated genes reflect a splice form found in tumors and lack normal records for many cancer-associated genes reflect a splice form found in tumors and lack normal splice forms. **Nucleic Acids Res** 2005; 33:5026-33.

Roychoudhury P, Paul RR, Chowdhury R, Chaudhuri K. HnRNP E2 is downregulated in human oral cancer cells and the overexpression of hnRNP E2 induces apoptosis. **Mol Carcinog** 2007; 46:198-207.

Sakabe NJ. ORESTES are enriched in rare exon usage variants affecting the encoded proteins. **C R Biol** 2003; 326:979-85.

Sampath J, Long PR, Shepard RL, et al. Human SPF45, a splicing factor, has limited expression in normal tissues, is overexpressed in many tumors, and can confer a multidrug-resistant phenotype to cells. **Am J Pathol** 2003; 163:1781-90.

Sanford JR, Ellis J, Caceres JF. Multiple roles of arginine/serine-rich splicing factors in RNA processing. **Biochemical Society Transactions** 2005; 33:443-7.

Sanford JR, Longman D, Caceres JF. Multiple roles of the SR protein family in splicing regulation. **Prog Mol Subcell Biol** 2003; 31:33-58.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science** 1995; 270:467-70.

Schenkel H, Hanke S, De Lorenzo C, Schmitt R, Mechler BM. P elements inserted in the vicinity of or within the *Drosophila* snRNP SmD3 gene nested in the first intron of the Ornithine Decarboxylase Antizyme gene affect only the expression of SmD3. **Genetics** 2002; 161:763-72.

Schweinfest CW. CaSm: an Sm-like protein that contributes to the transformed state in cancer cells. **Cancer Res** 1997; 57:2961-5.

Schwerk C, Schulze-Osthoff K. Regulation of apoptosis by alternative pre-mRNA splicing. **Mol Cell** 2005; 19:1-13

Sharp PA. The discovery of split genes and RNA splicing. **Trends Biochem Sci** 2005; 30:279-81.

Shatkin AJ, Manley JL. The ends of the affair: capping and polyadenylation. **Nat Struct Biol** 2000; 7:838-42.

Shetty S. Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. **Mol Cell Biochem** 2005; 272:107-18.

Spraggon L, Dudnakova T, Slight J, et al. hnRNP-U directly interacts with WT1 and modulates WT1 transcriptional activation. **Oncogene** 2007; 26:1484-91.

Strausberg RL, Feingold EA, Grouse LH, et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. **Proc Natl Acad Sci U S A** 2002; 99:16899-903.

Swiss-Prot. Available from: <URL:<http://ca.expasy.org/sprot/>> [2006 abr 29]

Tacke R, Manley JL. Determinants of SR protein specificity. **Curr Opin Cell Biol** 1999; 11:358-62.

UCSC Genome Bioinformatics. **Genome browser**. Available from: <URL:<http://genome.ucsc.edu/>> [2006 abr 16]

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. **Science** 2003; 302:1212-5.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. **Science** 1995; 270:484-7.

Venables JP. Unbalanced alternative splicing and its significance in cancer. **BioEssays** 2006; 28:378-86.

Villa T, Pleiss JA, Guthrie C. Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? **Cell** 2002; 109:149-52.

Watermann DO. Splicing factor Tra2-beta1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene. **Cancer Res** 2006; 66:4774-80.

Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. **Nature** 2002; 420:520-62.

Watson JD, Crick FH. The structure of DNA. **Cold Spring Harb Symp Quant Biol** 1953; 18:123-31.

Wei-dong H, Ya-li Z, Qi L, et al. Inhibition of proliferation of human breast cancer MCF-7 cells by small interference RNA against LRP16 gene. **Chin J Cancer Res** 2004; 16:239-45.

Weighardt F, Cobiauchi F, Cartegni L, et al. A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock. **J Cell Sci** 1996; 112:1465-76.

Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. **Genome Biol** 2004; 5:R74.

Yugami M, Kabe Y, Yamaguchi Y, Wada T, Handa H. hnRNP-U enhances the expression of specific genes by stabilizing mRNA. **FEBS Lett** 2007; 581:1-7.

Zech VF, Dlaska M, Tzankov A, Hilbe W. Prognostic and diagnostic relevance of hnRNP A2/B1, hnRNP B1 and S100 A2 in non-small cell lung cancer. **Cancer Detect Prev** 2006; 30:395-402.

Zhang Z, Schwartz S, Wagner L, Miller W, et al. A greedy algorithm for aligning DNA sequences. **J Comput Biol** 2002; 7:203-14.

Zhou KH, Tuncali K, Silverman SG. Correlation and simple linear regression. **Radiology** 2003; 227:617-22.

Anexo 1 - Bibliotecas de MPSS de humanos.

libid	tissue_or_Cell_type	patient_sex	mpss_method	tumor
1	adrenal gland – normal	male/female	C	N
2	bladder – normal	male/female	C	N
3	bone marrow – normal	male/female	C	N
4	brain, amygdala – normal	male/female	C	N
5	brain, caudate nucleus – norma	male/female	C	N
6	brain, cerebellum – normal	male/female	C	N
7	brain, corpus callosum - norm	male/female	C	N
8	brain, hypothalamus – normal	male/female	C	N
9	brain, thalamus – normal	male/female	C	N
10	fetal brain, whole – normal	male/female	C	N
11	heart – normal	male	C	N
12	kidney – normal	male/female	C	N
13	lung – normal	male	C	N
14	mammary gland – normal	female	C	N
15	pituitary gland – normal	male/female	C	N
16	placenta – normal	female	C	N
17	pancreas – normal	male/female	C	N
18	prostate – normal	male	C	N
19	retina – normal	male/female	C	N
20	spinal cord – normal	male/female	C	N
21	salivary gland – normal	male/female	C	N
22	small intestine – normal	male/female	C	N
23	spleen - normal	male/female	C	N
24	stomach - normal	male/female	C	N
25	testis - normal	male	C	N
26	thymus - normal	male/female	C	N
27	trachea - normal	male/female	C	N
28	thyroid - normal	male/female	C	N
29	uterus - normal	female	C	N
30	colon transversum - normal		C	N
33	breast cancer, ER- cell line		C	Y
34	breast cancer, ER+ cell line		C	Y
49	Monocytes		C	N
55	peripheral blood lymphocytes		C	N
56	HB4A Normal sample		S	N
57	HB4A modified C5.2; ErbB2 expr		S	Y
58	Human Breast cancer		S	Y
59	Human lung cancer cells		S	Y
60	Human melanoma		S	Y
61	Melanoma Biopsies		S	Y
62	Human placenta		S	N
63	Human testis		S	N
64	Human normal epithelial cell		S	N
65	human melanocyte lightly pigme		S	N
66	colon		S	-
67	colon		S	-

Cont/ Anexo 1 - Bibliotecas de MPSS de humanos.

libid	tissue_or_Cell_type	patient_sex	mpss_method	tumor
68	colon		S	N
69	colon		S	-
37	liver		C	N
38	skeletal muscle		C	N
39	brain whole		C	N

Anexo 2 - Bibliotecas de MPSS de camundongos.

libid	tissue_or_Cell_type	sex	mpss_method	tumor
1	Bladder	Male	C	NULL
2	Esophagus	Male	C	NULL
3	Heart:ventricles and septum	Female	C	NULL
4	Kidney:medulla	Female	C	NULL
5	Kidney:contex	Male	C	NULL
6	Kidney:medulla	Male	C	NULL
7	White fat	Female	C	NULL
8	Brain:Thalamus	Female	C	NULL
9	Brain:Cerebellum	Male	C	NULL
10	Brain:Hippocampus	Male	C	NULL
11	Brain:Midbrain	Male	C	NULL
12	Liver:left lobe	Male	C	NULL
13	Skin:haryy, from back	Female	C	NULL
14	White fat	Male	C	NULL
15	Gl. adrenal	Female	C	NULL
16	Bladder	Female	C	NULL
17	Bone:Femur	Male	C	NULL
18	Brain:Hypothalamus/preoptic ar	Male	C	NULL
19	Brain:Amygdala	Male	C	NULL
20	Kidney:cortex	Female	C	NULL
21	Ovary	Female	C	NULL
22	Stomach	Female	C	NULL
23	Uterus:Pregnant E18	Female	C	NULL
24	Uterus	Female	C	NULL
25	Esophagus	Female	C	NULL
26	Mammary gland	Female	C	NULL
27	Spinal cord:entire	Female	C	NULL
28	Bone:Femur	Female	C	NULL
29	Brain:Olfactory Bulb	Female	C	NULL
30	Brain:Caudate,Putamen,Accumben	Male	C	NULL
31	Eye	Male	C	NULL
32	Skin:hairy, from back	Male	C	NULL
33	Spleen	Female	C	NULL
34	Thyroid/parathyroid	Female	C	NULL
35	Brain:OlfactoryTubercle,Prefro	Female	C	NULL
36	Heart: ventricles and septum	Male	C	NULL
37	Large intestine	Male	C	NULL
38	Liver:right lobe	Male	C	NULL
39	Thymus	Male	C	NULL
40	Heart:Aorta	Male	C	NULL
41	Brain:Hypothalamus/preoptic ar	Female	C	NULL
42	Brain:Thalamus	Male	C	NULL
43	Cartilage:Xiphoid	Male	C	NULL
44	Eye	Female	C	NULL
45	Lymph nodes: mesenteric	Male	C	NULL

Cont/ Anexo 2 - Bibliotecas de MPSS de camundongos.

libid	tissue_or_Cell_type	sex	mpss_method	tumor
46	Pituitary	Female	C	NULL
47	Pituitary	Male	C	NULL
48	Spinal cord:entire	Male	C	NULL
49	Thymus	Female	C	NULL
50	Thyroid/parathyroid	Male	C	NULL
51	Brain:OlfactoryTubercle,Prefro	Male	C	NULL
52	Cervix and vagina	Female	C	NULL
53	Cartilage:Xiphoid	Female	C	NULL
54	Embryo E18	Female	C	NULL
55	Heart: atria	Female	C	NULL
56	Heart:Atria	Male	C	NULL
57	Large intestine	Female	C	NULL
58	Lung	Female	C	NULL
59	Lung	Male	C	NULL
60	Small intestine	Female	C	NULL
61	Testis	Male	C	NULL
62	Gl. adrenal	Male	C	NULL
63	Heart:Aorta	Female	C	NULL
64	Brown Fat	Female	C	NULL
65	Brown Fat	Male	C	NULL
66	Brain:Caudate,Putamen,Accumben	Female	C	NULL
67	Lymph nodes: mesenteric	Female	C	NULL
68	Placenta - E18	Female	C	NULL
69	Brain:Cortical mantle	Female	C	NULL
70	Brain:Midbrain	Female	C	NULL
71	Brain:Cortical mantle	Male	C	NULL
72	Brain:Olfactory Bulb	Male	C	NULL
73	Brain:Amygdala	Female	C	NULL
74	Brain:Cerebellum	Female	C	NULL
75	Brain:Hippocampus	Female	C	NULL
76	Liver:right lobe	Female	C	NULL
77	Prostate	Male	C	NULL
78	Small intestine	Male	C	NULL
79	SkeletalMuscle:Thigh	Female	C	NULL
80	SkeletalMuscle:Thigh	Male	C	NULL
81	Spleen	Male	C	NULL

Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
U1 + U2 snRNP	P09012	U1 small nuclear ribonucleoprotein A (U1 snRNP A protein)	Hs.466775
	-	U1 small nuclear ribonucleoprotein A (U1 snRNP protein A) (U1A protein) (U1-A).	Mm.386890
	P08579	U2 small nuclear ribonucleoprotein B"	Hs.280378
	-	U2 small nuclear ribonucleoprotein B"	Mm.1323
	P09234	U1 small nuclear ribonucleoprotein C (U1-C)	Hs.1063
	Q62241	U1 small nuclear ribonucleoprotein 1C (Snrp1c)	Mm.308514
	P08621	U1 small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa) (snRNP70) (U1-70K)	Hs.467097
	Q62376	U1 small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa) (snRNP70) (U1-70K).	Mm.216386
	Q14776	Formin binding protein 3 (Huntingtin yeast partner A) (Huntingtin-interacting protein HYPA/FBP11) (Fas-ligand associated factor 1) (NY-REN-6 antigen) (HSPC225)	Hs.591637
	-	formin binding protein 3 [Camundongos].	Mm.257474
	Q8NCZ1	Hypothetical protein DKFZp434O1520	Hs.33104
	-	PRP40 pre-mRNA processing factor 40 homolog B (yeast) (Prpf40b)	Mm.358668
	P09661	U2 small nuclear ribonucleoprotein A' (U2 snRNP-A')	Hs.528763
	-	U2 small nuclear ribonucleoprotein polypeptide A' [Camundongos].	Mm.821
	Q15459	Splicing factor 3 subunit 1 (Spliceosome associated protein 114) (SAP 114) (SF3a120)	Hs.406277
	Q9R0I7	Sf3a1: Splicing factor 3a, subunit 1	Mm.156914
	Q15428	Splicing factor 3A subunit 2 (Spliceosome associated protein 62) (SAP 62) (SF3a66)	Hs.115232
	-	splicing factor 3a, subunit 2 [Camundongos].	Mm.358633
	Q12874	Splicing factor 3A subunit 3 (Spliceosome associated protein 61) (SAP 61) (SF3a60)	Hs.77897

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	S3A3_MOUSE Splicing factor 3A subunit 3	<u>Mm.25779</u>
	<u>Q75533</u>	Splicing factor 3B subunit 1 (Spliceosome associated protein 155) (SAP 155) (SF3b155) (Pre-mRNA splicing factor SF3b 155 kDa subunit)	<u>Hs.471011</u>
	-	Splicing factor 3b, subunit 1 (Sf3b1)	<u>Mm.279736</u>
	<u>Q13435</u>	Splicing factor 3B subunit 2 (Spliceosome associated protein 145) (SAP 145) (SF3b150) (Pre-mRNA splicing factor SF3b 145 kDa subunit)	<u>Hs.406423</u>
	-	Splicing factor 3b, subunit 2 (Sf3b2)	<u>Mm.196532</u>
	<u>Q15393</u>	Splicing factor 3B subunit 3 (Spliceosome associated protein 130) (SAP 130) (SF3b130) (Pre-mRNA splicing factor SF3b 130 kDa subunit)	<u>Hs.514435</u>
	-	Splicing factor 3b, subunit 3 (Sf3b3)	<u>Mm.236123</u>
	<u>Q15427</u>	Splicing factor 3B subunit 4 (Spliceosome associated protein 49) (SAP 49) (SF3b50) (Pre-mRNA splicing factor SF3b 49 kDa subunit)	<u>Hs.516160</u>
	<u>P29341</u>	Splicing factor 3b, subunit 4 (Sf3b4)	<u>Mm.219671</u>
	<u>Q9Y3B4</u>	Pre-mRNA branch site protein p14 (CGI-110) (HSPC175) (Ht006)	<u>Hs.177861</u>
	-	splicing factor 3B, 14 kDa subunit [Camundongos].	<u>Mm.102627</u>
	<u>O43395</u>	U4/U6 small nuclear ribonucleoprotein Prp3 (Pre-mRNA splicing factor 3) (U4/U6 snRNP 90 kDa protein) (hPrp3)	<u>Hs.11776</u>
U4/U6 snRNP	-	PRP3 pre-mRNA processing factor 3 homolog (yeast) (Prpf3)	<u>Mm.279872</u>
	<u>O43172</u>	U4/U6 small nuclear ribonucleoprotein Prp4 (U4/U6 snRNP 60 kDa protein) (WD splicing factor Prp4) (hPrp4).	<u>Hs.374973</u>
	-	PRP4 pre-mRNA processing factor 4 homolog (yeast) (Prpf4)	<u>Mm.30660</u>
	<u>O43447</u>	Peptidyl-prolyl cis-trans isomerase H (EC 5.2.1.8) (PPIase H) (Rotamase H) (U-snRNP-associated cyclophilin SnuCyp-20) (USA-CYP) (Small nuclear ribonucleoprotein particle-specific cyclophilin H) (CypH)	<u>Hs.256639</u>
	-	peptidylprolyl isomerase D [Camundongos].	<u>Mm.304080</u>

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i>.
U5 snRNP	Q95320	U5 snRNP-specific 40 kDa protein	Hs.33962
	Q9WV18	WD repeat domain 57 (U5 snRNP specific) (Wdr57)	Mm.228018
	Q75643	U5 small nuclear ribonucleoprotein 200 kDa helicase (U5 snRNP-specific 200 kDa protein) (U5-200KD)	Hs.246112
	-	Activating signal cointegrator 1 complex subunit 3-like 1 (Ascc311)	Mm.215860
	Q15029	116 kDa U5 small nuclear ribonucleoprotein component (U5 snRNP- specific protein, 116 kDa) (U5-116 kDa)	Hs.151787
	Q08810	U5S1_MOUSE 116 kDa U5 small nuclear ribonucleoprotein component	Mm.873
	Q94906	U5 snRNP-associated 102 kDa protein (U5-102 kDa protein)	Hs.31334
	-	Prpf6: PRP6 pre-mRNA splicing factor 6 homolog (yeast)	Mm.292001
	Q43188	Prp28, U5 snRNP 100 kDa protein	Hs.130098
	-	Ddx23: DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	Mm.45725
	Q14834	Spliceosomal U5 snRNP-specific 15 kDa protein (DIM1 protein homolog) (Thioredoxin-like U5 snRNP protein U5-15kD)	Hs.465498
	-	N-acetyltransferase ARD1 [Camundongos].	Mm.172411
	Q14547	PRP8 protein	Hs.181368
	-	Prpf8: Pre-mRNA processing factor 8	Mm.3757
U4/U6.U5 tri-snRNP	Q43290	SART-1 (Squamous cell carcinoma antigen RECOGNISED BY T cells) (U4/U6.U5 TRI-snRNP-associated 110 kDa protein)	Hs.502883
	P05143	Squamous cell carcinoma antigen recognized by T-cells 1 (Sart1)	Mm.34562
	P55769	NHP2-like protein 1 (High mobility group-like nuclear protein 2 homolog 1) ([U4/U6.U5] tri-snRNP 15.5 kDa protein) (OTK27)	Hs.182255
	-	NHP2 non-histone chromosome protein 2-like 1 (<i>S. cerevisiae</i>) (Nhp211)	Mm.299312
	Q96RK9	U4/U6.U5 tri-snRNP-associated 65 kDa protein	Hs.469173

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	<u>O88623</u>	Ubiquitin specific peptidase 39 (Usp39)	<u>Mm.281900</u>
U11 + U12 snRNP	<u>Q9UDW3</u>	Hypothetical protein (U11/U12 snRNP 20K) (Em:AC005529.5 protein) (LOC55954 protein)	<u>Hs.38628</u>
	-	Zinc finger, matrin type 5 (Zmat5)	<u>Mm.271056</u>
	<u>Q9BV90</u>	U11/U12 snRNP 25K protein (Minus-99 protein)	<u>Hs.15277</u>
	-	RIKEN cDNA 3300001G02 gene (3300001G02Rik)	<u>Mm.29952</u>
	<u>Q96TA6</u>	MADP-1 protein (U11/U12 snRNP 31K)	<u>Hs.496279</u>
	-	Zinc finger CCHC-type and RNA binding motif 1 (Zcrb1)	<u>Mm.293181</u>
	<u>Q16560</u>	U1-snRNP binding protein homolog (U11/U12 snRNP 35K, isoform a).	<u>Hs.528306</u>
	-	U1 small nuclear ribonucleoprotein 70 kDa (U1 SNRNP 70 kDa)	<u>Mm.156035</u>
	<u>Q6IEG0</u>	U11/U12 snRNP 48K	<u>Hs.13366</u>
	-	U11/U12 snRNP 48K [Camundongos].	<u>Mm.250783</u>
	<u>Q96LT9</u>	Hypothetical protein FLJ25070 (U11/U12 snRNP 65K) (RNA recognition protein) (Novel protein)	<u>Hs.512635</u>
	-	RNA-binding region (RNP1, RRM) containing 3 (Rnpc3)	<u>Mm.316928</u>
Sm	<u>P14678</u>	Small nuclear ribonucleoprotein associated proteins B and B' (snRNP-B) (Sm protein B/B') (Sm-B/Sm-B') (SmB/SmB')	<u>Hs.83753</u>
	-	Snrbp: Small nuclear ribonucleoprotein B	<u>Mm.88216</u>
	<u>P14648</u>	Small nuclear ribonucleoprotein associated protein N (snRNP-N) (Sm protein N) (Sm-N) (SmN) (Sm-D) (Tissue-specific splicing protein)	<u>Hs.564847</u>
	-	Small nuclear ribonucleoprotein N (Snrpn)	<u>Mm.274995</u>
	<u>P13641</u>	Small nuclear ribonucleoprotein Sm D1 (snRNP core protein D1) (Sm-D1) (Sm-D autoantigen)	<u>Hs.464734</u>
	-	Small nuclear ribonucleoprotein D1 (Snrpd1)	<u>Mm.603</u>
	<u>P43330</u>	Small nuclear ribonucleoprotein Sm D2 (snRNP core protein D2) (Sm-D2)	<u>Hs.515472</u>
	-	Small nuclear ribonucleoprotein D2 (Snrpd2)	<u>Mm.29135</u>

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	P43331	Small nuclear ribonucleoprotein Sm D3 (snRNP core protein D3) (Sm-D3)	Hs.356549
	-	Small nuclear ribonucleoprotein D3 (Snrpd3)	Mm.45151
	P08578	Small nuclear ribonucleoprotein E (snRNP-E) (Sm protein E) (Sm-E) (SmE)	Hs.334612
	P08578	Small nuclear ribonucleoprotein E (Snrpe)	Mm.249110
	Q15356	Small nuclear ribonucleoprotein F (snRNP-F) (Sm protein F) (Sm-F) (SmF)	Hs.105465
	-	Small nuclear ribonucleoprotein polypeptide F (Snrpf)	Mm.350851
	Q15357	Small nuclear ribonucleoprotein G (snRNP-G) (Sm protein G) (Sm-G) (SmG)	Hs.516076
	-	Small nuclear ribonucleoprotein polypeptide G (Snrpg)	Mm.276802
	Q15116	U6 snRNA-associated Sm-like protein LSm1	Hs.425311
	-	LSM1 homolog, U6 small nuclear RNA associated (<i>S. cerevisiae</i>) (Lsm1)	Mm.30198
	Q9Y333	U6 snRNA-associated Sm-like protein LSm2 (Small nuclear ribonuclear protein D homolog) (G7b) (SnRNP core SM-like protein SM-x5)	Hs.103106
	-	LSM2 homolog, U6 small nuclear RNA associated (<i>S. cerevisiae</i>) (Lsm2)	Mm.165735
	Q9Y4Z1	U6 snRNA-associated Sm-like protein LSm3 (MDS017)	Hs.111632
	-	LSM3 homolog, U6 small nuclear RNA associated (<i>S. cerevisiae</i>) (Lsm3)	Mm.246693
	Q9Y4Z0	U6 snRNA-associated Sm-like protein LSm4 (Glycine-rich protein) (GRP)	Hs.515255
	Q9QXA5	LSM4_MOUSE U6 snRNA-associated Sm-like protein LSm4	Mm.248188
	Q9Y4Y9	U6 snRNA-associated Sm-like protein LSm5	Hs.424908
	-	LSM5 homolog, U6 small nuclear RNA associated (<i>S. cerevisiae</i>) (Lsm5)	Mm.25642
	Q9Y4Y8	U6 snRNA-associated Sm-like protein LSm6	Hs.190520
	-	LSM6 homolog, U6 small nuclear RNA associated (<i>S. cerevisiae</i>) (Lsm6)	Mm.28694
	Q9UK45	U6 snRNA-associated Sm-like protein LSm7	Hs.512610
	-	LSM7 homolog, U6 small nuclear RNA associated (<i>S. cerevisiae</i>) (Lsm7)	Mm.379101
	Q95777	U6 snRNA-associated Sm-like protein LSm8	Hs.592275

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	Lsm8: LSM8 homolog, U6 small nuclear RNA associated (S. cerevisiae)	Mm.275158
	Q969L4	U7 snRNA-associated Sm-like protein LSm10	Hs.471768
	-	Serine/threonine kinase 40 (Stk40)	Mm.41865
	Q8N4M0	Hypothetical protein	Hs.565094
	-	LSM domain containing 1 (Lsmd1)	Mm.45683
U2AF	P26368	Splicing factor U2AF 65 kDa subunit (U2 auxiliary factor 65 kDa subunit) (U2 snRNP auxiliary factor large subunit) (hU2AF(65))	Hs.528007
	-	U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 2 (U2af2)	Mm.360389
	Q01081	Splicing factor U2AF 35 kDa subunit (U2 auxiliary factor 35 kDa subunit) (U2 snRNP auxiliary factor small subunit)	Hs.365116
	-	U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 1 (U2af1)	Mm.379289
	Q8WU68	U2 AUXILIARY FACTOR 26	Hs.351558
	-		-
	Q15695	U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit related-protein 1	Hs.567353
	-	U2 small nuclear RNA auxiliary factor 1-like 4 (U2af1l4)	Mm.34790
	Q15696	U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit related-protein 2	Hs.171909
	-	U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 1, related sequence 2 (U2af1-rs2)	Mm.180953
SR	Q9UQ35	RNA binding protein	Hs.433343
	-	Serine/arginine repetitive matrix 2 (Srrm2)	Mm.5222
	Q15410	Nucleic acid binding protein (Fragment).	Hs.54649
	-	RIKEN cDNA 2610209M04 gene (2610209M04Rik)	Mm.182650
	Q01130	Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35) (Splicing component, 35 kDa) (PR264 protein)	Hs.584801
	-	Splicing factor, arginine/serine-rich 2 (SC-35) (Sfrs2)	Mm.21841

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	<u>Q9BRL6</u>	Similar to splicing factor, arginine/serine-rich 2 (SC-35) (SRp46 splicing factor)	<u>Hs.476680</u>
	-		-
	<u>P23152</u>	Splicing factor, arginine/serine-rich 3 (Pre-mRNA splicing factor SRP20) (X16 protein).	<u>Hs.405144</u>
	-	Splicing factor, arginine/serine-rich 3 (SRp20) (Sfrs3)	<u>Mm.358634</u>
	<u>Q16629</u>	Splicing factor, arginine/serine-rich 7 (Splicing factor 9G8)	<u>Hs.309090</u>
	-	Splicing factor, arginine/serine-rich 7 (Sfrs7)	<u>Mm.292016</u>
	<u>Q13242</u>	Splicing factor, arginine/serine-rich 9 (Pre-mRNA splicing factor SRp30C)	<u>Hs.369624</u>
	-	Splicing factor, arginine/serine rich 9 (Sfrs9)	<u>Mm.287826</u>
	<u>Q07955</u>	Splicing factor, arginine/serine-rich 1 (pre-mRNA splicing factor SF2, P33 subunit) (Alternative splicing factor ASF-1)	<u>Hs.68714</u>
	-	Splicing factor, arginine/serine-rich 1 (ASF/SF2) (Sfrs1)	<u>Mm.45645</u>
	<u>Q13243</u>	Splicing factor, arginine/serine-rich 5 (Pre-mRNA splicing factor SRP40) (Delayed-early protein HRS)	<u>Hs.166975</u>
	-	Splicing factor, arginine/serine-rich 5 (SRp40, HRS) (Sfrs5)	<u>Mm.43331</u>
	<u>Q13247</u>	Splicing factor, arginine/serine-rich 6 (Pre-mRNA splicing factor SRP55)	<u>Hs.6891</u>
	-	Splicing factor, arginine/serine-rich 6 (Sfrs6)	<u>Mm.24042</u>
	<u>Q08170</u>	Splicing factor, arginine/serine-rich 4 (Pre-mRNA splicing factor SRP75) (SRP001LB)	<u>Hs.469970</u>
	-	Splicing factor, arginine/serine-rich 4 (SRp75) (Sfrs4)	<u>Mm.2478</u>
	<u>Q05519</u>	Splicing factor arginine/serine-rich 11 (Arginine-rich 54 kDa nuclear protein) (p54)	<u>Hs.479693</u>
	-	Splicing factor, arginine/serine-rich 11 (Sfrs11)	<u>Mm.223946</u>
	<u>Q8WXA9</u>	Splicing factor, arginine/serine-rich 12 (Serine-arginine-rich splicing regulatory protein 86) (SRrp86) (Splicing regulatory protein 508) (SRrp508)	<u>Hs.519347</u>
	-	Splicing factor, arginine/serine-rich 12 (Sfrs12)	<u>Mm.33908</u>

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i>.
	Q13595	Transformer-2 protein homolog (TRA-2 alpha)	Hs.592175
	-	splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila) [Camundongos].	Mm.196598
	Q15815	Arginine/serine-rich splicing factor 10 (Transformer-2-beta) (HTRA2- beta) (Transformer 2 protein homolog) (Silica-induced protein 41) (RA301)	Hs.533122
	-	Splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila) (Sfrs10)	Mm.210352
	Q9UNR9	Topoisomerase I-binding RS protein	Hs.589962
	-	Topoisomerase I binding, arginine/serine-rich (Topors)	Mm.251548
hnRNP	Q13151	Heterogeneous nuclear ribonucleoprotein A0 (hnRNP A0)	Hs.96996
	-	Heterogeneous nuclear ribonucleoprotein A0 (Hnrpa0)	Mm.390606
	P09651	Heterogeneous nuclear ribonucleoprotein A1 (Helix-destabilizing protein) (Single-strand binding protein) (hnRNP core protein A1)	Hs.546261
	-	Heterogeneous nuclear ribonucleoprotein A1 (Hnrpa1)	Mm.299367
	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2 / hnRNP B1)	Hs.487774
	-	Heterogeneous nuclear ribonucleoprotein A2/B1 (Hnrpa2b1)	Mm.155896
	P51991	Heterogeneous nuclear ribonucleoprotein A3 (hnRNP A3) (D10S102)	Hs.516539
	-	Heterogeneous nuclear ribonucleoprotein A3 (Hnrpa3)	Mm.379375
	P07910	Heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1 / hnRNP C2).	Hs.508848
	-	Heterogeneous nuclear ribonucleoprotein C (Hnrpc)	Mm.274690
	Q60812	Heterogeneous nuclear ribonucleoprotein C-like dJ845O24.4 (hnRNP core protein C-like)	Hs.502617
	-		-
	Q9UKM9	RNA-binding protein Raly (hnRNP associated with lethal yellow homolog) (Autoantigen p542)	Hs.136947

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	hnRNP-associated with lethal yellow [Camundongos].	Mm.221440
	Q8N1C2	LOC138046 protein	Hs.121663
	-	RIKEN cDNA 0710005M24 gene (0710005M24Rik)	Mm.121014
	Q14103	Heterogeneous nuclear ribonucleoprotein D0 (hnRNP D0) (AU-rich element RNA-binding protein 1)	Hs.480073
	-	Heterogeneous nuclear ribonucleoprotein D (Hnrpd)	Mm.384474
	Q99729	Heterogeneous nuclear ribonucleoprotein A/B (hnRNP A/B) (APOBEC-1 binding protein 1) (ABBP-1)	Hs.248746
	-	Heterogeneous nuclear ribonucleoprotein A/B (Hnrpab)	Mm.256875
	O14979	JKTBP2 (Heterogeneous nuclear ribonucleoprotein D-like) (Hypothetical protein) (HNRPD protein)	Hs.527105
	-	Heterogeneous nuclear ribonucleoprotein D-like (Hnrpdl)	Mm.195310
	Q43347	WUGSC:H_166H1.2 protein (Musashi)	Hs.158311
	-	Musashi homolog 1 [Camundongos].	Mm.5077
	Q96DH6	Musashi 2, isoform a.	Hs.585782
	-	Musashi homolog 2 [Camundongos].	Mm.270331
	Q15365	Poly(rC)-binding protein 1 (Alpha-CP1) (hnRNP-E1) (Nucleic acid binding protein SUB2.3)	Hs.2853
	-	poly(rC) binding protein 1 [Camundongos].	Mm.274146
	Q15366	Poly(rC)-binding protein 2 (Alpha-CP2) (hnRNP-E2)	Hs.546271
	-	Poly(rC) binding protein 2 (Pcbp2)	Mm.236513
	P57721	Poly(rC)-binding protein 3 (Alpha-CP3)	Hs.474049
	-	Poly(rC) binding protein 3 (Pcbp3)	Mm.272803
	P57723	Poly(rC)-binding protein 4 (Alpha-CP4)	Hs.20930
	-	poly(rC) binding protein 4; poly(rC)-binding protein 4 [Camundongos].	Mm.286394
	P52597	Heterogeneous nuclear ribonucleoprotein F (hnRNP F)	Hs.808

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	Inversin (Inversion of embryo turning protein) (Nephrocystin-2).	Mm.317706
	P31943	Heterogeneous nuclear ribonucleoprotein H (hnRNP H)	Hs.202166
	-	heterogeneous nuclear ribonucleoprotein H1; heterogeneous nuclear ribonucleoprotein H [Camundongos]	Mm.21740
	P55795	Heterogeneous nuclear ribonucleoprotein H' (hnRNP H') (FTP-3)	Hs.432485
	-	Ribosomal protein L36a (Rpl36a)	Mm.286408
	P31942	Heterogeneous nuclear ribonucleoprotein H3 (hnRNP H3) (hnRNP 2H9)	Hs.591357
	-	heterogeneous nuclear ribonucleoprotein H2; heterogeneous nuclear ribonucleoprotein H' [Camundongos]	Mm.390303
	Q12849	G-rich sequence factor-1 (GRSF-1)	Hs.309763
	-	heterogeneous nuclear ribonucleoprotein H1; heterogeneous nuclear ribonucleoprotein H [Camundongos]	Mm.332474
	P38159	Heterogeneous nuclear ribonucleoprotein G (hnRNP G) (RNA binding motif protein, X chromosome) (Glycoprotein p43)	Hs.380118
	-	RNA binding motif protein, X chromosome [Camundongos]	Mm.28275
	Q75526	Testes specific heterogenous nuclear ribonucleoprotein G-T.	Hs.121605
	-	RNA binding motif protein, X chromosome [Camundongos]	Mm.128134
	Q14011	Cold-inducible RNA-binding protein (Glycine-rich RNA-binding protein CIRP) (A18 hnRNP)	Hs.501309
	-	Cold inducible RNA binding protein (Cirbp)	Mm.17898
	P98179	Putative RNA-binding protein 3 (RNA binding motif protein 3) (RNPL)	Hs.301404
	-	RNA binding motif protein 3 (Rbm3)	Mm.128512
	P26599	Polypyrimidine tract-binding protein 1 (PTB) (Heterogeneous nuclear ribonucleoprotein I) (hnRNP I) (57 kDa RNA-binding protein PPTB-1)	Hs.172550
	-	Polypyrimidine tract binding protein 1 (Ptbpl)	Mm.265610
	Q969N9	PTB-like protein L (Polypyrimidine tract binding protein 2)	Hs.591430

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	polypyrimidine tract binding protein 2 [Camundongos]	Mm.29966
	Q95758	Rod1	Hs.269988
	-	ROD1 regulator of differentiation 1 (S. pombe) (Rod1)	Mm.331640
	Q07244	Heterogeneous nuclear ribonucleoprotein K (hnRNP K) (DC-stretch binding protein) (CSBP)	Hs.522257
	-	heterogeneous nuclear ribonucleoprotein K [Camundongos]	Mm.142872
	P14866	Heterogeneous nuclear ribonucleoprotein L (hnRNP L).	Hs.589594
	-	Heterogeneous nuclear ribonucleoprotein L (Hnrpl)	Mm.9043
	Q8WVV9	Hypothetical protein	Hs.445497
	-	Heterogeneous nuclear ribonucleoprotein L-like (Hnrpll)	Mm.64579
	P52272	Heterogeneous nuclear ribonucleoprotein M (hnRNP M)	Hs.465808
	-	Heterogeneous nuclear ribonucleoprotein M (Hnrpm)	Mm.311439
	Q9H922	Myelin gene expression factor	Hs.591108
	-	Myelin basic protein expression factor 2, repressor (Myef2)	Mm.18535
	Q60506	Gry-rbp (hnRNP Q3)	Hs.571177
	-	NS1-associated protein 1-like; RRM RNA binding protein GRY-RBP [Camundongos]	Mm.260545
	Q43390	Heterogeneous nuclear ribonucleoprotein R (hnRNP R)	Hs.573762
	-	Heterogeneous nuclear ribonucleoprotein R (Hnrpr)	Mm.31051
	Q00839	Heterogenous nuclear ribonucleoprotein U (hnRNP U) (Scaffold attachment factor A) (SAF-A)	Hs.166463
	-	Heterogeneous nuclear ribonucleoprotein U (Hnrpu)	Mm.2115
	Q76022	E1B-55kDa-associated protein	Hs.155218
	-	Heterogeneous nuclear ribonucleoprotein U-like 1 (Hnrpul1)	Mm.254223
	Q8N3B3	Hypothetical protein DKFZp762N1910 (Fragment)	Hs.406377
	-	Heterogeneous nuclear ribonucleoprotein U-like 2 (Hnrpul2)	Mm.347805

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
TIA	P31483	Nucleolysin TIA-1 (RNA-binding protein TIA-1) (p40-TIA-1) [Contains: Nucleolysin TIA-1 isoform p15 (p15-TIA-1)]	Hs.516075
	-	TIA1 cytotoxic granule-associated RNA binding protein (TIA1)	Mm.274425
	Q01085	Nucleolysin TIAR (TIA-1 related protein)	Hs.585488
	-	Nucleolysin TIAR (TIA-1-related protein).	Mm.242072
CELF/ CUG-BP	Q92879	CUG triplet repeat RNA-binding protein 1 (CUG-BP1) (RNA-binding protein BRUNOL-2) (Deadenylation factor CUG-BP) (50 kDa Nuclear polyadenylated RNA-binding protein) (EDEN-BP)	Hs.269944
	-	CUG triplet repeat, RNA binding protein 2; elav-type RNA-binding protein 3; CUG triplet repeat, RNA-binding protein 2 [Camundongos]	Mm.29495
	Q92950	Apoptosis-related RNA binding protein (ETR-3)	Hs.309288
	-	CUG triplet repeat, RNA binding protein 2 (Cugbp2)	Mm.147091
	Q9BZC0	CUG-BP and ETR-3 like factor 5	Hs.567561
	-	Bruno-like 5, RNA binding protein (Drosophila) (Brunol5)	Mm.152689
	Q9BZC1	CUG-BP and ETR-3 like factor 4	Hs.435976
	-	Bruno-like 4, RNA binding protein (Drosophila) (Brunol4)	Mm.266435
	Q9BZC2	CUG-BP and ETR-3 like factor 3	Hs.26047
	-	CUG triplet repeat, RNA binding protein 2 [Camundongos].	Mm.44292
	Q96J87	BRUNO-like 6 RNA-binding protein (RNA-binding protein CELF6)	Hs.348342
	-	Bruno-like 6, RNA binding protein (Drosophila) (Brunol6)	Mm.265415
CLK	P49759	Protein kinase CLK1 (EC 2.7.1.-) (CLK)	Hs.433732
	-	CLK1_MOUSE Protein kinase CLK1	Mm.1761
	P49760	Protein kinase CLK2 (EC 2.7.1.-) (CDC-like kinase 2)	Hs.73986
	-	Dual specificity protein kinase CLK2 (CDC-like kinase 2).	Mm.288098
	P49761	Protein kinase CLK3 (EC 2.7.1.-) (CDC-like kinase 3)	Hs.584748

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	Dual specificity protein kinase CLK3 (CDC-like kinase 3).	<u>Mm.25720</u>
	<u>Q9HAZ1</u>	Protein serine threonine kinase Clk4	<u>Hs.406557</u>
	-	Dual specificity protein kinase CLK4 (CDC-like kinase 4).	<u>Mm.239354</u>
SRPK	<u>Q96SB4</u>	SRPK1a protein kinase	<u>Hs.443861</u>
	-	Serine/arginine-rich protein specific kinase 1 (Srp1)	<u>Mm.15252</u>
	<u>P78362</u>	Serine kinase SRPK2	<u>Hs.285197</u>
	-	serine/arginine-rich protein specific kinase 2 [Camundongos]	<u>Mm.288728</u>
	<u>Q9UPE1</u>	Serine/threonine-protein kinase 23 (EC 2.7.1.37) (Muscle-specific serine kinase 1) (MSSK-1)	<u>Hs.104865</u>
	-	serine/threonine kinase 23; muscle-specific serine kinase 1 [Camundongos]	<u>Mm.111904</u>
hprp4	<u>Q13523</u>	Serine/threonine-protein kinase PRP4 homolog (EC 2.7.1.37) (PRP4 pre-mRNA processing factor 4 homolog) (PRP4 kinase)	<u>Hs.159014</u>
	-	PRP4 pre-mRNA processing factor 4 homolog B (yeast) (Prpf4b)	<u>Mm.10027</u>
CRK7	<u>Q9NYV4</u>	Cell division cycle 2-related protein kinase 7 (EC 2.7.1.37) (CDC2-related protein kinase 7) (CrkRS).	<u>Hs.416108</u>
	-	Cdc2-related kinase, arginine/serine-rich (Crkrs)	<u>Mm.260516</u>
Skip	<u>Q13573</u>	Nuclear protein SkiP (Ski-interacting protein) (SNW1 protein) (Nuclear receptor coactivator NCoA-62).	<u>Hs.546550</u>
	-	SNW domain containing 1 (Snw1)	<u>Mm.271174</u>
NOVA	<u>P51513</u>	RNA-binding protein Nova-1 (Neuro-oncological ventral antigen 1) (Onconeural ventral antigen-1) (Paraneoplastic Ri antigen) (Ventral neuron-specific protein 1)	<u>Hs.592335</u>
	-	Neuro-oncological ventral antigen 1 (Nova1)	<u>Mm.247195</u>
	<u>Q9UNW9</u>	RNA-binding protein Nova-2 (Neuro-oncological ventral antigen 2) (Astrocytic NOVA1-like RNA-binding protein)	<u>Hs.375439</u>
	-		-
DEAD	<u>P17844</u>	Probable RNA-dependent helicase p68 (DEAD-box protein p68) (DEAD-box protein 5)	<u>Hs.279806</u>

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5 (Ddx5)	Mm.220038
	Q92841	Probable RNA-dependent helicase p72 (DEAD-box protein p72) (DEAD-box protein 17)	Hs.528305
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 17 (Ddx17)	Mm.29644
	Q9H5Z1	Probable ATP-dependent helicase DHX35 (DEAH-box protein 35)	Hs.444520
	-	DEAH (Asp-Glu-Ala-His) box polypeptide 35 (Dhx35)	Mm.315652
	Q96EI0	DEAD (Asp-Glu-Ala-Asp) box polypeptide 46	Hs.533245
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 46 (Ddx46)	Mm.202725
	O00571	DEAD-box protein 3 (Helicase-like protein 2) (HLP2) (DEAD-box, X isoform)	Hs.380774
	-	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3, X-linked (Ddx3x)	Mm.289662
	Q15523	DEAD-box protein 3, Y-chromosomal	Hs.99120
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked (Ddx3y)	Mm.302938
	P38919	Probable ATP-dependent helicase DDX48 (DEAD-box protein 48) (Eukaryotic initiation factor 4A-like NUK-34) (Nuclear matrix protein 265) (hNMP 265) (Eukaryotic translation initiation factor 4A isoform 3)	Hs.389649
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 48 (Ddx48)	Mm.391989
	Q6IPS3	DDX26B protein	Hs.496829
	-	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26; Notch2-like [Camundongos]	Mm.72753
	Q9UL03	Candidate tumor suppressor protein DICE1 (DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26) (OTTHUMP00000018439)	Hs.439440
	-	Integrator complex subunit 6 (Ints6)	Mm.319684
	Q94894	KIAA0801 protein (Mesma proteína Q96EI0)	Hs.533245
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 46 (Ddx46)	Mm.202725
	Q9UJV9	DEAD-box protein abstrakt homolog (DEAD-box protein 41)	Hs.484288

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 41 (Ddx41)	<u>Mm.205045</u>
	<u>Q43143</u>	Putative pre-mRNA splicing factor RNA helicase (DEAH box protein 15) (ATP-dependent RNA helicase #46)	<u>Hs.5683</u>
	-	DEAH (Asp-Glu-Ala-His) box polypeptide 15 (Dhx15)	<u>Mm.993</u>
	<u>Q13838</u>	Spliceosome RNA helicase BAT1 (DEAD-box protein UAP56) (56 kDa U2AF65 associated protein) (ATP-dependent RNA helicase p47) (HLA-B associated transcript-1)	<u>Hs.254042</u>
	-	Histocompatibility 2, K1, K region (H2-K1)	<u>Mm.33263</u>
	<u>O00148</u>	ATP-dependent helicase DDX39 (DEAD-box protein 39) (Nuclear RNA helicase URH49)	<u>Hs.311609</u>
	-	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39 (Ddx39)	<u>Mm.28222</u>
	<u>Q14562</u>	ATP-dependent helicase DHX8 (RNA helicase HRH1) (DEAH-box protein 8)	<u>Hs.463105</u>
	-	DEAH (Asp-Glu-Ala-His) box polypeptide 8 (Dhx8)	<u>Mm.28186</u>
	<u>O60231</u>	Putative pre-mRNA splicing factor RNA helicase (ATP-dependent RNA helicase #3) (DEAH-box protein 16)	<u>Hs.485060</u>
	-	DEAH (Asp-Glu-Ala-His) box polypeptide 16 (Dhx16)	<u>Mm.390986</u>
	<u>Q92620</u>	Pre-mRNA splicing factor ATP-dependent RNA helicase PRP16 (EC 3.6.1.-) (ATP-dependent RNA helicase DHX38) (DEAH-box protein 38)	<u>Hs.570079</u>
	-	DEAH (Asp-Glu-Ala-His) box polypeptide 38 (Dhx38)	<u>Mm.23705</u>
	<u>Q08211</u>	ATP-dependent RNA helicase A (Nuclear DNA helicase II) (NDH II) (DEAH-box protein 9)	<u>Hs.191518</u>
	-	DEAH (Asp-Glu-Ala-His) box polypeptide 9 (Dhx9)	<u>Mm.20000</u>
	<u>P42285</u>	KIAA0052 protein	<u>Hs.274531</u>
	-	superkiller viralicidic activity 2-like [Camundongos].	<u>Mm.291029</u>
Cyclophilin	<u>Q96BP3</u>	Hypothetical protein KIAA0073 (EC 5.2.1.8) (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase)	<u>Hs.121432</u>
	-	Peptidylprolyl isomerase domain and WD repeat containing 1 (Ppwd1)	<u>Mm.98910</u>
	<u>Q9H2H8</u>	Cyclophilin-like protein PPIL3b	<u>Hs.121076</u>

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i>.
	-	Peptidylprolyl isomerase (cyclophilin)-like 3 (Ppil3)	Mm.340195
	Q9Y3C6	Peptidyl-prolyl cis-trans isomerase like 1 (EC 5.2.1.8) (PPIase) (Rotamase) (CGI-124) (UNQ2425/PRO4984)	Hs.27693
	-	Peptidylprolyl isomerase (cyclophilin)-like 1 (Ppil1)	Mm.328928
	Q13356	Peptidyl-prolyl cis-trans isomerase like 2 (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin-60) (Cyclophilin-like protein Cyp-60)	Hs.438587
	-	Peptidylprolyl isomerase (cyclophilin)-like 2 (Ppil2)	Mm.253614
	Q9UNP9	Peptidyl-prolyl cis-trans isomerase E (EC 5.2.1.8) (PPIase E) (Rotamase E) (Cyclophilin E) (Cyclophilin 33)	Hs.524690
	-	Peptidylprolyl isomerase E (cyclophilin E) (Ppie)	Mm.126873
HeatShock	P08107	Heat shock 70 kDa protein 1 (HSP70.1) (HSP70-1/HSP70-2)	Hs.520028
	-	Heat shock 70 kDa protein 1B (HSP70.1).	Mm.6388
	P11142	Heat shock cognate 71 kDa protein	Hs.180414
	-	heat shock 70kD protein 8; heat shock protein cognate 70 [Camundongos]	Mm.290774
	P11021	78 kDa glucose-regulated protein precursor (GRP 78) (Immunoglobulin heavy chain binding protein) (BiP) (Endoplasmic reticulum lumenal Ca(2+) binding protein grp78)	Hs.522392
	-	78 kDa glucose-regulated protein precursor (GRP 78) (Immunoglobulin heavy chain-binding protein) (BiP).	Mm.330160
p52/p75	Q9UER6	Small nuclear RNA activating complex, polypeptide 3 (Snape3)	Hs.493516
	-	PC4 and SFRS1 interacting protein 1 (Psp1)	Mm.271985
ELAV	P26378	ELAV-like protein 4 (Paraneoplastic encephalomyelitis antigen HuD) (Hu-antigen D)	Hs.568556
	-	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 4 (Hu antigen D) (Elavl4)	Mm.3970
	Q14576	ELAV-like protein 3 (Hu-antigen C) (HuC) (Paraneoplastic cerebellar degeneration-associated antigen) (Paraneoplastic limbic encephalitis antigen 21)	Hs.1701

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	ELAV-like protein 3 (Hu-antigen C) (HuC).	Mm.390167
	Q12926	ELAV-like protein 2 (Hu-antigen B) (HuB) (ELAV-like neuronal protein 1) (Nervous system-specific RNA binding protein Hel-N1)	Hs.166109
	-	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B) (Elav12)	Mm.318042
	Q15717	ELAV-like protein 1 (Hu-antigen R) (HuR)	Hs.184492
	-	Transcribed locus, strongly similar to NP_034615.2 ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R) [Camundongos]	Mm.119162
P52/P100	P23246	Splicing factor, proline-and glutamine-rich (Polypyrimidine tract- binding protein-associated splicing factor) (PTB-associated splicing factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit)	Hs.355934
	-	Splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated) (Sfpq)	Mm.257276
	Q15233	54 kDa nuclear RNA- and DNA-binding protein (p54(nrb)) (p54nrb) (55 kDa nuclear protein) (NMT55) (Non-POU domain-containing octamer- binding protein) (DNA-binding P52/P100 complex, 52 kDa subunit)	Hs.533282
	-	Non-POU-domain-containing, octamer binding protein (Nono)	Mm.280069
FUSE	Q92945	KSRP	Hs.568331
	-	KH-type splicing regulatory protein (Khsrp)	Mm.34296
	Q92946	FUSE binding protein 3 (Fragment)	Hs.98751
	-	Far upstream element (FUSE) binding protein 3 (Fubp3)	Mm.207261
ColdShock	P16989	DNA-binding protein A (Cold shock domain protein A) (Single-strand DNA binding protein NF-GMB)	Hs.221889
	-	Cold shock domain protein A (Csda)	Mm.299604

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	<u>P16991</u>	Nuclease sensitive element binding protein 1 (Y box binding protein-1) (Y-box transcription factor) (YB-1) (CCAAT-binding transcription factor I subunit A) (CBF-A) (Enhancer factor I subunit A) (EFI-A) (DNA-binding protein B) (DBPB)	<u>Hs.473583</u>
	-	nuclease sensitive element binding protein 1 [Camundongos].	<u>Mm.258204</u>
FBP	<u>Q9NZA0</u>	FBP-interacting repressor (Siah binding protein 1, FBP interacting repressor, pyrimidine tract binding splicing factor, Ro ribonucleoprotein-binding protein 1)	<u>Hs.521924</u>
	-	Polyadenylate-binding protein 1 (Poly(A)-binding protein 1) (PABP 1).	<u>Mm.29965</u>
P32	<u>Q07021</u>	Complement component 1, Q subcomponent binding protein, mitochondrial precursor (Glycoprotein gC1qBP) (GC1q-R protein) (Hyaluronan-binding protein 1) (p32) (p33)	<u>Hs.555866</u>
	-	Complement component 1, q subcomponent binding protein (C1QBP)	<u>Mm.30049</u>
SNP70	<u>Q9Y2W2</u>	SH3 domain-binding protein SNP70 (NPW38-binding protein NPWBP) (Similar to WW domain binding protein 11)	<u>Hs.569122</u>
	-	WW domain binding protein 11 (Wbp11)	<u>Mm.141197</u>
CBP	<u>P52298</u>	Nuclear cap binding protein subunit 2 (20 kDa nuclear cap binding protein) (NCBP 20 kDa subunit) (CBP20) (NCBP interacting protein 1) (NIP1)	<u>Hs.591671</u>
	-	Nuclear cap binding protein subunit 2 (Ncbp2)	<u>Mm.290027</u>
	<u>Q09161</u>	80 kDa nuclear cap binding protein (NCBP 80 kDa subunit) (CBP80)	<u>Hs.591907</u>
	-	S50082 nuclear cap binding protein - human	<u>Mm.389536</u>
ALY	<u>Q43672</u>	THO complex subunit 4 (Tho4) (Ally of AML-1 and LEF-1) (Transcriptional coactivator Aly/REF) (bZIP enhancing factor BEF)	<u>Hs.534385</u>
	-	THO complex 4 (Thoc4)	<u>Mm.1886</u>
SLU7	<u>Q95391</u>	Step II splicing factor SLU7	<u>Hs.435342</u>
	-	DNA segment, Chr 11, ERATO Doi 730, expressed (D11Ert730e)	<u>Mm.28200</u>

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
PRP18	Q99633	Pre-mRNA splicing factor 18 (PRP18 homolog).	Hs.161181
	-	PRP18 pre-mRNA processing factor 18 homolog (yeast) (Prpf18)	Mm.38529
CA150	Q14776	Putative transcription factor CA150	Hs.443465
	-	Transcription elongation regulator 1 (CA150) (Tcerg1)	Mm.270511
RDP	P18615	Negative elongation factor E (NELF-E) (RD protein).	Hs.423935
	-	RD RNA-binding protein (Rdbp)	Mm.279907
ACIN	Q9UKV3	Apoptotic chromatin condensation inducer in the nucleus (Acinus).	Hs.124490
	-	Apoptotic chromatin condensation inducer 1 (Acin1)	Mm.297078
ILF3	Q12906	Interleukin enhancer-binding factor 3	Hs.465885
	-	interleukin enhancer binding factor 3 [Camundongos]	Mm.325205
CRN	Q9BZJ0	Crooked neck-like protein 1 (Crooked neck homolog)	Hs.171342
	-	Crn, crooked neck-like 1 (Drosophila) (Crnk11)	Mm.248755
WTAP	Q15007	Wilms' tumor 1-associating protein (WT1-associated protein) (Putative pre-mRNA splicing regulator female-lethal(2D) homolog)	Hs.446091
	-	Wilms' tumour 1-associating protein (Wtap)	Mm.275521
PRP17	O60508	Pre-mRNA splicing factor PRP17 (hPRP17) (Cell division cycle 40 homolog) (EH-binding protein 3)	Hs.428147
	-	Cell division cycle 40 homolog (yeast) (Cdc40)	Mm.46063
Others	Q9ULR0	KIAA1160 protein	Hs.512661
	-	RIKEN cDNA 5830446M03 gene (5830446M03Rik)	Mm.241546
	Q75937	DNAJC8 protein	Hs.433540
	-	DnaJ (Hsp40) homolog, subfamily C, member 8 (Dnajc8)	Mm.29685
	P35637	RNA-binding protein FUS (Oncogene FUS) (Oncogene TLS) (Translocated in liposarcoma protein) (POMp75) (75 kDa DNA-pairing protein)	Hs.3530
	-	FUS interacting protein (serine-arginine rich) 1 (Fusip1)	Mm.10229

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	Q92804	TATA-binding protein associated factor 2N (RNA-binding protein 56) (TAFII68) (TAF(II)68)	Hs.402752
	-	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor (Taf15)	Mm.181050
	Q8N2M8	Splicing factor, arginine/serine-rich 16 (Suppressor of white-apricot homolog 2)	Hs.466917
	-	Splicing factor, arginine/serine-rich 16 (suppressor-of-white-apricot homolog, Drosophila) (Sfrs16)	Mm.20913
	Q14498	RNA-binding region containing protein 2 (Hepatocellular carcinoma protein 1) (Splicing factor HCC1)	Hs.282901
	-	RNA-binding region (RNPI, RRM) containing 2 (Rnpc2)	Mm.153895
	Q43934	ET putative translation product	Hs.73965
	-	RIKEN cDNA 2600014M03 gene (2600014M03Rik)	Mm.390345
	Q43670	Zinc finger protein 207	Hs.500775
	-	Zinc finger protein 207 (Zfp207)	Mm.102253
	Q9Y5S9	RNA-binding protein 8A (RNA binding motif protein 8A) (Ribonucleoprotein RBM8A) (RNA-binding protein Y14) (Binder of OVCA1-1) (BOV-1)	Hs.591455
	-	RNA binding motif protein 8a (Rbm8a)	Mm.261972
	Q8IYB3	Ser/Arg-related nuclear matrix protein	Hs.18192
	-	Serine/arginine repetitive matrix 1 (Srrm1)	Mm.1963
	Q15637	Splicing factor 1 (Zinc finger protein 162) (Transcription factor ZFM1) (Zinc finger gene in MEN1 locus) (Mammalian branch point binding protein mBBP) (BBP)	Hs.502829
	-	Splicing factor 1 (Sf1)	Mm.256422
	Q15042	Hypothetical protein KIAA0332 (U2-associated SR140 protein)	Hs.529577
	-	PAB1_MOUSE Polyadenylate-binding protein 1	Mm.292742
	P52756	RNA-binding protein 5 (RNA binding motif protein 5) (Putative tumor suppressor LUCA15) (G15 protein)	Hs.439480
	-	RNA binding motif protein 5 (Rbm5)	Mm.259197

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	Q16630	HPBRII-4 mRNA (HPBRII-7 protein)	Hs.369606
	-	Cleavage and polyadenylation specific factor 6 (Cpsf6)	Mm.288682
	Q96T58	Msx2-interacting protein (SMART/HDAC1 associated repressor protein)	Hs.558463
	-	Msx2 interacting nuclear target protein [Camundongos]	Mm.299906
	Q96T37	Putative RNA-binding protein 15 (RNA binding motif protein 15) (One-twenty two protein)	Hs.435947
	-	RNA binding motif protein 15 (Rbm15)	Mm.27966
	Q9P2S7	Cisplatin resistance-associated overexpressed protein	Hs.130293
	-	RIKEN cDNA 3300001P08 gene (3300001P08Rik)	Mm.30927
	Q06787	Fragile X mental retardation 1 protein (Protein FMR-1) (FMRP)	Hs.103183
	-	Fragile X mental retardation syndrome 1 homolog (Fmr1)	Mm.3451
	O00425	Putative RNA binding protein KOC (Koc1)	Hs.432616
	-	Insulin-like growth factor 2, binding protein 3 (Igf2bp3)	Mm.281018
	P29558	Single-stranded DNA-binding protein MSSP-1 (RNA binding motif, single-stranded interacting protein 1)	Hs.470412
	-	RNA binding motif, single stranded interacting protein 1 (Rbms1)	Mm.259667
	O15355	Protein phosphatase 2C gamma isoform (EC 3.1.3.16) (PP2C-gamma) (Protein phosphatase magnesium-dependent 1 gamma) (Protein phosphatase 1C)	Hs.17883
	-	Protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform (Ppm1g)	Mm.14501
	O95926	Hypothetical protein DKFZp564O2082 (GCIP-interacting protein p29)	Hs.20013
	-	SYF2 homolog, RNA splicing factor (S. cerevisiae) (Syf2)	Mm.29989
	Q96I25	Splicing factor 45 (45kDa splicing factor) (RNA binding motif protein 17)	Hs.498548
	-	RNA binding motif protein 17 (Rbm17)	Mm.182769
	Q14152	Eukaryotic translation initiation factor 3 subunit 10 (eIF-3 theta) (eIF3 p167) (eIF3 p180) (eIF3 p185) (eIF3a)	Hs.523299

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i>.
	-	Eukaryotic translation initiation factor 3, subunit 10 (theta) (Eif3s10)	<u>Mm.2238</u>
	<u>Q9P013</u>	HSPC148	<u>Hs.503597</u>
	-	RIKEN cDNA 0610040D20 gene (0610040D20Rik)	<u>Mm.245938</u>
	<u>Q9BQ61</u>	Hypothetical protein	<u>Hs.515155</u>
	-	RIKEN cDNA 2310036O22 gene (2310036O22Rik)	<u>Mm.196005</u>
	<u>Q9BXP5</u>	Arsenite-resistance protein 2	<u>Hs.111801</u>
	-	arsenate resistance protein 2 [Camundongos]	<u>Mm.387734</u>
	<u>Q9BRD0</u>	Hypothetical protein	<u>Hs.437341</u>
	-	Ser/Arg-related nuclear matrix protein; plenty-of-prolines-101; serine/arginine repetitive matrix protein 1 [Camundongos]	<u>Mm.32648</u>
	<u>P20042</u>	Eukaryotic translation initiation factor 2 subunit 2 (Eukaryotic translation initiation factor 2 beta subunit) (eIF-2-beta)	<u>Hs.429180</u>
	-	Eukaryotic translation initiation factor 2, subunit 2 (beta) (Eif2s2)	<u>Mm.377134</u>
	<u>Q9NW64</u>	Hypothetical protein FLJ10290	<u>Hs.591253</u>
	-	RNA binding motif protein 22 (Rbm22)	<u>Mm.275106</u>
	<u>O75940</u>	Survival of motor neuron-related splicing factor 30 (SMN-related protein) (30 kDa splicing factor SMNrp) (Survival motor neuron domain containing protein 1)	<u>Hs.79968</u>
	-	Survival motor neuron domain containing 1 (Smndc1)	<u>Mm.313687</u>
	<u>O75229</u>	R31449_3	<u>Hs.128425</u>
	-	Collagen alpha-1(II) chain precursor [Contains: Chondrocalcin].	<u>Mm.87628</u>
	<u>O95400</u>	CD2 antigen cytoplasmic tail-binding protein 2	<u>Hs.202677</u>
	-	Zinc finger protein 553 (Zfp553)	<u>Mm.18</u>
	<u>O43719</u>	HIV TAT specific factor 1	<u>Hs.204475</u>
	-	HIV TAT specific factor 1 (Htatsf1)	<u>Mm.2152</u>
	<u>Q9HCE1</u>	Potential helicase MOV-10 (EC 3.6.1.-) (Moloney leukemia virus 10 protein)	<u>Hs.514941</u>

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	Moloney leukemia virus 10 (Mov10)	Mm.1597
	Q9HCS7	XPA-binding protein 2 (HCNP protein) (PP3898)	Hs.9822
	-	XPA binding protein 2 (Xab2)	Mm.23739
	Q9H5H0	Hypothetical protein FLJ23445	Hs.288151
	-	Ngg1 interacting factor 3 like 1 binding protein 1 isoform 1	Mm.295875
	Q8WYA6	Beta-catenin-like protein 1 (Nuclear associated protein) (NAP) (NYD-SP19) (PP8304)	Hs.472667
	-	Catenin, beta like 1 (Ctnnb1)	Mm.45193
	Q8WWY3	U4/U6 snRNP-associated 61 kDa protein	Hs.515598
	-	PRP31 pre-mRNA processing factor 31 homolog (yeast) (Prpf31)	Mm.246863
	Q96DF8	DGCR14 protein (DiGeorge syndrome critical region 14) (ES2 protein)	Hs.517407
	-	Expressed sequence 2 embryonic lethal (Es2el)	Mm.256480
	Q9Y6A4	Transcription factor IIB (EVORF)	Hs.532755
	-	Gene trap locus 3 (Gtl3)	Mm.2080
	P43243	Matrin 3	Hs.268939
	-	Matrin 3 (Matr3)	Mm.215034
	Q92973	Transportin 1 (Importin beta-2) (Karyopherin beta-2) (M9 region interaction protein) (MIP)	Hs.482497
	-	Transportin 1 (Tnp1)	Mm.173286
	P61326	Mago nashi protein homolog	Hs.421576
	-	MGN_HUMAN Mago nashi protein homolog	Mm.808
	Q43684	Mitotic checkpoint protein BUB3	Hs.418533
	-	BUB3_MOUSE MITOTIC CHECKPOINT PROTEIN BUB3	Mm.927
	P55081	Microfibrillar-associated protein 1	Hs.61418
	-	microfibrillar-associated protein 1 [Camundongos]	Mm.393415
	Q8NI27	THO complex subunit 2 (Tho2)	Hs.592243
	-	THO complex 2 (Thoc2)	Mm.259498

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	Q9Y5B6	GC-rich sequence DNA-binding factor homolog	Hs.143835
	-		-
	O60306	KIAA0560 protein	Hs.510958
	-	regulator of nonsense transcripts 1 [Camundongos].	Mm.390048
	Q13123	Red protein (RER protein) (IK factor) (Cytokine IK)	Hs.421245
	-	IK cytokine (Ik)	Mm.30234
	Q12905	NF45 protein	Hs.75117
	-	Interleukin enhancer binding factor 2 (Ilf2)	Mm.227258
	Q99974	Pombe Cdc5-related protein (CDC5 cell division cycle 5-like) (S.pombe) (CDC5-like)	Hs.485471
	-	Cell division cycle 5-like (S. pombe) (Cdc5l)	Mm.28270
	P41223	G10 protein homolog (EDG-2)	Hs.380233
	-	Pentatricopeptide repeat domain 1 (Ptcd1)	Mm.277413
	P05455	Lupus La protein (Sjogren syndrome type B antigen) (SS-B) (La ribonucleoprotein) (La autoantigen)	Hs.445603
	-	LA_MOUSE Lupus LA protein homolog	Mm.10508
	Q9HCG8	KIAA1604 protein	Hs.311363
	-	simple repeat sequence-containing transcript [Camundongos]	Mm.288151
	Q969P6	DNA topoisomerase I, mitochondrial precursor (EC 5.99.1.2) (TOP1mt)	Hs.528574
	-	DNA topoisomerase 1, mitochondrial (Top1mt)	Mm.182401
	O75934	Putative spliceosome associated protein (DAM1 protein) (Breast carcinoma amplified sequence 2)	Hs.22960
	-	breast carcinoma amplified sequence 2 [Camundongos]	Mm.104919
	Q9UBB9	Tuftelin-interacting protein 11 (HSPC006)	Hs.20225
	-	tuftelin-interacting protein, 39 kD; tuftelin-interacting protein 33 [Camundongos]	Mm.172947
	O43809	Pre-mRNA cleavage factor I 25 kDa subunit (Cleavage and polyadenylation specific factor 5, 25 kD subunit)	Hs.528834

Cont/ Anexo 3 - Anotações gênicas dos fatores de *splicing*.

Grupo Funcional	Identificador SwissProt	Descrição	Unigene Clusters: <i>Homo sapiens</i> <i>Mus musculus</i> .
	-	cleavage and polyadenylation specific factor 5, 25 kD subunit; RIKEN cDNA 3110048P04 gene [Camundongos]	Mm.28961
	Q9P2B8	KIAA1429 protein	Hs.202238
	-	retinitis pigmentosa GTPase regulator interacting protein 1; 0610005A07Rik [Camundongos]	Mm.331487
	O43660	Pleiotropic regulator 1	Hs.249996
	-	Pleiotropic regulator 1, PRL1 homolog (Arabidopsis) (Plrg1)	Mm.286349
	Q9UMS4	Nuclear matrix protein NMP200 (PRP19/PSO4 homolog)	Hs.502705
	-	PRP19/PSO4 pre-mRNA processing factor 19 homolog (S. cerevisiae) (Prpf19)	Mm.358657
	Q96J01	THO complex subunit 3 (Tho3)	Hs.548868
	-	THO complex 3 (Thoc3)	Mm.292487
	Q9BU59	Homolog of C. elegans smu-1	Hs.591093
	-	Smu-1 suppressor of mec-8 and unc-52 homolog (C. elegans) (Smu1)	Mm.289929
	Q96FV9	THO complex subunit 1 (Tho1) (Nuclear matrix protein p84)	Hs.592342
	-	THO complex 1 (Thoc1)	Mm.219648
	Q86W42	MGC2655 protein	Hs.412304
	-	THO complex 6 homolog (Drosophila) (Thoc6)	Mm.328831
	P49768	Presenilin 1 (PS-1) (S182 protein)	Hs.592324
	-	Presenilin 1 (Psen1)	Mm.998
	P04720	Elongation factor 1-alpha 1 (EF-1-alpha-1) (Elongation factor 1 A-1) (eEF1A-1) (Elongation factor Tu) (EF-Tu)	Hs.439552
	-	EF11_MOUSE Elongation factor 1-alpha 1	Mm.311918
	P11940	Polyadenylate-binding protein 1 (Poly(A)-binding protein 1) (PABP 1)	Hs.387804
	-	PAB1_MOUSE Polyadenylate-binding protein 1	Mm.371570

Anexo 4 - Conjunto de 211 *clusters* gênicos ortólogos curados: fatores de *splicing* e proteínas envolvidas no processo de *splicing* alternativo.

211 CLUSTERS GÊNICOS ORTÓLOGOS	
HUMANOS	CAMUNDONGOS
Hs.159014	Mm.10027
Hs.500775	Mm.102253
Hs.3530	Mm.10229
Hs.177861	Mm.102627
Hs.22960	Mm.104919
Hs.445603	Mm.10508
Hs.104865	Mm.111904
Hs.121663	Mm.121014
Hs.524690	Mm.126873
Hs.301404	Mm.128512
Hs.280378	Mm.1323
Hs.569122	Mm.141197
Hs.522257	Mm.142872
Hs.17883	Mm.14501
Hs.443861	Mm.15252
Hs.567561	Mm.152689
Hs.282901	Mm.153895
Hs.487774	Mm.155896
Hs.528306	Mm.156035
Hs.406277	Mm.156914
Hs.514941	Mm.1597
Hs.103106	Mm.165735
Hs.465498	Mm.172411
Hs.20225	Mm.172947
Hs.433732	Mm.1761
Hs.501309	Mm.17898
Hs.171909	Mm.180953
Hs.528574	Mm.182401
Hs.54649	Mm.182650
Hs.498548	Mm.182769
Hs.591108	Mm.18535
Hs.527105	Mm.195310
Hs.515155	Mm.196005
Hs.18192	Mm.1963
Hs.406423	Mm.196532

Hs.592175	Mm.196598
Hs.533245	Mm.202725
Hs.484288	Mm.205045
Hs.466917	Mm.20913
Hs.533122	Mm.210352
Hs.166463	Mm.2115
Hs.268939	Mm.215034
Hs.204475	Mm.2152
Hs.246112	Mm.215860
Hs.467097	Mm.216386
Hs.202166	Mm.21740
Hs.584801	Mm.21841
Hs.592342	Mm.219648
Hs.516160	Mm.219671
Hs.279806	Mm.220038
Hs.136947	Mm.221440
Hs.523299	Mm.2238
Hs.479693	Mm.223946
Hs.75117	Mm.227258
Hs.514435	Mm.236123
Hs.546271	Mm.236513
Hs.570079	Mm.23705
Hs.9822	Mm.23739
Hs.406557	Mm.239354
Hs.6891	Mm.24042
Hs.512661	Mm.241546
Hs.585488	Mm.242072
Hs.503597	Mm.245938
Hs.111632	Mm.246693
Hs.515598	Mm.246863
Hs.469970	Mm.2478
Hs.515255	Mm.248188
Hs.171342	Mm.248755
Hs.334612	Mm.249110
Hs.13366	Mm.250783
Hs.589962	Mm.251548
Hs.438587	Mm.253614
Hs.155218	Mm.254223
Hs.424908	Mm.25642
Hs.502829	Mm.256422
Hs.517407	Mm.256480

Hs.584748	Mm.25720
Hs.355934	Mm.257276
Hs.591637	Mm.257474
Hs.473583	Mm.258204
Hs.439480	Mm.259197
Hs.470412	Mm.259667
Hs.416108	Mm.260516
Hs.571177	Mm.260545
Hs.591455	Mm.261972
Hs.348342	Mm.265415
Hs.172550	Mm.265610
Hs.585782	Mm.270331
Hs.443465	Mm.270511
Hs.38628	Mm.271056
Hs.546550	Mm.271174
Hs.493516	Mm.271985
Hs.474049	Mm.272803
Hs.2853	Mm.274146
Hs.508848	Mm.274690
Hs.564847	Mm.274995
Hs.591253	Mm.275106
Hs.446091	Mm.275521
Hs.516076	Mm.276802
Hs.380233	Mm.277413
Hs.435947	Mm.27966
Hs.471011	Mm.279736
Hs.11776	Mm.279872
Hs.423935	Mm.279907
Hs.202677	Mm.28050
Hs.432616	Mm.281018
Hs.469173	Mm.281900
Hs.435342	Mm.28200
Hs.311609	Mm.28222
Hs.485471	Mm.28270
Hs.380118	Mm.28275
Hs.249996	Mm.286349
Hs.20930	Mm.286394
Hs.432485	Mm.286408
Hs.190520	Mm.28694
Hs.369624	Mm.287826
Hs.73986	Mm.288098

Hs.311363	Mm.288151
Hs.369606	Mm.288682
Hs.285197	Mm.288728
Hs.528834	Mm.28961
Hs.380774	Mm.289662
Hs.591093	Mm.289929
Hs.180414	Mm.290774
Hs.274531	Mm.291029
Hs.31334	Mm.292001
Hs.309090	Mm.292016
Hs.548868	Mm.292487
Hs.529577	Mm.292742
Hs.269944	Mm.29495
Hs.288151	Mm.295875
Hs.528305	Mm.29644
Hs.433540	Mm.29685
Hs.124490	Mm.297078
Hs.182255	Mm.299312
Hs.546261	Mm.299367
Hs.15277	Mm.29952
Hs.221889	Mm.299604
Hs.521924	Mm.29965
Hs.591430	Mm.29966
Hs.20013	Mm.29989
Hs.425311	Mm.30198
Hs.421245	Mm.30234
Hs.256639	Mm.304080
Hs.374973	Mm.30660
Hs.1063	Mm.308514
Hs.130293	Mm.30927
Hs.573762	Mm.31051
Hs.465808	Mm.311439
Hs.439552	Mm.311918
Hs.79968	Mm.313687
Hs.444520	Mm.315652
Hs.512635	Mm.316928
Hs.808	Mm.317706
Hs.166109	Mm.318042
Hs.439440	Mm.319684
Hs.465885	Mm.325205
Hs.437341	Mm.32648

Hs.412304	Mm.328831
Hs.27693	Mm.328928
Hs.522392	Mm.330160
Hs.202238	Mm.331487
Hs.254042	Mm.33263
Hs.519347	Mm.33908
Hs.121076	Mm.340195
Hs.568331	Mm.34296
Hs.103183	Mm.3451
Hs.502883	Mm.34562
Hs.405144	Mm.358634
Hs.502705	Mm.358657
Hs.33104	Mm.358668
Hs.528007	Mm.360389
Hs.387804	Mm.371570
Hs.181368	Mm.3757
Hs.429180	Mm.377134
Hs.512610	Mm.379101
Hs.365116	Mm.379289
Hs.516539	Mm.379375
Hs.480073	Mm.384474
Hs.161181	Mm.38529
Hs.466775	Mm.386890
Hs.111801	Mm.387734
Hs.510958	Mm.390048
Hs.591357	Mm.390303
Hs.73965	Mm.390345
Hs.96996	Mm.390606
Hs.485060	Mm.390986
Hs.471768	Mm.41865
Hs.184492	Mm.422763
Hs.33962	Mm.423019
Hs.166975	Mm.43331
Hs.26047	Mm.44292
Hs.356549	Mm.45151
Hs.472667	Mm.45193
Hs.68714	Mm.45645
Hs.565094	Mm.45683
Hs.130098	Mm.45725
Hs.433343	Mm.5222
Hs.464734	Mm.603

Hs.520028	Mm.6388
Hs.445497	Mm.64579
Hs.496829	Mm.72753
Hs.421576	Mm.808
Hs.528763	Mm.821
Hs.151787	Mm.873
Hs.128425	Mm.87628
Hs.589594	Mm.9043
Hs.418533	Mm.927
Hs.121432	Mm.98910
Hs.5683	Mm.993
Hs.592324	Mm.998

Anexo 5 - Conjunto final de 124 fatores de *splicing* ortólogos.

124 FATORES DE <i>SPLICING</i>	
CLK	Hs.433732
CLK	Hs.73986
CRN	Hs.171342
Cyclophilins	Hs.121076
Cyclophilins	Hs.27693
DEAD	Hs.254042
DEAD	Hs.274531
DEAD	Hs.279806
DEAD	Hs.311609
DEAD	Hs.439440
DEAD	Hs.484288
DEAD	Hs.485060
DEAD	Hs.496829
DEAD	Hs.570079
FBP	Hs.521924
FUSE	Hs.568331
hnRNP	Hs.136947
hnRNP	Hs.155218
hnRNP	Hs.166463
hnRNP	Hs.172550
hnRNP	Hs.202166
hnRNP	Hs.380118
hnRNP	Hs.432485
hnRNP	Hs.445497
hnRNP	Hs.465808
hnRNP	Hs.480073
hnRNP	Hs.508848
hnRNP	Hs.516539
hnRNP	Hs.546261
hnRNP	Hs.546271
hnRNP	Hs.571177
hnRNP	Hs.589594
hnRNP	Hs.808
hnRNP	Hs.96996
Others	Hs.18192
Others	Hs.20225

Others	Hs.249996
Others	Hs.421576
Others	Hs.466917
Others	Hs.498548
Others	Hs.502829
Others	Hs.510958
Others	Hs.515598
Others	Hs.548868
Others	Hs.9822
PRP18	Hs.161181
Skip	Hs.546550
SLU7	Hs.435342
Sm	Hs.103106
Sm	Hs.111632
Sm	Hs.190520
Sm	Hs.334612
Sm	Hs.356549
Sm	Hs.464734
Sm	Hs.512610
Sm	Hs.515255
Sm	Hs.516076
Sm	Hs.565094
SR	Hs.166975
SR	Hs.369624
SR	Hs.469970
SR	Hs.519347
SR	Hs.533122
SR	Hs.54649
SR	Hs.592175
SR	Hs.68714
SR	Hs.6891
SRPK	Hs.285197
U11+U12	Hs.15277
U11+U12	Hs.512635
U1+U2	Hs.1063
U1+U2	Hs.280378
U1+U2	Hs.33104
U1+U2	Hs.406277
U1+U2	Hs.406423
U1+U2	Hs.466775

U1+U2	Hs.467097
U1+U2	Hs.514435
U1+U2	Hs.516160
U1+U2	Hs.528763
U2AF	Hs.365116
U2AF	Hs.528007
U4/U6	Hs.256639
U4/U6	Hs.374973
U4/U6.U5	Hs.182255
U4/U6.U5	Hs.502883
U5	Hs.130098
U5	Hs.151787
U5	Hs.181368
U5	Hs.246112
U5	Hs.31334
U5	Hs.33962
CELF/CUG-BP	Hs.348342
CELF/CUG-BP	Hs.567561
CLK	Hs.406557
DEAD	Hs.444520
DEAD	Hs.528305
DEAD	Hs.533245
hnRNP	Hs.121663
hnRNP	Hs.2853
hnRNP	Hs.487774
hnRNP	Hs.527105
hnRNP	Hs.573762
hnRNP	Hs.591357
Others	Hs.128425
Others	Hs.20013
Others	Hs.282901
Others	Hs.485471
Others	Hs.502705
Others	Hs.512661
P52/P100	Hs.355934
Sm	Hs.424908
Sm	Hs.425311
Sm	Hs.564847
SR	Hs.405144
SR	Hs.479693

SR	Hs.584801
SRPK	Hs.443861
U11+U12	Hs.13366
U1+U2	Hs.177861
U1+U2	Hs.471011
U4/U6	Hs.11776
U4/U6.U5	Hs.469173
U5	Hs.465498

Anexo 6 - 88 fatores de *splicing* com expressões gênicas de MPSS com valores superiores em bibliotecas tumorais *versus* normais (mais do que 3 vezes)

hnRNP	Hs.480073
DEAD	Hs.311609
U1+U2	Hs.33104
U11+U12	Hs.13366
Others	Hs.421576
hnRNP	Hs.527105
hnRNP	Hs.166463
SR	Hs.533122
hnRNP	Hs.808
Sm	Hs.464734
Cyclophilins	Hs.27693
U1+U2	Hs.1063
Sm	Hs.425311
hnRNP	Hs.546271
U4/U6	Hs.11776
U5	Hs.31334
U1+U2	Hs.466775
Cyclophilins	Hs.121076
FBP	Hs.521924
Others	Hs.18192
P52/P100	Hs.355934
U4/U6	Hs.374973
U5	Hs.130098
SR	Hs.6891
Sm	Hs.516076
hnRNP	Hs.546261
SLU7	Hs.435342
Others	Hs.20013
DEAD	Hs.484288
SR	Hs.584801
Others	Hs.502705
hnRNP	Hs.589594
U5	Hs.246112
Sm	Hs.111632
SR	Hs.166975
U2AF	Hs.528007

FUSE	Hs.568331
CLK	Hs.433732
Others	Hs.466917
hnRNP	Hs.380118
SR	Hs.469970
U5	Hs.181368
Sm	Hs.103106
DEAD	Hs.533245
SR	Hs.54649
DEAD	Hs.274531
U1+U2	Hs.516160
SR	Hs.519347
Sm	Hs.356549
DEAD	Hs.254042
hnRNP	Hs.155218
Skip	Hs.546550
U1+U2	Hs.280378
CRN	Hs.171342
hnRNP	Hs.96996
U1+U2	Hs.406423
DEAD	Hs.528305
U5	Hs.33962
U1+U2	Hs.528763
SR	Hs.405144
hnRNP	Hs.2853
SR	Hs.479693
Sm	Hs.424908
CLK	Hs.73986
Sm	Hs.515255
Others	Hs.515598
U1+U2	Hs.177861
U4/U6.U5	Hs.469173
hnRNP	Hs.202166
DEAD	Hs.570079
hnRNP	Hs.573762
hnRNP	Hs.136947
hnRNP	Hs.571177
Others	Hs.282901
DEAD	Hs.279806
Others	Hs.249996

hnRNP	Hs.591357
U5	Hs.151787
Others	Hs.548868
U4/U6.U5	Hs.182255
U2AF	Hs.365116
hnRNP	Hs.487774
SRPK	Hs.443861
CLK	Hs.406557
hnRNP	Hs.516539
Others	Hs.485471
SR	Hs.592175
hnRNP	Hs.172550

Anexo 7 - 72 fatores de *splicing* com expressões gênicas de MPSS com valores superiores em bibliotecas tumorais *versus* normais (mais do que 5 vezes)

DEAD	Hs.311609
U1+U2	Hs.33104
Others	Hs.421576
hnRNP	Hs.527105
hnRNP	Hs.166463
hnRNP	Hs.808
Sm	Hs.464734
Cyclophilins	Hs.27693
hnRNP	Hs.546271
U4/U6	Hs.11776
U1+U2	Hs.466775
Cyclophilins	Hs.121076
Others	Hs.18192
P52/P100	Hs.355934
U4/U6	Hs.374973
U5	Hs.130098
Sm	Hs.516076
SLU7	Hs.435342
DEAD	Hs.484288
SR	Hs.584801
Others	Hs.502705
hnRNP	Hs.589594
Sm	Hs.111632
SR	Hs.166975
U2AF	Hs.528007
FUSE	Hs.568331
CLK	Hs.433732
Others	Hs.466917
SR	Hs.469970
Sm	Hs.103106
DEAD	Hs.533245
SR	Hs.54649
DEAD	Hs.274531
U1+U2	Hs.516160
SR	Hs.519347
Sm	Hs.356549
DEAD	Hs.254042

hnRNP	Hs.155218
Skip	Hs.546550
U1+U2	Hs.280378
CRN	Hs.171342
U1+U2	Hs.406423
DEAD	Hs.528305
U5	Hs.33962
U1+U2	Hs.528763
SR	Hs.405144
hnRNP	Hs.2853
SR	Hs.479693
Sm	Hs.424908
CLK	Hs.73986
Sm	Hs.515255
Others	Hs.515598
U1+U2	Hs.177861
U4/U6.U5	Hs.469173
hnRNP	Hs.202166
DEAD	Hs.570079
hnRNP	Hs.573762
hnRNP	Hs.136947
hnRNP	Hs.571177
Others	Hs.282901
DEAD	Hs.279806
Others	Hs.249996
hnRNP	Hs.591357
U5	Hs.151787
Others	Hs.548868
U4/U6.U5	Hs.182255
U2AF	Hs.365116
hnRNP	Hs.487774
SRPK	Hs.443861
hnRNP	Hs.516539
Others	Hs.485471
SR	Hs.592175

Anexo 8 - Fatores de *splicing* humanos com presença indicativa em padrões de eventos de *splicing* alternativos.

	Conjunto de <i>clusters</i> gênicos dos fatores de <i>splicing</i> presentes em eventos de <i>splicing</i> alternativos					
	Bibliotecas Normais			Bibliotecas Tumorais		
Glândula mamária	IR	ES	ASS	IR	ES	ASS
			Hs.498548	Hs.570079	Hs.502883	Hs.808
			Hs.181368	Hs.467097	Hs.406423	Hs.502829
				Hs.136947	Hs.516539	Hs.406423
				Hs.484288		Hs.502883
				Hs.519347		Hs.546261
				Hs.254042		Hs.508848
				Hs.487774		Hs.467097
						Hs.516539
						Hs.484288
						Hs.254042
Pulmão	IR	ES	ASS	IR	ES	ASS
	Hs.498548	Hs.161181	Hs.161181	Hs.808	Hs.546261	Hs.546261
	Hs.546271	Hs.508848	Hs.502705	Hs.508848	Hs.508848	Hs.181368
	Hs.565094	Hs.564847	Hs.130098	Hs.282901	Hs.465498	Hs.151787
	Hs.584801	Hs.570079	Hs.546261	Hs.480073	Hs.512610	Hs.589594
	Hs.465808	Hs.485471	Hs.508848	Hs.484288	Hs.445497	Hs.515598
	Hs.467097		Hs.564847	Hs.254042	Hs.285197	Hs.282901
	Hs.528007		Hs.570079	Hs.496829	Hs.496829	Hs.356549
	Hs.172550		Hs.514435			Hs.249996
	Hs.516160		Hs.565094			Hs.484288
	Hs.6891		Hs.528007			Hs.254042
	Hs.445497		Hs.479693			Hs.487774
	Hs.202166		Hs.256639			
	Hs.27693		Hs.20013			
	Hs.1063		Hs.282901			
			Hs.136947			
			Hs.445497			
			Hs.54649			
			Hs.202166			
			Hs.254042			
Fígado	IR	ES	ASS	IR	ES	ASS
	Hs.279806	Hs.498548	Hs.406423	Hs.808	Hs.466775	Hs.808
	Hs.479693	Hs.256639	Hs.546271	Hs.498548	Hs.516539	Hs.502705
	Hs.405144	Hs.487774	Hs.479693	Hs.546271	Hs.432485	Hs.406423
	Hs.485471		Hs.73986	Hs.151787		Hs.502829
	Hs.487774		Hs.282901	Hs.279806		Hs.502883
			Hs.31334	Hs.465808		Hs.546271
			Hs.433732	Hs.466917		Hs.546261
			Hs.249996	Hs.282901		Hs.33104
			Hs.274531	Hs.527105		Hs.546550
			Hs.405144	Hs.254042		Hs.151787
			Hs.487774	Hs.405144		Hs.279806
			Hs.285197	Hs.521924		Hs.467097
			Hs.496829	Hs.432485		Hs.589594

Cont/ Anexo 8 - Fatores de *splicing* humanos com presença indicativa em padrões de eventos de *splicing* alternativos.

Fígado	IR	ES	ASS	IR	ES	ASS
						Hs.256639
						Hs.406277
						Hs.516539
						Hs.485060
						Hs.425311
						Hs.374973
Próstata	IR	ES	ASS	IR	ES	ASS
	Hs.406423	Hs.508848	Hs.502705	Hs.151787	Hs.508848	Hs.508848
		Hs.564847	Hs.508848	Hs.9822	Hs.151787	Hs.584801
		Hs.466775	Hs.564847	Hs.172550	Hs.18192	Hs.465498
		Hs.31334	Hs.466917	Hs.480073	Hs.6891	Hs.512610
		Hs.528305	Hs.466775	Hs.487774		Hs.18192
			Hs.31334			Hs.166463
			Hs.365116			Hs.6891
			Hs.528305			Hs.487774
			Hs.484288			Hs.425311
Sistema nervoso	IR	ES	ASS	IR	ES	ASS
	Hs.546261	Hs.508848	Hs.502705	Hs.546261	Hs.161181	Hs.161181
	Hs.508848	Hs.546550	Hs.130098	Hs.528007	Hs.502705	Hs.502705
	Hs.528763	Hs.564847	Hs.33104	Hs.516160	Hs.130098	Hs.406423
	Hs.68714	Hs.528763	Hs.546261	Hs.166463	Hs.546550	Hs.546271
	Hs.9822	Hs.465808	Hs.546550	Hs.355934	Hs.508848	Hs.33104
	Hs.480073	Hs.73986	Hs.528763	Hs.282901	Hs.564847	Hs.546261
	Hs.484288	Hs.355934	Hs.564847	Hs.512661	Hs.466917	Hs.546550
	Hs.496829	Hs.528305	Hs.68714	Hs.527105	Hs.469970	Hs.508848
		Hs.516076	Hs.279806	Hs.480073	Hs.433732	Hs.514435
		Hs.54649	Hs.181368	Hs.202166	Hs.533122	Hs.466917
		Hs.533122	Hs.467097	Hs.484288	Hs.480073	Hs.515598
		Hs.527105	Hs.466917		Hs.533245	Hs.479693
		Hs.405144	Hs.465808		Hs.484288	Hs.166463
		Hs.254042	Hs.515255		Hs.521924	Hs.136947
			Hs.9822			Hs.31334
			Hs.515598			Hs.528305
			Hs.589594			Hs.54649
			Hs.479693			Hs.445497
			Hs.355934			Hs.433732
			Hs.280378			Hs.527105
			Hs.365116			Hs.484288
			Hs.182255			Hs.254042
			Hs.528305			Hs.487774
			Hs.406277			
			Hs.177861			
			Hs.533122			
			Hs.249996			
			Hs.527105			
			Hs.190520			

Cont/ Anexo 8 - Fatores de *splicing* humanos com presença indicativa em padrões de eventos de *splicing* alternativos.

Sistema nervoso	IR	ES	ASS	IR	ES	ASS
			Hs.202166			
			Hs.1063			
			Hs.103106			
			Hs.254042			
			Hs.571177			
			Hs.487774			
			Hs.521924			
			Hs.121663			
			Hs.432485			
Pele	IR	ES	ASS	IR	ES	ASS
	Hs.528007	Hs.546261	Hs.546271	Hs.151787	Hs.498548	Hs.406423
	Hs.433732	Hs.546271	Hs.546261	Hs.515598	Hs.130098	Hs.130098
	Hs.202166	Hs.172550	Hs.589594	Hs.155218	Hs.508848	Hs.546261
		Hs.334612	Hs.528007	Hs.166463	Hs.515598	Hs.508848
		Hs.136947	Hs.172550	Hs.282901	Hs.166463	Hs.151787
		Hs.246112	Hs.11776	Hs.480073	Hs.282901	Hs.181368
			Hs.479693	Hs.254042	Hs.444520	Hs.279806
			Hs.516076	Hs.487774	Hs.516539	Hs.465808
			Hs.246112		Hs.516076	Hs.515598
			Hs.202166		Hs.512661	Hs.466917
			Hs.254042		Hs.103106	Hs.528007
					Hs.487774	Hs.155218
					Hs.496829	Hs.466775
						Hs.311609
						Hs.166463
						Hs.136947
						Hs.282901
						Hs.20225
						Hs.516539
						Hs.516076
						Hs.527105
						Hs.487774
						Hs.521924
Cólon	IR	ES	ASS	IR	ES	ASS
		Hs.546261	Hs.546261	Hs.546261	Hs.502705	Hs.502705
			Hs.443861	Hs.166463	Hs.546261	Hs.502883
				Hs.136947	Hs.508848	Hs.546261
				Hs.519347	Hs.514435	Hs.546271
				Hs.202166	Hs.151787	Hs.508848
					Hs.311609	Hs.514435
					Hs.480073	Hs.570079
					Hs.254042	Hs.181368
						Hs.68714
						Hs.151787
						Hs.172550
						Hs.311609
						Hs.18192

Cont/ Anexo 8 - Fatores de *splicing* humanos com presença indicativa em padrões de eventos de *splicing* alternativos.

Cólon	IR	ES	ASS	IR	ES	ASS
						Hs.136947
						Hs.246112
						Hs.469173
						Hs.480073
						Hs.202166
						Hs.519347
						Hs.254042
						Hs.487774

IR – Retenção de íntron; ES – Uso alternativo do Éxon; ASS – Sítios doadores e/ou aceptores alternativos de *splicing*.