

**UNIVERSIDADE DE SÃO PAULO**  
**INSTITUTO DE QUÍMICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA**

**RODRIGO LOURO**

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE  
TRANSCRITOS HUMANOS: novas famílias de  
pequenas GTPases e novos Longos RNAs  
Intrônicos não-codificantes**

São Paulo  
Data do Depósito na SPG:  
24/10/2006

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

RODRIGO LOURO

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE  
TRANSCRITOS HUMANOS: novas famílias de  
pequenas GTPases e novos Longos RNAs  
Intrônicos não-codificantes**

*Tese apresentada ao Instituto de Química da  
Universidade de São Paulo para obtenção do  
Título de Doutor em Bioquímica*

*Orientador: Prof. Dr. Sergio Verjovski-Almeida*

São Paulo  
2006

Rodrigo Louro

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE TRANSCRITOS HUMANOS: novas famílias de pequenas GTPases e novos Longos RNAs Intrônicos não-codificantes.

*Tese apresentada ao Instituto de Química da  
Universidade de São Paulo para obtenção do  
Título de Doutor em Bioquímica*

Aprovado em: \_\_\_\_\_

**Banca Examinadora**

**Prof. Dr.** \_\_\_\_\_

Instituição: \_\_\_\_\_

Assinatura: \_\_\_\_\_

*Aos meus queridos pais, Percival e Izabel, que com amor extremo e incondicional proporcionam aos seus filhos TUDO para transformar sonhos em realidade.  
Devo tudo que sou a vocês.*

## **AGRADECIMENTOS**

Ao Prof. Dr. Sergio Verjovski-Almeida, e ao Prof. Dr. Eduardo Reis, pelo apoio, incentivo e orientação constantes desde o início de minhas atividades científicas. Ambos são os responsáveis pela minha formação como pesquisador.

À Profa. Dra. Adele Caterino-de-Araújo, minha madrinha, por servir como modelo de dedicação e sucesso nessa difícil, porém apaixonante, profissão.

Ao Prof. Dr. Bayardo Baptista, que abriu meus olhos para a beleza dos fenômenos bioquímicos, e que tem me inspirado a não deixar a importante função de ensinar em segundo plano.

À Profa. Dra. Aline Maria da Silva pelas importantes colaborações nesse trabalho, pela utilização dos equipamentos de seu laboratório, e pela fundamental orientação para formatação dessa tese.

Agradeço também à Profa. Dra. Mari Cleide Sogayar, importante colaboradora do grupo, em especial pelas críticas ao trabalho de análise de expressão gênica após estímulo por andrógeno.

Aos parceiros próximos de projeto (Helder Nakaya, Paulo Paiva Amaral e Fernanda Festa) com quem compartilhei diversos momentos de entusiasmo e de decepção proporcionados pela ciência nesses últimos anos.

Aos grandes amigos gaúchos, Tarik e Ângela, conquistados durante a pós-graduação. Agradeço pelos novos conhecimentos (científicos e culturais) adquiridos, e por serem ótimos companheiros de trabalho, de cafezinhos, de cervejinhas e de poker.

À Ana Paula Lopes Vidal que tem facilitado (e muito) minha relação com a FAPESP e com as empresas que fabricam e vendem reagentes para pesquisa. Sem ela a rotina de laboratório seria muito mais estressante.

Aos alunos de iniciação científica que contribuíram para a obtenção dos resultados apresentados aqui, em especial ao Rafael e Yuri (agora doutorandos), mas também ao Junio Cota Silva, ao Vinícius Coutinho e à Bárbara Camargo.

Aos demais colegas de laboratório (Glauber, Thiago, Renato, Jefferson, Camila, Giulliana, Kátia, Felipe, Lauren, Rogério, Ana Carolina Tahira, Ana Carolina Ayupe, Ana Cláudia, Denise, Adriana, Julio e Ricardo) e do departamento de bioquímica (em especial, Vanessa Falcão, Camila Carrião, e Osmar), que tive a sorte de conhecer e de dividir boas experiências tanto no trabalho quanto na vida “real”.

Ao Dr. Luiz Otavio Penalva, professor assistente da UTHSCSA, e seu grupo de pesquisa (Deepa, Patrícia Sanchez e Suzanne Burns) pela oportunidade de aprender novas técnicas e conceitos, de superar antigos temores, e pela agradável recepção no “inóspito” estado do Texas.

À Carina Victória e ao Carlos Alexandre, cuja amizade que se iniciou durante a “inocente” época de graduação, passou pela conturbada fase da pós-graduação e deve prosseguir pela vida toda. Obrigado pelo apoio nos momentos difíceis, pela manutenção das tradições, e pelas constantes discussões científicas livres de embasamento experimental.

À minha querida irmã Carolina Louro, ao meu grande irmão Alexandre Miranda, e aos demais amigos não-biólogos mais próximos (Paulão, Paulo “Feijão”, entre outros) que admiram meu trabalho mesmo achando que fazer pesquisa é a maior moleza.

Aos meus demais familiares, em especial ao meu padrinho Carlos, às minhas tias Marisa e Marilene e à minha falecida avó Amélia pelas constantes demonstrações de carinho e de confiança no meu potencial.

E finalmente, ao Morihei Ueshiba (*O-Sensei*), criador e fundador do Aikido. Praticando essa fantástica arte marcial consegui atingir um novo equilíbrio em minha vida, que evitou a concretização literal e integral da epígrafe contida na página seguinte.

*"I have put my heart and soul into this work...and lost my mind in the process."*

Vincent Van Gogh

## RESUMO

Louro, R. **Identificação e caracterização de transcritos humanos: novas famílias gênicas de pequenas GTPases e novos Longos RNAs Intrônicos não-codificantes.** 2006. 202 f. Tese (Doutorado) - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

Terminado o sequenciamento do genoma humano, as atenções se voltaram para a determinação do conjunto completo de transcritos humanos. Diversos trabalhos sugerem que enquanto apenas uma pequena fração de mRNAs codificantes para proteína não é conhecida, existe um grande número de RNAs não-codificantes (ncRNAs) ainda não caracterizados. Nesse contexto, o presente trabalho visou explorar as informações de expressão gênica contidas em ESTs para identificar e caracterizar novos transcritos humanos. A busca genômica por membros de famílias gênicas relacionadas com câncer levou a identificação de novas pequenas GTPases, destacando uma subfamília que deve apresentar função supressora tumoral em próstata. Uma classe de ncRNAs longos, sem *splicing*, expressos antisense a partir de regiões intrônicas foi descrita utilizando plataformas de *microarrays*, construídas pelo grupo, enriquecidas com seqüências sem anotação. O perfil de expressão de 23 ncRNAs intrônicos estava significativamente correlacionado com o grau de diferenciação de tumores de próstata (*Gleason Score*), e pode ser utilizado como candidato a marcador molecular de prognóstico. Um total de 39 ncRNAs intrônicos responderam à estimulação por andrógeno, apontando para um mecanismo regulatório da expressão intrônica por sinais fisiológicos hormonais. A biogênese da expressão intrônica parece ser complexa, pois uma fração não é transcrita pela RNA Polimerase II. A transcrição intrônica estava correlacionada com uso de exons em células tratadas com andrógeno.

Assinaturas de expressão intrônica conservadas em tecidos humanos e de camundongos, e interações de transcritos intrônicos com proteínas regulatórias foram observadas. Este trabalho contribui com novas e originais evidências que dão apoio ao papel postulado para esses ncRNAs no controle fino do programa transcricional humano.

Palavras-chave: Análise de expressão gênica, Pequenas GTPases monoméricas, Câncer de próstata, RNAs não-codificantes, Introns, Andrógeno, Controle da expressão gênica.

## ABSTRACT

Louro, R. **Identification and characterization of human transcripts: novel small GTPase gene families and novel Long Intronic non-coding RNAs**. 2006. 202 p. Thesis (Doctoral) - Biochemistry Graduate Program. Instituto de Química. Universidade de São Paulo.

With the completion of the human genome sequence, attention has shifted towards determining the complete set of human transcripts. Multiple lines of evidence suggest that while only a small fraction of protein-coding mRNAs remains to be described, there is a huge amount of uncharacterized non-coding RNAs (ncRNAs). In this context, the present work sought to explore the gene expression information provided by ESTs to identify and characterize new human transcripts. A genomic-wide search for cancer related gene family members identified novel small GTPase genes, and highlighted an uncharacterized subfamily that may have a tumor suppressor role in prostate cancer. A class of long unspliced ncRNAs, expressed antisense from introns of protein-coding genes was described using custom-designed microarray platforms enriched with unannotated sequences. The expression profile of 23 intronic ncRNAs was significantly correlated to the degree of prostate tumor differentiation (Gleason Score), and could be used as a candidate prognostic molecular marker. A total of 39 intronic ncRNAs were responsive to androgen stimulation, pointing to a mechanism of intronic expression regulation by physiological hormone signals. Intronic ncRNA biogenesis seems to be complex, since a fraction of them is not transcribed by RNA Polymerase II. Intronic transcription was correlated to exon usage in androgen treated cells. Tissue expression signatures of intronic transcription were conserved in human and mouse, and intronic transcripts were found to interact with regulatory

proteins. This work provides new and original contributions that support the postulated role of ncRNAs in the fine tuning of the human transcriptional program.

Keywords: Gene expression analysis, Small monomeric GTPases, Prostate cancer, Non-coding RNAs, Introns, Androgen, Gene expression control.

## LISTA DE ABREVIATURAS E SIGLAS

<b>ARE</b>	Elemento de Resposta a Andrógeno
<b>ATCC</b>	<i>American Type Culture Collection</i>
<b>BLAST</b>	<i>Basic Local Alignment Search Tool</i>
<b>bp</b>	Pares de bases
<b>CAGE</b>	Cooperação para Análise da Expressão Gênica
<b>CBC</b>	<i>Cap Binding Complex</i>
<b>CGAP</b>	<i>Cancer Genome Anatomy Project</i>
<b>ChIP</b>	Imunoprecipitação da cromatina
<b>C<sub>T</sub></b>	<i>Cycle Threshold</i>
<b>ESE</b>	<i>Exonic Splicing Enhancer</i>
<b>EST</b>	<i>Expressed Sequence Tag</i>
<b>FANTOM</b>	<i>Functional Annotation of the Mouse Genome</i>
<b>GEO</b>	<i>Gene Expression Omnibus</i>
<b>GO</b>	<i>Gene Ontology</i>
<b>GS</b>	<i>Gleason Score</i>
<b>GTF</b>	<i>General Transcription Factor</i>
<b>HCGP</b>	<i>Human Cancer Genome Project</i>
<b>IgG</b>	Imunoglobulina G
<b>IPTG</b>	<i>isopropyl-beta-D-thiogalactopyranoside</i>
<b>ISE</b>	<i>Intronic Splicing Enhancer</i>
<b>ISS</b>	<i>Intronic Splicing Silencer</i>
<b>kb</b>	Quilobase
<b>LCR</b>	<i>Locus Control Region</i>
<b>MPSS</b>	<i>Massively Parallel Signature Sequencing</i>
<b>mRNA</b>	RNA mensageiro
<b>NCBI</b>	<i>National Center for Biotechnology Information</i>
<b>ncRNA</b>	RNA não-codificante para proteínas
<b>nt</b>	Nucleotídeos
<b>ORF</b>	<i>Open reading frame</i>
<b>OSRT</b>	<i>Orientation Specific Reverse Transcription</i>
<b>PCR</b>	Reação em Cadeia da Polimerase
<b>PIC</b>	<i>Preiniciation Complex</i>
<b>PMSF</b>	Fluoreto de fenilmetilsulfonil
<b>qPCR</b>	PCR quantitativo

<b>RACE</b>	<i>Rapid Amplification of cDNA Ends</i>
<b>RBP</b>	<i>RNA Binding Protein</i>
<b>RefSeq</b>	<i>Reference Sequence</i>
<b>RNAP</b>	<i>RNA Polimerase</i>
<b>RNP</b>	<i>Ribonucleoprotein</i>
<b>SAGE</b>	<i>Serial Analysis of Gene Expression</i>
<b>SAM</b>	<i>Significance Analysis of Microarrays</i>
<b>SDS</b>	<i>Dodecilsulfato de sódio</i>
<b>TF</b>	<i>Transcription Factor</i>
<b>TFBS</b>	<i>Transcriptional Factor Binding Site</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>18</b>
1.1 Conteúdo transcricional humano .....	18
1.2 Os RNAs não-codificantes .....	24
1.3 Transcrição gênica e seus mecanismos de regulação .....	29
1.3.1 Fatores de transcrição .....	32
1.3.2 Modificações pós-transcricionais.....	35
1.3.4 Proteínas ligadoras de RNA.....	38
1.3.5 Envolvimento de RNAs não-codificantes .....	39
1.4 Genômica comparativa e seqüências não-codificantes .....	47
1.5 Expressão gênica e câncer.....	49
1.5.1 Parálogos e famílias gênicas .....	49
1.5.2 Marcadores moleculares de diagnóstico e prognóstico em câncer de próstata .....	54
<b>2 OBJETIVOS E RELEVÂNCIA</b> .....	<b>58</b>
<b>3 MATERIAIS E MÉTODOS</b> .....	<b>61</b>
3.1 Análises Informáticas.....	61
3.1.1 Identificação <i>in silico</i> de parálogos de genes relacionados com câncer ...	61
3.1.2 Análises dos produtos protéicos hipotéticos de <i>RASL11A</i> e <i>ARL-9</i> .....	62
3.1.3 Análises Estatísticas para experimentos de expressão gênica com <i>microarrays</i> .....	63
3.1.5 Análises de enriquecimento de categorias de ontologia gênica (GO).....	67
3.2 Amostras de tecido e linhagens celulares.....	68
3.2.1 Amostras de tecidos humanos .....	68
3.2.2 Amostras de tecidos de camundongos .....	69
3.3 Extração de RNA .....	71
3.3.1 Extração de RNA poli-(A)+ .....	71
3.3.2 Extração de RNA total.....	72
3.3.3 Extração de RNA co-imunoprecipitado com proteínas ligadoras de RNA..	72
3.3.4 Tratamento com DNase livre de RNase .....	73
3.3.5 Quantificação e verificação da qualidade dos RNAs extraídos .....	74

3.4	Abordagens de PCR para amplificação de transcritos completos .....	75
3.4.1	RACE-PCR .....	75
3.4.2	PCR com oligonucleotídeos iniciadores específicos .....	76
3.5	Estudo de expressão gênica por <i>microarrays</i> de cDNA.....	77
3.5.1	Plataforma de <i>microarrays</i> de cDNA do IQUSP .....	77
3.5.2	Obtenção de alvos fluorescentes .....	79
3.5.3	Hibridizações em lâminas de <i>microarrays</i> de cDNA.....	82
3.5.4	Desenho dos experimentos com <i>microarrays</i> de cDNA .....	83
3.5.5	Obtenção de imagens e valores de intensidade comparáveis .....	85
3.6	Hibridizações em <i>Northern Blots</i> .....	88
3.6.1	Detecção de mensagens utilizando sondas de fita dupla.....	89
3.6.2	Detecção de mensagens utilizando sondas de fita simples .....	90
3.7	Análise de expressão gênica por PCR .....	92
3.7.1	Determinação da orientação de mensagens utilizando cDNA transcrito reversamente com oligonucleotídeos iniciadores específicos.....	92
3.7.2	Avaliação da abundância relativa de mensagens utilizando PCR quantitativo.....	94
3.8	Experimentos de Imunoprecipitação.....	95
3.8.1	Ensaio de imunoprecipitação da cromatina.....	95
3.8.2	Imunoprecipitação de proteínas ligadoras de RNA (RBPs).....	96
3.9	Sistemas de expressão epissomal .....	98
3.9.1	Construção do sistema de expressão de Rasl11a/b em procarioto .....	98
3.9.2	Indução da expressão de Rasl11b em bactéria .....	100
3.9.3	Detecção de Rasl11b expresso em bactéria com anticorpo anti-histidinas. .....	101
3.9.4	Sistema de expressão de proteínas ligadoras de RNA em células mamífero.....	102
3.9.5	Detecção das RBPs expressas em 293T com anticorpo contra proteína A. .....	103
3.9.6	Sistema de expressão de transcritos intrônicos em células de mamífero .....	104
3.10	Experimentos para obtenção de anticorpos.....	106
3.10.1	Preparo de antígenos de Rasl11b para imunização de coelhos. ....	106
3.10.2	Imunização de coelhos com corpos de inclusão de Rasl11b. ....	107

3.11	Análise de expressão gênica por <i>microarrays</i> de oligonucleotídeos .....	108
3.11.1	Plataforma de <i>microarrays</i> de oligonucleotídeos.....	108
3.11.2	Marcação de alvos fluorescentes por amplificação de RNA .....	110
3.11.3	Hibridização dos experimentos com lâminas de <i>microarrays</i> de oligonucleotídeos .....	110
3.11.4	Obtenção de imagens e valores de intensidade comparáveis .....	111
<b>4</b>	<b>RESULTADOS.....</b>	<b>112</b>
4.1	Identificação e caracterização de membros representativos de novas famílias de pequenas GTPases monoméricas.....	112
4.1.1	<i>RASL11A</i> e <i>ARL-9</i> , dois novos genes codificantes para pequenas GTPases monoméricas.....	112
4.1.2	<i>RASL11A</i> e <i>ARL-9</i> constituem novas subfamílias de pequenas GTPases. ....	114
4.1.3	Produtos protéicos codificados por <i>RASL11A</i> e <i>ARL9</i> são muito conservados entre os eucariotos. ....	120
4.1.4	Níveis do transcrito <i>RASL11A</i> estão diminuídos em tumores de próstata. ....	122
4.1.5	Sistema de expressão de proteínas Rasl11 em procariotos. ....	124
4.1.6	Expressão de Rasl11b recombinante em corpos de inclusão para imunização de coelhos.....	126
4.2	Transcritos intrônicos marcadores de prognóstico em câncer de próstata....	128
4.2.1	Transcritos intrônicos são expressos em próstata .....	129
4.2.2	Níveis de transcritos intrônicos não codificantes são correlacionados ao grau de diferenciação tumoral em próstata .....	130
4.2.3	Caracterização de transcritos intrônicos relacionados com câncer.....	133
4.2.4	Validação por PCR quantitativo em tempo real das mensagens intrônicas de <i>RASSF1</i> .....	137
4.3	Transcritos intrônicos regulados pelo estímulo de andrógeno.....	139
4.3.1	Identificação de transcritos intrônicos regulados por andrógeno .....	139
4.3.2	RNAs intrônicos regulados por andrógeno mapeiam em genes envolvidos em transdução de sinais .....	143
4.3.3	Caracterização dos RNAs intrônicos regulados por andrógeno.....	143
4.3.4	Efeitos do andrógeno nos níveis dos RNAs intrônicos e dos transcritos codificantes de proteína correspondentes.....	147

4.3.5 Aumento da ligação de AR a um motivo ARE a 5' do <i>asMYO5A</i> .....	151
4.4 Transcritos intrônicos sintênicos com baixa conservação expressos diferencialmente em tecidos humanos e de camundongos. ....	153
4.4.1 Detecção de transcritos intrônicos em tecidos humanos e de camundongos.....	154
4.4.2 Expressão diferencial de transcritos intrônicos em tecidos humanos e de camundongos.....	155
4.4.3 Validação da existência e orientação transcricional de mensagens intrônicas sintênicas diferencialmente expressas em tecidos humanos e de camundongos.....	156
4.5 Interação de transcritos intrônicos com proteínas ligadoras de RNA. ....	159
4.5.1 Expressão das proteínas ligadoras de RNA em células 293T .....	159
4.5.2 Identificação, utilizando <i>microarrays</i> de cDNA, de RNAs longos intrônicos não-codificantes associados a proteínas ligadoras de RNA .....	160
4.6 Análise de expressão de RNAs intrônicos em células LNCaP tratadas com $\alpha$ - amanitina .....	161
4.6.1 Uma grande fração de RNAs intrônicos aparentemente não é transcrita pela RNAP II .....	162
4.6 Sistema para modulação da expressão de <i>asRASSF1</i> e <i>asPPHLN1</i> em células humanas .....	164
<b>5 DISCUSSÃO .....</b>	<b>165</b>
5.1 Identificação e caracterização parcial de <i>RASL11A</i> e <i>ARL-9</i> , membros representativos de novas famílias de pequenas GTPases monoméricas. ....	165
5.2 Transcritos intrônicos marcadores de prognóstico em câncer de próstata... ..	168
5.3 Transcritos intrônicos regulados pelo estímulo de andrógeno.....	171
5.4 Transcritos intrônicos diferencialmente expressos em tecidos humanos e de camundongos. ....	175
5.5 Interação de transcritos intrônicos com proteínas ligadoras de RNA. ....	178
5.6 Análise de expressão de RNAs intrônicos em células LNCaP tratadas com $\alpha$ - amanitina .....	179
<b>6 CONCLUSÕES E CONSIDERAÇÕES FINAIS .....</b>	<b>181</b>
<b>7 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>184</b>
<b>8 LISTA DE ANEXOS.....</b>	<b>201</b>

# 1 INTRODUÇÃO

## 1.1 Conteúdo transcricional humano

Com a conclusão do projeto Genoma Humano (*Human Genome Project*) no começo de 2001 e a publicação simultânea dos rascunhos da seqüência de bases completa que compõe nosso código genético, produzidos pelas iniciativas pública (Lander *et al.*, 2001) e privada (Venter *et al.*, 2001), tornou-se tanto factível quanto desejável a determinação do conjunto completo de transcritos humanos. Contudo, rapidamente ficou evidente que a simples utilização de algoritmos *in silico* para buscas de regiões exônicas na seqüência completa de DNA genômico era uma abordagem difícil e bastante limitada, além de ser totalmente enviesada, pois definia apenas as regiões codificantes para proteína.

Assim, diversas estratégias complementares têm sido utilizadas por pesquisadores no mundo todo para atingir os objetivos tanto de identificação do repertório transcricional humano total, quanto de caracterização individual das mensagens que o compõe. Nesse contexto foi fundamental o aperfeiçoamento de técnicas de análise de expressão gênica em larga escala. Abordagens de estudo por *microarrays* de DNA (Shoemaker *et al.*, 2001; Kapranov *et al.*, 2002; Rinn *et al.*, 2003; Bertone *et al.*, 2004; Kampa *et al.*, 2004), assim como através de sequenciamento de fragmentos expressos (Williamson, 1999; Dias-Neto *et al.*, 2000; Strausberg *et al.*, 2000; Camargo *et al.*, 2001; Saha *et al.*, 2002; Ota *et al.*, 2004; Jongeneel *et al.*, 2005) utilizando amostras de diversos tecidos, têm sido largamente empregadas.

Análises de estruturas gênicas realizadas a partir da seqüência completa do genoma humano revelaram que os introns compreendem em média 95% dos genes codificantes de proteína, os exons internos são relativamente pequenos, e os genes

aparentemente possuem muitas formas alternativas, criando dificuldades para os programas de predição baseados na busca de fases abertas de leitura (*Open Reading Frames* – ORFs) devido a pequena razão sinal-ruído (Lander *et al.*, 2001; Venter *et al.*, 2001). Dessa maneira utilizando abordagens complexas, compostas de análises *in silico* e evidências experimentais, foram estimadas cerca de 32.000 (Lander *et al.*, 2001) e 27.000 (Venter *et al.*, 2001) mensagens codificantes para proteína no genoma humano. Aproximadamente 700 RNAs não-codificantes (ncRNAs) estruturais clássicos, incluindo tRNAs, rRNAs, snRNAs e snoRNAs, e o *Xist* foram preditos no genoma através de buscas por similaridade (Lander *et al.*, 2001). Porém, estimativas anteriores baseadas no mapeamento de seqüências expressas (*Expressed Sequence Tags* – ESTs) apontavam para um número total de mensagens substancialmente maior, que variava desde 35.000 (Ewing & Green, 2000) até 120.000 (Liang *et al.*, 2000) transcritos.

De fato, nos esforços pioneiros onde foram utilizadas plataformas de *microarrays* contendo sondas para toda a seqüência genômica dos cromossomos 21 e 22, foi observado um número cerca de 10 vezes maior de regiões transcricionalmente ativas do que o predito anteriormente pelo mapeamento de regiões exônicas de genes humanos conhecidos (Kapranov *et al.*, 2002; Rinn *et al.*, 2003; Kampa *et al.*, 2004). Frações comparáveis de atividade transcricional foram detectadas nos cromossomos 21 e 22 a partir de regiões exônicas e de regiões intrônicas de genes funcionalmente anotados (Rinn *et al.*, 2003; Kampa *et al.*, 2004).

Ainda foi constatado que cerca da metade dos transcritos provenientes das regiões intrônicas são expressos antisenso em relação ao bem caracterizado intron do gene codificante para proteína de mesmo *locus*, tanto nas análises focadas nos cromossomos 21 e 22 (Rinn *et al.*, 2003; Kampa *et al.*, 2004) quanto em um estudo

com todo o genoma humano (Bertone *et al.*, 2004). Nesse último estudo apenas 13.889 regiões humanas transcricionalmente ativas foram detectadas devido principalmente à utilização de uma abordagem de análise que privilegiou a especificidade em sacrifício da sensibilidade (Bertone *et al.*, 2004). Entretanto, Bertone e colaboradores já puderam gerar um rascunho do mapeamento da expressão proveniente do genoma humano completo, através da observação de muitos transcritos conhecidos ou preditos, e de 10.595 novas seqüências, das quais 3.095 (22% do total) eram intrônicas, sendo cerca de metade dessas (1.529), provenientes de transcrição antisense em relação ao gene codificante para proteína de mesmo *locus* (Bertone *et al.*, 2004).

Evidências similares de atividade transcricional antisense ubíqua a partir de regiões intrônicas no genoma humano, assim como no genoma de camundongo, foram também obtidas por metodologias baseadas em sequenciamento de fragmentos expressos. Essas abordagens são ainda mais abrangentes, pois possuem como vantagens a possibilidade de detecção de transcritos provenientes de regiões repetitivas, e de avaliação da orientação das mensagens com menor probabilidade de artefatos. Análises de mapeamento de porções 3' de transcritos utilizando LongSAGE, uma modificação de SAGE (*Serial Analysis of Gene Expression*) (Saha *et al.*, 2002), de porções 5' (Ota *et al.*, 2004), e de MPSS (*Massively Parallel Signature Sequencing*) (Jongeneel *et al.*, 2005), a partir de amostras humanas; assim como análises semelhantes realizadas pelo projeto de anotação funcional do genoma de camundongo (*Functional Annotation of the Mouse Genome* - FANTOM) (Okazaki *et al.*, 2002; Carninci *et al.*, 2005), confirmaram as observações iniciais.

No projeto FANTOM, através de sequenciamento em larga escala e anotação manual de cDNAs completos foram caracterizadas 102.281 mensagens, incluindo 32.129 codificantes para proteína (2.222 para proteínas novas) e 34.030 ncRNAs (Carninci *et al.*, 2005). Ainda nesse projeto os pesquisadores também concluíram que o número total de transcritos seria pelo menos 10 vezes maior que o previsto pelas estimativas de predição informática de genes no genoma de camundongos (Carninci *et al.*, 2005), e notaram que apenas 40% dos ncRNAs já haviam sido identificados em coleções anteriores de cDNA FANTOM (Okazaki *et al.*, 2002), sugerindo que essa quantidade ainda poderia ser maior.

As análises de seqüências expressas humanas têm confirmado as observações que apontam para um elevado nível de diversidade transcricional em células de mamíferos. Os estudos de expressão por LongSAGE indicam cerca de 15.000 regiões 3' humanas ainda sem caracterização, sugerindo também que o número de novas isoformas ou mesmo de novos transcritos deva ser na ordem de milhares (Saha *et al.*, 2002). Depois do sequenciamento da região 5' de cerca de 1.000.000 de clones de cDNA, e da remoção dos referentes a genes conhecidos, Ota e colaboradores encontraram 7.000 novos transcritos humanos completos a partir de diversos tecidos e linhagens celulares humanos (Ota *et al.*, 2004). Apenas metade dessas mensagens parecem codificar para proteínas, sugerindo que tanto mRNAs quanto ncRNAs ainda permanecem a ser caracterizados (Ota *et al.*, 2004). Mais importante, a análise por MPSS realizada em uma coleção de 32 tecidos humanos, apontou para cerca de 65% de assinaturas transcricionais (total de 142.872) que não possuem sobreposição com mensagens anotadas (Jongeneel *et al.*, 2005). Dessas, 53.402 (37%) mapeiam em regiões intergênicas ou intrônicas de *loci* gênicos anotados, não descritas anteriormente como sendo expressas

individualmente (Jongeneel *et al.*, 2005). Aproximadamente metade (29.661 = 55,5%) dessas assinaturas de MPSS possuem orientação antisenso, mapeando na fita oposta de transcritos já conhecidos (Jongeneel *et al.*, 2005).

Além da expressão de mensagens a partir de porções intrônicas e intergênicas, a observação de ocorrência natural e simultânea de transcrição senso e antisenso também tem sido observada de forma muito mais comum do que antecipado, tanto por estratégias de hibridização em *microarrays* quanto por análises computacionais de seqüências expressas, no genoma humano e de camundongo (Shendure & Church, 2002; Kiyosawa *et al.*, 2003; Yelin *et al.*, 2003; Chen *et al.*, 2004; Kampa *et al.*, 2004). Utilizando bancos de EST e de seqüências completas foram identificados 5.880 *clusters* de transcritos humanos que contêm mensagens antisenso sobrepostas, sendo algumas dessas validadas por *Northern Blot* (Yelin *et al.*, 2003; Chen *et al.*, 2004). De 60.000 clones de cDNA de camundongo, ~2.400 pares senso–antisenso foram detectados (Kiyosawa *et al.*, 2003). E em experimentos utilizando *tiling arrays* foram detectados 11% de sondas para exons conhecidos com transcrição na fita oposta (Kampa *et al.*, 2004). Transcrição, tanto de ncRNAs quanto de mRNAs, a partir da fita oposta de *locus* codificantes para proteína já anotados, produzem RNAs que podem hibridizar com o DNA ou RNA e interferir com a transcrição, tradução ou estabilidade do produto senso (Reis *et al.*, 2005<sup>1</sup>; Mattick & Makunin, 2006), como será discutido mais adiante, no tópico apropriado.

Uma análise direcionada da extensão da transcrição intrônica em humanos foi realizada apenas recentemente, através de uma abordagem de “clusterização” e mapeamento de mais de 5 milhões ESTs depositadas em bancos de dados públicos

---

<sup>1</sup> Reis *et al.* 2005 (Artigo publicado). ANEXO 1 - incluso no CD-ROM suplementar a esta Tese.

(Nakaya *et al.*, 2006)<sup>2</sup>. Aproximadamente 55 mil unidades transcricionais (representadas por *clusters* de ESTs), sem evidência de *splicing*, totalmente intrônicas e 12.600 parcialmente intrônicas (com sobreposição com um éxon) e com tamanhos médios 4 vezes maiores que os tamanhos médios de exons foram identificadas (Nakaya *et al.*, 2006). Uma análise de potencial codificante indicou que cerca de 90% desses transcritos intrônicos provavelmente correspondem a ncRNAs (Nakaya *et al.*, 2006). Ainda, regiões intrônicas transcricionalmente ativas foram observadas em 74 % dos genes humanos RefSeq conhecidos, ressaltando a abundância dos ncRNAs intrônicos e apontando para alguma importância funcional (Nakaya *et al.*, 2006).

Estimativas baseadas nos trabalhos discutidos até aqui indicam que até 60% do genoma humano é transcrito, porém essa porcentagem ainda pode estar sendo subestimada devido ao número limitado de células e estados de diferenciação analisados (Willingham & Gingeras, 2006). Esse enorme número de transcritos aponta para uma arquitetura transcricional em humanos extremamente complexa, que inclui formas alternativas de *splicing* para a imensa maioria de mensagens codificantes de proteína, uma imensa quantidade de ncRNAs, e transcrição antisense disseminada (Figura 1) que provavelmente possui importantes funções biológicas mas ainda precisam ser determinadas (Willingham & Gingeras, 2006).

Finalmente, uma diversidade transcricional como essa pode responder como quantidades relativamente similares de genes codificantes para proteína estimadas para moscas (13.985; *BDGP release 4*), para vermes (21.009; *Wormbase release 150*), e para humanos (23.341; *NCBI release 36*) resultam nos fenótipos extremamente diferentes observados nessas espécies (Willingham & Gingeras,

---

<sup>2</sup> Nakaya *et al.* 2006 (Artigo submetido). ANEXO 2 - incluso no CD-ROM suplementar a esta Tese.

2006). Assim, a identificação sistemática e a caracterização das diversas classes de mensagens humanas expressas continuam sendo importantes desafios, que quando superados resultarão na ampliação da compreensão da biologia de nossa espécie.

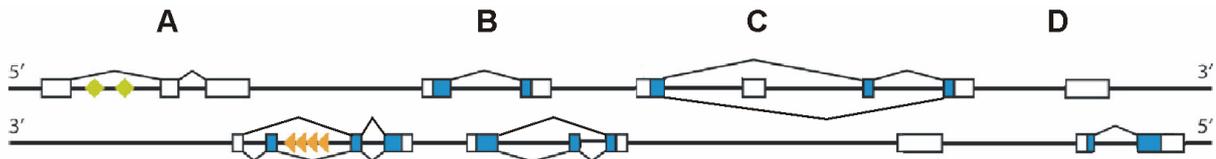


Figura 1 – **Representação gráfica da complexidade transcricional humana.** As caixas brancas representam seqüências não-codificantes e as caixas azuis as regiões codificantes. A utilização de exons no transcrito maduro é representada pelas linhas diagonais. Losangos verdes representam snoRNAs e os triângulos laranja os microRNAs. Em (A) estão representados transcritos codificantes e não-codificantes para proteína e as regiões de transcrição de snoRNAs e microRNAs. Em (B) transcritos antisense com exons entrelaçados. Em (C) transcritos totalmente intrônicos senso e antisense. Em (D) transcritos parcialmente intrônicos. Figura adaptada de Mattick, J. S. & Makunin, I. V. (2006). Non-coding RNA. *Hum Mol Genet*, **15** suppl 1, R17-29.

## 1.2 Os RNAs não-codificantes

Potencialmente existem muitas explicações para a atividade transcricional fora de regiões exônicas conhecidas: seriam novas mensagens, ou novos exons alternativos codificantes para proteína, RNAs não-codificantes (ncRNAs), ou apenas simples ruído biológico. Apesar do conjunto central de genes curados e bem caracterizados (*RefSeq*) ser concordante com a maioria das predições para o número de genes humanos, a comparação das estimativas iniciais realizadas pelas diferentes montagens do genoma humano (Lander *et al.*, 2001; Venter *et al.*, 2001) indicava pouca sobreposição de genes não-*RefSeq* (Hogenesch *et al.*, 2001), sugerindo um maior número de mensagens codificantes para proteína. De fato alguns trabalhos, em especial baseados em sequenciamento de cDNA, indicam que ainda existem genes humanos codificantes para proteína a serem descobertos, na maioria mensagens com baixos níveis de expressão, específicas de tecidos ou

estágios de desenvolvimento, ou ainda com propriedades de seqüência atípicas que dificultaram sua detecção (Ota *et al.*, 2004). Porém a observação que de um conjunto de 21.243 clusters não-redundantes de transcritos, menos de 9% eram novos e apresentavam um potencial codificante significativo (Ota *et al.*, 2004), indica uma tendência clara de queda na taxa de descobrimento desse tipo de mensagem, sugerindo que dificilmente a grande maioria dos transcritos humanos sejam codificantes para proteína. Na realidade, atualmente já existem fortes indícios que a expressão de ncRNAs, provenientes principalmente de regiões humanas intrônicas com orientação transcricional senso ou antisenso em relação ao mRNA do mesmo *locus* gênico e de regiões intergênicas ativas, é um fenômeno bastante geral (Mattick & Makunin, 2006).

Um importante argumento a favor da expressão funcional de ncRNAs foi provido por estudos que combinaram ensaios de imunoprecipitação da cromatina (*Chromatin Immunoprecipitation Assays - ChIP*) e *tiling arrays*. Essa estratégia permitiu um mapeamento aleatório, e de alta resolução ao longo do genoma humano, de sítios de ligação para fatores de transcrição (*Transcription Factor Binding Sites – TFBS*) e para outros complexos proteicos envolvidos na transcrição. Foi encontrada ligação destes fatores não só na região promotora a montante de genes codificantes para proteína, mas também em regiões intrônicas e intergênicas do genoma, o que deu apoio a existência disseminada de transcrição (Martone *et al.*, 2003; Cawley *et al.*, 2004; Euskirchen *et al.*, 2004; Kim *et al.*, 2005). Os TFBS para RelA/p65, membro da família NF- $\kappa$ B, foram mapeados no cromossomo 22, sendo que 22% dos sítios encontravam-se a mais de 50kb distantes de genes codificantes para proteína, e 40% em regiões intrônicas de genes anotados (Martone *et al.*, 2003). O mapeamento dos TFBS para Sp1, cMyc, e p53 nos cromossomos 21 e 22

revelou que cerca da metade desses está correlacionada com a presença de ncRNAs, sendo que uma fração considerável desses transcritos respondeu a um tratamento com ácido retinóico (Cawley *et al.*, 2004). Um número significativo de sítios de ligação de CREB no cromossomo 22 foi também encontrado em regiões sem anotação, mas transcricionalmente ativas, como introns e seqüências intergênicas (Euskirchen *et al.*, 2004). Ainda, o mapeamento no genoma completo do sítios de ligação para o complexo de pré-iniciação da RNA polimerase II em fibroblastos humanos identificou cerca de 10.000 promotores ativos, sendo aproximadamente 13% correspondentes a regiões intrônicas ou intergênicas sem anotação (Kim *et al.*, 2005).

Outros fortes argumentos são obtidos a partir das evidências de relação funcional de ncRNAs com importantes processos celulares. Até recentemente, em uma época anterior a essa revolução genômica que foi parcialmente ilustrada no tópico anterior (item 1.1), um número restrito de ncRNAs foi descrito como participando de funções relativamente genéricas nas células. Os RNAs ribossomais (rRNAs) e RNAs transportadores (tRNAs) atuando na tradução de RNAs mensageiros (mRNAs), os pequenos RNAs nucleares (snRNAs) envolvidos em *splicing*, e os pequenos RNAs nucleolares (snoRNAs) participando principalmente na modificação dos rRNAs (Mattick & Makunin, 2006). Aparentemente a grande maioria desses outros ncRNAs desempenham papéis definidos no programa transcricional humano, assim como apresentado nos próximos parágrafos. Contudo, a expressão desse tipo de transcrito ocorre em níveis mais baixos (2 a 3 vezes menores) quando comparados com os níveis dos transcritos codificantes de proteína conhecidos (Kampa *et al.*, 2004), sendo essa a razão mais provável para sua ampla existência ter sido reconhecida apenas a partir do aperfeiçoamento das técnicas de análise

transcricional em larga escala. Provavelmente essa característica também responde pela escassez ainda hoje de informações sobre a funcionalidade dessa classe de transcritos.

Dentre os ncRNAs melhor caracterizados encontram-se os microRNAs, descobertos na década anterior em *C. elegans*, sendo atualmente reconhecidos como uma grande família conservada de RNAs regulatórios com 22 nucleotídeos gerados a partir de percussores maiores, e que aparentemente são essenciais para uma variedade de processos celulares (Bartel, 2004; Esquela-Kerscher & Slack, 2006). Estimativas apontam para a existência de centenas de microRNAs no genoma humano, e essas análises computacionais ainda sugerem que mais de 20%–30% dos mRNAs codificantes para proteína humanos sejam regulados por essa classe de ncRNAs. Ilustrando a importância desses transcritos existe a observação que assinaturas de expressão de microRNAs, específicas de tecidos humanos, têm sido utilizadas para classificação molecular de câncer (Lu *et al.*, 2005), refletindo de maneira surpreendentemente mais acurada os estados de diferenciação tumoral do que perfis gerados com mRNAs.

Até mesmo os snoRNAs, que compreendem uma classe de ncRNAs que apresentam tamanho de 60 a 300 nucleotídeos, descritos inicialmente como funcionais apenas na modificação de rRNAs durante a biogênese dos ribossomos, hoje são vistos como mensagens mais participativas no programa transcricional de mamíferos tendo como alvos outros pequenos ncRNAs e até mesmo mRNAs (Kiss, 2002). Apesar da maioria delas, por serem derivados de introns de mRNAs, serem transcritas concomitantemente com outras mensagens (Kiss, 2002), aparentemente algumas são derivadas de unidades transcricionais independentes, pois apresentam *caps* de metilguanossina em suas porções 5' (Tycowski *et al.*, 2004). Uma fração

desses snoRNAs também exibe especificidade de tecidos e regulação durante o desenvolvimento (Cavaille *et al.*, 2000; Cavaille *et al.*, 2001; Rogelj & Giese, 2004), o que é um forte indicativo de papéis regulatórios. Funções regulatórias para snoRNA também são evidenciadas pelo seu papel no desenvolvimento da síndrome de Prader–Willi (Kishore & Stamm, 2006).

Outra classe de pequenos ncRNAs que apresenta potenciais papéis funcionais específicos no desenvolvimento foi recentemente descrita. A subfamília Piwi de proteínas Argonata tem sido demonstrada como importante no desenvolvimento de linhagens germinativas (Cox *et al.*, 1998). RNAs com tamanhos de 26–31 nucleotídeos foram identificados como associados a essas proteínas em amostras de testículos murinos e foram denominados piRNAs (Girard *et al.*, 2006). Mais uma vez a especificidade tecidual é um ótimo indício de funcionalidade.

Existem ainda milhares de outros ncRNAs pouco caracterizados. Experimentos utilizando tanto *tiling arrays* genômicos, quanto RACE (*Rapid Amplification of cDNA Ends*) têm demonstrado que alguns ncRNAs provenientes de regiões intrônicas são mensagens longas, contendo de 400 a 2.000 bases (Rinn *et al.*, 2003; Reis *et al.*, 2004; Kapranov *et al.*, 2005). Enquanto a existência de vários ncRNAs longos já foi descrita em mamíferos (Okazaki *et al.*, 2002; Yelin *et al.*, 2003), poucos têm sido caracterizados em detalhes, sendo o RNA de ~15 kb, denominado *Xist* e que está envolvido na inativação do cromossomo X, uma óbvia exceção (Penny *et al.*, 1996). O potencial papel funcional de longos ncRNAs é exemplificado também pela associação do transcrito com cerca de 7 kb, *MALAT-1*, com câncer de pulmão metastático (Ji *et al.*, 2003), e pelo poder de classificação de amostras de tumores de próstata, com diferentes graus de diferenciação histológica, observado para o perfil de expressão de um conjunto de 23 longos RNAs intrônicos não-

codificantes (Reis *et al.*, 2004)<sup>3</sup> As poucas funções específicas dessa e das demais classes de ncRNAs apresentadas aqui serão discutidas posteriormente no tópico apropriado.

Entretanto, é fato que pouco conhecimento encontra-se disponível com relação aos papéis biológicos precisos da enorme quantidade de RNAs, predominantemente sem potencial codificante de proteína, transcritos nas células humanas (Johnson *et al.*, 2005; Willingham & Gingeras, 2006). O mesmo pode ser dito com relação ao mecanismo celular de biogênese desses transcritos. Diversas abordagens de investigação indicam que os RNAs intrônicos estão envolvidos em processos de controle transcricional e pós-transcricional da expressão gênica (discutido no tópico 1.3.4). Assim, têm-se postulado que sua transcrição controlada constitui um avanço fundamental na complexa rede de regulação da expressão gênica, fundamental para a ontogenia e evolução dos eucariotos (Mattick & Makunin, 2006).

### **1.3 Transcrição gênica e seus mecanismos de regulação**

Tradicionalmente é reconhecido que os humanos, assim como todos os eucariotos, possuem três RNA polimerases dependentes de DNA (RNAP I, II e III), as quais apresentam particularidades e especificidades funcionais, e que seriam participantes centrais na produção de todos RNAs codificados pelo genoma nuclear (Howe, 2002). Uma quarta RNAP nuclear, que consiste de uma isoforma da RNAP mitocondrial humana, denominada spRNAP IV, foi descrita muito recentemente e também é responsável pela transcrição de uma população RNAs codificados pelo DNA do núcleo da célula (Kravchenko *et al.*, 2005). As RNAPs são facilmente

---

<sup>3</sup> Reis *et al.* 2004 (Artigo publicado). ANEXO 3 - incluso no CD-ROM suplementar a esta Tese. Os resultados desse trabalho são apropriadamente apresentados e discutidos nessa Tese.

discriminadas a partir de sua sensibilidade ao octapeptídeo  $\alpha$ -amanitina, que não afeta a atividade da RNAP I, inibe fortemente a RNAP II, apenas diminui a atividade da RNAP III (Jacob *et al.*, 1970; Kedinger *et al.*, 1970; Lindell *et al.*, 1970), e aparentemente até estimula a transcrição realizada pela spRNA IV (Kravchenko *et al.*, 2005).

A RNAP I e a RNAP III são em conjunto responsáveis por aproximadamente 80% da síntese de RNA em células em crescimento, sendo os rRNAs; e alguns pequenos RNAs como tRNAs, U6 snRNA, 7SL RNA, assim como o 5S rRNA, respectivamente transcritos por essas enzimas (Paule & White, 2000). Contudo, acredita-se que a RNAP II é a RNA polimerase responsável pela transcrição da grande maioria das regiões genômicas humanas, sendo sintetizados por ela os precursores dos RNAs mensageiros (pré-mRNAs), destinados a serem traduzidos em proteína no citoplasma pelo complexo ribossômico (Howe, 2002). Devido ao compartilhamento de características estruturais, como *splicing*, adição de *cap 5'* e poliadenilação 3', observado em parte do enorme repertório de ncRNAs e que são comuns aos mRNAs (Numata *et al.*, 2003), e pela correlação da existência de ncRNAs com a presença no genoma de sítios de ligação de proteínas e complexos protéicos necessários a transcrição pela RNAP II (Martone *et al.*, 2003; Cawley *et al.*, 2004; Kim *et al.*, 2005), assumiu-se que a expressão de ncRNAs também deva ser função da RNAP II. Entretanto, para mRNAs já existem fortes evidências de transcrição também pela recém-descoberta spRNAP IV (Kravchenko *et al.*, 2005). O envolvimento da spRNAP IV com a transcrição de ncRNAs ainda não foi estudado.

A transcrição é um processo cíclico com diversos passos intimamente regulados. Nos seus estágios iniciais o complexo de pré-iniciação (*Preinitiation Complex* - PIC) da RNA polimerase se liga ao promotor e a síntese de uma cadeia

de ribonucleotídeos começa (Lee & Young, 2000). A iniciação é rapidamente seguida pela liberação do promotor, passo onde os contatos entre a polimerase e a região promotora são rompidos (Dvir, 2002). Essa etapa é intermediária entre a alongação, onde a maior parte do transcrito é sintetizado, e a terminação onde acontece a liberação da polimerase do molde de DNA e do transcrito nascente, e inicia-se um novo ciclo pela enzima em outra região genômica (Howe, 2002).

O controle preciso do processo de transcrição, através da expressão gênica, é fundamental para fazer com que as informações contidas no genoma gerem os eventos biológicos, desde a proliferação celular até à diferenciação e desenvolvimento (Mata *et al.*, 2005). As células precisam integrar informações intrínsecas e ambientais e coordenar múltiplos mecanismos regulatórios, em diversos níveis, para poder exercer suas funções corretamente, sendo que erros ou defeitos nesse processo levam invariavelmente ao surgimento de doenças.

A etapa onde se acredita que ocorra a maior parte regulação da expressão gênica é a iniciação da transcrição. Os genes transcritos pela RNAP II contêm tipicamente duas famílias distintas de elementos regulatórios no DNA, que atuam em *cis*: um promotor, composto por uma região central e elementos regulatórios proximais; e outros elementos distais, que podem ser *enhancers*, *silencers*, *insulators*, ou regiões de controle no *locus* (*Locus Control Regions* - LCR) (Figura 2) (Maston *et al.*, 2006). Esses elementos em *cis* no DNA contêm sinais de reconhecimento e ligação para fatores de transcrição que atuam em *trans*, que funcionam tanto como repressores quanto estimuladores da transcrição (Maston *et al.*, 2006). Aparentemente cerca de 1.850 fatores de transcrição (TF) ligadores de DNA existem (Venter *et al.*, 2001), e é a presença desses múltiplos elementos e a

ligação combinatória desses TF que são responsáveis pela geração de padrões de expressão específicos (Maston *et al.*, 2006).

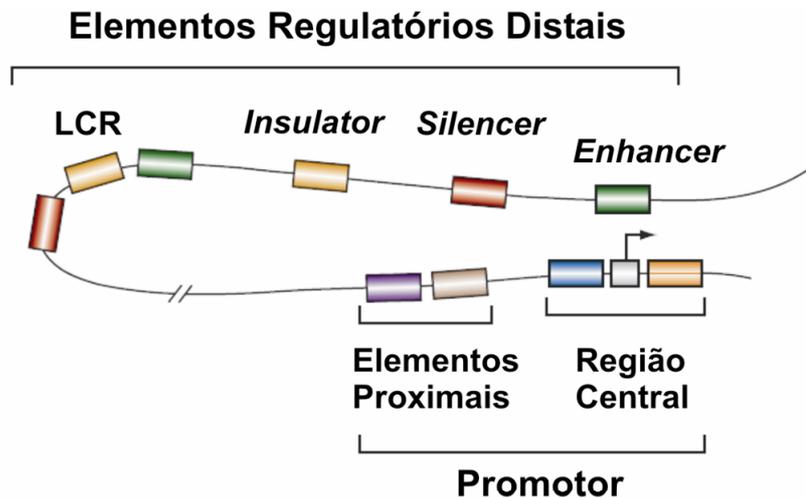


Figura 2 – **Esquema de uma região gênica regulatória típica.** O promotor é composto por uma região central e elementos regulatórios proximais. Outros elementos regulatórios distais podem ser *enhancers*, *silencers*, *insulators*, ou regiões de controle no locus (*Locus Control Regions* - LCR). Figura adaptada de Maston, G. A. *et al.* (2006). *Transcriptional Regulatory Elements in the Human Genome. Annu Rev Genomics Hum Genet.* 7:29–59.

### 1.3.1 Fatores de transcrição

Fatores proteicos envolvidos na acurada transcrição gênica pela RNAP II podem ser classificados em três grupos: fatores de transcrição gerais (*General Transcription Factors* - GTFs), ativadores, e coativadores (Figura 3) (Maston *et al.*, 2006). Os GTFs, que incluem a própria RNAP II e uma variedade de componentes auxiliares como TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, e TFIIH, e um complexo proteico denominado Mediador, se ligam a região central do promotor de maneira ordenada para formação do PIC, que direciona a maquinaria para o local de iniciação da transcrição (Maston *et al.*, 2006).

Após a ligação dos GTFs a transcrição já pode ocorrer em níveis basais, mas sua taxa é bastante aumentada após a ligação dos ativadores. Os ativadores são proteínas ligadoras de DNA que possuem especificidade de seqüência e suas

regiões de reconhecimento ficam próximas à porção central do promotor (Ptashne & Gann, 1997). Comumente denominados apenas como fatores de transcrição, os ativadores existem em diversas famílias que discriminam domínios específicos de ligação no DNA, por exemplo: ativadores do tipo *zinc finger* ricos em cisteínas, *homeobox*, HLH (*helix-loop-helix*), *zipper* de leucina básico (bZIP), *forkhead*, ETS, e POU (Pit-Oct-Unc) (Pabo & Sauer, 1992). Os ativadores podem atuar aumentando a formação do PIC, na iniciação e alongação da transcrição, e ainda podem recrutar atividades que modificam a estrutura da cromatina facilitando a interação da maquinaria transcritcional com o promotor (Maston *et al.*, 2006).

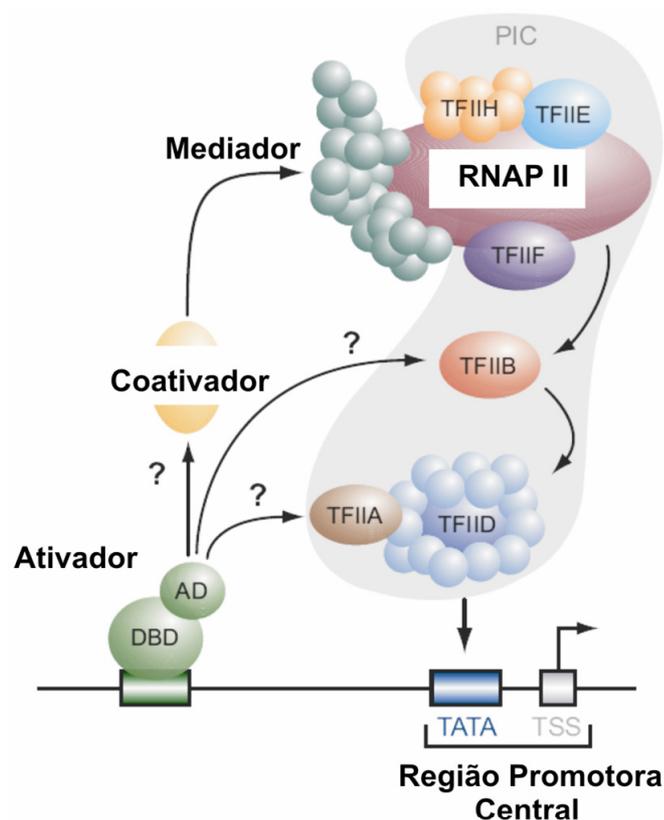


Figura 3 – **Maquinaria transcritcional eucariótica**. Grupos de fatores envolvidos na transcrição pela RNAP II: fatores de transcrição gerais, ativadores, e coativadores. Figura adaptada de Maston, G. A. *et al.* (2006). *Transcriptional Regulatory Elements in the Human Genome. Annu Rev Genomics Hum Genet*, 7:29–59.

Os sítios no DNA para ligação dos ativadores, denominados de sítios de ligação para fatores de transcrição (*Transcription Factor Binding Sites* - TFBSs) são em geral pequenos com 6–12 bp, a especificidade é normalmente devido a não mais que 4–6 posições, em geral são bastante degenerados e descritos por uma seqüência consenso (Maston *et al.*, 2006). Essas diferenças nos TFBSs para cada ativador determinam a afinidade da ligação, o que tem muito impacto em sua atividade regulatória na taxa de transcrição (Maston *et al.*, 2006). A atividade dos ativadores pode ainda ser modulada pelo terceiro grupo de fatores protéicos envolvidos na transcrição gênica, os coativadores. Esses fatores não tem propriedades de ligação ao DNA, apresentando a propriedade de ligação direta aos ativadores (Lonard & O'malley, 2005).

Um dos exemplos de fatores de transcrição mais estudados são os receptores nucleares de hormônios, que são proteínas ativadoras que dependem de ligantes e incluem receptores para andrógenos, estrógenos, glicocorticóides, mineralocorticóides, progesterona e vitamina D (Claessens & Gewirth, 2004). O controle transcricional realizado pelos receptores nucleares envolve a ligação do hormônio, direcionamento para o núcleo e interação do complexo formado com regiões específicas do DNA denominadas elementos de resposta hormonal (Claessens & Gewirth, 2004). Tanto a seqüência de reconhecimento do elemento de resposta, quanto as vias reguladas são, em diferentes níveis, específicas para cada hormônio e seu receptor nuclear (Claessens & Gewirth, 2004). Em especial o andrógeno, e seu receptor nuclear, tem sido muito estudado devido sua implicação central na regulação de diversos processos fisiológicos em células normais e neoplásicas de próstata. Esse hormônio é reconhecido por promover divisão e proliferação celular, modular morte celular programada, quiescência celular e regular

o metabolismo (Isaacs *et al.*, 1992; Geck *et al.*, 1997; Swinnen & Verhoeven, 1998; Maffini *et al.*, 2002). A região de ligação do complexo andrógeno/receptor de andrógeno ao DNA é identificada pelo consenso palindrômico AGAACAnnnTGTTCT, denominado de elemento de resposta ao andrógeno (*Androgen Response Element – ARE*) (Shaffer *et al.*, 2004). A figura 4 apresenta um esquema ilustrativo desse processo.

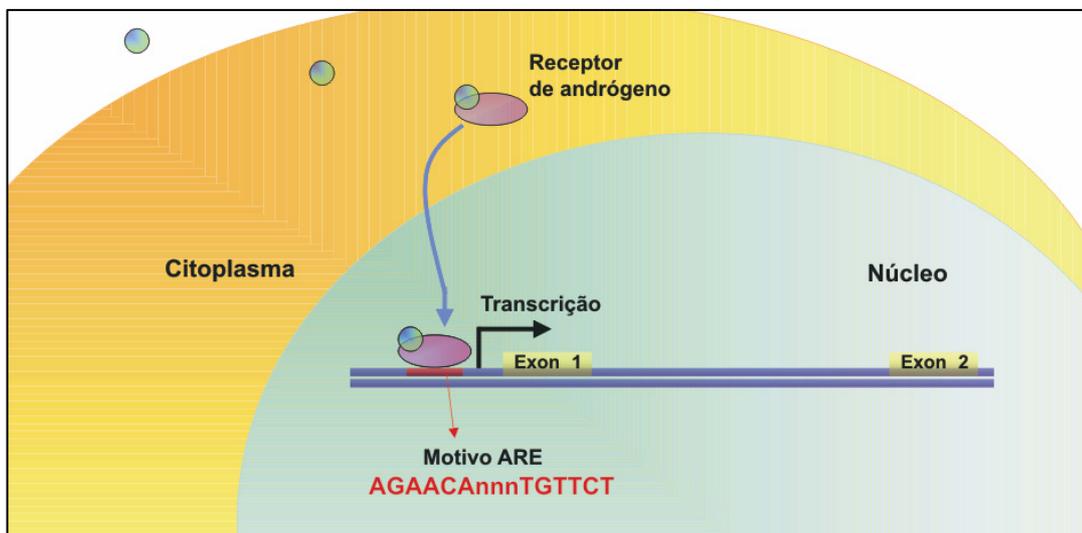


Figura 4 – **Hormônio andrógeno e o seu receptor nuclear.** Diagrama esquemático do controle da transcrição gênica realizado pelo fator de transcrição formado pelo complexo andrógeno/receptor de andrógeno. O local de ligação ao DNA é identificado pelo consenso palindrômico AGAACAnnnTGTTCT, denominado de elemento de resposta ao andrógeno (ARE).

### 1.3.2 Modificações pós-transcricionais

A maturação dos transcritos primários da RNAP II no núcleo requer diversos eventos enzimáticos que incluem o processamento das porções terminais 5' e 3', e a remoção dos introns (Howe, 2002). E apesar desses eventos regulatórios serem classicamente denominados como modificações pós-transcricionais, tem sido demonstrado um acoplamento físico entre esses processos e a síntese da cadeia de ribonucleotídeos pela RNAP II, o que responde em parte pela especificidade dessas alterações nos mRNAs (Howe, 2002).

Um desses processos de maturação é a adição do *cap* 5', que estabiliza o mRNA contra degradação exonucleolítica 5'-3', promove a retirada dos introns e o processamento da região 3' terminal, facilita o transporte ao citoplasma, e auxilia na tradução (Shatkin & Manley, 2000). Essa modificação envolve 3 atividades enzimáticas: uma trifosfatase 5' que remove o fosfato  $\gamma$  do primeiro nucleotídeo transcrito, uma guanililtransferase que adiciona uma guanosina através de uma ligação trifosfato 5'-5', e uma 7 metiltransferase que modifica a guanina terminal pela adição de um radical metil (Shatkin & Manley, 2000). As proteínas envolvidas nessas reações formam o complexo de ligação do *cap* (*Cap Binding Complex* - CBC), que mantêm sua associação ao mRNA durante a transcrição, processamento e transporte para o citoplasma (Howe, 2002). De fato, essa modificação ocorre no pré-mRNA nascente logo após 25–30 bases do RNA terem sido polimerizadas e depois que a região 5' terminal já estiver fora do sítio de ligação de RNA da RNAP II (Howe, 2002).

Outra modificação é a remoção dos introns, o *splicing*, que é catalizado por um complexo ribonucleoprotéico (RNP) denominado spliceossomo (Howe, 2002). Brevemente, a formação do complexo é iniciada pelo reconhecimento no pré-mRNA nascente de pequenos elementos em *cis* conservados, os *splice sites* e o *branchpoint*, localizados próximos ao final dos introns (Howe, 2002). Após o reconhecimento desses sinais, os *splice sites* 5' e 3' interagem através de RNPs, e diversos outros fatores (proteicos e RNPs) são adicionados de maneira ordenada formando o spliceossomo que após várias alterações conformacionais e reações químicas realiza o *splicing* (Howe, 2002). Essas etapas de reconhecimento envolvem ainda outras proteínas reguladoras que apresentam capacidade de se ligar ao pré-mRNA para regular a correta utilização de exons, os fatores de *splicing*,

o que possibilita diversas combinações de uso de exons (Howe, 2002). O uso de exons alternativos surge quando diferentes *splice sites* são reconhecidos e pareados para gerar transcritos maduros com distintos conteúdos de informação (Howe, 2002). Realmente, os *splicing* alternativos provêm um importante controle adicional na expressão gênica, permitindo as células tanto modular os níveis das proteínas expressas, quanto produzir múltiplas isoformas de proteínas a partir do mesmo *locus* gênico (Smith & Valcarcel, 2000). Mais uma vez esse processos ocorrem ainda durante a transcrição, o que parece ser fundamental para a acurácia e eficiência dessa modificação (Howe, 2002). A figura 5 ilustra e sumariza os processos descritos nesse parágrafo.

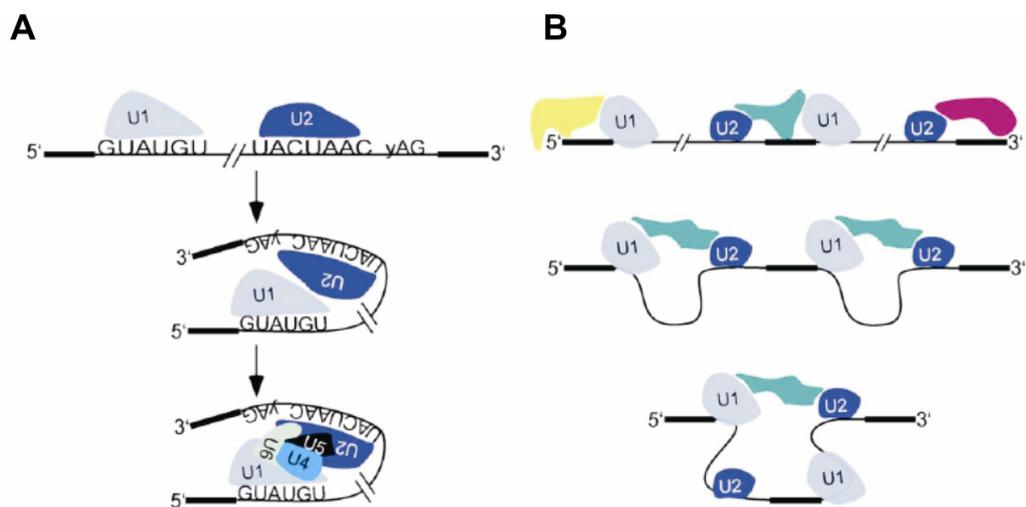


Figura 5 – **Processamento dos pré-mRNAs pelo spliceossomo.** (A) Reconhecimento e pareamento dos *splice sites* e do *branchpoint* (consenso para leveduras mostrado) através de interações com ribonucleoproteínas (RNPs) (em tons de azul e preto) para formação do spliceossomo. (B) Ligação de fatores de *splicing* (em amarelo, verde e roxo) ao pré-mRNA para regular a utilização de exons, permitindo os diversos padrões de *splicing*. Figura adaptada de Howe, K. J. (2002). RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta*, **1577**(2):308-324.

A modificação das regiões 3' terminais dos pré-mRNAs nascentes ocorre por um mecanismo de múltiplos passos que é iniciado pelo reconhecimento do sítio de poliadenilação por fatores de processamento (Howe, 2002). Duas atividades

químicas, mediadas por proteínas evolutivamente conservadas, são necessárias: quebra endonucleolítica do transcrito nascente e adição de uma cauda de poli-(A) por uma polimerase na região 3' clivada (Howe, 2002). O processo se inicia pelo reconhecimento dos sinais de poliadenilação (em geral AAUAAA), e pela clivagem de alguns nucleotídeos após essa região para criar o local de adição da cauda poli-(A) (Howe, 2002). Nessa modificação uma importante proteína ligadora de RNA, PABP (*Poly(A)-binding protein*), apresenta um papel estimulatório, e ainda regula o tamanho da cauda (Mangus *et al.*, 2003). De maneira similar a adição do *cap* 5', o processamento 3' aumenta a resistência dos transcritos a exonucleases e tem papéis funcionais no transporte e tradução dos mRNAs (Howe, 2002).

#### **1.3.4 Proteínas ligadoras de RNA**

Atualmente tem sido demonstrado que diversas proteínas ligadoras de RNA (*RNA Binding Proteins* – RBPs), assim como a PABP apresentam papéis centrais nos processos de regulação pós-transcricionais, representando para esses processos o mesmo que os fatores de transcrição representam para a transcrição (Sanchez-Diaz & Penalva, 2006). As RBPs estão envolvidas em todas as etapas que conectam a transcrição dos mRNAs a sua tradução em proteínas, determinando o destino dos pré-mRNAs e RNAs aos quais se ligam alterando padrões de *splicing*, modificando sua estabilidade, seus níveis de tradução e sua localização celular (Siomi & Dreyfuss, 1997).

As RBPs são definidas pela presença de domínios de ligação de RNA característicos (RRM ou RBD, KH e dsRBD). Alguns exemplos intensamente analisados funcionalmente são: PABP (*Poly(A)-Binding Protein*), proteína que se liga

a cauda poli-(A) de mRNAs. É necessária para a síntese da cauda poli-(A) no núcleo, regulando seu tamanho e estimulando a maturação do mRNA (Mangus *et al.*, 2003). Alguns mRNAs também precisam se associar a PABP para serem exportados do núcleo (Mangus *et al.*, 2003); hTRA2 (*Human Transformer 2*), proteína ativadora de *splicing*. Esta RBP pode se ligar a elementos conhecidos como amplificadores exônicos de *splicing* (*Exonic Splicing Enhancer* - ESE) (Kondo *et al.*, 2004); EIF4E (*Eukaryotic Translation Initiation Factor 4E*), proteína envolvida no direcionamento dos ribossomos a estrutura do *cap* 5' dos mRNAs. EIF4E é o componente limitante do aparato de tradução nos eucariotos, sendo por esse motivo de fundamental importância para a regulação espacial e temporal da síntese protéica (Richter & Sonenberg, 2005); e HuD (*HU-Antigen D*), esta RBP se liga e estabiliza uma grande quantidade de mRNA neuronais específicos. Ela também está implicada na exportação do RNA sintetizado no núcleo e na tradução de mRNAs específicos que contêm elementos ricos em AU (*AU-rich elements*) na porção 3' não traduzida (Saito *et al.*, 2004).

### **1.3.5 Envolvimento de RNAs não-codificantes**

Até recentemente apenas os ncRNAs infraestruturais, rRNAs e tRNAs tinham suas funções celulares conhecidas. Mas de uma forma geral outros ncRNAs já tiveram sua funcionalidade determinada, sendo demonstrados como participantes ativos de praticamente todos os níveis de regulação da expressão gênica em eucariotos (Mattick & Makunin, 2006). Especula-se que esses transcritos, particularmente os expressos a partir de introns de genes codificantes para proteína, representem um avanço fundamental na regulação gênica dos organismos superiores, por definirem como e quando certas trajetórias de diferenciação e

desenvolvimento celular seriam tomadas, assim como por responder por diferenças de complexidade entre espécies com programa de expressão protéico muito similares (Mattick, 2004). Porém, os mecanismos específicos que respondem por essa ainda desconhecida rede de regulação não foram totalmente elucidados.

Exemplos na literatura de regulação gênica por transcritos não-codificantes ainda são esparsos, mas têm sido coletados e reunidos em excelentes revisões atuais (Mattick, 2004; Reis *et al.*, 2005; Mattick & Makunin, 2006; Willingham & Gingeras, 2006), o que contribui muito para estimular estudos com ncRNAs, e facilita a geração de hipóteses e experimentos para a determinação da funcionalidade da grande maioria do produto transcricional humano. Alguns desses exemplos, separados por categorias funcionais, serão apresentados nos próximos parágrafos.

### **Pequenos ncRNAs regulatórios, snoRNAs e microRNAs**

Parte do produto transcricional humano apresenta funcionalidade regulatória como pequenos ncRNAs, que pertencem principalmente a duas classes: snoRNAs e miRNAs/siRNAs. Os papéis na regulação da expressão gênica dos snoRNAs ainda são pouco estudados, pois assim como já mencionado, acreditava-se que essas mensagens tinham suas funções restritas a biogênese do ribossomo, através de modificações no rRNA (Mattick & Makunin, 2006). Mas aos poucos o envolvimento desses transcritos na síntese de DNA telomérico (Kiss, 2002), e em especial no processo de *splicing* de mRNAs codificantes de proteína (Kishore & Stamm, 2006), tem sido reconhecido. Essa última função contribui de forma importante na síndrome Prader-Willi (Kishore & Stamm, 2006). Aparentemente a perda de expressão do snoRNA *HBII-52*, que é transcrito a partir da região silenciada do cromossomo 15 classicamente reconhecida como a causa dessa doença congênita,

causa um processamento aberrante do mRNA que codifica o receptor de serotonina *5-HT2CR*, que tem influência na síndrome (Kishore & Stamm, 2006). O *HBII-52* exibe complementaridade de seqüência com a região do elemento de silenciamento do exon Vb do mRNA de *5-HT2CR*, causando a excisão dessa porção da mensagem (Kishore & Stamm, 2006).

Provavelmente os microRNAs são os ncRNAs que tem recebido maior atenção nos últimos anos. Esses transcritos são funcionais como seqüências contendo aproximadamente 22 nucleotídeos derivados de moléculas precursoras maiores, tanto RNAs dupla-fita, quanto RNAs que formam estruturas tipo *hairpin* (Bartel, 2004). As funções regulatórias dos microRNAs envolvem a supressão da tradução através de pareamento parcial na região 3' não traduzida (UTR), no geral de 6 a 8 nucleotídeos (Lewis *et al.*, 2005), com mRNAs alvo codificantes de proteína, ou degradação de mRNAs através de pareamento perfeito e direcionamento ao complexo RISC, pelo fenômeno reconhecido como interferência de RNA (Du & Zamore, 2005).

Até pouco tempo, apenas a atuação na inibição da tradução era reconhecida em humanos, na verdade nos vertebrados de uma forma geral. Porém em 2004 foi descrita a associação da supressão da expressão genica de *HOX* por uma degradação do seu mRNA dirigida pelo pareamento perfeito com um microRNA também na porção 3'-UTR (Mansfield *et al.*, 2004; Yekta *et al.*, 2004). Outro exemplo de regulação por esse mecanismo de atuação tem sido reportado para o *locus* que sofre *imprinting*, *Rtl1/Peg11* (Davis *et al.*, 2005). A mensagem transcrita maternalmente antisenso-*Peg11* é processada em diversos microRNAs, que causam degradação mediada por RISC do mRNA paternalmente expresso *Rtl1/Peg11*, após ligação em um segmento na região codificante (Davis *et al.*, 2005).

A expressão dos microRNAs é regulada, e alguns desses transcritos tem sido demonstrados como centrais em vários processos do desenvolvimento, como expressão temporal controlada, proliferação celular, apoptose, metabolismo de gorduras, morfogênese cerebral, diferenciação muscular e divisão de células tronco (Mattick & Makunin, 2006). Assim, alterações aberrantes provocadas por microRNAs têm sido reconhecidas como causais em diversas doenças humanas como o câncer (Mattick & Makunin, 2006). O banco de dados miRBase (<http://microna.sanger.ac.uk/>) apresenta uma lista com mais de 300 microRNAs identificados experimentalmente em humanos, e também uma lista com os transcritos alvos preditos para esses pequenos ncRNAs (Griffiths-Jones *et al.*, 2006).

Outras classes de ncRNAs, em especial os longos, sem processamento (*splicing* ou excisão similar a que ocorre em alguns pequenos ncRNAs como os microRNAs) e transcritos a partir da região intrônica da fita oposta de *loci* que contêm mRNAs codificantes de proteína, são reconhecidas. Tem sido postulado que a grande maioria desses ncRNAs tem papel na expressão gênica, atuando diretamente nos mRNAs de mesmo *loci* em *cis* (Reis *et al.*, 2005). E apesar de não existirem evidências concretas, ainda não se pode excluir a possibilidade que essas mensagens apenas modulem eventos regulatórios transcricionais ou pós-transcricionais após serem processadas em transcritos menores, e que atuem em *trans*, como os microRNAs. Nos itens seguintes será principalmente discutida a funcionalidade desse tipo ainda pouco estudado de ncRNA, e a figura 6 apresenta uma ilustração dos mecanismos postulados.

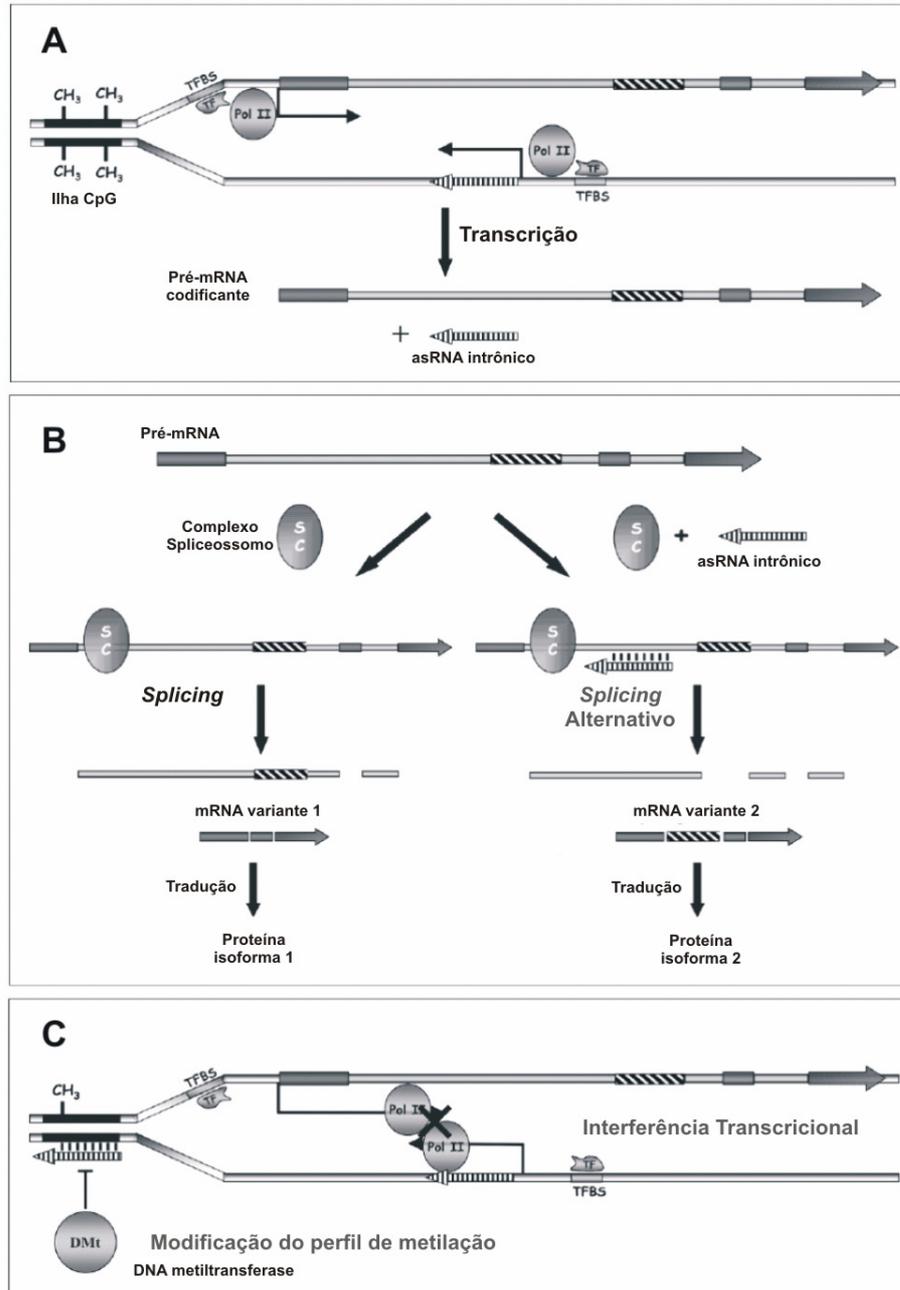


Figura 6 – Mecanismos propostos de regulação da expressão gênica envolvendo Longos RNAs Intrônicos não-codificantes antisense. (A) Transcrição de pré-mRNA codificante de proteína (alvo) e da mensagem intrônica não codificante antisense (asRNA). (B) Modulação da abundância de isoforma alternativa pelo controle do *splicing*. (C) Controle do perfil de metilação de ilhas CpG de regiões promotoras de genes alvos (**esquerda**) ou modulação da abundancia de RNAs por interferência transcricional (**direita**). Sítios de ligação para fatores de transcrição (TFBS), fatores de transcrição (TF), RNA polimerase II (Pol II), spliceossomo (SC), e DNA metiltransferase (DMt) estão indicados. Figura adaptada de Reis, E. M. *et al.* (2005). *As Antisense Gets Intronic*. *Omics*, **9**(1):2-12.

### **Longos RNAs Intrônicos não-codificantes antisense e *splicing* alternativo**

Assim como brevemente descrito no tópico 1.3.3, a correta excisão dos introns de um pré-mRNA e a religação dos exons é dirigida pelo spliceossomo, que reconhece sítios nas bordas de introns e exons, assim como certas seqüências regulatórias localizadas principalmente dentro de regiões do mRNA ainda não processado (Black, 2003). A ligação de fatores a elementos regulatórios denominados ISE (*intronic splicing enhancer*) e ISS (*intronic splicing silencer*), está envolvida na modulação do processamento do mRNA, favorecendo ou bloqueando a formação do spliceossomo (Akker *et al.*, 2001; Bruno *et al.*, 2004). A utilização de oligonucleotídeos antisense sintéticos, complementares a essas regiões intrônicas regulatórias é capaz de aumentar tanto a utilização de um exon do gene *FGFR1*, aberrantemente retirado em amostras de glioblastoma humano (Bruno *et al.*, 2004), quanto a exclusão do exon 7 no mRNA do gene *SMN* (Miyaso *et al.*, 2003).

Assim, ficou embasada uma hipótese de atuação de ncRNAs provenientes de regiões intrônicas, naturalmente transcritos com orientação antisense, em mecanismos envolvidos na alteração do padrão de *splicing* de mensagens codificantes. De fato, o trabalho recente de Yan e colaboradores mostrou que após transfecção de células Jurkat e indução do aumento da expressão de *SAF*, um longo (~1,5kb) ncRNA intrônico antisense normalmente transcrito a partir do primeiro intron do gene que codifica para a proteína Fas, ocorrem modificações na freqüência do uso de exons do mRNA *FAS* (Yan *et al.*, 2005). A presença de *SAF* apresenta importantes conseqüências na homeostasia celular, pois as alterações desencadeadas no processamento de *FAS* ocorrem tanto no primeiro exon, que é adjacente ao intron que se sobrepõe ao ncRNA, quanto nos exons da região 3' final

(Yan *et al.*, 2005). Essas últimas mudanças levam a obtenção de mensagens que codificam para uma proteína Fas que não é mais capaz de cumprir seu papel no mecanismo de apoptose através da ativação de caspases, devido a ausência de uma porção carboxi-terminal na seqüência protéica primária que determina seu ancoramento a membrana plasmática (Yan *et al.*, 2005). A figura 6B apresenta uma esquema geral a respeito do mecanismo de atuação de ncRNAs antisense nessa etapa da regulação da expressão gênica

### **Longos ncRNAs intrônicos antisense e alterações no perfil de metilação de promotores de genes alvos**

Modificações na estrutura da cromatina são processos que regulam a expressão gênica através da reorganização de nucleossomos e de alterações químicas nos resíduos de nucleotídeos do DNA (Huang *et al.*, 2003). Esses processos já foram bastante estudados, porém nos últimos anos surgiram evidências do envolvimento direto de moléculas de ncRNA nesses eventos (Akhtar *et al.*, 2000; Tufarelli *et al.*, 2003; Imamura *et al.*, 2004).

Proteínas que realizam transferência de grupos acetil de histonas apresentam um papel fundamental nos mecanismos de rearranjo da cromatina (Van Der Vlag & Otte, 1999; Gebuhr *et al.*, 2000). Nessas e em outras proteínas envolvidas em processos de remodelagem similares, é possível identificar um domínio conservado com propriedades de ligação a RNA, o *chromodomain* (Akhtar *et al.*, 2000; Jones *et al.*, 2000). Esse domínio conservado na histona acetil transferase MOF de *Drosophila sp* interage com o ncRNA *roX*, contribuindo para sua interação ao cromossomo X do macho durante a compensação de dose (Akhtar *et al.*, 2000). Acredita-se que fenômeno similar ocorra em humanos, mas existem poucos estudos

de inativação do cromossomo X e *imprinting* genômico, associados a metilação/desmetilação do DNA, que sugerem um mecanismo dependente de RNA (Lee *et al.*, 1999; Sleutels *et al.*, 2000; Sleutels *et al.*, 2002).

Evidências diretas em humanos de metilação dependente de ncRNAs foram apenas obtidas recentemente. Tufarelli e colaboradores observaram em um paciente afetado com alfa-talassemia um silenciamento não esperado do gene *HBA2* através da hipermetilação de seu promotor, desencadeada pela transcrição aberrante de um longo ncRNA antisense, *LUC7L*, a partir do mesmo *locus* gênico (Tufarelli *et al.*, 2003). De forma similar, o ncRNA intrônico antisense *KHPS1a* modula a metilação da ilha CpG do promotor de *SPHK1*, gene codificante de proteína de mesmo *locus* (Imamura *et al.*, 2004). Esse tipo de função está esquematizado na figura 6C (direita).

### **Interferência transcricional e ncRNAs antisense**

Quando dois RNAs são concomitantemente transcritos a partir de fitas opostas do mesmo *locus* gênico pode ocorrer o fenômeno de interferência transcricional devido a colisão das maquinarias de transcrição (Prescott & Proudfoot, 2002). Durante as observações iniciais desse fenômeno em *Saccharomyces cerevisiae* foi notado que a abundância de ambos RNAs é reduzida pela própria restrição espacial da elongação pela RNAP II (Prescott & Proudfoot, 2002).

Evidências da ocorrência natural desse fenômeno foram obtidas para os *loci* do receptor tipo 2 para fator de crescimento similar a insulina (*Insulin-like growth factor type-2 receptor*) *Igf2r*, e do RNA antisense *AIR* (Wutz *et al.*, 1997), do fator eucariótico de iniciação 2a, *eIF2a* (Silverman *et al.*, 1992; Noguchi *et al.*, 1994), e do colágeno alfa I (Farrell & Lukens, 1995). Assim, um dos possíveis mecanismos de

atuação dos ncRNAs pode ser regulação através de interferência transcricional, ilustrado na figura 6C (esquerda).

#### **1.4 Genômica comparativa e seqüências não-codificantes**

Abordagens utilizando genômica comparativa têm sido largamente utilizadas para inferir conservação e funcionalidade para seqüências de DNA de múltiplas espécies. Esse tipo de análise realizada em larga escala tornou-se possível após o aperfeiçoamento das técnicas de sequenciamento que permitiram obter os genomas completos de mais de 80 organismos microscópicos, e de muitas espécies eucarióticas, incluindo uma levedura (*Saccharomyces cerevisiae*), um nematodo (*Caenorhabditis elegans*), uma mosca (*Drosophila melanogaster*), algumas plantas como *Arabidopsis thaliana*, e uma variedade de arroz (*Oryza sativa japonica*), e de alguns animais tais como o *Fugu ribripes* e humanos (*Homo sapiens*) (Frazer *et al.*, 2003). Nesse arsenal de informações genômicas podem ainda ser incluídos o camundongo (*Mus musculus*), o rato (*Rattus norvegicus*), recém sequenciados, e outros eucariotos como chimpanzés, e peixe zebra cujas seqüências de DNA completa devem ser obtida nos próximos anos, o que ampliou significativamente a possibilidade de realização de estudos comparativos de conservação entre espécies (Frazer *et al.*, 2003).

Baseados na premissa que as regiões mais conservadas que a taxa basal de mutação poderiam estar sobre algum grau de seleção positiva para sua manutenção, uma considerável quantidade de estudos tem usado estratégias comparativas para identificar novas regiões funcionais nos genomas (Makalowski *et al.*, 1996; Hardison *et al.*, 1997; Loots *et al.*, 2000; Desilva *et al.*, 2002; Margulies *et al.*, 2003). A enorme maioria desses estudos comparativos é direcionada para a

identificação de transcritos codificantes para proteína e longas regiões regulatórias conservadas no DNA, utilizando critérios bastante específicos para a determinação de conservação e inferência de funcionalidade. Abordagens de genômica comparativa utilizando o enorme repertório de RNAs não-codificantes (ncRNAs) identificados como transcritos a partir do genoma humano (apresentado no item 1.2) tornou-se um grande desafio ainda não completamente enfrentado.

Aplicando os critérios classicamente utilizados em genômica comparativa, como tamanho mínimo e grau de identidade, diversos trabalhos têm identificado regiões não-codificantes altamente conservadas, e tem postulado funcionalidade para elas. Assim, termos como *Ultra Conserved Regions* (UCR) (Sandelin *et al.*, 2004), *Highly Conserved Regions* (HCR) (Duret *et al.*, 1993), *Multi-species Conserved Sequences* (MCS) (Margulies *et al.*, 2003), *Conserved Non Coding Elements* (CNE) (Woolfe *et al.*, 2005), e seqüências *Conserved Non-Genic* (Dermitzakis *et al.*, 2005) têm sido utilizados para descrever essas regiões genômicas possivelmente funcionais. Estima-se que cerca de 65% dessas seqüências estejam em regiões intergênicas e 35% em regiões intrônicas do genoma humano (Dermitzakis *et al.*, 2005).

Entretanto apesar desses critérios de conservação de identidade terem sido considerados bons para anotação e validação de seqüências funcionais, devido ao viés mencionado para transcritos codificantes, eles podem falhar para identificar novos ncRNAs funcionais. De fato, entre os ncRNAs já é reconhecido que grandes níveis de conservação evolutiva de identidade de seqüência primária não são pré-requisitos para funcionalidade (Numata *et al.*, 2003; Ravasi *et al.*, 2006), já que estes trabalhos indicam um número potencialmente elevado de seqüências não-

codificantes com papéis celulares, expressas em diferentes organismos e possuindo baixa conservação entre si.

## **1.5 Expressão gênica e câncer**

### **1.5.1 Parálogos e famílias gênicas**

A elucidação da seqüência genômica (Lander *et al.*, 2001; Venter *et al.*, 2001), e os esforços para determinação do programa transcricional completo humano (discutido no item 1.1) sugerem que ainda existam mensagens codificantes para proteína a serem descobertas, na maioria transcritos com baixos níveis de expressão, mas que provavelmente são específicos de tecidos ou estágios de desenvolvimento (Ota *et al.*, 2004). E mesmo entre os cerca de 25.000 transcritos codificantes já anotados (She *et al.*, 2004), existe uma considerável fração pouco estudada funcionalmente. Abordagens em escala genômica para entender os processos celulares que envolvem as proteínas codificadas no genoma humano têm sido desenvolvidas, principalmente no contexto de doenças como o câncer, para determinação de novos alvos potenciais para o tratamento com drogas (Harris & Stevens, 2006).

Uma interessante estratégia vem sendo utilizada por diversos pesquisadores nos dias de hoje. *Chemogenomics* é um novo conceito cunhado a luz da nova era nas pesquisas científicas iniciada com o conhecimento do genoma humano, que almeja a identificação e descrição de todos os possíveis alvos genômicos associados às diversas patologias, e suas drogas relacionadas (Caron *et al.*, 2001; Harris & Stevens, 2006). Uma das metodologias utilizadas para alcançar esses ambiciosos objetivos é o estudo de famílias gênicas, definidas pela homologia entre os genes pertencentes, ao nível da estrutura primária da proteína, e que

preferencialmente apresentem propriedades funcionais similares (Caron *et al.*, 2001; Harris & Stevens, 2006). Os genes pertencentes à mesma família gênica podem ser parálogos, quando derivados de um ancestral comum, mesmo que não apresentem necessariamente as mesmas funções celulares (Jensen, 2001). Com isso o estudo de famílias gênicas para identificação de possíveis parálogos de genes envolvidos em doenças humanas, utilizando análise informática de seqüências e posterior caracterização bioquímica, é muito útil para o desenvolvimento de drogas atuantes em múltiplos alvos da mesma família, ou de várias drogas similares efetivas apenas em cada parálogo (Caron *et al.*, 2001; Harris & Stevens, 2006). Um exemplo bem sucedido de resultado obtido com esse tipo de abordagem é exemplificado a partir dos trabalhos realizados com a família das proteínas quinases (Cohen, 1999; Garcia-Echeverria *et al.*, 2000). Muitos inibidores foram desenvolvidos para suprimirem as atividades de várias proteínas dessa família gênica; serinas, treoninas e tirosinas quinases (Cohen, 1999). Embasado nesses estudos, Garcia-Echeverria e colaboradores desenvolveram inibidores específicos para cada tipo de proteína quinase (Garcia-Echeverria *et al.*, 2000).

No desenvolvimento de medicamentos, geralmente uma porcentagem significativa das drogas criadas para agir em um membro da família é ativa em outros membros, cujas funções celulares são freqüentemente diferentes, enquanto outra é apenas efetiva em certos membros de uma família, que possuem funções essencialmente semelhantes (Caron *et al.*, 2001; Harris & Stevens, 2006). Assim a identificação de novos membros parálogos de famílias gênicas envolvidas no surgimento e progressão de doenças pode também elucidar os motivos, e apontar soluções, para a existência de certos pacientes com características clínicas

semelhantes, mas que são refratários aos tratamentos convencionais normalmente efetivos.

Especificamente no surgimento de neoplasias humanas, uma família gênica tem chamado atenção pelo envolvimento de muitos membros em câncer de praticamente todas as origens teciduais, a superfamília Ras de pequenas GTPases monoméricas. De fato, os genes dessa família codificam para proteínas que modulam uma grande variedade de processos celulares, desde o controle do crescimento, rearranjos no citoesqueleto, sobrevivência e apoptose celular, até regulação do tráfego intracelular de vesículas (Bos, 1997; Pasqualato *et al.*, 2002a). Mais de 100 membros da superfamília têm sido descritos em células eucarióticas, e são tradicionalmente classificados em seis famílias (Ras, Rho, Rab, Arf, Ran, e Rad) devido a similaridade de seqüência primária com os membros bem caracterizados (Paduch *et al.*, 2001; Oxford & Theodorescu, 2003). Apesar de apresentarem múltiplas funções celulares, todas as pequenas GTPases compartilham um domínio G estruturalmente conservado (Figura 7), que responde pelo mecanismo molecular de ativação (Vetter & Wittinghofer, 2001b). Essas proteínas passam por estados ativos e inativos quando ligadas a GTP ou GDP, respectivamente (Vetter & Wittinghofer, 2001b), e apesar de normalmente possuírem atividade GTPásica intrínseca, suas funções celulares são reguladas por diversos efetores (Figura 8), GAPs (*GTPase-Activating Proteins*), GEFs (*Guanine Nucleotide Exchange Factors*), e GDIs (*Guanine Nucleotide Dissociation Inhibitors*) (Paduch *et al.*, 2001). As funções biológicas das pequenas GTPases também são bastante dependentes de modificações pós-traducionais em sua cadeia de aminoácidos (Paduch *et al.*, 2001). Por exemplo, as famílias Ras, Rho e Rab contêm uma seqüência C-terminal que sinaliza e permite modificações que são essenciais tanto para sua localização

celular, quanto para interação com seus efetores (Paduch *et al.*, 2001) Existem muitas possibilidades de alterações na região C-terminal, mas as mais importantes são: ligação covalente de lipídios, metilação e quebra proteolítica (Paduch *et al.*, 2001).

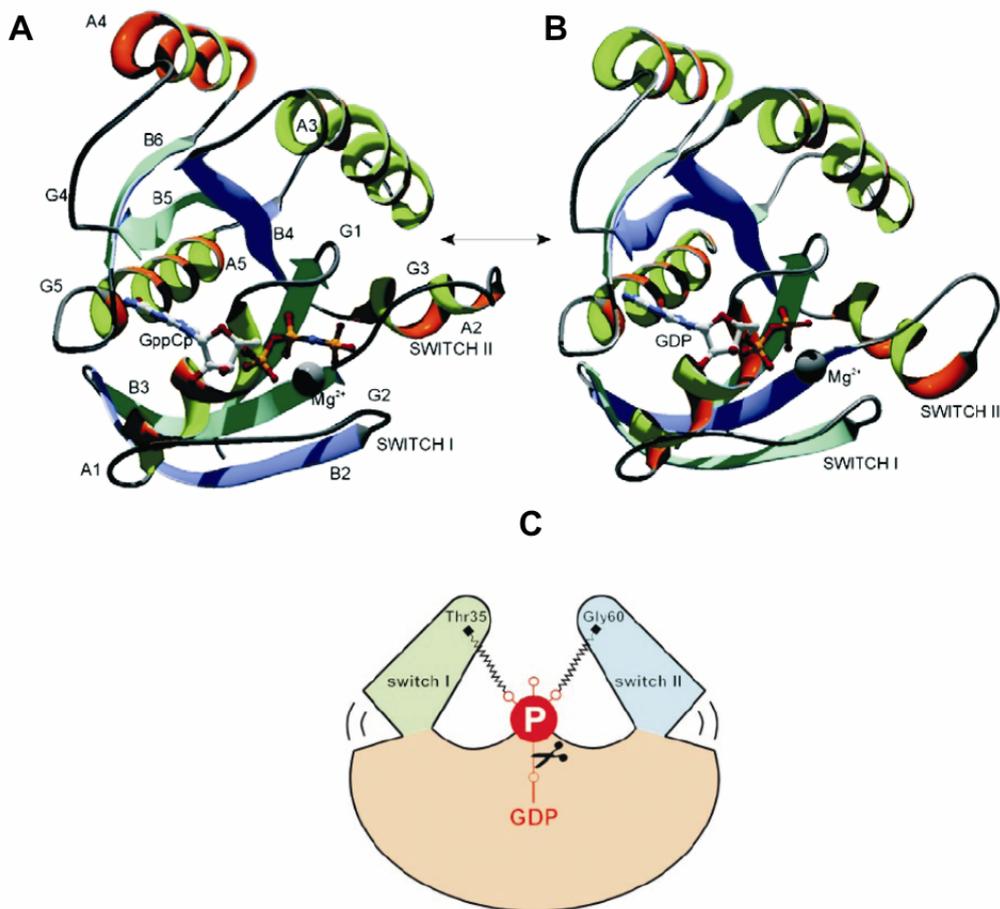


Figura 7 – **Diagrama esquemático do domínio G.** As cinco  $\alpha$ -hélices (A1–A5), seis folhas  $\beta$  (B1–B6) e seis voltas de polipeptídeos (G1–G5) que formam o domínio G estão indicadas. Em **(A)** conformação típica das proteínas da superfamília Ras de pequenas GTPases monoméricas quando ligadas a GTP. Em **(B)** conformação após a hidrólise a GDP. As maiores alterações conformacionais ocorrem nas regiões do *switch I* e *switch II*. Em **(C)** é mostrado um esquema do mecanismo de ativação, onde os domínios *switch I* e *switch II* estão ligados ao fosfato  $\gamma$  através dos grupos NH da cadeia principal dos resíduos conservados Thr e Gly. A hidrólise do GTP permite que as regiões dos *switches* relaxem, alterando a conformação da enzima. Painéis A e B adaptados de Paduch, M. *et al.* 2001. Structure of small G proteins and their regulators. *Acta Biochim Pol*, **48**(4):829-850. Painel C adaptado de Vetter, I. R. & Wittinghofer, A. 2001. The guanine nucleotide-binding switch in three dimensions. *Science*, **294**(5545):1299-1304.

Como já mencionado, muitas proteínas da superfamília Ras têm sido associadas ao câncer. É documentado que mutações nos genes *RAS* (*H-RAS*, *N-RAS* ou *K-RAS*), protótipos da família, ocorrem em aproximadamente 30% de todos os cânceres humanos (Adjei, 2001). As Rho GTPases também têm sido bastante implicadas em processos que levam a progressão do câncer (Boettner & Van Aelst, 2002). Ainda, alguns membros das famílias Ras, Ral, Rit (Reuther & Der, 2000b), e Rho (Zohn *et al.*, 1998) apresentam comportamentos oncogênicos. Outros como Noey2/Arh1 (Yu *et al.*, 1999a; Luo *et al.*, 2003), Rap1a (Wani *et al.*, 1997b), Rerg (Finlin *et al.*, 2001), e Rig (Ellis *et al.*, 2002a) têm sido descritas como possuindo atividade de supressor tumoral.

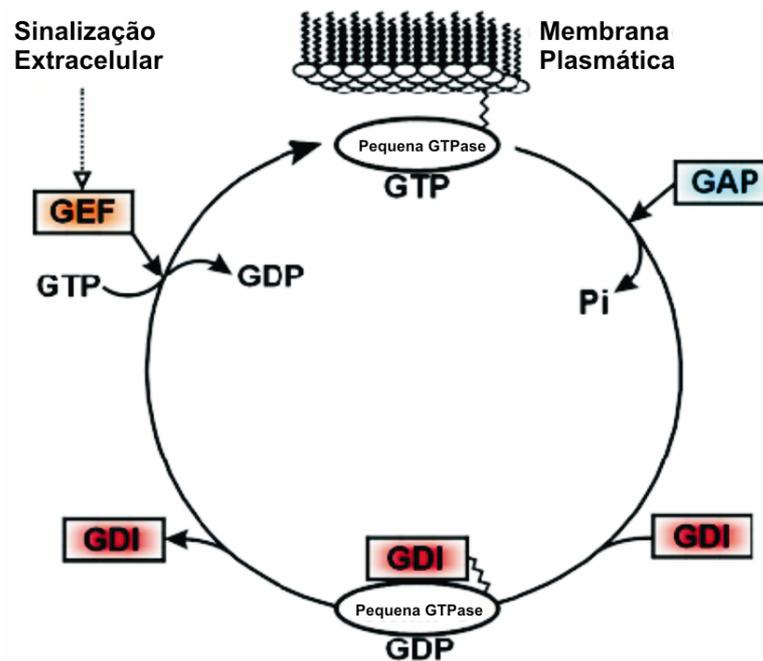


Figura 8 – **Modo de ativação das pequenas GTPases monoméricas.** Proteínas GAP estimulam a baixa atividade GTPásica intrínseca para promover a conversão da forma ligada a GTP para a ligada a GDP. Proteínas GDI afetam a dissociação do GDP e mantém as enzimas no estado inativo. Proteínas GEF catalizam a troca do GDP por GTP. Figura adaptada de Paduch, M. *et al.* (2001). Structure of small G proteins and their regulators. *Acta Biochim Pol*, 48(4):829-850.

Naturalmente, o envolvimento das GTPases relacionadas a Ras no desenvolvimento de tumores coloca essas proteínas como potenciais alvos para tratamentos. A estratégia principal tem como alvo impedir a associação dessas proteínas com membrana citoplasmática. Como mencionado no parágrafo anterior, certas proteínas dessa família estudadas funcionalmente apresentam modificações covalentes em sua estrutura primária, que são importantes para sua funcionalidade. E de fato, são algumas dessas alterações que possibilitam a interação com a face interna da membrana plasmática (Paduch *et al.*, 2001). Assim, estratégias para impedir essa associação têm sido intensamente avaliadas, principalmente através de inibidores de farnesilação que já estão em fases de testes clínicos (Adjei, Croghan *et al.*, 2003; Adjei, Mauer *et al.*, 2003; Rao *et al.*, 2004). Outras abordagens têm como objetivo interromper a via de transdução de sinal dessas proteínas inibindo Raf e MEK, inibir a sinalização autócrina (inibidor do receptor EGF), ou ainda a própria expressão das proteínas relacionadas a Ras, como H-Ras e c-Raf-1 (Cox & Der, 2002). Contudo, a existência de vias bioquímicas alternativas, provavelmente com envolvimento de outros membros similares da família não caracterizados, torna certos pacientes insensíveis aos tratamentos desenvolvidos, sendo uma das principais limitações para o desenvolvimento de metodologias de tratamento mais efetivas.

### **1.5.2 Marcadores moleculares de diagnóstico e prognóstico em câncer de próstata**

O câncer de próstata (PrCa) é uma doença com alta prevalência e é a principal causa de morte por câncer em homens no ocidente (Calvo *et al.*, 2005). O surgimento e o desenvolvimento do PrCa é de fato um processo complexo que inclui

alterações cromossômicas e na expressão gênica, levando a proliferação celular, inibição da apoptose, invasão, angiogênese e metástase (Calvo *et al.*, 2005). Uma revisão detalhada da biologia molecular do PrCa pode ser encontrada no artigo (Abate-Shen & Shen, 2000).

O diagnóstico precoce de câncer de próstata vem sendo realizado principalmente utilizando uma combinação do exame clínico de toque retal e a análise dos níveis de PSA sérico, levando em conta principalmente a idade do paciente (Potter *et al.*, 2001). Para o prognóstico de evolução da doença, tem sido largamente utilizado o método histológico denominado *Gleason Grading System*, onde vários aspectos da arquitetura do tecido, que indicam o nível de diferenciação celular, são analisados e recebem valores de 1 a 5, conforme a diferenciação é perdida (Gleason, 1977). Essa análise é realizada em duas regiões do tecido, o primeiro e o segundo padrão que ocupem as maiores áreas da amostra, e a soma dos dois valores compõem o *Gleason Score* (GS) do tumor do paciente (Gleason, 1977). Amostras de pacientes que apresentam GS de 2 a 4 indicam pouca agressividade e um risco mínimo de morte por câncer de próstata, amostras com valores de GS de 5 e 6 apresentam baixa agressividade com um bom prognóstico, amostras com GS de 8 a 10 são muito agressivas e o prognóstico é ruim, e finalmente valores de GS 7 indicam tumores com agressividade intermediária e prognóstico bastante heterogêneo, de difícil predição (Albertsen *et al.*, 1998). A figura 9 apresenta um esquema do *Gleason Grading System*.

Entretanto, existem severas limitações tanto na metodologia de diagnóstico precoce do PrCa, quanto na utilização do GS como fator de prognóstico. Já é bastante documentado que os níveis sanguíneos de PSA também podem variar devido a fatores não neoplásicos (Oesterling *et al.*, 1997; Carroll, 2001). Como

qualquer procedimento histológico, a avaliação do GS é bastante subjetiva. E por sua vez, a utilização do GS como fator de prognóstico está sujeita a erros amostrais e só pode ser realizada após a remoção cirúrgica radical da próstata. Assim, a busca e identificação de marcadores moleculares pode gerar importantes contribuições para complementar a avaliação clínica dessa doença.

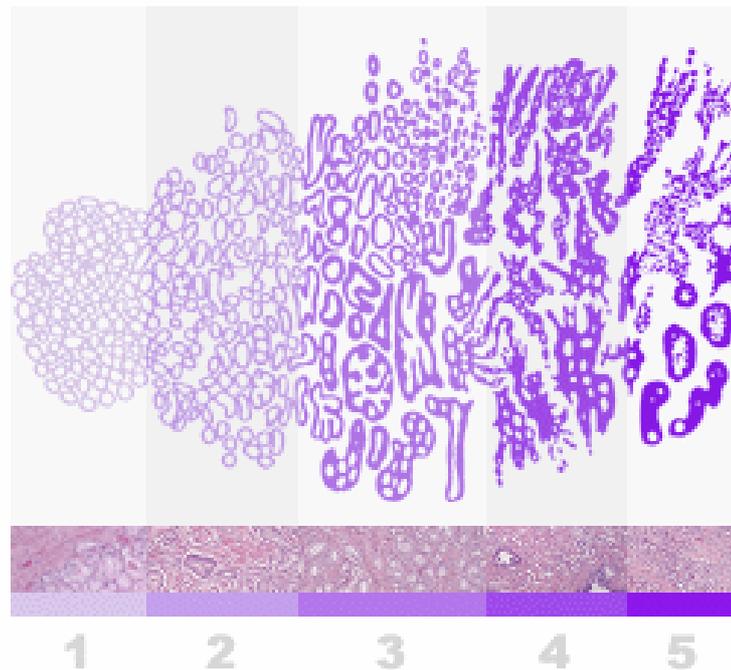


Figura 9 – **Classificação de diferenciação tumoral em próstata por *Gleason Grading System*.** O *Gleason Score* (GS) é dado pela soma dos valores (de 1 a 5) referentes as duas maiores porções do tumor, segundo a característica da arquitetura celular do tecido. Amostras que apresentam GS de 2 a 4 indicam pouca agressividade e um risco mínimo de morte do paciente por câncer de próstata, amostras com valores de GS de 5 e 6 apresentam baixa agressividade com um bom prognóstico, amostras com GS de 8 a 10 são muito agressivas e o prognóstico é ruim, e finalmente valores de GS 7 indicam tumores com agressividade intermediária e prognóstico bastante heterogêneo, de difícil predição. Figura obtida de [www.psc.edu/science/Wetzel/images/gleason.png](http://www.psc.edu/science/Wetzel/images/gleason.png).

Nos últimos anos diversas alterações em transcritos envolvidos em vias de sinalização hormonal, na expressão de oncogenes e supressores tumorais, de fatores de crescimento e seus receptores, de moléculas de adesão e reguladoras de angiogênese têm sido identificadas utilizando a abordagem de análise de expressão

gênica em larga escala por *microarrays* (Calvo *et al.*, 2005). Diferentes trabalhos têm demonstrado a existência de mensagens transcritas cuja expressão diferencial discrimina entre amostras de carcinoma de próstata (PrCa) e amostras com hiperplasia benigna (Dhanasekaran *et al.*, 2001); entre amostras de PrCa e amostras normais de próstata (Luo *et al.*, 2002; Rhodes *et al.*, 2002); e entre amostras localizadas de PrCa e amostras metastáticas (Varambally *et al.*, 2002; Ramaswamy *et al.*, 2003). Além de um potencial uso futuro como marcadores de diagnóstico, as listas de mensagens diferencialmente expressas geradas por essas abordagens podem ajudar na compreensão da biogênese desse processo neoplásico (Calvo *et al.*, 2005).

Similarmente, perfis de expressão gênica também são capazes de classificar subgrupos de pacientes com PrCa cujos tumores apresentam distintos graus de diferenciação histológica dado pelo GS (Singh *et al.*, 2002), assim como de pacientes que apresentam diferentes respostas as terapias atuais (Glinsky *et al.*, 2004). Esse tipo de abordagem pode ter importantes implicações por possibilitar a identificação de mensagens cuja expressão diferencial pode servir como um fator determinante de prognóstico. O prognóstico já poderia ser avaliado durante o diagnóstico, pois o perfil de expressão distingue molecularmente tanto tumores mais agressivos dos menos agressivos, quanto os pacientes que seriam refratários dos pacientes que responderam a terapia utilizada (Calvo *et al.*, 2005).

## 2 OBJETIVOS E RELEVÂNCIA

No Brasil, no período de 1998 a 2000, o Ministério da Saúde apontava para as neoplasias malignas como a terceira principal causa de morte, suplantada apenas pelas doenças cardiovasculares e por causas externas. Segundo as estimativas realizadas pelo Instituto Nacional de Câncer (INCA) em 2003 (<http://www.inca.org.br/estimativas/2003>), cerca de 400.000 brasileiros desenvolveram algum tipo de câncer, com aproximadamente 130.000 óbitos. Desse total, quase 40.000 seriam novos casos de câncer de próstata, com a morte de aproximadamente 9.000 homens, fazendo com que esse tipo de câncer se apresente com a maior taxa de incidência e a segunda maior taxa de morte por câncer em homens no Brasil, ficando atrás apenas dos tumores malignos de pulmão (Figura 10). Esses dados estatísticos ainda mostravam que, desde 1979, a taxa de mortalidade por câncer de próstata aumentou 141%.

Assim, inserido no contexto pós-genômico amplamente discutido na introdução desse texto, o estudo que foi desenvolvido para a obtenção do título de doutor em bioquímica teve como objetivo inicial a identificação de novos transcritos humanos potencialmente interessantes como marcadores de malignidade para o câncer próstata. O trabalho fez parte do projeto “Cooperação para Análise da Expressão Gênica – sub-área Câncer”, financiado pela Fapesp (Processo nº 99/07390-0), que foi desenvolvido sob coordenação do Prof. Dr. Sergio Verjovski-Almeida para explorar o potencial de informação gerado durante o projeto de sequenciamento em larga escala de ESTs humanas presentes em mais de 20 tipos diferentes de tumores e tecidos normais (HCGP, *Human Cancer Genome Project*). A partir de 2002, esse estudo passou a fazer parte do projeto temático para

“Identificação de marcadores moleculares para diagnóstico e prognóstico em câncer utilizando *microarrays* de DNA” (Processo nº 02/13283-6).

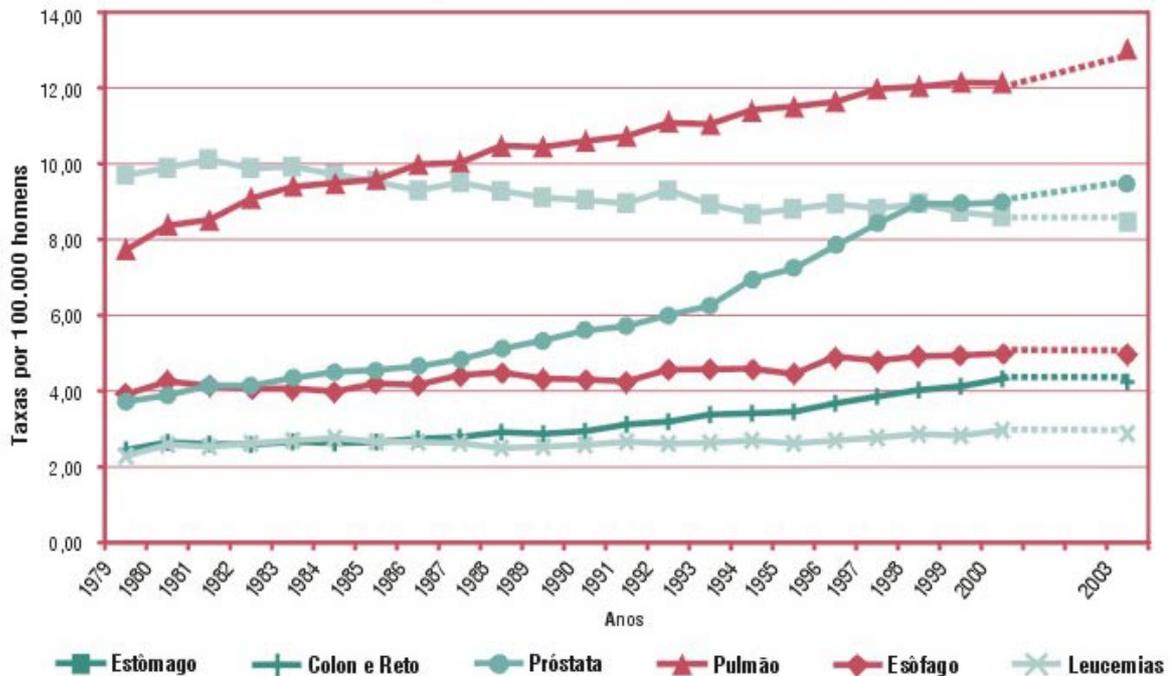


Figura 10 – **Câncer de próstata no Brasil no período de 1998 a 2003.** Taxa de morte em homens brasileiros. Dados computados pelo Instituto Nacional de Câncer (INCA) no período de 1998 a 2000, e estimativas realizadas em 2003 (<http://www.inca.org.br/estimativas/2003>).

Posteriormente uma caracterização mais aprofundada de alguns dos transcritos interessantes identificados pelas diferentes abordagens utilizadas, tanto no contexto de câncer quanto em situações normais de homeostase, foram incluídos entre os objetivos principais, descritos a seguir:

- Identificação e caracterização da expressão de novos transcritos potencialmente interessantes no contexto do câncer de próstata. Entre os novos genes estudados incluem-se parálogos de genes já descritos na literatura como envolvidos em câncer, e mensagens sem anotação funcional, mas diferencialmente expressas em tumores de próstata tanto em relação ao tecido prostático normal quanto em distintos graus de diferenciação

histológica. Espera-se que esses novos transcritos possam ser marcadores de malignidade (diagnóstico e prognóstico) e/ou alvos para o desenvolvimento, no futuro, de drogas contra o câncer;

- Caracterização da biogênese e do envolvimento em eventos regulatórios transcricionais e pós-transcricionais da nova classe de Longos RNAs não-codificantes, derivados de regiões intrônicas de *loci* gênicos que também expressam mensagens codificantes de proteína, identificados pela abordagem de *microarrays*. Para esse objetivo, células humanas foram estimuladas com o hormônio andrógeno, foram transcricionalmente inibidas pela adição de  $\alpha$ -amanitina, e transfectadas com vetores para expressão episomal de proteínas ligadoras de RNA;
- E finalmente, o início de uma caracterização funcional, da especificidade tecidual e da conservação de expressão em outras espécies para os Longos RNAs Intrônicos não-codificantes. Espera-se que o entendimento da funcionalidade desses ncRNAs ajude na compreensão do complexo programa transcricional humano.

## 3 MATERIAIS E MÉTODOS

### 3.1 Análises Informáticas

#### 3.1.1 Identificação *in silico* de parálogos de genes relacionados com câncer

Para o estudo de famílias gênicas foram utilizadas cerca de três milhões de etiquetas de seqüências transcritas (ESTs) contidas no GenBank, depositadas por dois projetos independentes de sequenciamento em larga escala de tumores, o CGAP (*Cancer Genome Anatomy Project*), lançado em 1997 pelo *National Cancer Institute*, e o HCGP (*Human Cancer Genome Project*), lançado pela FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo) e pelo Instituto Ludwig de São Paulo em 1999. Foi realizada uma busca computacional por novos membros de famílias gênicas relacionadas à tumorigênese, e para isso foi usada uma lista de 1.127 genes com envolvimento reconhecido em câncer, curada manualmente durante um *workshop* realizado em São Paulo em agosto de 2001. A lista contém genes codificantes de proteínas participantes de processos de angiogênese, crescimento celular, sinalização celular, transporte intracelular, entre outros. Dentre os genes dessa lista 36 membros da superfamília Ras de pequenas GTPases monoméricas estavam representados.

As seqüências protéicas codificadas pelos genes da lista foram utilizadas em uma busca por similaridade contra as ESTs do HCGP-CGAP traduzidas nas seis fases de leitura possíveis, utilizando a ferramenta tBLASTn (NCBI). Candidatos a novos parálogos foram identificados com base na conservação dos resíduos de aminoácido. As ESTs que apresentassem no mínimo 50% de identidade, com uma cobertura de pelo menos 50 resíduos, foram selecionadas, propiciando a distinção

entre proteínas de uma mesma família e proteínas que apenas compartilham pequenos domínios comuns (Futreal *et al.*, 2001). A seqüência de nucleotídeos das ESTs que respeitam o critério acima ainda foram comparadas com um banco contendo cerca de 100 mil transcritos humanos já conhecidos utilizando BLASTn (NCBI), para filtrar as ESTs que se alinhassem com 100 % de identidade de bases contra seqüências deste banco, o que indicaria que representavam os próprios genes presentes na lista inicial, e não parálogos. Um esquema completo da metodologia de busca de novos parálogos de genes humanos envolvidos com processos de tumorigênese está apresentado na figura 11.

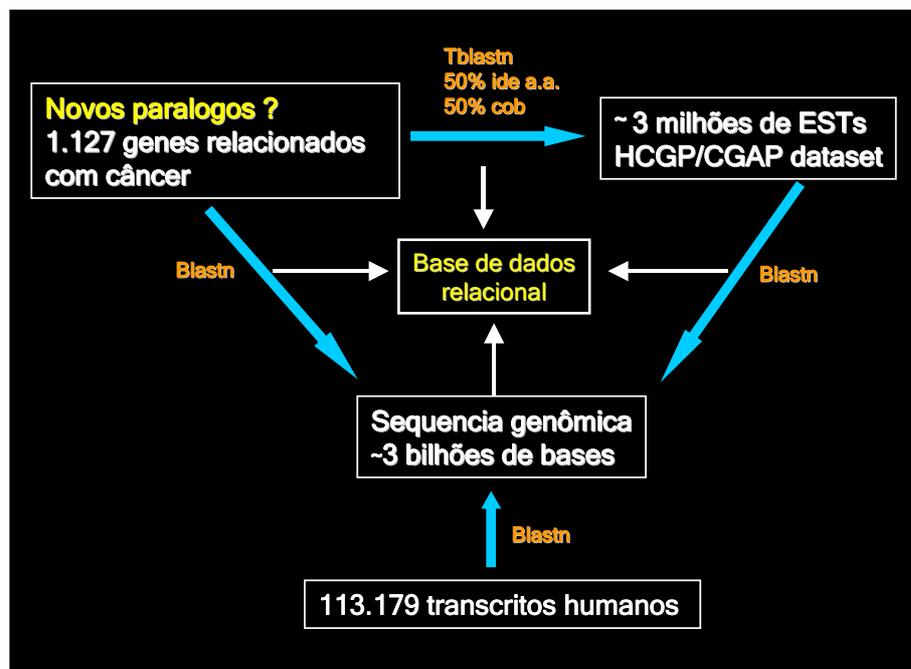


Figura 11 – **Identificação *in silico* de novos parálogos de genes humanos relacionados com câncer.** Diagrama esquemático dos processos utilizados para identificação de parálogos. Tradução nas 6 fases de leitura possíveis, do conjunto de cerca de 3 milhões de ESTs, comparação com as 1.127 proteínas relacionadas com câncer, (tBLASTn, 50% de identidade, em cerca de 50 aminoácidos), e filtragem das ESTs derivadas dos próprios genes utilizados na busca (BLASTn).

### 3.1.2 Análises dos produtos protéicos hipotéticos de *RASL11A* e *ARL-9*

Os produtos protéicos hipotéticos, Rasl11a e Arl9, foram utilizados em buscas por proteínas similares presentes no banco de dados de proteínas do GenBank,

utilizando BLASTp (NCBI). Arquivos *multifasta* foram construídos a partir das saídas individuais do BLASTp (NCBI), contendo as novas seqüências e as 50 proteínas humanas mais similares. Para uma análise de ortólogos, arquivos similares foram gerados contendo apenas seqüências provenientes de outros organismos. Esses arquivos foram manualmente inspecionados para redução da redundância e também para possibilitar a adição de alguns membros representativos de famílias menos relacionadas durante a análise das proteínas humanas. As listas finais foram utilizadas na geração de alinhamentos múltiplos de proteína e na reconstrução árvores filogenéticas da superfamília de pequenas GTPases monoméricas.

Todos os alinhamentos e a construção de árvores filogenéticas foram realizados através do programa ClustalX v 1.83 (Thompson *et al.*, 1997), usando os parâmetros padrão. As árvores filogenéticas não enraizadas foram construídas pelo método de *Neighbor Joining*, excluindo as regiões não alinháveis (*gaps*), a confiança dos ramos foi determinada através de permutações (1.000 amostragens), e finalmente, essas foram visualizadas utilizando o programa Tree View v. 1.6.6 (Page, 1996).

### **3.1.3 Análises Estatísticas para experimentos de expressão gênica com *microarrays***

#### **Marcadores de prognóstico de tumores de próstata**

Os valores de intensidade de expressão de alta qualidade (ver item 3.5.5) foram inicialmente analisados utilizando algoritmos de agrupamento para verificação se as amostras diferentes também possuíam perfis globais de expressão distintos. Agrupamentos não supervisionados dos dados foram realizados tanto por *Self-*

*Organizing Map* (SOM) usando o pacote de análise *GeneCluster2* (Tamayo *et al.*, 1999), ou por *Hierarchical clustering* (Sanchez-Carbayo *et al.*, 2002), através de distância Euclidiana e *Ward linkage* utilizando o *Spotfire Decision Site In Function Genomics* (Spotfire Inc.). Amostragens por permutações foram utilizadas para confirmar a robustez dos agrupamentos observados, onde 10.000 cópias dos dados foram geradas por adição de ruído Gaussiano aos dados originais (Sanchez-Carbayo *et al.*, 2002). Um dendograma consenso foi construído, mostrando a relação entre os nós mais freqüentes nas 10.000 amostragens.

A seleção supervisionada de mensagens cujos perfis de expressão eram significativamente correlacionados ao grau de diferenciação tumoral dado pelo *Gleason Score* (GS) (Gleason, 1977), foi realizada por análise de correlação de Pearson, de maneira similar a descrita anteriormente (Singh *et al.*, 2002). A significância estatística dessa análise foi assegurada pelo cálculo das correlações após 50.000 permutações randômicas dos rótulos de GS, sendo computada a freqüência na qual a correlação medida no grupo original foi também observada nos dados permutados. O intervalo de confiança escolhido foi de  $p < 0,001$ , onde número esperado de mensagens identificadas ao acaso seria apenas de sete em uma distribuição bicaudal. A aplicação do classificador baixo/alto GS ao perfil de expressão de dezesseis amostras com grau intermediário de diferenciação tumoral (GS 7) foi também realizada por um método supervisionado similar ao utilizado anteriormente (Van 'T Veer *et al.*, 2002).

A robustez do classificador baixo/alto GS foi avaliada por uma abordagem de *leave-one-out* (Singh *et al.*, 2002; Best *et al.*, 2003; Lapointe *et al.*, 2004). Essencialmente, cada paciente foi removido da análise e um novo grupo de mensagens correlacionadas ( $p < 0,001$ ) foi determinado com as amostras restantes.

### **Transcritos regulados pelo tratamento com andrógeno**

As hibridizações para busca de transcritos regulados por andrógeno foram realizadas com uma amostra referência marcada com Cy3 (ver item 3.5.4). Entretanto apenas os valores de intensidade provenientes das amostras de LNCaP marcadas com Cy5 foram utilizadas para determinar as razões entre tratamento (andrógeno) e controle (etanol) durante a análise de expressão diferencial.

A determinação dos transcritos com mudanças de expressão significativa em pelo menos três tempos consecutivos de tratamento com andrógeno foi realizada pela metodologia de *Significance Analysis of Microarray* (SAM) (Tusher *et al.*, 2001). Os parâmetros utilizados nessa abordagem foram: *fold-change*  $\geq 1,5$ , *K-Nearest Neighbors Imputer* e taxa de falsa descoberta (FDR)  $< 0,05$ .

### **Transcritos expressos em tecidos humanos e de camundongos**

Foram analisados três diferentes tecidos, como descrito adiante. Os transcritos expressos em pelo menos um tecido segundo os critérios utilizados (ver item 3.5.5), foram submetidos a uma análise de expressão diferencial utilizando uma abordagem complexa. O conjunto foi primeiro analisado pelo método *Significance Analysis of Microarray* (SAM) (Tusher *et al.*, 2001), resposta com múltiplas classes, 1.000 permutações, *K-Nearest Neighbors Imputer*, e taxa de falsa descoberta (FDR)  $< 0,1$ . Os considerados diferencialmente expressos entre tecidos por SAM, foram também analisados por ANOVA (*Analysis of Variance*), usando-se o limite de significância de  $p \leq 0.001$ .

Os transcritos considerados expressos diferencialmente pelos dois métodos foram selecionados para prosseguir no estudo.

## **Transcritos afetados pela inibição da RNA polimerase II**

Para determinação dos transcritos diferencialmente expressos durante o tratamento com  $\alpha$ -amanitina, molécula inibidora da RNA polimerase II, foi utilizada uma análise estatística combinada. Os valores normalizados das duas réplicas biológicas de LNCaP tratadas ou controle foram analisados pelo método *Significance Analysis of Microarray* (SAM) (Tusher *et al.*, 2001), resposta de duas classes não pareadas, com 1.000 permutações, *K-Nearest Neighbors Imputer*, e taxa de falsa descoberta (FDR) de 0,002 a 0,035; e de *Signal-to-Noise Ratio* (SNR) com 10.000 permutações ( $p < 0,05$ ) assim como em estudo similar (Golub *et al.*, 1999).

O conjunto de transcritos com diferenças significativas de expressão após o tratamento com  $\alpha$ -amanitina foi constituído apenas pelas mensagens identificadas simultaneamente por SAM e por SNR.

### **3.1.4 Identificação *in silico* de elementos de resposta ao andrógeno (ARE)**

Cerca de 3kb de seqüência genômica, a 5' dos sítios postulados para o início de transcrição das 39 mensagens intrônicas identificadas pela análise de *microarray* como reguladas por andrógeno, foram utilizados em uma busca pelo consenso de ligação do receptor de andrógeno. Para os transcritos estendidos por RACE-PCR, o novo sítio de início de transcrição foi utilizado.

A seqüência palindrômica AGAACAnnnTGTTCT de elementos de resposta ao andrógeno (*Androgen Response Elements* – ARE) (Shaffer *et al.*, 2004) foi obtida do banco de dados de elementos de DNA regulatórios em eucariotos, o TRANSFAC (Matys *et al.*, 2003) para utilização nessa análise. Os motivos ARE identificados

foram alinhados utilizando o CLUSTAL W (Thompson *et al.*, 1994), usando os parâmetros padrão, e uma representação visual gráfica do consenso foi gerado pelo WEBLOGO (Crooks *et al.*, 2004).

### **3.1.5 Análises de enriquecimento de categorias de ontologia gênica (GO)**

Buscas por categorias enriquecidas de ontologia gênica (GO) entre os transcritos identificados como diferencialmente expressos nas análises de *microarray* foram realizadas utilizando os programas EASE, *Expression Analysis Systematic Explorer* (Dennis *et al.*, 2003) ou eGOn, *Explore Gene Ontology* (<http://www.genetools.microarray.ntnu.no/egon/>). As comparações foram sempre realizadas com relação a uma lista mestra contendo a abundância de transcritos com as mesmas características de mapeamento representados na plataforma de *microarray* em cada categoria de GO. Para mensagens exônicas a comparação foi feita contra uma lista contendo 2.292 transcritos, enquanto a lista mestra de intrônicos possuía 822 transcritos.

Tanto para as mensagens exônicas quanto para intrônicas foi utilizado o número de acesso no GenBank de depósito da EST correspondente à sonda presente no *microarray* para atribuição dos GOs via UniGene (NCBI). Foi utilizada anotação de GO dessa maneira inclusive para os transcritos intrônicos, pois se postula uma atuação desses sobre os próprios transcritos codificantes de proteína de mesmo *locus* gênico (Reis *et al.*, 2005).

## 3.2 Amostras de tecido e linhagens celulares

### 3.2.1 Amostras de tecidos humanos

As amostras de tecidos tumorais e normais foram obtidas de pacientes que sofreram remoção cirúrgica total dos órgãos, e são provenientes de convênios com hospitais brasileiros. A utilização dessas nos estudos de expressão gênica descritos aqui foi previamente aprovada pelo comitê de ética das instituições de saúde que as forneceram, e foi formalmente consentida pelos próprios pacientes. Os tecidos foram congelados em nitrogênio líquido até 10 minutos após terem sido removidos dos pacientes, ficando assim armazenados até uma macro-dissecção, ou utilização direta para extração de RNA. Essa inspeção histológica, quando realizada visou definir e separar as porções tumorais e normais de cada peça.

As amostras de tecido prostático foram fornecidas pelo Hospital do Câncer/RJ e o Hospital Sirio-Libanês/SP. No total foram recebidas e armazenadas no banco de tecidos 55 pares de amostras de próstata (o tecido tumoral e seu normal adjacente) provenientes do Hospital do Câncer/RJ, e 7 amostras tumorais e 5 normais do Hospital Sirio-Libanês/SP. Apenas as amostras que apresentaram um mínimo de 80% de células malignas foram consideradas tumorais para utilização no estudo. As amostras que continham quantidade de tecido insuficiente para a extração e purificação do RNA foram eliminadas. Assim, prosseguiram nos experimentos 20 amostras normais adjacentes ao tumor e 27 amostras tumorais com características histológicas de diferenciação celular diversas: uma com *Gleason Score* (GS) igual a 5, cinco com GS igual a 6, dezesseis com GS igual a 7, uma com GS igual a 9, e quatro com GS igual a 10.

Os Hospital do Câncer/RJ também forneceu as 5 amostras de rim humano normal adjacente ao tecido tumoral, e 5 amostras de fígado humano normal foram obtidas do Hospital das Clínicas da USP, de pacientes que receberam transplante.

### **3.2.2 Amostras de tecidos de camundongos**

Amostras de fígado, próstata e rim foram extraídas de cinco camundongos (*Mus musculus*) machos adultos da linhagem BALB/c. Os órgãos dos animais também foram congelados em nitrogênio líquido até 10 minutos após terem sido removidos cirurgicamente, ficando assim armazenados até a extração de RNA.

### **3.2.3 Linhagens e culturas celulares**

As linhagens celulares de carcinoma de próstata LNCaP, DU145 e PC3, a linhagem celular tumoral de tecido renal 293T, e a HepG2, de tumor de fígado humano foram obtidas da ATCC (*American Type Tissue Collection*) e mantidas com os meios sugeridos suplementados com 10% (vol/vol) de soro fetal bovino (FCS), 3mM L-glutamina, 100µg/ml estreptomicina e 100U/ml penicilina. As células LNCaP foram cultivadas em RPMI 1640 enquanto DU145, PC3, 293T e HepG2 em DMEM, para o crescimento em estufa de CO<sub>2</sub> a 37°C. As culturas foram realizadas em placas de petri de 150mm de diâmetro contendo 25ml de meio de cultura completo, atingindo geralmente 60-80% de confluência para replicação, extração de RNA ou realização de manipulações específicas de cada experimento.

Para os experimentos de resposta ao andrógeno as células LNCaP foram cultivadas por 24 horas em meio privado de hormônios, RPMI 1640 com 5% (wt/vol) soro fetal bovino absorvido com carvão (Life Technologies). Após 24h, as células foram plaqueadas em novo meio RPMI 1640 com 10% (wt/vol) soro fetal bovino

absorvido com carvão, com adição de 1nM de andrógeno sintético R1881 (NEN - Life Sciences Products) ou volume equivalente do veículo (etanol). As células foram coletadas após 0, 6, 9, 12, 18, 24, e 48 horas de tratamento para extração de RNA total, e após 4 horas para extração de DNA genômico.

Para os experimentos de imunoprecipitação com proteínas ligadoras de RNA (*RNA Binding Proteins* – RBPs), o crescimento das células 293T foi realizado em placas contendo poli-lisina para garantir uma melhor aderência celular devido as manipulações necessárias. Após atingirem a confluência desejada (60%) as células foram transfectadas com cerca de 20µg de vetor pKALLU com as construções de cada RBP, ou com GFP no controle negativo (ver item 3.9.4), por placa, usando o reagente *GeneJammer* (Stratagene) seguindo estritamente as condições descritas pelo fabricante. Brevemente, em 1ml de DMEM (Invitrogen) sem soro e sem antibióticos, para não prejudicar a formação das vesículas lipossômicas, foi adicionado *GeneJammer* (Stratagene). Após breve incubação à temperatura ambiente por 10 minutos, os vetores foram adicionados e a mistura foi incubada novamente por 10 minutos. Nas placas semiconfluentes com 293T, com 10ml de novo meio de cultura completo, foi gotejado o volume total da solução de lipossomos com o vetor, permanecendo dessa maneira por cerca de 3 horas, quando o volume de meio completo foi elevado até 25ml em todas as placas com células transfectadas.

Para os experimentos de inibição da RNA polimerase II, as células LNCaP foram cultivadas por dois dias (até atingir cerca de  $8 \times 10^5$  células) quando os meios de cultura iniciais foram substituídos por novos contendo 50µg/mL  $\alpha$ -amanitina (Roche) ou apenas veículo. Após 24 horas as células foram coletadas e

armazenadas a  $-80\text{ }^{\circ}\text{C}$  até a extração do RNA. Os procedimentos foram repetidos para obtenção de uma réplica biológica.

### **3.3 Extração de RNA**

As amostras de tecidos foram pulverizadas utilizando mortar e pistilo de porcelana para extração de RNA. Esse procedimento foi realizado em baixa temperatura, utilizando gelo seco e nitrogênio líquido. Todas as culturas de células foram lavadas uma vez com PBS, antes do início do protocolo de extração de RNA.

#### **3.3.1 Extração de RNA poli-(A)+**

A extração de RNA mensageiro (mRNA) foi realizada utilizando tanto tecidos pulverizados quanto células. As amostras foram lisadas por adição de tampão de lise na proporção de 40mg de tecido ou  $1 \times 10^7$  células para cada 1ml, e a população de moléculas de mRNA foi purificada utilizando o  *$\mu$ MACS mRNA Isolation Kit* (Miltenyi Biotec). Esse procedimento foi realizado seguindo estritamente o protocolo do fabricante e baseia-se no pareamento das bases de oligonucleotídeos poli-dT, modificados com uma porção magnética, com as bases da cauda poli-(A) dos mRNAs eucarióticos. A separação é feita pela interação magnética dos complexos mRNA/oligonucleotídeos poli-dT com um ímã. Após diversas lavagem que eliminam proteínas, DNA e outros RNAs, os mRNAs são eluídos pela adição de  $\text{H}_2\text{O}$  a  $65^{\circ}\text{C}$ , que rompe as pontes de hidrogênio formadas entre as bases das caudas poli-A e as bases dos oligonucleotídeos poli-dT. Em média foram obtidos cerca de 5 $\mu\text{g}$  de mRNA poli-(A)+ a partir de cada 100mg de tecido processado.

### **3.3.2 Extração de RNA total**

A extração de RNA total foi realizada utilizando tanto tecidos pulverizados quanto células. As amostras foram lisadas e o RNA extraído por adição de Trizol (Invitrogen) na proporção de 50mg de tecido ou  $5 \times 10^6$  células para cada 1ml. Todo o procedimento seguiu o protocolo padrão do fabricante. O Trizol é uma solução ácida, monofásica de fenol e isotiocianato de guanidina que possibilita a extração de RNA através dos mesmos princípios do método original desenvolvido por Chomczynski e Sacchi (Chomczynski & Sacchi, 1987). Para cada 50mg de tecido ou  $5 \times 10^6$  células, foram obtidas de 10 a  $30 \mu\text{g}$  RNA total.

Para os experimentos de expressão gênica após estímulo com andrógeno, RNA total de células LNCaP, tratadas ou controles, DU145 e PC3, foi extraído utilizando gradiente de CsCl (Chirgwin *et al.*, 1979), após lise com 4M isotiocianato de guanidina; 0,1M  $\beta$ -mercaptoetanol; e 25mM citrato de sódio pH 7,0. RNA total de células da linhagem tumoral de fígado humano HepG2 também foi extraído por gradiente de CsCl. A partir de uma placa de 150mm de diâmetro com de 50 a 80% de confluência, foi obtido de 70-100 $\mu\text{g}$  RNA total.

### **3.3.3 Extração de RNA co-imunoprecipitado com proteínas ligadoras de RNA**

Os complexos de proteínas e RNA imunoprecipitados obtidos (ver item 3.8.2) foram tratados com proteinase K (600 $\mu\text{l}$  de uma solução livre de RNases contendo: 50mM Tris pH 7,5; 5mM EDTA; 25mM NaCl; 0,5% SDS e 500 $\mu\text{g}$  proteinase K) a 50°C por 30 minutos, em agitação constante para degradação proteica. Após uma centrifugação a 1.000 $\times\text{g}$  por 3 min a 4°C, os RNAs co-imunoprecipitados foram

extraídos do sobrenadante com adição de 1 volume de fenol:clorofórmio, centrifugação a 13.000×g por 5 min a 4°C, adição de um volume de clorofórmio, e finalmente, outra centrifugação a 13.000×g por 5 min a 4°C.

As populações de RNA obtidas foram precipitadas usando 0,8 volumes de isopropanol, 80µl de 4M acetato de sódio ou amônia, 3µl de 1M MgCl<sub>2</sub>, e 8µl de 20mg/ml glicogênio (Roche), e incubação a -80°C. As amostras de RNA precipitado foram centrifugadas a 13.000×g por 30 min a 4°C, lavadas com 200µl de etanol, secas por 10 min a temperatura ambiente e ressuspendidas em H<sub>2</sub>O livre de RNAses.

#### **3.3.4 Tratamento com DNase livre de RNase**

Contaminações com DNA genômico foram minimizadas nas preparações de RNA total e mRNA. Cerca 0,5µg de RNA poli-(A)<sup>+</sup>, foi tratado com 1 unidade (U) de *RQ1 RNase-free DNase* (Promega) por 45 minutos a 37°C, em um volume de 15µl na presença de 20U de *RNasin* (Promega), inibidor de RNase. Após inativação enzimática por adição da solução de parada com EDTA (*Stop Solution*) e aquecimento a 65°C por 15 minutos, as amostras de mRNA foram utilizadas diretamente nos experimentos subseqüentes.

As amostras de RNA total foram tratadas com a *RNase-free DNase I* (Qiagen). Cerca de 100µg de RNA total foram tratados e purificados seguindo estritamente o protocolo de digestão de DNA genômico nas colunas do *RNeasy* (Qiagen).

### 3.3.5 Quantificação e verificação da qualidade dos RNAs extraídos

Para determinação da qualidade e quantidade dos RNAs extraídos, todas as amostras tiveram suas absorbâncias medidas nos comprimentos de ondas 260 e 280nm no espectrofotômetro *ND-1000* (NanoDrop Technologies), e foram separadas por eletroforese utilizando o *2100 BioAnalyzer* (Agilent Technologies).

A quantificação do RNA foi feita utilizando o valor de sua densidade ótica (DO) em 260nm (40µg/mL RNA possui uma  $DO_{260} = 1$ ). A razão 260nm/ 280nm foi utilizada para estimar a porcentagem de contaminação de proteína nas amostras de mRNA. Razões entre 1,6-2,0 foram consideradas empiricamente satisfatórias, devido a observação de reações de marcação por transcrição reversa com eficiência similar utilizando RNAs dentro dessa faixa.

As amostras de RNA foram também submetidas a uma eletroforese capilar para determinação simultânea da integridade da população de RNAs e da ausência de grande contaminação por DNA genômico. Essas análises baseiam-se na observação dos perfis dos eletroferogramas, e na comparação com o padrão *RNA 6000 ladder* (Ambion Inc.). A integridade das amostras de mRNA de próstata foi estimada pela maior concentração de moléculas de RNA de maior peso molecular. Para amostras de RNA total a razão da quantidade dos RNAs ribossomais 28S e 18S foi utilizada como parâmetro de qualidade, sendo aceitáveis valores acima de 1,1 para amostras de tecido e acima de 1,5 para amostras derivadas de linhagens celulares.

Para cada tecido prostático no estudo a quantidade de mRNA obtida foi em média de 1,0-2,5µg para cada 50mg processado, sendo que a maioria das amostras possuía no mínimo 100mg de tecido. Como saldo final, para as análises de expressão gênica foram obtidas amostras de mRNA de alta qualidade para 47

tecidos, 20 pares de tumores e suas respectivas regiões normais adjacentes e 7 amostras tumorais não pareadas. As amostras de RNA total provenientes de tecidos humanos ou de camundongo geraram cerca de 10-30 $\mu$ g para cada 50mg processado, enquanto o rendimento de RNA das amostras de linhagem celular foi de cerca de 50 $\mu$ g para cada  $1 \times 10^7$  células processadas.

### 3.4 Abordagens de PCR para amplificação de transcritos completos

#### 3.4.1 RACE-PCR

A metodologia para amplificação de extremidades desconhecidas de fragmentos de cDNA, RACE-PCR (*Rapid Amplification of cDNA Ends*), foi utilizada para obtenção das seqüências completas de parte dos novos transcritos humanos identificados. Amplificações por RACE das extremidades 5' e 3' de moléculas de cDNA foram realizadas utilizando bibliotecas comerciais *Marathon Ready cDNA* (BD Biosciences) geradas a partir de RNA poli-(A)+ de uma linhagem de adenocarcinoma de próstata (*Human XG Prostatic Adenocarcinoma*), de testículo (*Human Testis*), de pulmão (*Human Lung*), ou de uma biblioteca feita de RNA poli-(A)+ de colon, clonada direcionamente no vetor lambda gt11 (*Lambda gt11* - BD Biosciences).

Dois pares de oligonucleotídeos iniciadores específicos complementares as seqüências das ESTs candidatas a novos transcritos humanos foram desenhados ancorados em suas pontas, orientados para o centro da mensagem, e de forma a possibilitar a utilização de cada par em reações de PCR seqüenciais (*Nested*), que aumentam a especificidade dos produtos gerados. As reações de RACE-PCR com as bibliotecas *Marathon* foram realizadas em 50 $\mu$ l, com: 2 $\mu$ l de molde de cDNA; 0,2mM de dNTP; 1X *Advantage 2 PCR Buffer*; 1X *Advantage 2 Polimerase Mix* (BD Biosciences); 0,2 $\mu$ M de oligonucleotídeos iniciadores específicos antisense ou sense;

e 0,2 $\mu$ M do oligonucleotídeo iniciador universal (*Adaptor Primer*) complementar aos adaptadores existentes nas porções 5' e 3' de cada molécula de cDNA. As reações com as bibliotecas *Lambda gt11* foram similares, exceto pela adição de 0,2 $\mu$ M do oligonucleotídeo iniciador complementar a região do vetor próxima a porção 5' ou complementar a região próxima a porção 3' das mensagens clonadas, o que permitiu a determinação da orientação transcricional das mensagens amplificadas. Todas as reações foram incubadas no termociclador *GeneAmp PCR System 9700* (Applied Biosystems), utilizando as seguintes condições de ciclagem: 40 repetições de 95°C por 30 segundos para desfazer as duplas fitas, e de 68°C por 4 minutos para anelamento e extensão, finalizando com 68°C por mais 10 minutos.

Os produtos obtidos foram ligados no *pGEM-T Easy* (Promega) para sequenciamento dideoxi (MegaBace – Armeshan) utilizando oligonucleotídeos iniciadores complementares as regiões promotoras das RNA polimerases T7 e SP6 presentes na seqüência do vetor, flanqueando o inserto. Confirmadas as identidades das porções 5' e 3' obtidas, através de mapeamento no genoma humano, esses foram unidos com uma única reação de amplificação em condições idênticas as descritas acima, porém utilizando ambos fragmentos em quantidades equimolares como molde, e um novo par de oligonucleotídeos iniciadores ancorados nas extremidades do novo transcrito humano. Procedimentos similares de clonagem e sequenciamento foram repetidos com as mensagens completas.

### **3.4.2 PCR com oligonucleotídeos iniciadores específicos**

Para caso particulares onde a metodologia de RACE-PCR não produziu resultados interpretáveis, foram utilizadas abordagens de PCR com oligonucleotídeos iniciadores específicos desenhados utilizando a informação

contida em predições genômicas para novos genes humanos codificantes de proteína, ou em RNAs já depositados no GenBank. Novamente, as bibliotecas comerciais *Marathon Ready cDNA* (BD Biosciences) geradas a partir de RNA poli-(A)+ de uma linhagem de adenocarcinoma de próstata (*Human XG Prostatic Adenocarcinoma*), de testículo (*Human Testis*), e de pulmão (*Human Lung*) foram utilizadas para esse propósito.

As condições de amplificação também foram idênticas às relatadas para RACE-PCR (ver item 3.4.1), exceto na adição simultânea de 0,2 $\mu$ M do oligonucleotídeos iniciadores específicos antisenso e senso, e na utilização de uma temperatura independente e mais baixa para anelamento (65°C) durante os ciclos. Os produtos obtidos foram também ligados no *pGEM-T Easy* (Promega) para sequenciamento dideoxi (MegaBace – Armeshan) utilizando oligonucleotídeos iniciadores complementares as regiões promotoras das RNA polimerases T7 e SP6 presentes na seqüência do vetor que flanqueia os insertos.

### **3.5 Estudo de expressão gênica por *microarrays* de cDNA**

#### **3.5.1 Plataforma de *microarrays* de cDNA do IQUSP**

Parte dos estudos de expressão gênica descritos aqui fizeram parte do projeto CAGE (Cooperação para Análise da Expressão Gênica), sub-área de câncer, financiado pela FAPESP (Processo n° 99/07390-0) (<http://bioinfo.iq.usp.br/cagecancer/>). Com esse financiamento foi construído um *microarray* de cDNA cujas características encontram-se publicadas no banco de dados *Gene Expression Omnibus* (GEO – <http://www.ncbi.nlm.nih.gov/geo/>), sob o número de acesso GPL3985.

Brevemente, a plataforma foi construída após seleção de sondas a serem depositados nas lâminas por análises bioinformáticas com as 1.016.501 seqüências expressas (ESTs) obtidas pela metodologia ORESTES (Dias-Neto *et al.*, 2000) a partir de diversos tipos de câncer durante o Projeto Genoma Humano do Câncer (HCGP - <http://www.ludwig.org.br/ORESTES>). Essa plataforma contém duplicatas de 4.608 sondas de cDNA provenientes dos produtos amplificados das ESTs do HCGP selecionadas, com um total de 9.216 elementos depositados em lâminas de vidro *Type 7 STAR* (Amersham Biosciences Inc.) utilizando o *Generation-III Array Spotter* (Molecular Dynamics). Dentre as sondas depositadas encontram-se 1.301 produtos amplificados de clones cuja EST associada não apresenta similaridade com seqüências completas humanas anotadas. Um conjunto de 2.385 produtos derivados de ESTs similares a seqüências relacionadas com desenvolvimento de câncer, e 922 produtos de genes constitutivos (*housekeeping*) e controles negativos e positivos para os procedimentos de hibridização, também foram depositados. Dentre os controles, 384 cDNAs são provenientes do *Lucidea Microarray ScoreCard* (Amersham Bioscience), onde além de controles positivos e negativos (cDNA de planta), existe um outro conjunto de fragmentos de DNA intergênico de levedura que não hibridizam com cDNAs humanos.

Do conjunto de transcritos humanos sem similaridade com transcritos conhecidos encontram-se 822 transcritos derivados de ESTs que mapeiam em regiões intrônicas de mensagens anotadas, 241 derivados de seqüências que mapeiam em regiões sem mensagens conhecidas (intergênicas). Detectamos também 238 transcritos contaminantes de outros organismos, que durante a seleção não foram detectados pelos filtros informáticos para seqüências, e agora foram eliminados das análises.

Nas lâminas as sondas estão organizadas em 12 *sub-arrays*, cada um contendo 12 linhas e 32 colunas (384 sondas/*sub-array*). Em cada *sub-array*, as linhas de 1 a 5 contêm sondas para mensagens humanas sem similaridade com transcritos conhecidos; as linhas de 6 a 10 contêm as sondas para mensagens humanas conhecidas; e finalmente, as linhas 11 e 12 de cada *sub-array* contêm sondas para os 32 controles do *ScoreCard* e 32 sondas para conhecidos marcadores de próstata e genes *housekeeping*.

### **3.5.2 Obtenção de alvos fluorescentes**

#### **Marcação direta para geração de alvos fluorescentes**

Os alvos fluorescentes foram obtidos por marcação direta usando o *CyScribe First-Strand cDNA Labelling Kit* (Amershan Biosciences), através de transcrição reversa na presença de dCTP modificado com Cy3, fluoróforo que emite uma fluorescência verde (550nm de absorção e 570nm de emissão) ou Cy5, que emite uma fluorescência vermelha (650nm de absorção e 670nm de emissão).

Brevemente, 15µg de RNA total combinado apenas com oligo-dT, ou 1µg de RNA poli-(A)+ com oligo-dT e nonâmeros randômicos (9 bases) foram utilizados para síntese de cDNA, seguindo estritamente o protocolo do fabricante, usando-se 100 unidades de transcriptase reversa (*CyScript reverse transcriptase*) a 42°C, durante 90 minutos. Os moldes de RNA foram removidos por hidrólise alcalina com 2µl de 2,5M NaOH a 37°C por 15 minutos, e em seguida as reações foram neutralizadas com 10µl de 2M MOPS. Cinco volumes de tampão de purificação (5,3M Guanidina-HCl; 150mM KOAc) foram adicionados às amostras para promover a eliminação dos nucleotídeos modificados não incorporados usando *96-well*

*Multiscreen Filter Plates* (Millipore). Quatro lavagens com etanol 80% foram realizadas por centrifugação das placas a 3.000 x g por 5 minutos, e uma última centrifugação por 5 minutos foi feita para eliminação de etanol residual. Os alvos marcados foram eluídos das placas com 10mM Tris pH 8,5 através de centrifugação a 3.000×g por 5 minutos, desidratados totalmente em uma *SpeedVac* (Eppendorf) e mantidos a -20°C protegidos da luz até o uso.

Em todas as reações de marcação realizadas foram incluídos 2,5µl da mistura de mRNAs controle contidos no *Lucidea Microarray ScoreCard* (Amersham Bioscience). Esses mRNAs são complementares ao grupo de 32 seqüências de levedura depositadas no *microarray* de cDNA (detalhes no item 3.5.1).

### **Marcação indireta para geração de alvos fluorescentes**

Os alvos fluorescentes foram obtidos por marcação indireta usando *CyScribe Post-Labeling Kit Protocol* (Amersham Biosciences), e *SuperScript™ Indirect cDNA Labeling System* (Invitrogen) através de transcrição reversa utilizando aminoallyl-dUTP, e acoplamento a esses resíduos de uracila modificados, respectivamente, com Cy5 ou *AlexaFluor 647* (Molecular Probes). Ambas moléculas emitem uma fluorescência vermelha (650nm de absorção e 670nm de emissão).

As reações com o *CyScribe Post-Labeling Kit Protocol* (Amersham Biosciences) utilizaram 10-15µg de RNA total combinado apenas com oligo-dT para síntese de cDNA seguindo estritamente o protocolo do fabricante, usando-se 100 unidades da transcriptase reversa (*CyScript reverse transcriptase*) a 42°C, durante 90 minutos. Os moldes de RNA foram removidos por hidrólise alcalina com 15µl de 1N NaOH a 70°C por 10 minutos, em seguida as reações foram neutralizadas com 15µl de 1N HCl, e 20µl de 3M Acetato de Sódio pH 5,2 foi adicionado a mistura. Os

nucleotídeos não incorporados foram eliminados das amostras através do protocolo *SNAP Column* (Invitrogen). As amostras de cDNA purificado foram incubadas a -20°C por cerca de 16 horas após adição de 10µl de 3M Acetato de Sódio pH5,2; 2µl de 20mg/ml de glicogênio; e 300µl de 100% etanol gelado. Os produtos precipitados foram centrifugados a 14.000×g por 20 minutos a 4°C, lavados com 250µl de etanol 75% gelado (14.000×g por 2 minutos a 4°C), e secos por 10 minutos a temperatura ambiente. Quinze microlitros de H<sub>2</sub>O foram utilizados para ressuspender as amostras onde foram também adicionados 15µl de Cy5 (*Cy5™ Dye Vials* – Amershan Biosciences) reconstituído com 0,1M bicarbonato de sódio pH 9. Após 2 horas à temperatura ambiente e protegidas da luz, as reações de acoplamento foram finalizadas por adição de 15µl 4M hidroxilamina e incubação a 15min à temperatura ambiente. Os alvos marcados e purificados foram obtidos após repetição do protocolo das *SNAP Column* (Invitrogen), sendo esses mantidos a -20°C e protegidos da luz até o uso.

As reações utilizando o *SuperScript™ Indirect cDNA Labeling System* (Invitrogen) foram feitas com oligo-dT mais hexâmeros randômicos (6 bases) e toda população de RNA imunoprecipitada em conjunto com as proteínas ligadoras de RNA (ver item 3.8.2) para síntese de cDNA, seguindo estritamente o protocolo do fabricante, usando-se 800 unidades da transcriptase reversa (*SuperScript™ III RT*) a 46°C, durante 3 horas. As etapas de hidrólise alcalina, purificação do cDNA com aminoallyl, e purificação do cDNA acoplado com *AlexaFluor 647* (Molecular Probes), foram idênticas as descritas acima para as reações feitas com *CyScribe Post-Labeling Kit Protocol* (Amershan Biosciences). Entretanto, as reações de acoplamento foram feitas com adição de 5µl *2X Coupling Buffer* as amostras de cDNA purificadas, adição de 5µl *AlexaFluor 647* (Molecular Probes) ressuspendido

com 2 $\mu$ l de DMSO e 3 $\mu$ l H<sub>2</sub>O, e incubação por 2 horas a temperatura ambiente e protegidas da luz.

Apenas nas reações realizadas com o *CyScribe Post-Labeling Kit Protocol* (Amersham Biosciences) foram incluídos 2,5 $\mu$ l da mistura de mRNAs controle contidos no *Lucidea Microarray ScoreCard* (Amersham Bioscience). Esses mRNAs são complementares ao grupo de 32 seqüências de levedura depositadas no *microarray* de cDNA (detalhes no item 3.5.1).

### **3.5.3 Hibridizações em lâminas de *microarrays* de cDNA**

Cada lâmina foi hibridizada com no mínimo 12pmol de cDNA marcado com Cy3 e/ou Cy5. Os alvos marcados foram ressuspendidos em 250 $\mu$ l de tampão de hidrização, composto por 50% formamida, 25% de H<sub>2</sub>O, e 25% de *Microarray Hybridization Buffer Version 2* (Amersham Biosciences); desnaturados por 2 minutos a 92°C; e centrifugados a 14.000 $\times$ g por 5 minutos.

Todos os passos de processamento das lâminas (bloqueio, hibridização, lavagens) foram realizados no *Automated Slide Processor* (Amersham Biosciences). As lâminas foram incubadas por 16 horas a 42°C e posteriormente lavadas a temperatura ambiente com 1xSSC/0,2%SDS por 5 minutos; com 0,1xSSC/0,2%SDS por 5 minutos; e com 0,1XSSC por 3 minutos. Finalmente, lâminas prontas para obtenção das imagens por excitação com laser foram obtidas após um jato de isopropanol, e um jato de ar a 42°C para secá-las.

### **3.5.4 Desenho dos experimentos com *microarrays* de cDNA**

#### **Marcadores de prognóstico de tumores de próstata**

Vinte pares de amostras tumorais e suas respectivas regiões normais adjacentes marcadas respectivamente com Cy3-dCTP e Cy5-dCTP, foram hibridizadas em 20 lâminas. Um novo conjunto de hibridizações com as amostras tumorais e normais marcadas com fluoróforos invertidos (*dye-swap*) foi realizado para facilitar a correção de diferenças de incorporação e de emissão características de cada molécula fluorescente utilizada. As 7 amostras tumorais não pareadas foram marcadas duas vezes, uma com Cy3 outra com Cy5, e hibridizadas em uma mesma lâmina.

Todas a hibridizações, assim como os procedimentos de lavagem foram realizados automaticamente no *Automated Slide Processor* (Amershan Biosciences), como descrito (ver item 3.5.3).

#### **Transcritos regulados pelo tratamento com andrógeno**

Cada amostra proveniente de células LNCaP tratadas com andrógeno ou controle apenas com etanol foram marcadas com Cy5-dCTP, e hibridizadas duas vezes contras as lâminas de *microarray* de cDNA. Uma referência universal composta por quantidades equimolares de RNA proveniente de células LNCaP, DU145 e PC3, foi marcada com Cy3-dCTP, e hibridizada em conjunto com cada amostra experimental, totalizando 26 lâminas.

Todas a hibridizações, assim como os procedimentos de lavagem foram realizados automaticamente no *Automated Slide Processor* (Amershan Biosciences), como descrito (ver item 3.5.3).

### **Transcritos expressos em tecido humanos e de camundongo**

Os RNAs extraídos de cada tecido (rim, próstata e fígado) humano e de camundongos provenientes de 5 espécimes diferentes foram unidos em amostras únicas para a síntese de cDNA e posterior marcação com Cy5 (ver item 3.5.2).

Os alvos marcados de cada tecido (humano e camundongo) foram divididos em três alíquotas iguais para serem hibridizados em três diferentes lâminas de *microarray* de cDNA. Ao todo, 18 experimentos foram realizados, sendo 9 com material proveniente de tecido humano e 9 de tecido de camundongo.

Todas as hibridizações, assim como os procedimentos de lavagem foram realizados automaticamente no *Automated Slide Processor* (Amersham Biosciences), como descrito (ver item 3.5.3).

### **Transcritos co-imunoprecipitados com proteínas ligadoras de RNA**

As imunoprecipitações com cada proteína ligadora de RNA (RBP) foram realizadas em duplicata com quantidades similares de RBP ou da proteína controle GFP, normalizadas pela quantidade total de proteínas presentes nos extratos de 293T (ver item 3.9.5), e todo o RNA obtido foi transformado em cDNA marcado com *Alexa Fluor 647* (Molecular Probes). Cada amostra foi hibridizada sem referência em uma lâmina de *microarray* de cDNA, gerando um conjunto de 10 experimentos.

Todas as hibridizações, assim como os procedimentos de lavagem foram realizados automaticamente no *Automated Slide Processor* (Amersham Biosciences), assim como descrito (ver item 3.5.3).

### 3.5.5 Obtenção de imagens e valores de intensidade comparáveis

Após os procedimentos de hibridização e lavagens, as lâminas de *microarray* foram excitadas com laser nos devidos comprimentos de onda utilizando o *Gen III Confocal Scanner* (Molecular Dynamics) ou *GenePix 4000B* (Axon Instruments), para obtenção das imagens relativas a cada experimento. Os seguintes parâmetros foram utilizados como padrão nesse procedimento: resolução de 10 $\mu$ m por pixel; comprimento de onda de excitação de 532nm (para alvos marcados com Cy3) ou 633nm (para alvos marcados com Cy5 ou *AlexaFluor 647*); e tensão PMT de 350 volts (para alvos marcados com Cy3) e 600 volts (para alvos marcados com Cy5 ou *AlexaFluor 647*).

Os valores de intensidade de cada um dos canais (Cy3, e Cy5 ou *AlexaFluor 647*) para cada sonda depositada foram obtidos a partir das imagens geradas, utilizando o programa *ArrayVision* (Imaging Research Inc.), com a abordagem de *Artifact-Removed* (ARM). O valor de ARM representa a média de todos os valores de pixels restantes provenientes daquela sonda após a retirada dos valores que excedem quatro desvios absolutos da mediana (MADs). Esse tratamento minimiza artefatos provenientes, por exemplo, de partículas de poeira. Os valores brutos de intensidade foram obtidos após subtração dos valores de hibridização inespecífica na lâmina (*background*) nos arredores da sonda (distância de 2 pixels com 3 pixels de largura) dos valores de ARM calculados.

Quando possível, os arquivos de saída com os valores brutos de intensidade foram importados para o programa do *Lucidea Microarray ScoreCard* (Amersham Bioscience), onde os valores de intensidade dos genes controles contidos no *microarray* foram usados na avaliação da qualidade das hibridizações.

### **Marcadores de prognóstico de tumores de próstata**

Os valores brutos de intensidade obtidos de experimentos com boa qualidade para determinação de marcadores de prognóstico de tumores de próstata foram normalizados pelo algoritmo *LOWESS* (*locally weighted linear regression algorithm*), disponível no pacote R (<http://www.r-project.org/>), para redução dos erros sistemáticos nos valores de intensidade gerados pela diferença de incorporação e emissão de fluorescência de Cy3 e Cy5 (Quackenbush, 2002). Esses valores normalizados passaram depois por um tratamento estatístico para eliminação de dados artefatuais, ou pouco reprodutíveis, no programa *ArrayStat* (Imaging Research Inc.). Esse programa constrói um histograma de frequências a partir da média de intensidade de cada sonda, além de gerar um *scatterplot* do desvio padrão dos dados em relação à mediana, revelando que o erro dos dados varia de acordo com a faixa de intensidade de sinal dos genes. Isso cria um modelo estatístico em cada experimento para eliminação de valores com desvios acima do desvio-padrão observado em cada faixa de intensidade, através das curvas ajustadas de erro (Nadon *et al.*, 2001). Com esses tratamentos finalizados, foram obtidos dados de intensidade de expressão de alta qualidade para as análises de expressão diferencial.

### **Transcritos regulados pelo tratamento com andrógeno**

Os valores brutos provenientes de cada sonda presente no *microarray* foram comparados ao valor médio de intensidade mais três desvios-padrão dos controles negativos do *Lucidea Microarray Scorecard* (G1p, um gene de planta) presentes em cada *sub-array*, para determinação do limite mínimo de detecção de cada hibridização. Os transcritos representados pelas sondas que possuíam intensidade

acima deste limite mínimo foram considerados acessos naquele lado da lâmina. Cada um dos lados foi considerado individualmente, e a média dos valores de intensidade, retirando-se 20% das sondas com maior e 20% das sondas com menor intensidade de sinal, de cada réplica foi utilizada como fator de normalização, para tornar os experimentos comparáveis entre si.

### **Transcritos expressos em tecidos humanos e de camundongos**

Os valores brutos provenientes de cada sonda gerados pelos experimentos de hibridização com amostras humanas e de camundongo foram normalizados pela intensidade dos genes *housekeeping* presente nas diferentes lâminas (ver item 3.5.1). A média dos valores de intensidade, retirando-se 20% dessas sondas com maior e 20% das sondas com menor intensidade de sinal, de cada réplica foi utilizada como fator de normalização para tornar os experimentos comparáveis.

Os transcritos considerados como expressos em cada tecido passaram por um critério bastante conservador, visto que a hibridização de mensagens de camundongo em uma plataforma com sondas humanas poderia ser bastante criticada. Os diferentes transcritos foram considerados acesos em um dos lados da lâmina se sua intensidade estivesse igual ou acima da intensidade do controle negativo do *Lucidea Microarray Scorecard* (G1p, um gene de planta) presente no mesmo *sub-array*, acrescido de três desvios-padrão da média dos valores em toda a lâmina. Cada transcrito só foi considerado aceso em uma lâmina se os valores de intensidade das sondas de ambos os lados satisfizessem o critério anterior. E finalmente, um transcrito só foi considerado como expresso no tecido, se tivesse sido considerado aceso em, no mínimo, duas das três lâminas hibridizadas.

### **Transcritos co-imunoprecipitados com proteínas ligadoras de RNA**

Os valores brutos provenientes de cada sonda gerados pelo experimento de imunoprecipitação com proteínas ligadoras de RNA (RBPs) foram comparados com a intensidade média dos 12 controles negativos do *Lucidea Microarray Scorecard* (G1p, um gene de planta) presente em cada *sub-array* da lâminas. Foram considerados como acessas as mensagens cujas intensidades excederam o valor da média dos controles negativos mais 3 desvios-padrão, em no mínimo 3 dos 4 valores gerados (dois de cada uma das lâminas, réplicas técnicas).

Contudo, só foram considerados como especificamente imunoprecipitados os transcritos que estavam acessos nas amostras imunoprecipitadas com as RBPs (PABP, EIF4E, hTRA2 e HUD), cujos valores de intensidade ainda eram no mínimo 3 desvios-padrão acima de seu valor médio observado nas amostras imunoprecipitadas com GFP. Esse último filtro minimizou a presença de resultados indicativos de RNAs imunoprecipitados inespecificamente ligados a proteína GFP, que não tem domínios de ligação a RNAs conhecidos, ou a região comum de proteína A presente em todas produtos expressos pelo sistema do vetor pKALLU (ver item 3.9.4).

### **3.6 Hibridizações em *Northern Blots***

A expressão em diferentes tecidos, o tamanho, e a orientação transcricional de algumas das novas mensagens identificadas foi investigada através da hibridização de sondas radioativas em *Northern Blots*. Para esses procedimentos foram utilizadas tanto a membrana de nitrocelulose comercial *Human Multiple Tissue Northern (MTN) Blot II* (Clontech Laboratories Inc.), onde encontram-se transferidos 2 $\mu$ g de RNA poli-(A)+ de alta qualidade, isolados de 9 tecidos humanos diferentes:

baço, timo, próstata, testículo, ovário, intestino delgado, colon e leucócitos sanguíneos periféricos, quanto membranas construídas no laboratório.

Cerca de 4µg de RNA poli-(A)+ isolados de 3 linhagens celulares diferentes de tumores de próstata, PC3, DU145 e LNCaP (ver item 3.3.1 para detalhes da extração) foram utilizadas na construção de membranas de *Northern Blot*. Os RNAs foram separados por eletroforese em gel desnaturante com 1,5% de agarose; 20mM ácido bórico, pH 8,3; 0,25mM de EDTA; e 6,5% de formaldeído (Merck). Antes de aplicadas no gel, as amostras foram desnaturadas a 60°C por 10 minutos em uma solução contendo formamida, formaldeído e tampão de corrida (20% glicerol; 20mM ácido bórico, pH 8,3; 0,25mM EDTA; 50% formamida; 16% formaldeído; e azul de bromofenol). Após 5 minutos no gelo, as amostras foram aplicadas no gel e submetidas à eletroforese a 70 volts em tampão de corrida durante cerca de 4 horas.

O gel foi lavado em 4x SSC (0,6 M NaCl, 60 mM citrato de sódio, pH 6,45) e 10x SSC (1,5 M NaCl, 0,15 M citrato de sódio, pH 6,45) durante 20 minutos para eliminação do excesso de formaldeído. As amostras foram transferidas por capilaridade a uma membrana de nylon (Hybond-N) com 10x SSC durante 16 horas, e os RNA foram fixados na membrana por radiação ultravioleta utilizando o *UV Crosslinker* (Biorad).

### **3.6.1 Detecção de mensagens utilizando sondas de fita dupla**

A marcação das sondas com [ $\alpha^{32}\text{P}$ ]-dCTP seguiu o protocolo do *Rediprime II* (Amershan Pharmacia), usando como molde fragmentos gerados a partir de RACE-PCR (descrito no item 3.2.6), e um produto com cerca de 500bp proveniente da região codificante do gene da  $\beta$ -actina (*ACTB*). Os procedimentos de hibridização foram realizados de acordo com as condições especificadas no manual fornecido

pelo fabricante das membranas *Multiple Tissue Northern (MTN) Blot II* (Clontech Laboratories Inc.). A detecção foi realizada no *Storm Phosphoimager* (Molecular Dynamics) após dois dias de exposição das membranas no cassete *Storm Phosphor Screen* (Molecular Dynamics).

### 3.6.2 Detecção de mensagens utilizando sondas de fita simples

Duas abordagens distintas foram utilizadas para gerar sondas radioativas que fossem específicas para definir a orientação transcricional de mensagens intrônicas em hibridizações com membranas de Northern Blot. Uma foi baseada no uso de oligonucleotídeos marcados com  $^{32}\text{P}$  em suas regiões 5'. A marcação dessas foi realizada através de reações de fosforilação com  $[\gamma^{32}\text{P}]\text{-ATP}$  e uma polinucleotídeo quinase (T4 PNK – Invitrogen). A hibridização dos oligonucleotídeos marcados foi realizada em uma membrana comercial de nylon contendo  $2\mu\text{g}$  de RNA poli-(A)+ isolado de vários tecidos humanos (*Multiple Tissue Northern blot II* - Clontech Inc.), seguindo as especificações recomendadas pelo fabricante.

Na outra abordagem a síntese das sondas radioativas foi realizada através de transcrição *in vitro* usando a T7 RNA polimerase na presença de  $[\alpha^{32}\text{P}]\text{-dCTP}$  (Riboprobe - Promega). Produtos gerados pelos experimentos de RACE-PCR (descrito no item 3.2.6) foram utilizados como molde após sua linearização orientada por reações de PCR com um oligonucleotídeo de seqüência igual à região anterior ao sítio promotor da RNA polimerase (T7 ou SP6) e outro complementar a região do inserto com a orientação desejada. A figura 12 ilustra esse procedimento. A geração das sondas de RNA seguiu as especificações do fabricante. As hibridizações ocorreram em membranas de nylon construídas com  $4\mu\text{g}$  de poli-(A)+ RNA isoladas

de 3 linhagens celulares diferentes de tumores de próstata, PC3, DU145 e LNCaP (ver item 3.6).

Em ambas abordagens, a detecção foi realizada no *Storm Phosphor Screen* (Molecular Dynamics) após dois dias de exposição da membrana marcada no cassete apropriado.

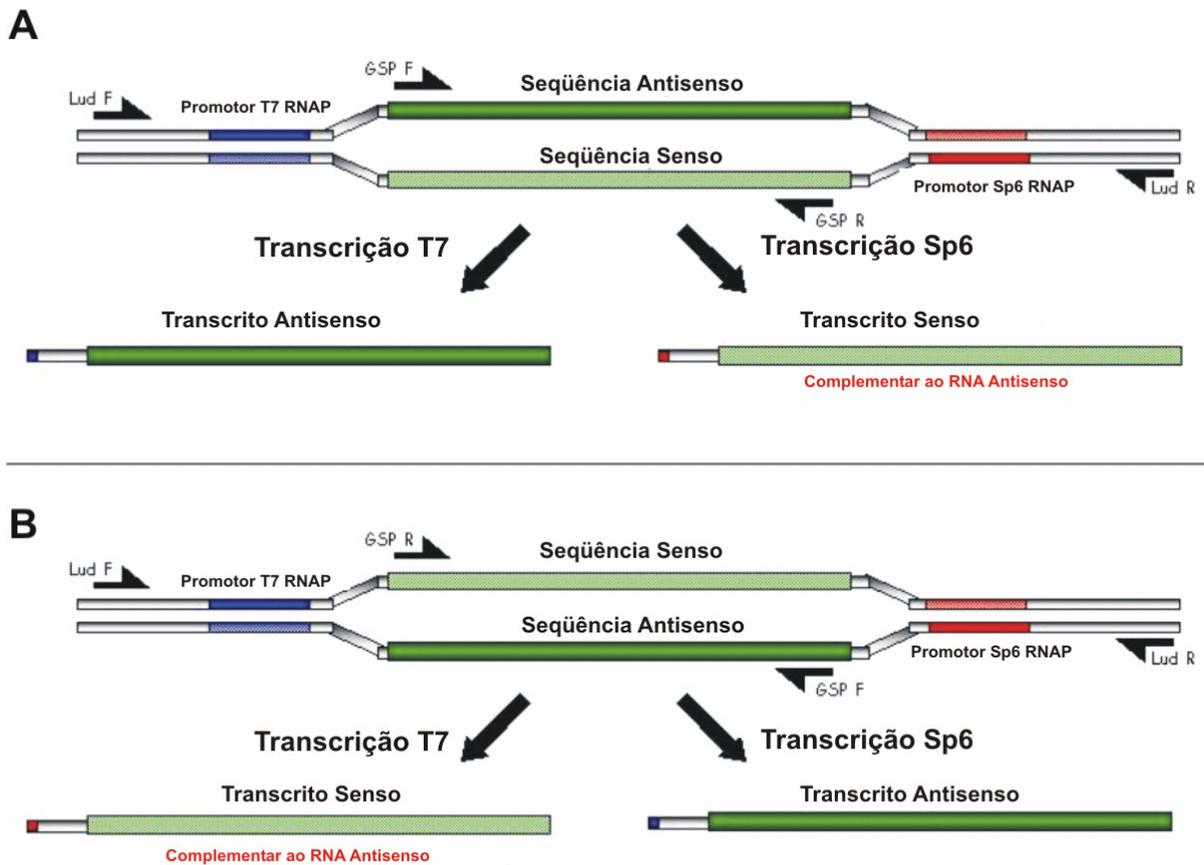


Figura 12 – **Transcrição *in vitro* para geração de sondas fita-específicas.** (A) Molde linearizado por PCR usando os oligonucleotídeos **GSP\_F** e **Lud\_R** é utilizado para gerar sondas complementares à um RNA com orientação antisense usando-se SP6 RNA polimerase. (B) Molde linearizado por PCR usando os oligonucleotídeos **Lud\_F** e **GSP\_F** é utilizado para gerar sondas complementares à um RNA com orientação antisense usando-se a T7 RNA polimerase. As denominações senso e antisense se referem ao mRNA codificante de proteína. Em azul consenso do promotor da T7 RNA polimerase, em vermelho da SP6, e em verde as regiões do transcrito alvo. As setas cortadas indicam os oligonucleotídeos de linearização.

### 3.7 Análise de expressão gênica por PCR

#### 3.7.1 Determinação da orientação de mensagens utilizando cDNA transcrito reversamente com oligonucleotídeos iniciadores específicos

Transcrições reversas com orientação específica (*Orientation-Specific Reverse Transcription*) seguidas de amplificação por PCR (OSRT-PCR) foram também utilizadas para determinação da natureza transcricional de mensagens intrônicas. Utilizando 100ng de mRNA ou 2-4 $\mu$ g de RNA total, 200U de transcriptase reversa *SuperscriptII* (Invitrogen) mais 0,9 $\mu$ M de oligonucleotídeos iniciadores complementares a fita de RNA antisenso ou a fita senso de cada região intrônica, as reações de transcrição reversa foram mantidas a 50°C por 60 minutos (Shendure & Church, 2002; Cawley *et al.*, 2004). Controles para a ausência de *RNA self-priming* e de contaminação com DNA genômico, foram realizados com reações sem adição de oligonucleotídeos iniciadores e de transcriptase reversa, respectivamente.

Após a transcrição reversa, quantidades de cDNA equivalentes a 10ng de RNA reversamente transcrito (utilizando uma estimativa de 2% de RNA reversamente transcrito na população de RNA total, quando necessário) foram utilizadas como molde para amplificações das mensagens intrônicas senso e antisenso, por 37-40 ciclos no *GenAmp 5700 Sequence Detector* (Applied Biosystems) utilizando o *SYBR<sup>®</sup> Green PCR Master Mix Kit* (Applied Biosystems), como meio de reação. As amplificações ocorreram com 0,8 $\mu$ M de oligonucleotídeos desenhados internamente aos usados durante a transcrição reversa, de modo a garantir uma maior especificidade nas reações. Finalmente, os produtos obtidos foram separados por eletroforese em gel com 3% de agarose. A figura 13 apresenta um diagrama esquemático do processo.

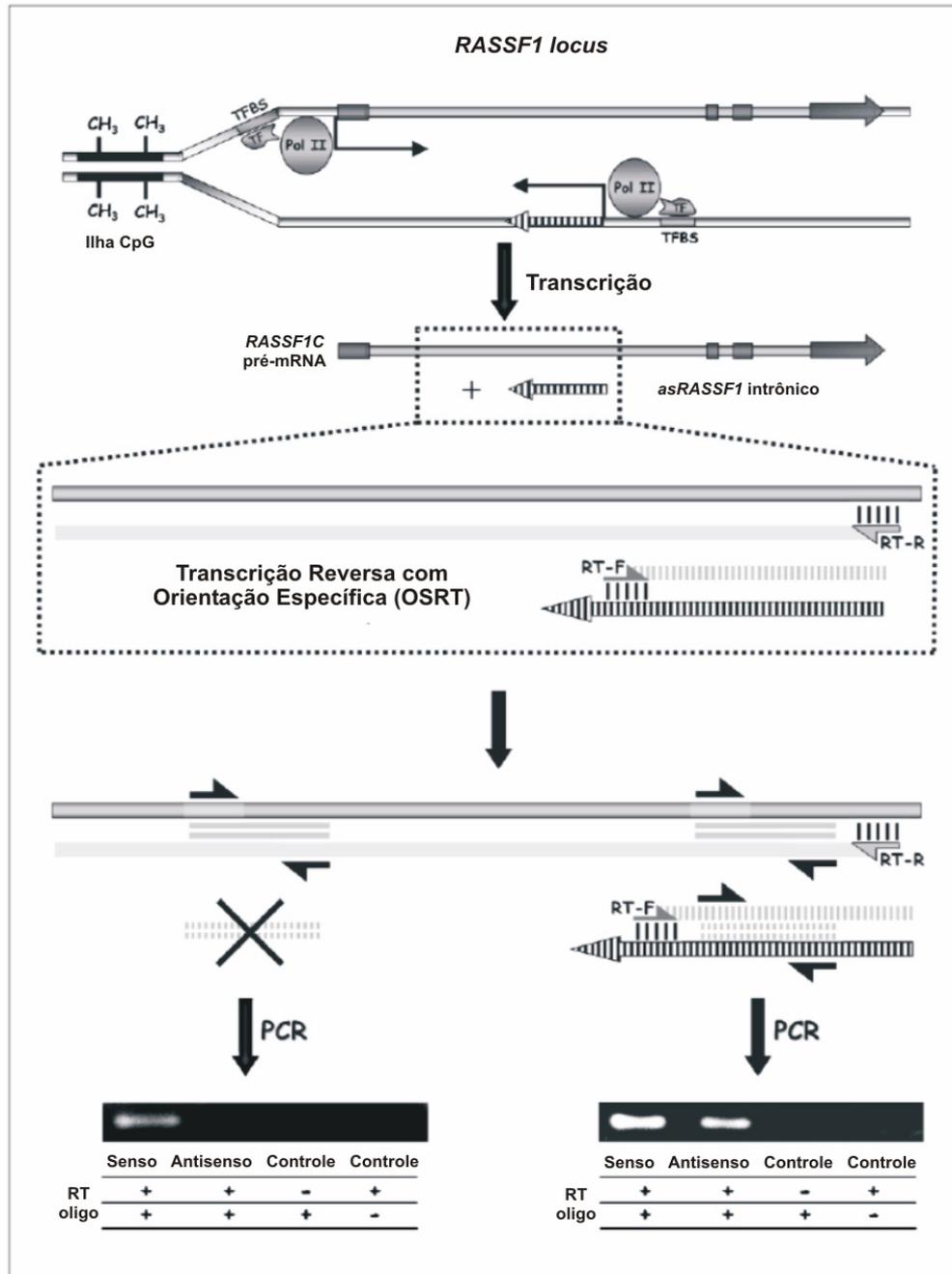


Figura 13 – **Transcrições reversas com orientação específica seguidas de amplificação por PCR.** A figura apresenta um esquema da metodologia utilizando o locus de *RASSF1* como exemplo (**superior**). Um cDNA com orientação definida é sintetizado por reações de transcrição reversa com oligonucleotídeos (RT-F e RT-R) complementares a RNAs antisense, ou senso (pré-mRNA no exemplo). Gel de 3% de agarose para visualização dos produtos amplificados por PCR (**inferior**) utilizando oligonucleotídeos internos para a seqüência de sobreposição das mensagens (**direita**) ou para a seqüência que só existe com orientação senso (**esquerda**), realizado aqui como controle extra. Controles para ausência de contaminação com DNA genômico (- RT) e de auto-anelamento do RNA (- oligo) durante a transcrição reversa são fundamentais para interpretação dos resultados. Figura adaptada de Reis, E. M. *et al.* (2005). *As Antisense Gets Intronic. Omics*, **9**(1):2-12.

### **3.7.2 Avaliação da abundância relativa de mensagens utilizando PCR quantitativo**

Algumas diferenças de expressão observadas nas análises por *microarrays* de cDNA foram confirmadas através de PCR quantitativo, utilizando a técnica de Real Time PCR com o aparelho no *GenAmp 5700 Sequence Detector* (Applied Biosystems) e o *SYBR<sup>®</sup> Green PCR Master Mix Kit* (Applied Biosystems). Precedente às validações, foram necessárias otimizações das condições para garantir eficiências comparáveis de amplificação. A concentração ideal dos pares de oligonucleotídeos iniciadores foi determinada realizando reações idênticas de amplificação, porém com quantidades crescentes desses, de 200nM-800nM, através da identificação da situação na qual não ocorre a formação de dímeros e onde a amplificação é mais robusta. Outro fator importante atestado foi se as eficiências de amplificação dos transcritos alvos e do transcrito referência, em função da concentração de RNA utilizado, são aproximadamente iguais.

### **Avaliação da abundância de mensagens utilizando PCR quantitativo em tempo real**

Amostras contendo 5µg de RNA total ou 500ng de mRNA, extraídas das condições a serem comparadas, foram reversamente transcritas em cDNA utilizando o protocolo padrão da *SuperScript<sup>™</sup> II First-Strand Synthesis System for RT-PCR* (Invitrogen), ou a metodologia para obter transcrições reversa com orientação específica (*Orientation-Specific Reverse Transcription*) descrita anteriormente (ver item 3.7.1). A quantificação foi baseada na observação da abundância dos produtos no ciclo de PCR onde as reações estão em fase exponencial de amplificação. O

acúmulo dos produtos de PCR nessa fase foi detectado em tempo real, através da ligação do *SYBR<sup>®</sup> Green*, que é uma molécula cuja emissão de fluorescência aumenta várias vezes quando ligada a DNA dupla-fita. Através da definição arbitrária da região exponencial de amplificação, foram determinados os *Cycle Threshold* ( $C_T$ ) de cada transcrito. O valor relativo da expressão foi dado pela diferença ( $\Delta C_T$ ) entre o  $C_T$  do gene alvo e o  $C_T$  do gene de referência, tubulina (*TUBA6*) (um gene estrutural que normalmente não apresenta variação de expressão entre as amostras estudadas). Os resultados finais foram expressos como *fold-change*, usando a fórmula  $2^{-\Delta C_T}$  (Pfaffl, 2001). A significância estatística das diferenças de expressão observadas (*p-values*) foi determinada usando *Student's t-test*, assumindo uma distribuição bicaudal e um modelo de variância heterocedástico.

### **3.8 Experimentos de Imunoprecipitação**

#### **3.8.1 Ensaio de imunoprecipitação da cromatina**

Ensaio de imunoprecipitação da cromatina (*Chromatin Immunoprecipitation - ChIP*) foram desenhados para detectar a ligação do complexo do receptor de andrógeno à regiões promotoras dos transcritos intrônicos. Nessas regiões foi predita por análise informática a existência de um ARE, elemento de resposta a andrógeno (ver item 3.1.4). As condições e os reagentes para realização dos experimentos de ChIP foram do *Chromatin Immunoprecipitation Assay Kit* (Upstate Biotechnology Inc.).

Brevemente, células LNCaP tratadas com andrógeno ou as células controle foram fixadas com 1 % de formaldeído por 10 minutos a temperatura ambiente para ligação covalente das proteínas ao DNA, e foram lisadas com tampão contendo SDS. Esses extratos celulares foram sonicados para quebra mecânica do DNA em

tamanhos de 200-1000bp usando o *Sonic Dismembrator Model 500* (Fischer Scientific). As imunoprecipitações foram realizadas com 3µg do anticorpo policlonal contra receptor de andrógeno (anti-AR) denominado *H-280* (Santa-Cruz) ou 1µg de IgG de camundongo (controle negativo - Upstate) usando 100µl de lisados contendo aproximadamente  $2 \times 10^6$  células. Após reversão das ligações entre DNA e proteínas, a cromatina imunoprecipitada foi tratada com RNase A e proteinase K, e os fragmentos de DNA foram purificados utilizando colunas de purificação de produtos de PCR *GFX* (Amershan Biosciences).

A detecção das regiões promotoras identificadas *in silico* foi realizada por reações de PCR contendo os oligonucleotídeos iniciadores para as regiões do ARE predito para o transcrito intrônico antisense do locus *MYO5A* (TTTGCTGGATAAGGATTCCCA, AGTTCCTTAACTTTAGCTTAGAAAAAGGA); para o AREIII do *PSA* (controle positivo) (GCTCAGCCTTTGTCTCTGATGA, TGCAAGATGATATCTCTCTCAGATCC); e para uma região arbitrária do transcrito de *TUBA6* (controle negativo) (GACTACAGGTGTGCGCCATCAT, TGCCGTGTTCCAGGCAGTAG).

### **3.8.2 Imunoprecipitação de proteínas ligadoras de RNA (RBPs)**

Volumes apropriados de cada lisado de 293T transfectado com pKALLU contendo as RBPs, ou contendo GFP (controle negativo), normalizado pela quantidade total de proteína (ver item 3.9.5) foram centrifugados a  $13.000 \times g$  por 10 minutos a 4°C e os sobrenadantes foram transferidos para tubos novos contendo 10 volumes de NT2 buffer (50mM Tris; pH 7,4; 150 mM NaCl; 1 mM MgCl<sub>2</sub>; 0,05% NP40) suplementado com 25mM EDTA, 2mM DTT, 100U/ml RNaseOUT, e 0,4 mM *Vanadyl Ribonucleoside Complex*, VRC. Para cada lisado foi adicionado cerca de

200 $\mu$ l de *Sepharose 6B beads* (Sigma), e a mistura foi incubada por 30 minutos a 4°C, com agitação constante. Essa primeira incubação visou minimizar a quantidade de proteínas ou RNAs que podem ser inespecificamente precipitados através de interação direta com os *Sepharose beads*. Ainda para minimizar interações inespecíficas, as alíquotas de *IgG Sepharose 6 fast flow beads* (GE Health Care) que foram usados para imunoprecipitar as RBPs através da interação da IgG com a proteína A, foram também incubados por 30 minutos com agitação constante a 4°C, com o mesmo tampão descrito acima, porém além dos suplementos já mencionados foi adicionado 5% BSA.

As amostras contendo os *Sepharose beads* com os extratos das células que expressaram as RBP, e os *IgG Sepharose 6 fast flow beads* em solução com BSA foram então centrifugadas a 1.000 $\times$ g por 5 minutos a 4°C. Os *IgG Sepharose 6 fast flow beads*, após retirada do sobrenadante por sucção, foram lavados 4 vezes com o NT2 suplementado. O sobrenadante com as RBPs foi transferido para os tubos contendo cerca de 200 $\mu$ l de *IgG Sepharose 6 fast flow beads* lavados. Essa nova mistura foi incubada por no mínimo 3 horas, a 4°C com agitação constante.

Finalizada essa última incubação, os *IgG beads* com as RBPs e seus RNAs alvo ligados foram centrifugados novamente a 1.000 $\times$ g por 5 minutos a 4°C, e foram lavados por 4 vezes com cerca de 5 volumes (referente aos *beads*) do tampão NT2. Após esse procedimento, os RNA co-imunoprecipitados com as RBPs puderam ser extraídos (ver item 3.3.3) para os experimentos de expressão gênica por *microarrays* de cDNA.

### 3.9 Sistemas de expressão epissomal

#### 3.9.1 Construção do sistema de expressão de Rasl11a/b em procaríoto

A estratégia escolhida para obtenção de proteínas das novas GTPases monoméricas da subfamília Rasl11 foi a expressão em bactérias através de estímulo de promotores ativados por IPTG (*isopropyl-beta-D-thiogalactopyranoside*). No vetor de expressão *pET21* (Novagen) foram clonadas as regiões codificantes dos genes de interesse *RASL11A* e *RASL11B*. A metodologia de clonagem foi baseada nas seqüências únicas de corte por enzimas de restrição presentes no sítio de clonagem do referido vetor, que não possuísem correspondência nas regiões codificantes de *RASL11A/B*. Assim, foram desenhados pares de oligonucleotídeos iniciadores complementares ao começo e final das fases abertas de leitura dos genes. Esses oligos foram modificados para apresentarem tanto as seqüências de reconhecimento pela enzima de restrição *NdeI* para a região 5' inicial, quanto de *XhoI*, mas com substituição de uma base para cancelamento do códon de parada de tradução da região 3' terminal. Essa última característica foi acrescentada aos insertos para que a região codificante ficasse em fase com uma repetição de nucleotídeos que codificam para histidina. A presença dessa porção final leva a expressão da proteína dos membros da subfamília Rasl11 em fusão com uma cauda de seis histidinas. A presença dessa cauda em proteínas recombinantes permite sua purificação através de cromatografia de troca iônica usando colunas de níquel, e sua detecção em *Western Blots* ou *Dot Blots* com anticorpos gerados contra um peptídeo contendo múltiplas histidinas.

Os insertos amplificados da forma descrita acima, foram clonados primeiramente em *pGEM-T Easy* (Promega), tanto para a confirmação da identidade das seqüências obtidas, por sequenciamento dideoxi (Megabace – Amershan

Pharmacia), quanto para facilitar o reconhecimento dos sítios de digestão pelas enzimas a serem utilizadas durante a transferência para o vetor de expressão *pET21*.

Confirmadas as identidades, os insertos foram transferidos para o vetor de expressão *pET21*. Quantidades referentes a 200µg de produtos de mini-preparações de DNA plasmidial (*Plasmid Mini Kit* Qiagen) de *pET21* vazio (sem inserto) e 100µg de *pGEM-T-Ras11a/b* foram utilizados em reações simultâneas de digestão com as enzimas *NdeI* e *XhoI* (New England Biolabs) seguindo as recomendações do fabricante quanto ao número de unidades utilizadas, tampão de reação, temperatura e tempo de incubação. A seguir, 50µg de vetor digerido foram adicionados a uma quantidade dos insertos (em mol) relativa a três vezes a quantidade do vetor digerido, para uma reação de ligação com *T4 DNA Ligase* (Promega) por 12-16 horas a 6°C. As ligações após dialisadas para a diminuição da concentração de sal, foram utilizadas para transformar por eletroporação células de *E. coli* da linhagem DH10B, destinadas à propagação das construções *pET21-Ras11a/b*. Nessa etapa os insertos foram mais uma vez seqüenciados para assegurar a expressão de proteínas com total ausência de mutações.

As colônias transformantes foram crescidas em meio LB líquido com ampicilina para novas preparações de DNA plasmidial (*Plasmid Mini Kit* - Qiagen), a serem utilizadas nos experimentos de indução da expressão em células de *E. coli* da linhagem ORIGAMI. A presença de um gene para resistência a ampicilina no vetor *pET21*, assim como no *pGEM-T*, e o acréscimo de ampicilina no meio foram utilizados para selecionar as bactérias transformantes em todas as etapas.

### 3.9.2 Indução da expressão de Ras11b em bactéria

A construção pET21-Ras11b, e em paralelo o vetor pET21 vazio (sem inserto), foi utilizada para transformação de células de *E. coli* da linhagem para expressão ORIGAMI. A cada experimento de indução de expressão uma fração da preparação de DNA plasmidial das construções (descrito no item 3.9.1) foi utilizada para nova transformação, pois após um breve período de armazenamento em glicerol, os clones de bactérias ORIGAMI transformadas perdem a capacidade de expressão heteróloga.

Uma colônia de bactérias transformantes com pET21-Ras11b foi pré-inoculada em 5ml de LB contendo ampicilina por aproximadamente 1 hora a 37°C, até atingirem densidades ópticas em 600nm ( $DO_{600}$ ) de 0,5-1,0. Estando as bactérias crescendo em fase exponencial, esses 5ml foram transferidos para inóculos contendo 200ml de LB com ampicilina, para crescimento por cerca de 3 horas a 37°C, até novamente atingirem  $DO_{600}$  de 0,5-1,0. Em seguida a expressão das proteínas foi induzida pelo IPTG adicionado ao meio em concentração final de 1mM, e incubando as células a 37°C por até 6 horas. Através de centrifugação por 15 minutos a 5.000×g (4°C) as bactérias foram isoladas do meio de cultura para congelamento a -80°C até serem lisadas para verificação da expressão. As  $DO_{600}$  finais foram medidas para normalização da quantidade de tampão de lise utilizado, visando garantir quantidades de proteína total por ml de solução relativamente semelhantes.

Os aglomerados de bactérias induzidas foram ressuspensos em um tampão de lise contendo 20mM de Tris-Cl pH 8,0 e 1mM PMSF (*Phenylmethylsulfonyl fluoride*, um potente inibidor de proteases) no volume apropriado, e lisados em uma *French Press* através de três passagens a pressão de 2.000 psi. Uma fração

referente ao lisado total foi armazenada e o restante foi centrifugado a 20.000×g por 30 minutos (4°C) para separação das frações solúvel (sobrenadante) e insolúvel (precipitado). Para visualização das proteínas expressas, quantidades normalizadas pelas DO<sub>600</sub> finais, foram separadas por SDS-PAGE com 12% de poliacrilamida, ao lado de um padrão de peso molecular de proteínas *Prestained Protein Marker* (Fermentas), corado posteriormente por *Coomassie Blue*.

### **3.9.3 Detecção de Rasl11b expresso em bactéria com anticorpo anti-histidinas.**

Para certificação da natureza da proteína expressa, foi utilizada a cauda de seis histidinas fusionada à região carboxi-terminal de Rasl11b para realização de um *Dot Blot*. Para isso, após as etapas descritas acima (ver item 3.1.6), uma diluição seriada do lisado total, das frações solúvel e insolúvel, referentes à construção pET21-Rasl11b, pET21 vazio (controle negativo), e 6xHis-Ca<sup>+2</sup>-ATPase (controle positivo), foram aplicadas em uma membrana de nitrocelulose utilizando o *Filtration Manifold System* (Gibco).

As membranas foram bloqueadas em 10ml de Blotto (5% de leite em pó desnatado; 0,03% de azida sódica) durante 3-6 horas. Em seguida, foram incubadas com 10µl de anticorpo primário anti-histidinas (*Anti-His Antibody* - Amershan Biosciences) em nova solução de Blotto por 16 horas, sendo em seguida lavada com PBS Tween 0,1%, uma vez por 15 minutos e duas vezes por 5 minutos. Usando 10ml de Blotto sem azida, a membrana foi então incubada com 10µl de anticorpo secundário (*Anti-mouse IgG, Horseradish Peroxidase* - Amershan Biosciences), por 1 hora, sendo as lavagens descritas repetidas em seguida. A revelação foi realizada

por quimioluminescência (*ECL Western Blotting Detection Reagents* - Amershan Biosciences), seguindo o protocolo do fabricante.

#### **3.9.4 Sistema de expressão de proteínas ligadoras de RNA em células mamífero**

Para investigar se algumas das proteínas ligadoras de RNA (*RNA Binding Proteins* – RBPs) envolvidas em mecanismos de regulação pós-transcricional ligam-se aos longos RNAs intrônicos, foi utilizada uma abordagem de imunoprecipitação, similar a já utilizada em estudos anteriores (Penalva & Keene, 2004; Penalva *et al.*, 2004), para posterior análise *microarrays* de cDNA.

As RBPs de interesse foram expressas em um sistema composto pelo vetor pKALLU, baseado no vetor comercial *pcDNA3* (Invitrogen), previamente construído (Penalva *et al.*, 2004), e pela linhagem celular tumoral de tecido renal 293T, obtida da ATCC (*American Type Tissue Collection*). A escolha dessa linhagem foi baseada em sua velocidade de crescimento e em sua eficiência de transfecção, cerca de 90%. O vetor utilizado apresenta como principais características uma região codificante para proteína A, proveniente do *pEZZ* (Amershan Biosciences) para ser expressa em fusão com a RBP de interesse; um promotor de CMV para expressão constitutiva; e genes de resistência a antibióticos.

O estudo foi realizado com cinco construções em pKALLU: um controle negativo com uma proteína reconhecidamente não ligadora de RNA, GFP (*Green Fluorescent Protein*); e quatro RBPs, PABP (*Poly(A)-Binding Protein*), hTRA2 (*Human Transformer 2*), EIF4E (*Eukaryotic Translation Initiation Factor 4E*), e HuD (*HU-Antigen D*).

### 3.9.5 Detecção das RBPs expressas em 293T com anticorpo contra proteína A.

Quarenta e oito horas após a transfecção com pKALLU contendo as RBPs descritas acima ou GFP (ver item 3.2.3), as células foram coletadas e tratadas de modo a ser possível a extração de RNAs íntegros associados às proteínas ligadoras de RNA em estudo. As células foram lavadas duas vezes com PBS gelado, coletadas usando um *scraper* e centrifugadas a 2.000×g por 5 minutos a 4°C. O precipitado celular foi ressuscitado e homogeneizado em 2 volumes de tampão de lise (100mM KCl, 25mM EDTA; 5mM MgCl<sub>2</sub>; 10mM HEPES, pH 7,0; 0,5% Nonidet® P-40 (NP40); 10% glicerol; 2mM DTT; 0,4 mM *Vanadyl Ribonucleoside Complex*, VRC, da New England Biolabs; 1 tablete/50 ml de *Complete Proteinase Inhibitor* da Roche; e 100U/ml *RNaseOUT* da Invitrogen). Os lisados celulares foram centrifugados a 13.000×g por 10 minutos a 4°C e o sobrenadante foi transferido para um tubo novo. O precipitado celular recebeu novamente o mesmo tratamento descrito, mas dessa vez usando apenas 1 volume de tampão de lise. Ambos sobrenadantes foram centrifugados e armazenados a -80°C em alíquotas de 1ml para a realização das imunoprecipitações (ver item 3.8.2), e uma alíquota menor com cerca de 100µl para quantificação protéica através da metodologia de Bradford (Harlow & Lane, 1998) e para a verificação da expressão das RBP específicas.

A verificação da expressão específica das RBPs foi realizada por *Western Blot*. Um volume de lisado celular referente à cerca de 50µg de proteína total de cada proteína expressa foi separado por SDS-PAGE, 10% de acrilamida. As proteínas foram transferidas para uma membrana de nitrocelulose, e essa membrana foi bloqueada com 5% de leite desnatado em PBS com 0,1% Tween® 20 por cerca de 16 horas. Essa membrana foi incubada com uma diluição em PBS de

1:1000 de anti-proteína A por 2 horas, seguida de uma incubação com o anticorpo secundário anti-IgG de coelho (1:1000) conjugado com peroxidase (*horseadish*) por 1 hora. A detecção foi feita por quimioluminescência (*ECL Western Blotting Detection Reagents* - Amershan Biosciences), seguindo o protocolo do fabricante.

### **3.9.6 Sistema de expressão de transcritos intrônicos em células de mamífero**

Os longos RNAs intrônicos selecionados para posterior caracterização funcional foram clonados no sistema de expressão em células de mamífero *pBudCE4.1* (Invitrogen). Esse vetor foi escolhido para o início desse estudo devido à existência de duas unidades de expressão distintas, uma sob o controle do promotor CMV, e outra do promotor EF-1  $\alpha$ .

A estratégia constou em clonar em uma das unidades o gene que codifica para a proteína fluorescente verde (GFP), que facilitará a realização de ensaios transientes através de seleção das células transfectantes pelo *sorter* do citômetro de fluxo; e na outra os transcritos intrônicos a serem estudados. Devido a baixa eficiência de transfecção da linhagem celular de próstata tumoral, LNCaP, escolhida como modelo para esses estudos funcionais, a co-expressão de GFP é muito útil para rápida seleção das células. Entretanto, como recurso adicional poderá também ser realizada seleção clonal das células onde ocorrer integração do vetor ao seu genoma, utilizando o antibiótico *Zeocin* (Invitrogen), letal tanto para células eucarióticas quanto procarióticas.

Através de procedimentos de PCR, sítios para enzimas de restrição de corte único, *HindIII* e *BamHI*, foram adicionados às regiões 5' e 3', respectivamente, de um fragmento de cDNA contendo toda a porção codificante de proteína de GFP. Da

mesma maneira, sítios de restrição para as enzimas *NotI* e *XhoI* foram adicionados às pontas 5' e 3', respectivamente, de transcritos intrônicos correlacionados com o grau de diferenciação tumoral em próstata cuja seqüência completa foi obtida por RACE (Reis *et al.*, 2004). Uma clonagem intermediária dos produtos obtidos em *pGEM-T* (Promega) foi realizada, tanto para confirmação da natureza e integridade das mensagens obtidas através de sequenciamento dideoxi, quanto para facilitar o reconhecimento das enzimas de restrição ao sítio criado.

Os clones confirmados foram crescidos em LB com ampicilina para extração de DNA plasmidial com o *Plasmid Mini Kit* (Qiagen). Através de reações de digestão seqüencial (GFP), e de dupla digestão (transcritos intrônicos), dos plasmídeos extraídos, seguidas de procedimentos de purificação com o *Gel Extraction Kit* (Qiagen), foram obtidos os insertos finais para ligação ao vetor de expressão. Em primeiro lugar foi gerada a construção pBud-GFP, através da inserção do fragmento de GFP modificado ao vetor digerido com *HindIII* e *BamHI*, na região localizada bem em frente ao promotor CMV. Além de ser o vetor base para as clonagens dos transcritos a serem estudados, o pBud-GFP será utilizado como um dos controles negativos para determinação se os efeitos desencadeados nas células são específicos para as mensagens intrônicas. Seguindo procedimentos similares foram finalmente geradas as construções, utilizando o pBud-GFP como base, com os transcritos intrônicos sob o controle do promotor EF-1  $\alpha$ .

Todos os sistemas construídos foram amplificados por propagação em bactérias, em quantidades suficientes para uma preparação média de DNA plasmidial com o *Plasmid Midi Kit* (Qiagen). Os produtos finais a serem utilizados para modulação da expressão de mensagens intrônicas estão representados esquematicamente na figura 14, que também resume a estratégia de clonagem.

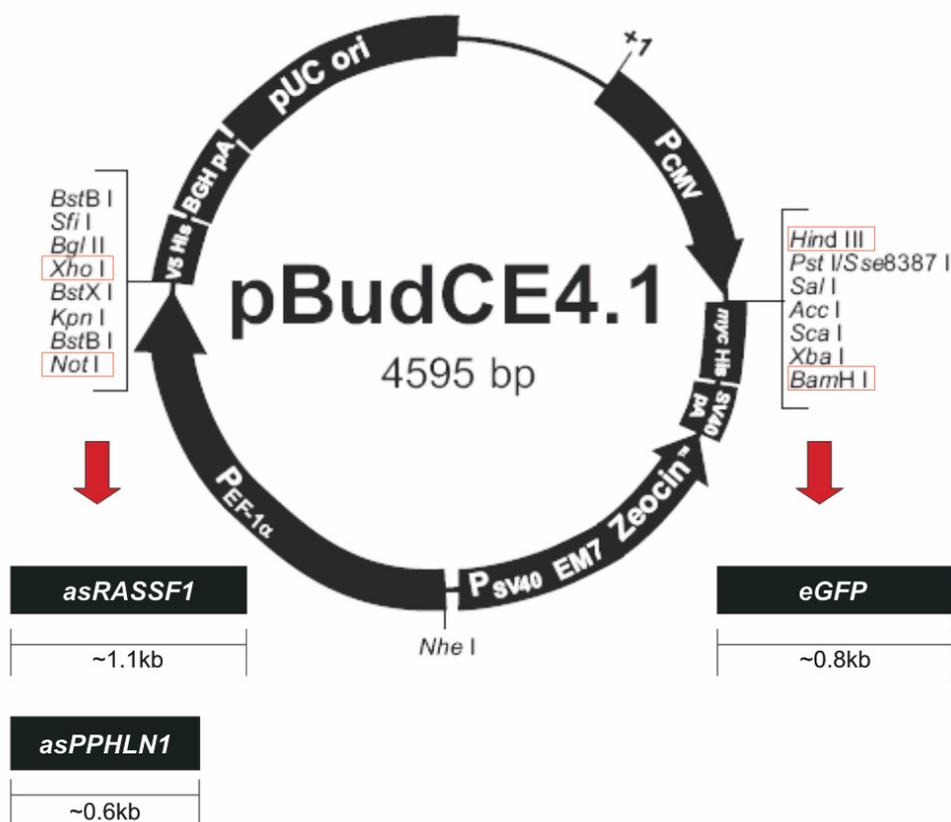


Figura 14 – Sistema de expressão de Longos RNAs Intrônicos não-codificantes antisense em células de mamífero. Esquema do sistema construído para superexpressão em células de mamífero dos ncRNAs intrônicos *asRASSF1* e *asPPHLN1* utilizando o vetor comercial pBudCE4.1 (Invitrogen) como base. Foram geradas as construções pBud-GFP, pBud-GFP-*asRASSF1* e pBud-GFP-*asPPHLN1* a partir de uma estratégia de clonagem utilizando enzimas de restrição (destacadas na caixas vermelhas). Os promotores CMV e EF-1 a, e o gene de resistência ao antibiótico Zeocin (Invitrogen) estão indicados.

### 3.10 Experimentos para obtenção de anticorpos

#### 3.10.1 Preparo de antígenos de Ras11b para imunização de coelhos.

Os lisados totais de bactérias transformadas com pET21 Ras11b, obtidos da mesma maneira relatada anteriormente, foram fracionados por uma centrifugação em menor velocidade, 10.000×g, e por um tempo inferior, 15 minutos, visando minimizar o efeito de “co-purificação” de proteínas bacterianas. Em seguida os corpos de inclusão foram lavados 2 vezes com solução contendo uréia (50mM Tris-Cl pH 8,0 com 1M Uréia) e finalmente ressuspensos em tampão para corpos de inclusão (50mM Tris-Cl pH 8,0 com 100mM NaCl e 1mM EDTA) para quantificação

protéica seguindo a metodologia de Bradford, a partir de uma curva padrão de albumina sérica (Harlow & Lane, 1998). Quantidades suficientes de proteína para as imunizações foram ainda centrifugadas (20.000×g por 30 minutos) para a obtenção da preparação de proteína-adjuvante. Essa preparação corresponde a uma mistura 1:1 da solução do tampão para corpos de inclusão com adjuvante de Freund completo (Difco), na dose inicial, e incompleto no reforço.

### **3.10.2 Imunização de coelhos com corpos de inclusão de Rasl11b.**

Foram utilizadas duas fêmeas jovens da linhagem *New Zealand*, cuja saúde foi atestada após um período de observação (15 dias) anterior ao início do experimento. Cada fêmea recebeu doses subcutâneas em vários pontos do dorso (4-8 pontos) da preparação de proteína-adjuvante (descrito anteriormente) totalizando cerca de 100µg de proteína por espécime. Quatro semanas após, foi dado um reforço contendo as mesmas quantidades de proteína. Foram coletadas amostras de sangue, até 20ml, proveniente da artéria central da orelha do animal para a obtenção de: soro pré-imune (antes da primeira inoculação), o soro imune da primeira dose (logo antes do reforço) e o soro imune final, para serem testados.

A obtenção dos soros foi conseguida através de coagulação do sangue coletado a 37°C por 1 hora com um palito de madeira no centro do tubo. Após completa coagulação ao redor do palito, o tubo foi armazenado a 4°C por 12 horas para contração do coágulo. O líquido ao redor do coágulo foi reservado, e o coágulo foi centrifugado a 1.500×g a 4°C, por 10 minutos, para obtenção do soro residual. Todo soro extraído foi então centrifugado a 10.000×g a 4°C, por 10 minutos para retirada das células e material insolúvel. Finalmente, os soros foram aliquotados, identificados e armazenados a -80°C até seu uso.

### 3.11 Análise de expressão gênica por *microarrays* de oligonucleotídeos

#### 3.11.1 Plataforma de *microarrays* de oligonucleotídeos

Alguns dos estudos de expressão gênica descritos aqui fizeram parte do projeto temático para “Identificação de marcadores moleculares para diagnóstico e prognóstico em câncer utilizando *microarrays* de DNA”, financiado pela FAPESP (Processo nº 02/13283-6). Com parte desse financiamento foi desenhado pelo grupo um *microarray* de oligonucleotídeos, cujas características se encontram publicadas sob o número de acesso GPL4051, no banco de dados *Gene Expression Omnibus* (GEO – <http://www.ncbi.nlm.nih.gov/geo/>).

Brevemente, sondas de oligonucleotídeos foram desenhadas para 7.715 RNAs totalmente intrônicos e para 5.289 mensagens parcialmente intrônicas (com uma região de sobreposição com exons de *RefSeq*). Esses transcritos foram escolhidos randomicamente a partir de listas de *contigs* de ESTs sem *splicing* geradas pelo projeto de mapeamento em larga escala que estudou o conteúdo transcricional intrônico humano (Nakaya *et al.*, 2006).

Para cada transcrito intrônico foram obtidas sondas de 60 bases que satisfizessem uma série de critérios (Hughes *et al.*, 2001). As sondas deveriam ter menos de oito bases derivadas de regiões repetitivas do genoma, e menos de sete bases de segmentos com baixa complexidade (homopoliméricos); as sondas deveriam ter um conteúdo de GC de 35–55%; e uma temperatura de desnaturação calculada ( $T_m = \text{melting temperature}$ ) de 68–76°C. Para minimizar a possibilidade de hibridizações cruzadas, cada sonda de 60 bases foi submetida a uma busca com BLASTn (NCBI) contra um banco de dados específico contendo todas as regiões genômicas humanas onde existe evidência de transcrição por mRNAs ou ESTs. O

melhor *hit* de cada sonda contra este banco é exato, ou seja, tem por definição 100% de identidade ao longo das 60 bases, o que corresponde a um BLAST *bit-score* de 119. Um *bit-score* de 42.1 corresponde a um alinhamento máximo de 21 bases sem *gap*. As sondas cujo segundo melhor *hit* contra esse banco de dados tivessem um *bit-score* mais baixo que 42.1 foram consideradas com baixa probabilidade de reação cruzada, ou seja, seriam sondas específicas, e prosseguiram para o próximo passo de seleção.

Os oligonucleotídeos considerados específicos foram mapeados em seus respectivos transcritos e a sonda mais próxima a porção 3' de cada alvo foi selecionada, sempre assumindo uma orientação antisenso para mensagem intrônica em relação ao transcrito exônico codificante de proteína correspondente. Para cada alvo, uma segunda sonda com orientação oposta foi obtida utilizando o reverso complementar da primeira seqüência de 60 bases selecionada, sendo que para os transcritos parcialmente intrônicos essa segunda sonda corresponde ao exon do *RefSeq* onde a mensagem intrônica se sobrepõe.

Para possibilitar a detecção em paralelo dos níveis transcricionais das mensagens codificantes de proteína de mesmo *loci* dos RNAs intrônicos, foram incluídos 14.393 elementos correspondentes a sondas desenhadas pela própria *Agilent* para a plataforma comercial *Whole Human Genome Oligo Microarray*. Foram incluídas também 2.256 sondas para controles positivos e negativos presentes nessa plataforma comercial. As lâminas desta plataforma *Intronic 37k microarray*, desenhada pelo grupo como descrito acima, foram construídas por encomenda pela *Agilent Technologies*.

### 3.11.2 Marcação de alvos fluorescentes por amplificação de RNA

As amostras de RNA provenientes de duas réplicas biológicas de células LNCaP tratadas com  $\alpha$ -amanitina e de células controle foram marcadas através de amplificação do RNA seguindo as recomendações do *Agilent Low RNA Input Fluorescent Linear Amplification Kit* (Agilent Technologies). Brevemente, cRNAs marcados com Cy5 e Cy3 foram obtidos usando 500ng de RNA total como molde para amplificação de RNAs poli-(A)+ pela T7 RNA polimerase. Cada amostra foi marcada um vez com cada fluoróforo, sendo incluídos os mRNAs sintetizados *in vitro* *Agilent RNA Spike-In Kit* (Agilent Technologies) para serem utilizados como controle.

A marcação utilizando cRNA amplificado por transcrição *in vitro* com a T7 RNA polimerase apresenta uma grande vantagem em relação a abordagem de marcação do cDNA com transcrição reversa. A orientação das sondas geradas é sempre preservada, enquanto estruturas secundárias nos mRNA podem gerar artefatos durante a transcrição reversa que causam interpretações incorretas da orientação das moléculas.

### 3.11.3 Hibridização dos experimentos com lâminas de *microarrays* de oligonucleotídeos

Hibridizações com 750ng cada de cRNA marcado com Cy3 e Cy5 foram realizadas seguindo as recomendações do *Agilent In situ Hybridization Kit-plus* (Agilent Technologies). Foram utilizadas 4 lâminas no total, sendo que em cada par foi hibridizada uma das réplicas biológicas com cada uma das amostras marcadas com cada fluoróforo (*dye-swap*).

As lâminas foram lavadas e processadas de acordo com o protocolo *Agilent 60-mer Oligo Microarray* e as imagens foram obtidas por excitação com laser utilizando o *GenePix 4000B* (Molecular Devices). Os seguintes parâmetros foram utilizados como padrão para esse procedimento: resolução de 10 $\mu$ m por pixel; comprimento de onda de excitação de 532nm (para alvos marcados com Cy3) ou 633nm (para alvos marcados com Cy5); e tensão PMT de 350 volts (para alvos marcados com Cy3) e 600 volts (para alvos marcados com Cy5).

#### **3.11.4 Obtenção de imagens e valores de intensidade comparáveis**

Seguindo a rotina de tratamento de dados estabelecida de nosso laboratório durante as análises com as lâminas de cDNA, os valores de intensidade foram extraídos das imagens obtidas com as lâminas de oligonucleotídeos utilizando o *ArrayVision 8.0* (Imaging Research Inc.), e os valores de intensidade de réplicas marcadas com Cy5 e Cy3 foram corrigidos usando algoritmo *LOWESS* (*locally weighted linear regression algorithm*), disponível no pacote R (<http://www.r-project.org/>). O processo de normalização para deixar os diferentes experimentos comparáveis foi utilizar como fator de normalização a média das 300 sondas presentes no *microarray* complementares aos mRNAs controle que foram adicionados durante a marcação, removidos os 20% dos valores de intensidade extremos, mais altos e mais baixos.

## **4 RESULTADOS**

### **4.1 Identificação e caracterização de membros representativos de novas famílias de pequenas GTPases monoméricas.**

Uma das estratégias utilizadas para contribuir na exploração do conhecimento potencial gerado durante o HCGP (*Human Cancer Genome Project*) baseou-se em uma busca informática por mensagens codificantes para novos parálogos, isto é, produtos novos porém similares a proteínas já conhecidas, e que além disso tivessem envolvimento demonstrado em processos de tumorigênese (detalhes no item 3.1.1). Com essa abordagem um conjunto de candidatos foi selecionado, e desses, dois novos transcritos humanos similares a mensagens codificantes para proteínas membros da superfamília de pequenas GTPases foram identificados e parcialmente caracterizados. Os resultados descritos a seguir mostram os passos desse estudo que apontou para evidências mais concretas do envolvimento de um dos transcritos em câncer, e ainda estimulou o início de uma caracterização mais aprofundada do seu produto protéico.

#### **4.1.1 *RASL11A* e *ARL-9*, dois novos genes codificantes para pequenas GTPases monoméricas.**

Por comparação do potencial codificante do conjunto de cerca de três milhões de ESTs geradas a partir de RNA proveniente de tumores humanos presentes no GenBank com a lista de 1.127 proteínas humanas com função já relacionada com câncer, foram encontradas 16 ESTs candidatas a parálogos de genes humanos. O conjunto de informações referentes aos transcritos codificantes com relacionados com o câncer, assim como das ESTs candidatas a parálogos estão na Tabela 1.

Tabela 1 – Novos parálogos: dados de alinhamento das seqüências parálogas preditas e de seu gene relacionado com câncer (CRG) correspondente.

CRG	CRG Locus	CRG mapeamento	CRG tamanho	CRG estrutura	CRG domínios	Parálogo tamanho	Parálogo estrutura	Parálogo mapeamento	Parálogo domínios	% Similaridade Local (amino ácidos)	% Identidade Local (amino ácidos)	% Cobertura CRG (aminoácidos)	% Cobertura Parálogos (aminoácidos)	Gaps	Evidência experimental
Talin 1	<a href="#">TLN1</a>	9p13	2.541 aa	57 exons	<a href="#">smart00746</a> <a href="#">smart00707</a> <a href="#">smart00307</a>	2.777 aa	59 exons	15q22.2	<a href="#">smart00286</a> <a href="#">smart00307</a> <a href="#">pfam01808</a>	85	76	95	100	20	EST, mRNA
Growth arrest-specific 2	<a href="#">GAS2</a>	11p14.3-15.2	313 aa	6-7 exons	<a href="#">LOAD_gas2arao</a> <a href="#">pfam02187</a> <a href="#">smart00743</a> <a href="#">smart00333</a>	182 aa	8 exons	12q23.1	<a href="#">pfam02187</a> <a href="#">smart00243</a> <a href="#">LOAD_gas2arao</a> <a href="#">smart00333</a>	71	62	50	86	2	EST
Calcineurin-modulin-dependent protein kinase 1	<a href="#">CAMK1</a>	3p25.2	370 aa	11 exons	<a href="#">smart00720</a> <a href="#">pfam00069</a> <a href="#">smart00719</a> <a href="#">LUPU_kin</a> <a href="#">smart00391</a> <a href="#">pfam01429</a>	366 aa	12 exons	4q31.23	<a href="#">smart00720</a> <a href="#">pfam00069</a> <a href="#">smart00719</a> <a href="#">smart00219</a>	62	47	74	80	7	EST
Methyl-CpG binding domain protein 2	<a href="#">MBD2</a>	18q21	411 aa	3 exons	<a href="#">smart00220</a> <a href="#">pfam00069</a> <a href="#">smart00719</a> <a href="#">pfam00639</a>	194 aa	1 exon	19p13.2	<a href="#">smart00195</a>	59	42	44	94	2	EST, mRNA
Cytokine-inducible kinase	<a href="#">CNK</a>	1p34.1	607 aa	15 exons	<a href="#">smart00173</a> <a href="#">smart00176</a> <a href="#">smart00174</a> <a href="#">smart00176</a>	338 aa	14 exons	19p13.3	<a href="#">smart00220</a> <a href="#">pfam00069</a>	56	42	61	99	13	EST, mRNA (FLJ30246)
Dual specificity phosphatase 3 (vaccinia virus phosphatase \4H1-related)	<a href="#">DUSP3</a>	17q21	185 aa	3 exons	<a href="#">smart00195</a>	1024 aa	3 exons	1q24.1	<a href="#">smart00195</a>	55	41	97	19	3	EST
Ras-related protein 2A	<a href="#">RAP2A</a>	13q34	183 aa	2 exons	<a href="#">smart00173</a> <a href="#">smart00176</a> <a href="#">smart00174</a> <a href="#">smart00176</a>	242 aa	4 exons	13q12.2	<a href="#">smart00173</a> <a href="#">smart00176</a> <a href="#">smart00174</a> <a href="#">pfam01926</a>	55	37	89	70	3	EST
Cytochrome P450, subfamily 11B (phenobarbital-inducible), polypeptide 6	<a href="#">CYP2B6</a>	19q13.2	481 aa	9 exons	<a href="#">pfam00067</a>	388 aa	5 exons	11p15.2	<a href="#">pfam00067</a>	54	38	76	97	6	EST
GTP binding protein	<a href="#">ARL3</a>	10q23.3	182 aa	5 exons	<a href="#">pfam00029</a> <a href="#">smart00177</a> <a href="#">smart00044</a>	123 aa	2 exons	4q12	<a href="#">pfam00029</a> <a href="#">smart00177</a> <a href="#">smart00172</a>	51	33	60	93	2	EST
Phosphatase and tensin homolog	<a href="#">PTEN</a>	10q23.3	403 aa	9 exons	<a href="#">pfam01102</a> <a href="#">smart00186</a> <a href="#">smart00194</a>	254 aa	6 exons	7p13	<a href="#">smart00404</a> <a href="#">pfam01102</a>	51	38	52	83	6	EST
Butyrylcholinesterase	<a href="#">BCHE</a>	3q26.1-q26.2	602 aa	3 exons	<a href="#">pfam01135</a>	666 aa	5 exons	17p13.1	<a href="#">pfam01135</a>	50	37	74	69	12	EST, mRNA (KIAA1386, partial cds)
Cyclin-dependent kinase 2	<a href="#">CDK2</a>	12q13	298 aa	7 exons	<a href="#">smart00720</a> <a href="#">pfam00069</a> <a href="#">smart00719</a>	341 aa	2 exons	1p36.11	<a href="#">smart00720</a> <a href="#">pfam00069</a> <a href="#">smart00719</a> <a href="#">smart00339</a>	50	35	96	97	17	EST, mRNA
Forkhead protein FREAC-1	<a href="#">FOXF1</a>	16q24	354 aa	2 exons	<a href="#">smart00339</a> <a href="#">pfam00651</a> <a href="#">smart00612</a> <a href="#">smart00612</a>	570 aa	6 exons	2p11.1	<a href="#">pfam00260</a> <a href="#">pfam00691</a> <a href="#">smart00612</a> <a href="#">LOAD_kelch</a> <a href="#">pfam01344</a>	47	38	92	65	19	EST
PIG10	<a href="#">ENC1</a>	5q12-q13.3	516 aa	1 exon	<a href="#">smart00720</a> <a href="#">smart00023</a> <a href="#">smart00005</a>	649 aa	9 exons	2q37.3	<a href="#">smart00612</a> <a href="#">LOAD_kelch</a> <a href="#">pfam01344</a>	47	31	93	83	17	EST
DAP-kinase 1	<a href="#">DAPK1</a>	9q34.1	1401 aa	24 exons	<a href="#">smart00720</a> <a href="#">smart00005</a>	171 aa	5 exons	2q11.1	<a href="#">pfam00023</a>	45	31	12	84	7	EST
XPG-complementing protein	<a href="#">ERCC5</a>	13q22	1.186 aa	15 exons	<a href="#">pfam00762</a> <a href="#">pfam00887</a>	687 aa	7 exons	2p24.22	<a href="#">smart00485</a> <a href="#">pfam00887</a>	37	24	50	94	24	EST

Análise de similaridade/identidade Local: BestFit (pacote GCG). Matrix de comparação: Blossum 35, Parâmetros: -GAPweight=10 -LENGTHweight=1 -ENDWeight

Dentre as seqüências identificadas pela abordagem, duas aparentemente codificavam para proteínas similares a membros da superfamília Ras de pequenas GTPases monoméricas, cuja representatividade na lista inicial de genes relacionados com câncer era de 36 membros. Devido a grande participação de muitos membros dessa família em processos relacionados à tumorigênese (discutido no item 1.5.1), o estudo se concentrou na clonagem e caracterização parcial desses dois novos transcritos humanos.

O primeiro gene candidato parece codificar para um produto que possui 55% de similaridade com a proteína Rap2a, e foi posteriormente denominado *RASL11A* (*RAS-like, family 11, member A*). O segundo, denominado *ARL-9* (*ADP-ribosylation factor-like 9*) parece codificar para um produto que possui 51% de similaridade com a proteína Arl-3. Os transcritos de *RASL11A* (Figura 15A) e *ARL-9* (Figura 15B) amplificados a partir de cDNA de próstata, mapeiam separados, nas regiões cromossômicas 13q12.2 e 4q12, respectivamente, e cada um possui quatro exons distintos. Como representado na Figura 15A, o transcrito completo de *RASL11A*, obtido por RACE-PCR, possui 1.543bp com uma fase aberta de leitura (*Open Reading Frame* – ORF) que codifica para uma proteína com 242 a.a., enquanto o fragmento de *ARL-9*, obtido por PCR com oligonucleotídeos específicos, possui 836bp com uma ORF que gera uma proteína com 187 a.a. (Figura 15B).

#### **4.1.2 *RASL11A* e *ARL-9* constituem novas subfamílias de pequenas GTPases.**

Utilizando as seqüências protéicas hipotéticas codificadas por essas novas GTPases monoméricas foi possível identificar no GenBank (descrito no item 3.1.2) outras proteínas similares a essas e alinhá-las para: i) identificar com quais

proteínas, e/ou subfamílias de pequenas GTPases, os novos genes seriam mais relacionados; ii) localizar os resíduos conservados presentes nos domínios característicos das proteínas da superfamília; e iii) inferir sobre novos motivos conservados nas proteínas mais relacionadas aos novos genes.

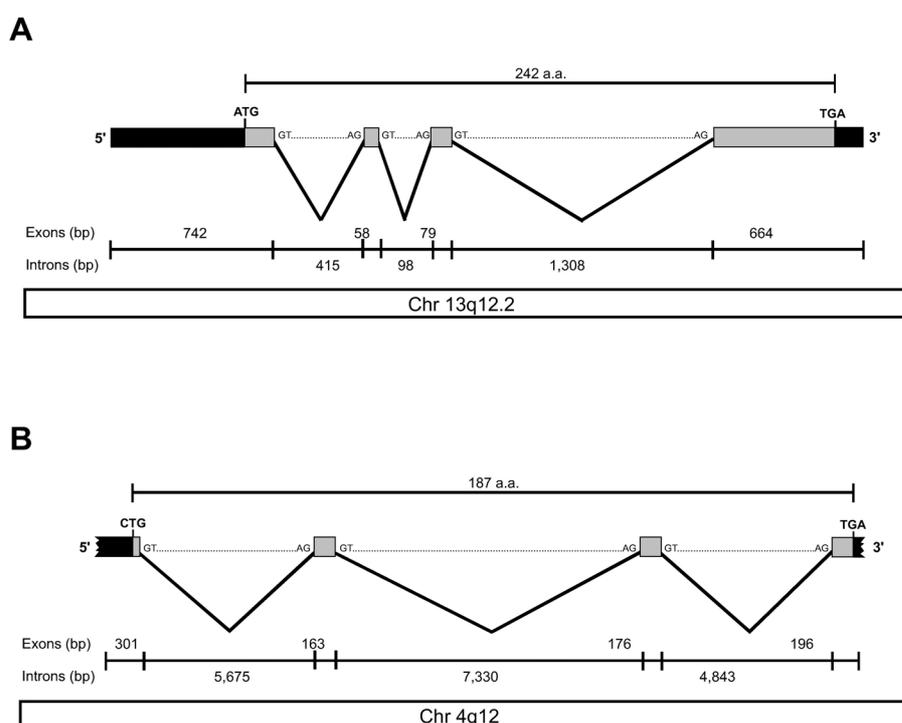


Figura 15 – *RASL11A* e *ARL-9* mapeiam nas regiões cromossômicas 13q12.2 e 4q12. Exons de *RASL11A* (A) e *ARL-9* (B) são mostrados como caixas preenchidas. Regiões codificantes para proteína são preenchidas em cinza e as não traduzidas (UTR) em preto. O tamanho das proteínas preditas está no topo de cada painel. Tamanhos dos exons e introns estão em pares de base (bp).

Com uma abordagem por comparação das seqüências primárias, foi determinado que entre as 10 proteínas mais similares a Rasl11a constam claramente membros representativos das famílias Ras/Rho/Rab (Figura 16A), enquanto entre as 10 mais similares a Arl-9 aparecem proteínas pertencentes às famílias Arf/Arl (Figura 16B). Com essa análise ainda foi possível identificar outros possíveis parálogos, também não descritos, para cada uma das novas mensagens. Com 74% de similaridade de seqüência primária com o produto protéico putativo de

*RASL11A*, foi identificado *RASL11B* (NP\_076429), anteriormente apenas anotado como codificante para uma proteína hipotética humana (Figura 16A). De forma semelhante, foi identificado *ARL-10* (NP\_775935), cujo produto protéico possui 62% de similaridade com Arl-9 (Figura 16B).

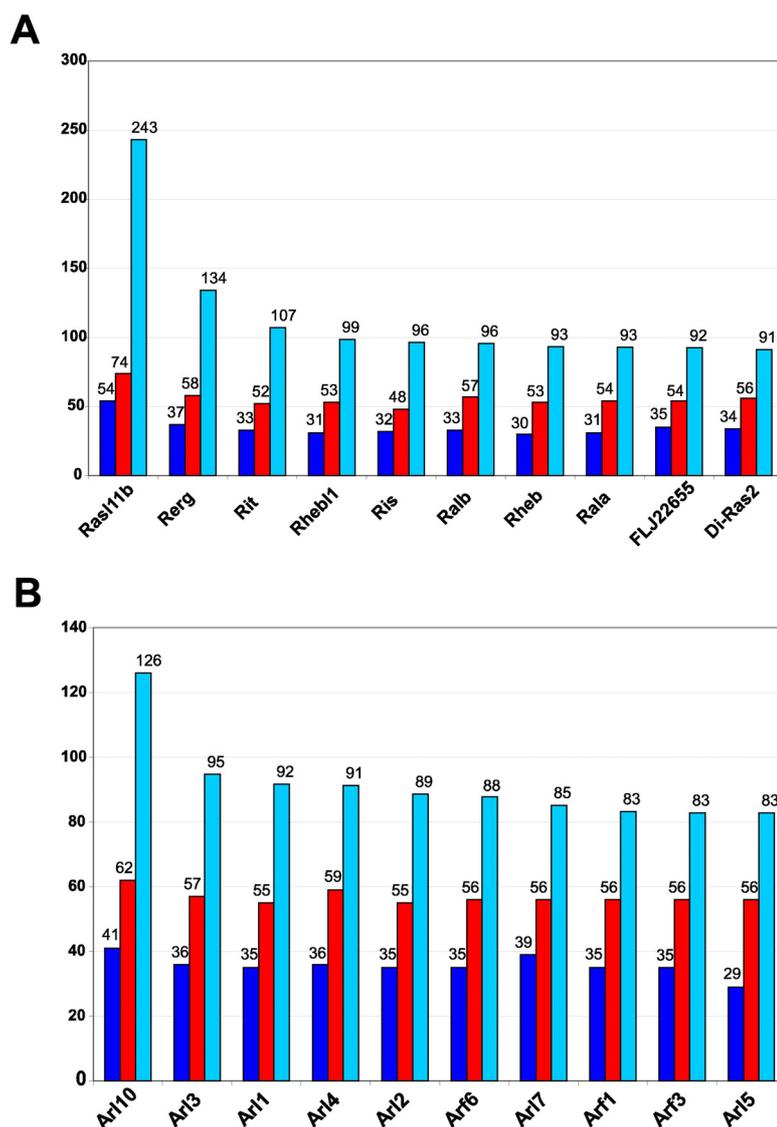


Figura 16 – Produtos protéicos hipotéticos de *RASL11A* e *ARL-9* possuem parálogos e são similares a membros de subfamílias distintas de pequenas GTPases. As seqüências de *Rasl11a* (A) e *Arl-9* (B) foram alinhadas com seqüências do banco de dados de proteína do NCBI, usando BLASTp. Os 10 melhores *hits* estão indicados no eixo X de cada painel. Todas as proteínas mais similares a *Rasl11a* pertencem a subfamílias Ras/Rho/Rab, incluindo seu parálogo *Rasl11b*, enquanto as de *Arl-9* e *Arl-10* pertencem a subfamílias Arf/Arl. Os parâmetros de alinhamento obtidos pelo BLASTp são *bit-score* (azul claro), porcentagem de identidade (azul escuro) e porcentagem de substituições conservativas ou similaridade (vermelho).

Análises filogenéticas foram também realizadas utilizando os produtos dos novos transcritos em conjunto com diversos membros já caracterizados da superfamília, reforçando e ampliando os resultados observados durante a análise de similaridade descrita acima. Rasl11a e Arl9 apresentam-se como representantes de duas novas subfamílias gênicas ainda não descritas, que codificam para pequenas GTPases monoméricas relacionadas às famílias Ras/Rho/Rab (Figura 17A) e Arf/Arl/Sar (Figura 17B), respectivamente. Rasl11a situa-se em um ramo separado, juntamente com Rasl11b (Figura 17A) e similarmente, Arl-9 está próxima a Arl-10 (Figura 17B).

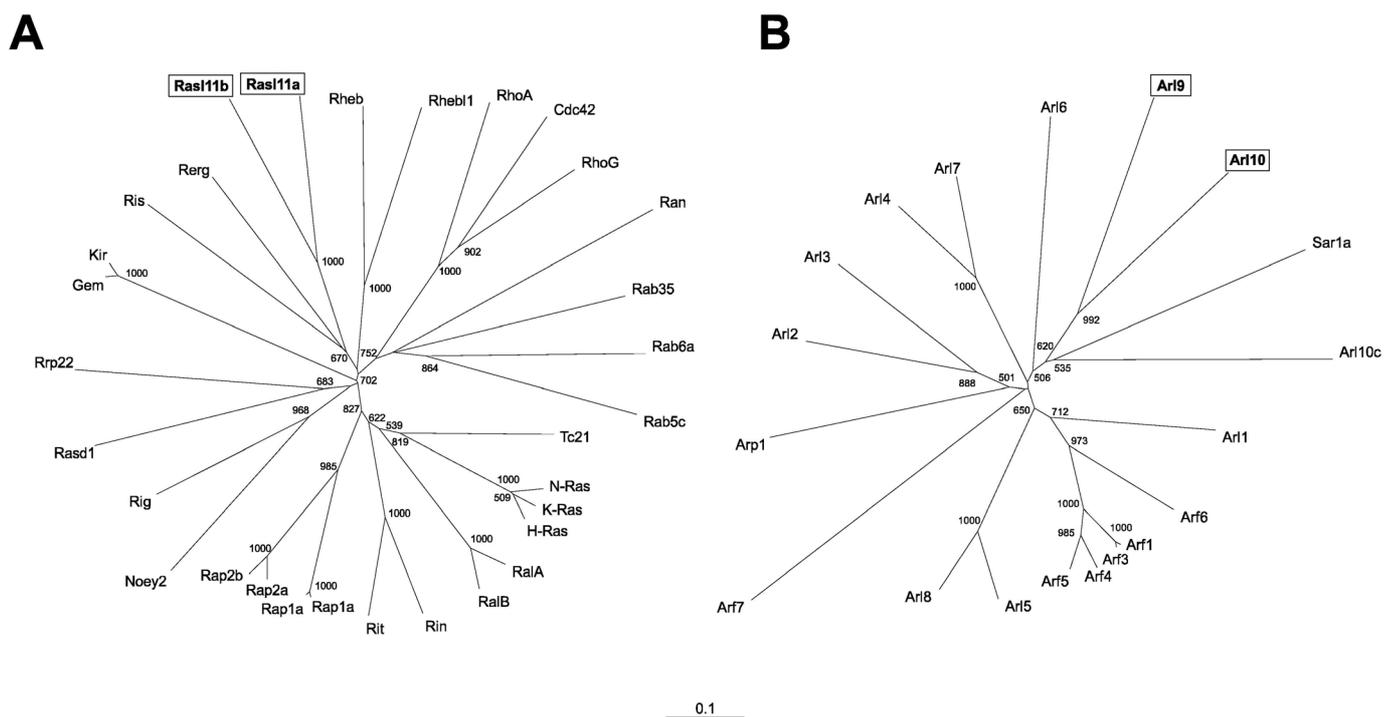


Figura 17 – **Rasl11a e Arl9 são membros de novas subfamílias de pequenas GTPases que pertencem a ramos distintos da superfamília.** Árvores filogenéticas para as proteínas das subfamílias Ras/Rho/Rab, incluindo as novas proteínas Rasl11a/b identificadas neste trabalho (**A**) e para Arf/Arl/Sar, incluindo as novas proteínas Arl-9/10 (**B**). Os números representam a confiança nos ramos (em 1.000 amostragens). Os valores abaixo de 500 não estão indicados.

As proteínas da subfamília Rasl11, filogeneticamente mais próximas da família Ras, compartilham a maioria dos resíduos conservados do domínio G das pequenas GTPases (Figura 18). Entretanto essas proteínas exibem uma substituição por Asn na posição correspondente ao resíduo Thr35 (relativo a H-Ras), situada no sub-domínio denominado *Switch 1* do domínio G (Figura 18) e conservado em todas as GTPases já caracterizadas funcionalmente (Vetter & Wittinghofer, 2001a), que está esquematizado na figura 7. Este achado sugere que o mecanismo de ativação possa ser estruturalmente diferente nessa nova subfamília de pequenas GTPases, assim como discutido posteriormente (ver item 5.1). Além disso, essas proteínas provavelmente não apresentam as modificações lipídicas em aminoácidos características da família, pois não possuem o motivo CaaX (Reuther & Der, 2000a) na região carboxi-terminal. Porém, apresentam agrupamentos de resíduos básicos que devem garantir sua ligação à membrana plasmática (Lee *et al.*, 1996), normalmente essencial para suas funções biológicas (Figura 18, resíduos em verde).

Arl-9 e Arl-10 apresentam além de resíduos do domínio G a assinatura canônica característica do *Interswitch Toggle* (wDvGGqxxxRxxW) da família Arf de pequenas GTPases (Pasqualato *et al.*, 2002b), com apenas algumas alterações conservativas (Figura 18) já verificadas em outras proteínas funcionais (Kuge *et al.*, 1994; Eckley & Schroer, 2003). Importante ressaltar que alguns desses resíduos estão apenas conservados em Arl-9 se usarmos como códon de iniciação de tradução CUG. O uso de desse tipo de iniciação é documentado em proto-oncogenes, fatores de transcrição e fatores de proliferação, e normalmente aparece em um contexto genômico de seqüência (Boeck & Kolakofsky, 1994) similar ao observado para o *ARL-9*.



#### **4.1.3 Produtos protéicos codificados por *RASL11A* e *ARL9* são muito conservados entre os eucariotos.**

Foram encontradas também diversas proteínas hipotéticas, entre os demais eucariotos, similares as proteínas da subfamília Rasl11 (Figura 19A e B) e Arl-9/Arl-10 (Figura 20A e B), mostrando uma grande conservação evolutiva dessas novas proteínas. Cada membro da subfamília humana Rasl11 (Figura 19A), assim como da subfamília Arl-9/Arl-10 (Figura 20A), situa-se em ramos separados, sendo que em cada ramo encontram-se ortólogos muito relacionados.

Ortólogos próximos para Rasl11a foram encontrados apenas em *Mus musculus* e em *Rattus norvegicus*, entretanto para Rasl11b foram encontrados também em *Danio rerio* e *Xenopus leavis* (Figura 19). Ainda foram encontrados ortólogos em *Drosophila sp* e *Anopheles gambiae*, ligeiramente menos relacionados à subfamília Rasl11 (Figura 19). Importante notar pelo múltiplo alinhamento que as proteínas Rasl11 são muito conservadas dentro dos metazoários e que todos os ortólogos encontrados apresentam a substituição do resíduo canônico Thr do *switch* 1, por um resíduo de Asn (Figura 19B).

Resultados bastante similares foram obtidos para Arl-9/Arl-10. Enquanto foram encontrados ortólogos próximos para Arl-9 apenas em *Mus musculus* e em *Rattus norvegicus*, para o Arl-10 foi encontrado também em *Xenopus leavis* (Figura 20A). O alinhamento múltiplo dessas proteínas mostra elevada conservação de aminoácidos, inclusive nas alterações observadas no consenso do *Interswitch Toggle* (Figura 20B).

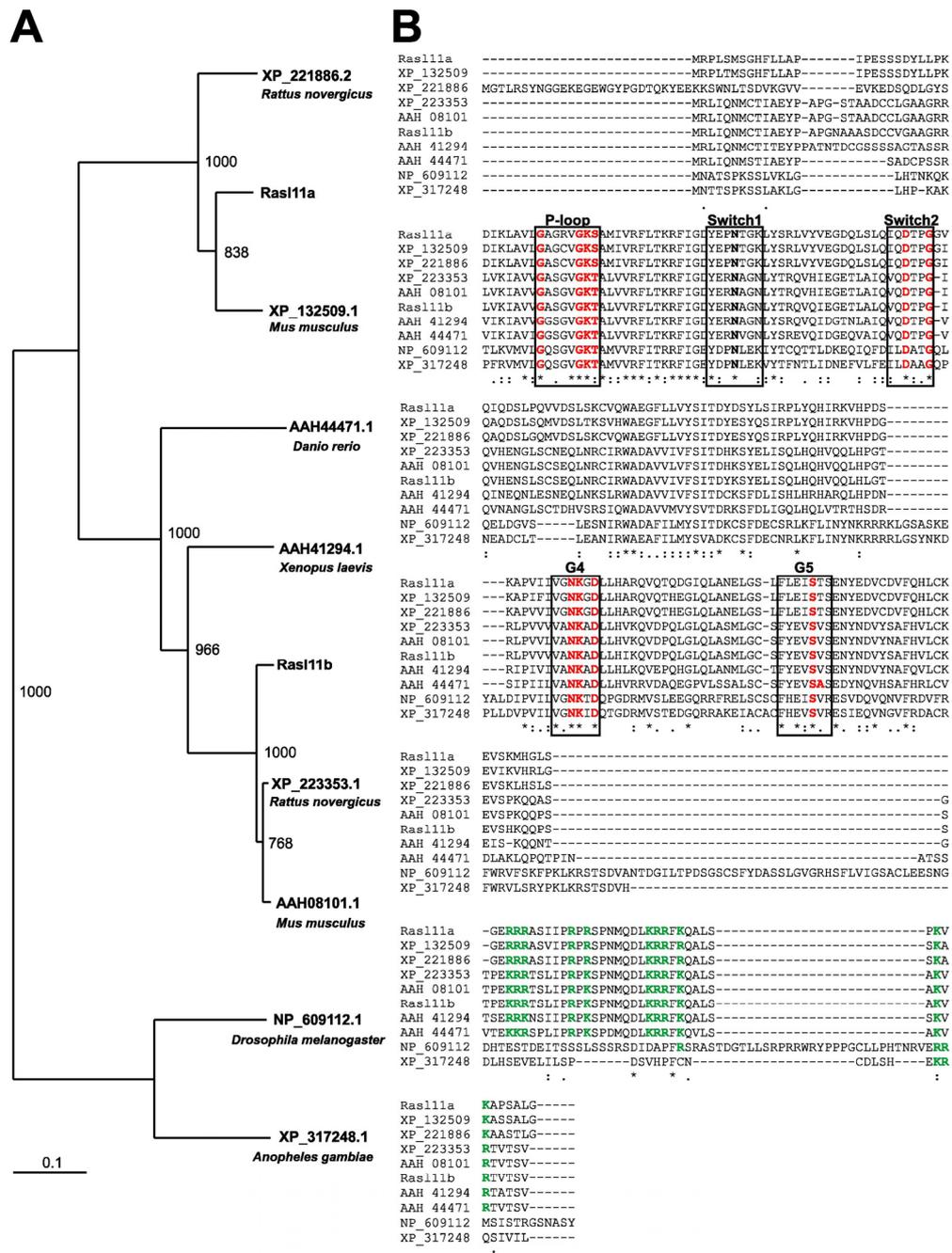


Figura 19 – As proteínas Ras11 são muito conservadas entre os metazoários. (A) Árvores filogenéticas da subfamília Ras11 incluindo os ortólogos encontrados no GenBank nessa análise. Os números representam a confiança nos ramos (em 1.000 amostragens). (B) Alinhamentos dos ortólogos de Ras11 mostrando as cinco regiões conservadas do domínio G entre as caixas e os resíduos invariantes da superfamília estão em vermelho. O resíduo Asn que está na posição da Thr no switch 1 está em negrito. Os resíduos básicos carboxi-terminais estão em verde.

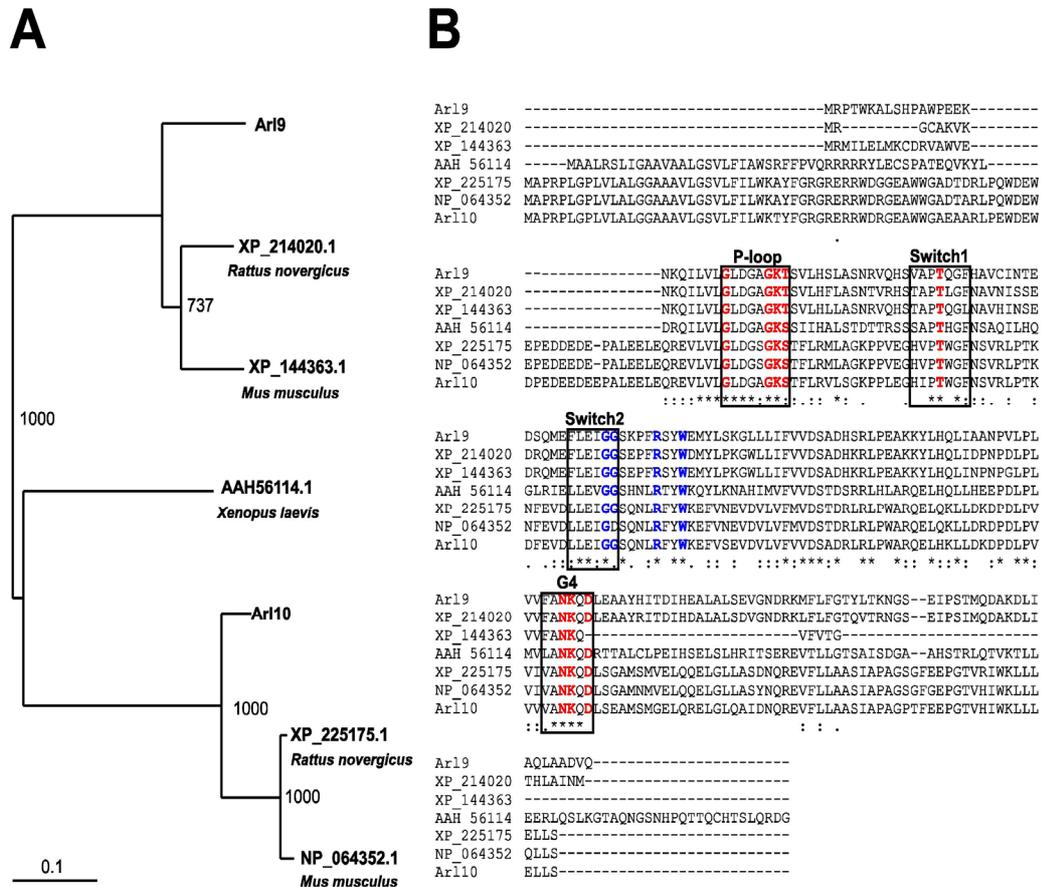


Figura 20 – As proteínas da subfamília Arl-9/Arl-10 são muito conservadas entre os metazoários. (A) Árvore filogenética da subfamília Arl-9/10 incluindo os ortólogos encontrados no GenBank nessa análise. Os números representam a confiança nos ramos (em 1000 amostragens). (B) Alinhamento dos ortólogos de Arl-9/10 mostrando as regiões conservadas do domínio G estão entre as caixas e os resíduos invariantes da superfamília em vermelho. Resíduos do *interswitch toogle* da família Arf estão em azul.

#### 4.1.4 Níveis do transcrito *RASL11A* estão diminuídos em tumores de próstata. A proteína funcionalmente já estudada mais semelhante a Rasl11A,

Rerg com 58% de similaridade e 37% de identidade (Figura 16A), foi recentemente descrita como possuindo atividade supressora tumoral em carcinomas de mama (Finlin *et al.*, 2001). Essa observação, o interesse particular do grupo em estudar

esse processo tumoral, e também a possibilidade de obter amostras de mRNA derivado de tecido prostático normal e neoplásico, motivaram a verificação do padrão de expressão de *RASL11A*. Utilizando PCR quantitativo em tempo real (ver item 3.7.2), verificamos que a abundância do transcrito *RASL11A* está significativamente ( $p=0,00065$ ) diminuída em amostras de adenocarcinoma de próstata (dos pacientes 4, 9, 525, 110) em relação ao tecido prostático normal adjacente a tumores (dos pacientes 6, 105, 106, 86, 88) (Figura 21A), sugerindo um possível papel de supressor tumoral também para este novo transcrito.

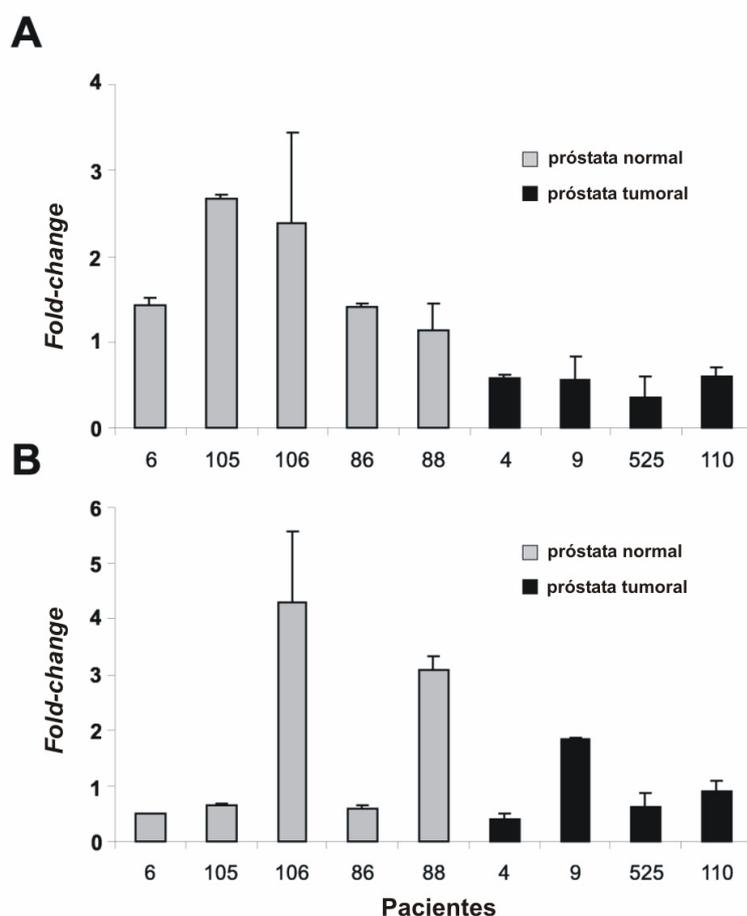


Figura 21 – ***RASL11A* tem expressão diminuída em tumores de próstata.** RT-PCR quantitativo em tempo real a partir de mRNA isolado de tumores (barras pretas) ou tecido normal adjacente ao tumor (barras cinzas) de próstata. Os níveis dos transcritos de *RASL11A* (A) e *ARL-9* (B) foram normalizados pelo mRNA de *beta-tubulina*. Valores de p para a expressão diferencial foram determinados para *RASL11A* ( $p=0,00065$ ) e *ARL-9* ( $p=0,52$ ).

A expressão em diferentes tecidos e o tamanho desse transcrito foram também investigados através da hibridização de sondas radioativas com membranas comerciais de nitrocelulose contendo mRNA de diversos tecidos (ver item 3.6.1). Através de Northern Blot foi observada a presença constitutiva de *RASL11A* em vários tecidos humanos. Duas formas alternativas para o transcrito de *RASL11A* (1.6 e 1.2 kb) foram claramente detectadas em bexiga, próstata, testículo e colón (Figura 22). A expressão de *RASL11A* parece ser bastante fraca em ovário e ausente em timo e leucócitos (Figura 22).

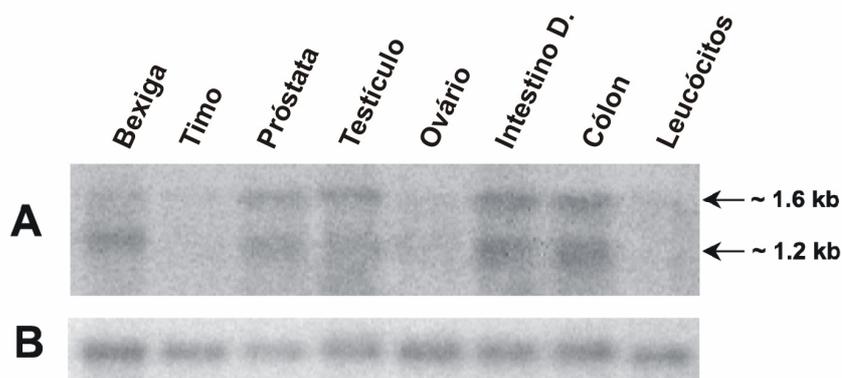


Figura 22 – **Duas isoformas de *RASL11A* são transcritas nos tecido humanos.** (A) Transcrição de *RASL11A* em diversos tecidos detectada por Northern blot. (B) A presença de quantidades comparáveis de mRNA em cada poço foi confirmada pela subsequente hibridização da membrana com sonda de *beta-actina*.

#### 4.1.5 Sistema de expressão de proteínas Rasl11 em procariotos.

Para possibilitar o início de uma caracterização mais aprofundada dessa nova subfamília de pequenas GTPases identificada nesse trabalho (Louro *et al.*, 2004)<sup>4</sup> com possível envolvimento em câncer de próstata foi construído um sistema de expressão heteróloga das proteínas Rasl11 em bactérias (descrito no item 3.9.1).

<sup>4</sup> Louro *et al.* 2004 (Artigo publicado). ANEXO 4 - incluso no CD-ROM suplementar a esta Tese.

Dessa forma, bactérias transformadas com o vetor pET21 vazio (sem inserto) e com a construção pET21-Ras11b foram induzidas a expressar as proteínas de interesse com adição de IPTG (descrito no item 3.9.2). Esses experimentos se iniciaram com a proteína Ras11b devido a problemas de contaminação durante a construção do pET21-Ras11a. Quando essa construção foi seqüenciada, identificou-se na verdade um outro clone da construção com Ras11b.

Na figura 23 pode-se visualizar um SDS-PAGE mostrando o acúmulo de proteína de cerca de 27kDa no lisado total das bactérias contendo pET21-Ras11b, comparado com o vetor vazio, durante os tempos de indução por IPTG. Observou-se que até 6 horas de indução a quantidade de proteína expressa no tamanho compatível com o esperado para as novas GTPases monoméricas da sub-família Ras11 (~27kDa) ainda está aumentando, proporcionalmente ao tempo de indução utilizado.

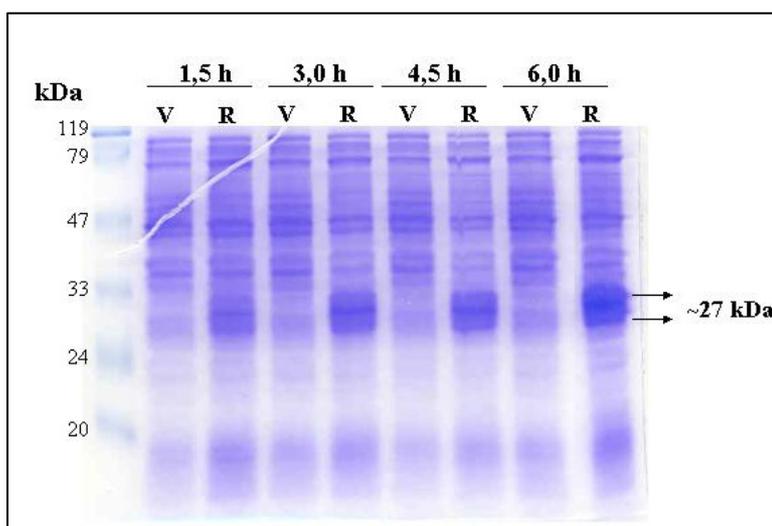


Figura 23 – **Indução da expressão de Ras11b por IPTG em bactéria.** SDS-PAGE com 12% de acrilamida mostrando a expressão heteróloga de Ras11b induzida por IPTG em diferentes tempos: 1,5h; 3h; 4,5h. Marcador de massa molecular de proteínas *Prestained Protein Marker* (Fermentas). V = vetor pET21b sem inserto (controle negativo) e R = pET21b com inserto da região codificante de Ras11b.

Devido à existência da cauda de seis histidinas fusionada às proteínas em estudo, foi possível assegurar a natureza do produto expresso através de *Dot Blot* com anticorpos anti-histidina (descrito no item 3.9.3). A figura 24 apresenta a confirmação da expressão de uma proteína com cauda de histidinas presente apenas nas bactérias transformadas com pET21-RasI11b.

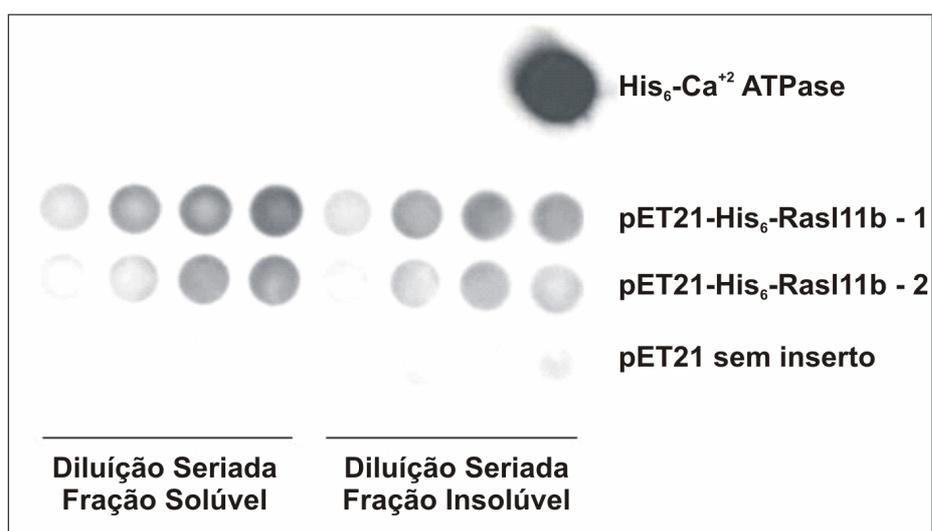


Figura 24– **Confirmação da expressão da proteína de fusão His<sub>6</sub>-RasI11b por Dot Blot.** Diluições seriadas de cada fração do extrato de bactérias aplicadas a uma membrana de nitrocelulose e incubadas com anticorpos contra poli-histidinas. His<sub>6</sub>-Ca<sup>+2</sup>ATPase adicionada como controle positivo e uma construção de pET21 sem inserto como controle negativo. Dois clones de His<sub>6</sub>-RasI11b estão mostrados.

#### **4.1.6 Expressão de RasI11b recombinante em corpos de inclusão para imunização de coelhos.**

Foi desenhada uma estratégia de imunização para geração de anticorpos policlonais a partir de uma preparação de corpos de inclusão obtidos da expressão heteróloga de RasI11b. A utilização dessa metodologia foi escolhida devido à

observação inicial, através de SDS-PAGE, que as condições testadas para indução da expressão de Ras11b levavam a um acúmulo de proteínas na fração insolúvel (Figura 25 – coluna 3). Provavelmente esses produtos não estão corretamente dobrados e foram armazenadas em corpos de inclusão devido a uma indução de expressão muito robusta e acelerada.

Esse resultado foi extremamente indesejado para o prosseguimento da caracterização funcional das GTPases monoméricas através de ensaios bioquímicos, pois além de purificação por cromatografia de troca iônica em coluna de níquel necessitaria de uma etapa adicional de renaturação estrutural das proteínas. Todavia, a proporção relativa da proteína induzida nesses corpos de inclusão é bastante elevada, sendo relatada na literatura como de no mínimo 70% (Harlow & Lane, 1998). Assim, essa característica foi utilizada para a obtenção dos antígenos utilizados em uma estratégia de imunização. É possível notar que quantidades baixas de proteínas contaminantes de bactéria que foram “co-purificadas” a partir de interações hidrofóbicas com os corpos de inclusão (Figura 25 - coluna 3), foram praticamente eliminadas após a lavagem do precipitado dos corpos de inclusão (ver item 3.10.1) (Figura 25 - coluna 6).

Esses antígenos foram então utilizados para imunização de coelhos, assim como descrito (item 3.10.2), entretanto durante esse doutorado não chegou a ser realizado um teste sistemático de reatividade dos soros.

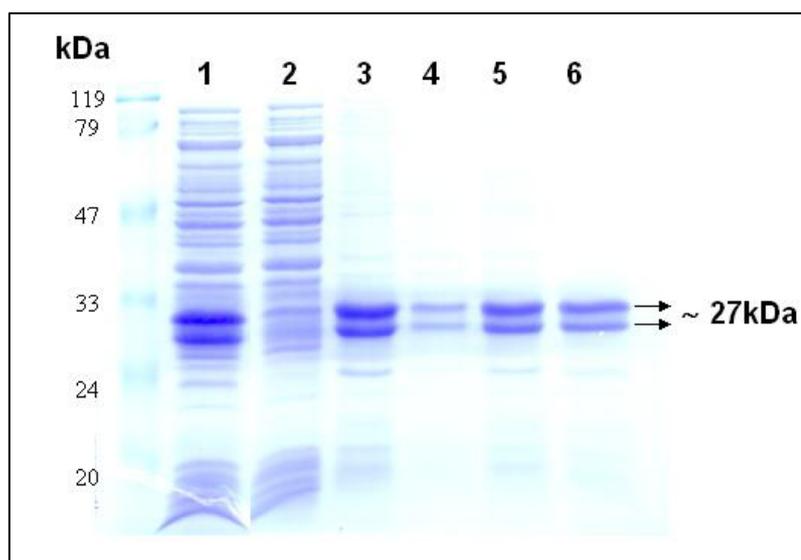


Figura 25 – **Corpos de inclusão com His<sub>6</sub>-Ras11b para imunização em coelhos.** SDS PAGE com 12% de acrilamida, mostrando a eliminação das proteínas contaminantes de bactéria “co-purificadas” com os corpos de inclusão, após o tratamento descrito no texto. Marcador de massa molecular de proteínas *Prestained Protein Marker* (Fermentas). **Poço 1** – Lisado total. **Poço 2** – Fração solúvel. **Poço 3** – Fração insolúvel com corpos de inclusão. **Poço 4** – Lavagem com uréia. **Poço 5** – Corpos de inclusão após a lavagem. **Poço 6** – Corpos de inclusão usados para imunização.

#### 4.2 Transcritos intrônicos marcadores de prognóstico em câncer de próstata.

Como parte das estratégias para aprofundar a exploração das informações geradas durante o projeto HCGP, o grupo de pesquisa coordenado pelo Prof. Dr. Sergio Verjovski-Almeida construiu uma plataforma de *microarray* enriquecida em mensagens novas, provenientes principalmente de regiões intrônicas de genes codificantes para proteína (detalhes da plataforma no item 3.5.1). Essa abordagem resultou na identificação de um conjunto de novas mensagens humanas, longas e sem *splicing*, sem potencial codificante para proteínas, transcritas a partir de introns de mRNAs já anotados, e mais importante com perfil de expressão correlacionado com o grau de diferenciação tumoral de próstata. O conjunto de resultados que levou a esse importante achado está descrito abaixo.

#### 4.2.1 Transcritos intrônicos são expressos em próstata

Inicialmente, para verificar qual fração dos novos transcritos é expressa em tecidos prostáticos foram realizadas hibridizações contra a plataforma intrônica de *microarrays* de cDNA utilizando mRNA poli-(A)+ derivado de 20 amostras normais adjacentes ao tumor, e 27 amostras tumorais provenientes de 27 pacientes. Como pode ser verificado na tabela 2, foram detectadas frações comparáveis de transcritos expressos em próstata que mapeiam em regiões intrônicas de mensagens manualmente curadas (*RefSeq* - *Reference Sequence*), em regiões não anotadas (intergênicas), assim como em exons conhecidos de genes humanos bem anotados (genes *RefSeq*).

Mapeamento do clone ORESTES	Número de regiões genômicas únicas	Número de clones ORESTES		
		Total na coleção	Total no <i>microarray</i>	Expresso em próstata
Introns de <i>RefSeq</i>	37.899	150.236	619	346 (56%)
Regiões não anotadas	45.400	136.628	1.019	539 (53%)
Exons conhecidos de <i>RefSeq</i>	29.444	390.263	2.048	1.369 (67%)

Tabela 2 – **Atividade transcricional em próstata de fragmentos de mensagens que mapeiam em diferentes regiões do genoma humano, e representação dos clones.** “Total na coleção” se refere ao número de clones na coleção gerada pelo *Human Cancer Genome Project* pela técnica ORESTES (Dias-Neto *et al.*, 2000). “Expresso em próstata” se refere aos fragmentos de mensagens que cujas intensidades passaram pelo filtro de expressão (3 desvios padrão acima da média das intensidades dos controles negativos da lâmina) em pelo menos 5 amostras de tecidos (total de 47). A porcentagem de fragmentos expressos está indicada entre parênteses. A anotação de mapeamento de cada fragmento foi realizada em 2004.

A observação de que grande parte dos transcritos totalmente intrônicos sem *splicing* representados na plataforma são expressos em próstata (Tabela 2) é um forte argumento contra abordagens informáticas conservadoras, que normalmente excluem de análises de produtos de transcrição esse tipo de mensagem intrônica, argumentando com a possibilidade de serem contaminantes de DNA genômico (Sorek & Safer, 2003). Além disso, esse resultado aponta para uma falha nas análises de expressão gênica para estudos de doenças, que utilizam plataformas exônicas comerciais, contendo apenas sondas para genes com *splicing* e codificantes para proteínas, previamente caracterizados. Com essa abordagem mais ortodoxa, usando somente sondas de genes conhecidos, não é possível detectar alterações nessas novas mensagens intrônicas, potencialmente regulatórias como discutido posteriormente, nas condições patológicas em estudo.

#### **4.2.2 Níveis de transcritos intrônicos não codificantes são correlacionados ao grau de diferenciação tumoral em próstata**

Com o objetivo de verificar se os novos transcritos expressos em próstata, principalmente os derivados de regiões intrônicas, possuem alguma significância biológica foram realizadas análises utilizando apenas os valores de expressão obtidos nas hibridizações com amostras tumorais de próstata. Já sem nenhuma análise estatística prévia, o perfil de expressão das 27 amostras apontava para a existência de uma assinatura gênica diferenciada para amostras com *Gleason Score* (GS) (Gleason, 1977) altos (GS 9 e 10). Um *Self-Organizing Map* e uma análise de *Hierarchical clustering* não supervisionada mostraram que todas amostras com alto-GS (9 e 10) agrupam-se juntas, apontando para um perfil de expressão bastante distinto nessas amostras (Figura 26A e B). De fato, alguns trabalhos já haviam

identificado genes com perfis de expressão altamente correlacionados ao GS (Singh *et al.*, 2002; Best *et al.*, 2003; Lapointe *et al.*, 2004).

Em seguida, utilizando a metodologia de análise de correlação de Pearson com permutações (detalhes no item 3.1.3) foi realizada uma busca por transcritos significativamente correlacionados com o GS, apenas entre as amostras de valores baixos (5 e 6) e de valores altos (9 e 10). Com essa abordagem foi identificado um conjunto classificador com 56 mensagens cujos níveis de expressão eram significativamente ( $p < 0,001$ ) correlacionados com a distinção entre baixo/alto GS (Figura 27A). Dentro desse, foram identificados um subgrupo 27 seqüências representativas de exons conhecidos de RefSeq, 23 transcritos intrônicos de RefSeq e 6 mensagens provenientes de regiões genômicas sem anotação.

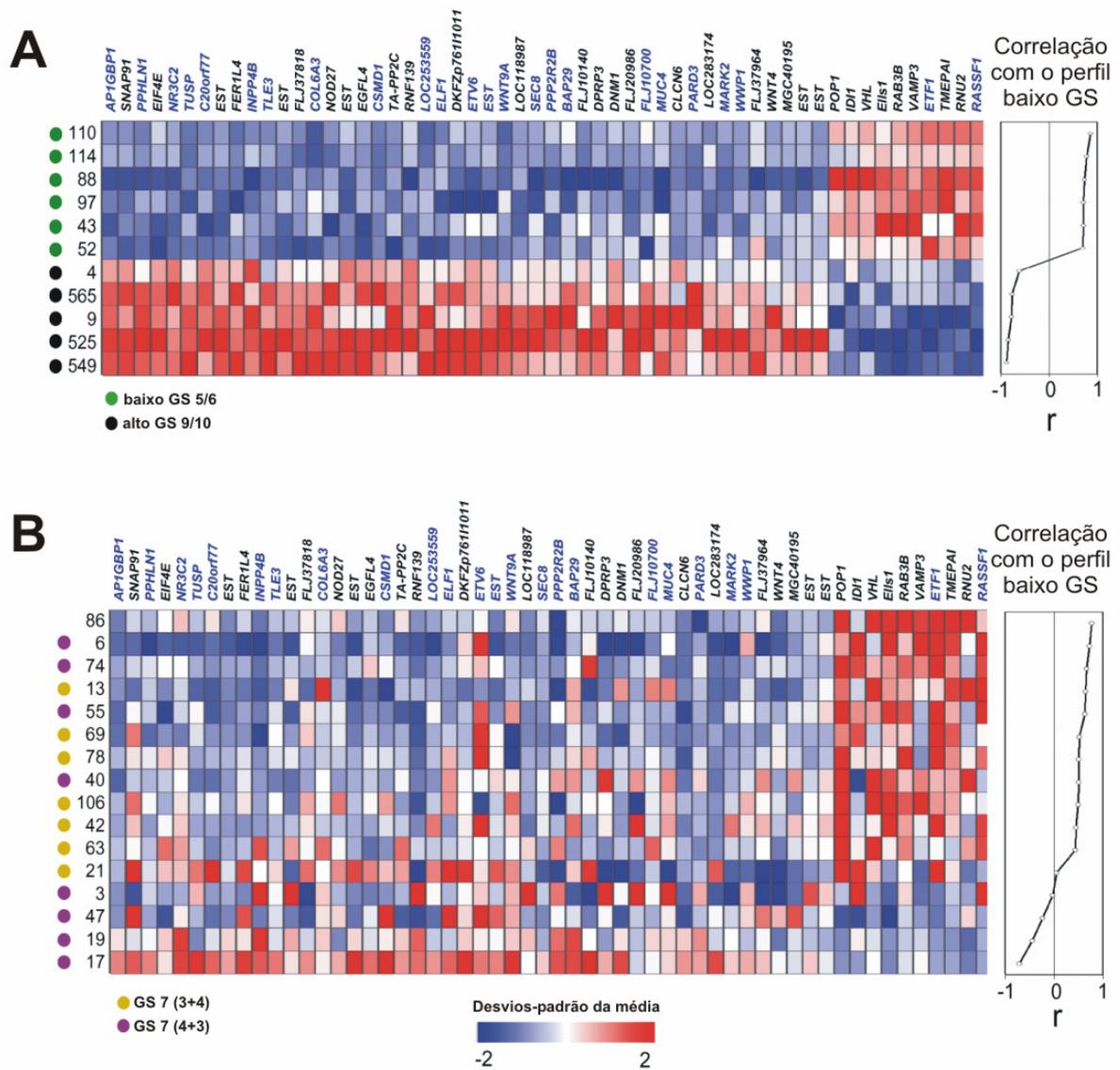


Figura 27 - **Transcritos correlacionados com o grau de diferenciação (GS) em tumores de próstata.** Os painéis mostram a matriz de expressão de 56 transcritos correlacionados com a distinção de classe baixo/alto GS ( $p \leq 0,001$ ) (A) e sua aplicação nos perfis de expressão dos pacientes de GS intermediário (GS 7) (B). Os transcritos (colunas) estão ordenados pela sua correlação de Pearson e os pacientes (linhas) pela sua correlação ( $r$ ) com o perfil médio das amostras de baixo GS, mostrada nos painéis da direita. Fragmentos intrônicos têm o nome do *locus* gênico em azul. Níveis de expressão em cada paciente estão representados pelo número de desvios-padrão acima (vermelho) ou abaixo (azul) da média do transcrito em todas as amostras.

A aplicação do classificador baixo/alto GS ao perfil de expressão de dezesseis amostras com grau intermediário de diferenciação tumoral (GS 7), revelou dois grupos com padrão de expressão distintos dentre pacientes com características

histológicas bastante similares (Figura 27B). Enquanto a maioria das amostras GS 7 exibiram uma correlação positiva com as amostras de baixo GS, quatro amostras (pacientes 17, 19, 47 e 3) apresentaram perfil de expressão mais correlacionado com amostras de alto GS. Esse resultado provavelmente está relacionado ao fato de que pacientes com tumores de próstata de GS 7 apresentam características clínicas bastante heterogêneas (Epstein *et al.*, 1996).

Do conjunto de transcritos correlacionados, a fração de mensagens mais robustas como classificadores de baixo/alto GS foi determinada por *leave-one-out* (detalhes no item 3.1.3). Transcritos que apareceram em pelo menos 60% dos classificadores gerados são mostrados na figura 28. Surpreendentemente, sete das doze mensagens mais robustas são transcritas a partir de regiões intrônicas de *RefSeq*. As cinco demais representam exons conhecidos de *RefSeq*, dos quais três genes, *EIF4E*, *TMEPA1* e *RNF139*, têm sido associados com a progressão de diferentes tipos de câncer (Gemmill *et al.*, 1998; Brunschwig *et al.*, 2003; Ramaswamy *et al.*, 2003).

#### **4.2.3 Caracterização de transcritos intrônicos relacionados com câncer**

Para caracterizar o tamanho e a orientação transcricional dos sete transcritos intrônicos identificados no classificador baixo/alto GS mostrados na figura 28, foram utilizadas abordagens de RACE-PCR (detalhes no item 3.4.1), Northern Blot (detalhes no item 3.6.2), e transcrições reversa com orientação específica (OSRT) seguida de PCR (detalhes no item 3.7.1).

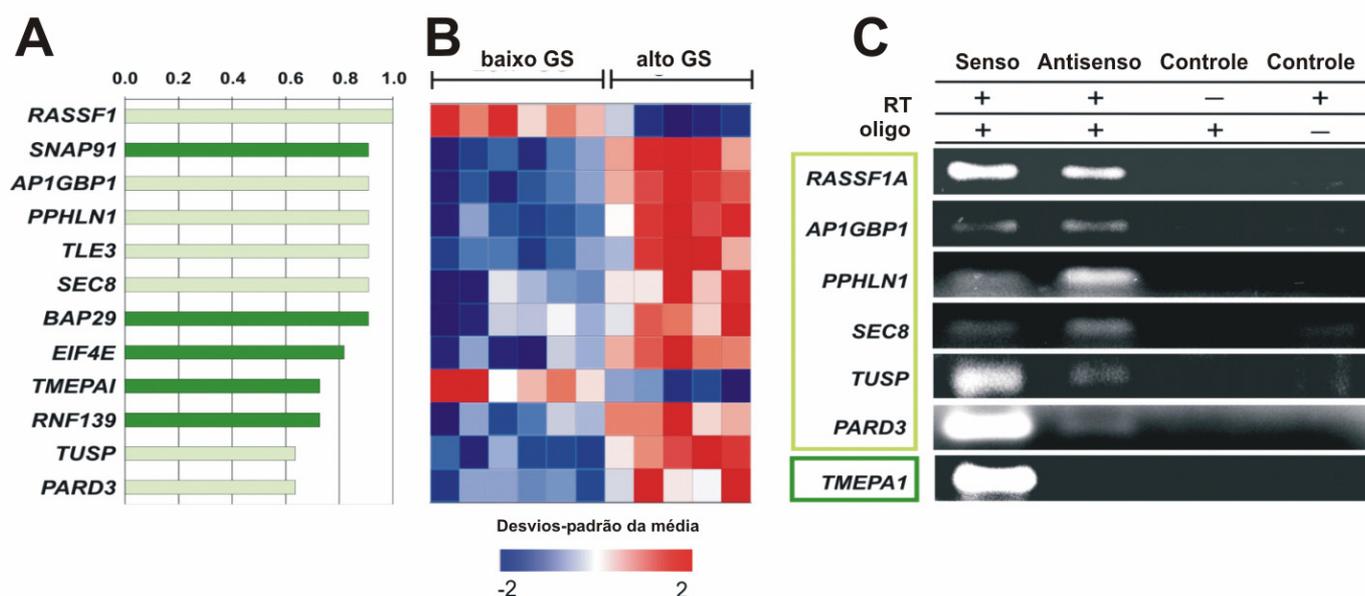


Figura 28 - **Transcritos intrônicos mais robustos na distinção de classe baixo/alto GS.** (A e B) 12 mensagens classificadoras de grau de Gleason. O painel (A) mostra os transcritos correlacionados ( $p \leq 0,001$ ) que aparecem em pelo menos 60% dos classificadores (fração indicada no eixo x), identificados por *leave-one-out*. Barras verde-escuras correspondem a exons conhecidos de *RefSeq*, e barras verde-claras são transcritos intrônicos. O painel (B) mostra o nível de expressão de cada transcrito (linha) para cada amostra de tumor (coluna). Níveis de expressão em cada paciente estão representados pelo número de desvios-padrão acima (vermelho) ou abaixo (azul) da média do transcrito em todas as amostras. (C) Mensagens intrônicas foram detectadas tanto com orientação senso quanto antisenso. Os controles para garantir a ausência de contaminação por DNA genômico (- RT) e de auto-anelamento do RNA (- oligo) estão indicados.

O resultado de RACE-PCR para a seqüência intrônica de *TLE3* mostrou que essa mensagem é na verdade um novo exon do transcrito de mesmo *locus* (Figura 29A), que codifica para uma proteína membro da via de sinalização Notch, envolvida em diferenciação epitelial (Liu *et al.*, 1996). O resultado do Northern blot utilizando sonda específica para esse novo exon mostrou a presença de uma banda com cerca de 5,5 kb, a qual é totalmente compatível com uma mensagem completa de *TLE3* com esse novo fragmento incluso (Figura 29B). A adição desse exon introduz um códon de terminação de tradução precoce que trunca o domínio WD-40 da porção carboxi-terminal, sendo observada a expressão dessas formas alternativas na maioria dos metazoários cuja função acredita-se estar envolvida na regulação negativa das proteínas Tle3 (Chen & Courey, 2000).

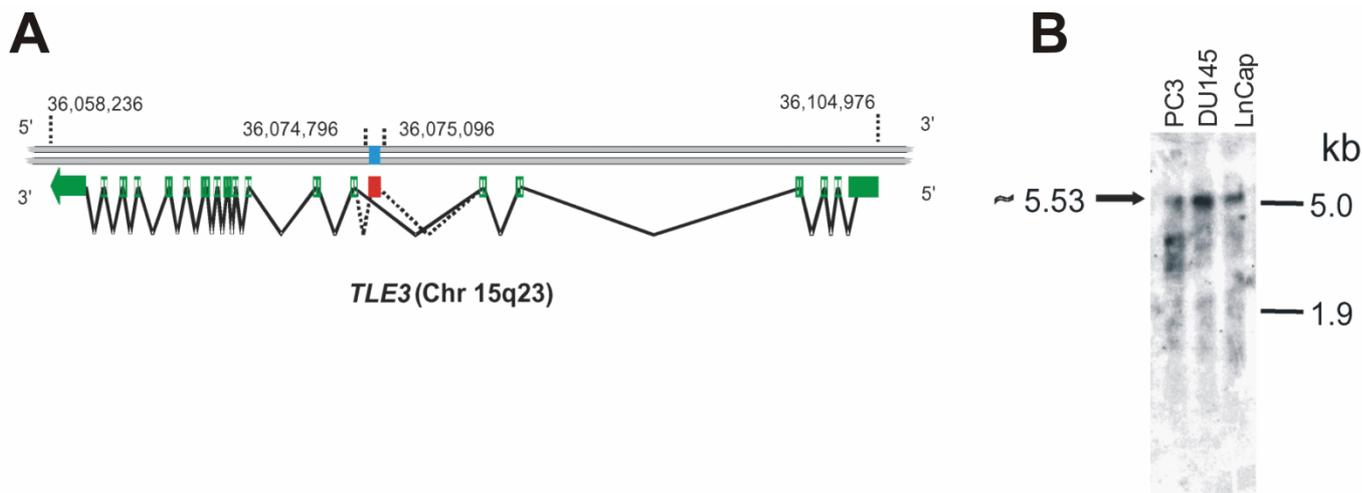


Figura 29 – **Novo exon no transcrito de TLE3.** O mapeamento genômico dos fragmentos de cDNA dupla-fita presentes no *microarray* (em azul) e as coordenadas cromossômicas estão indicadas. **(A)** Forma alternativa de *splicing* do gene *TLE3*, confirmadas por RACE-PCR (bloco vermelho). **(B)** Forma alternativa de *splicing* do gene *TLE3*, um transcrito com 5,53 kbp detectado por *Northern blot* em três linhagens derivadas de tumores de próstata, utilizando uma sonda complementar a fita senso do novo exon de *TLE3*.

Utilizando RACE-PCR com uma biblioteca comercial de pulmão foi determinado que as mensagens intrônicas de *RASSF1*, *AP1GBP1* e *PPHLN1* eram longas (0,6 a 1,1 kb), sem *splicing* e contidas totalmente nas regiões intrônicas dos RNAs codificantes de proteína de mesmo *loci* (Figuras 30A, B e C - caixas amarelas). Aparentemente essas mensagens estendidas não geram proteínas, pois análise utilizando *ESTScan* (Iseli *et al.*, 1999) não apontou nenhum potencial codificante. Para as mensagens intrônicas de *RASSF1* e *AP1GBP1*, um RACE-PCR utilizando a biblioteca de cDNA de cólon clonada com orientação determinou a presença de transcrição antisenso (Figuras 30A e B - setas vermelhas). Ainda, a orientação antisenso das mensagens intrônicas de *RASSF1* e *PPHLN1*, assim como seu tamanho, foi confirmado por Northern Blot com sondas de fita simples (Figuras 30D e E). Produtos de RACE-PCR para os 3 restantes transcritos intrônicos, com as bibliotecas utilizadas, não estenderam os respectivos fragmentos de cDNA depositados na plataforma de *microarray*.

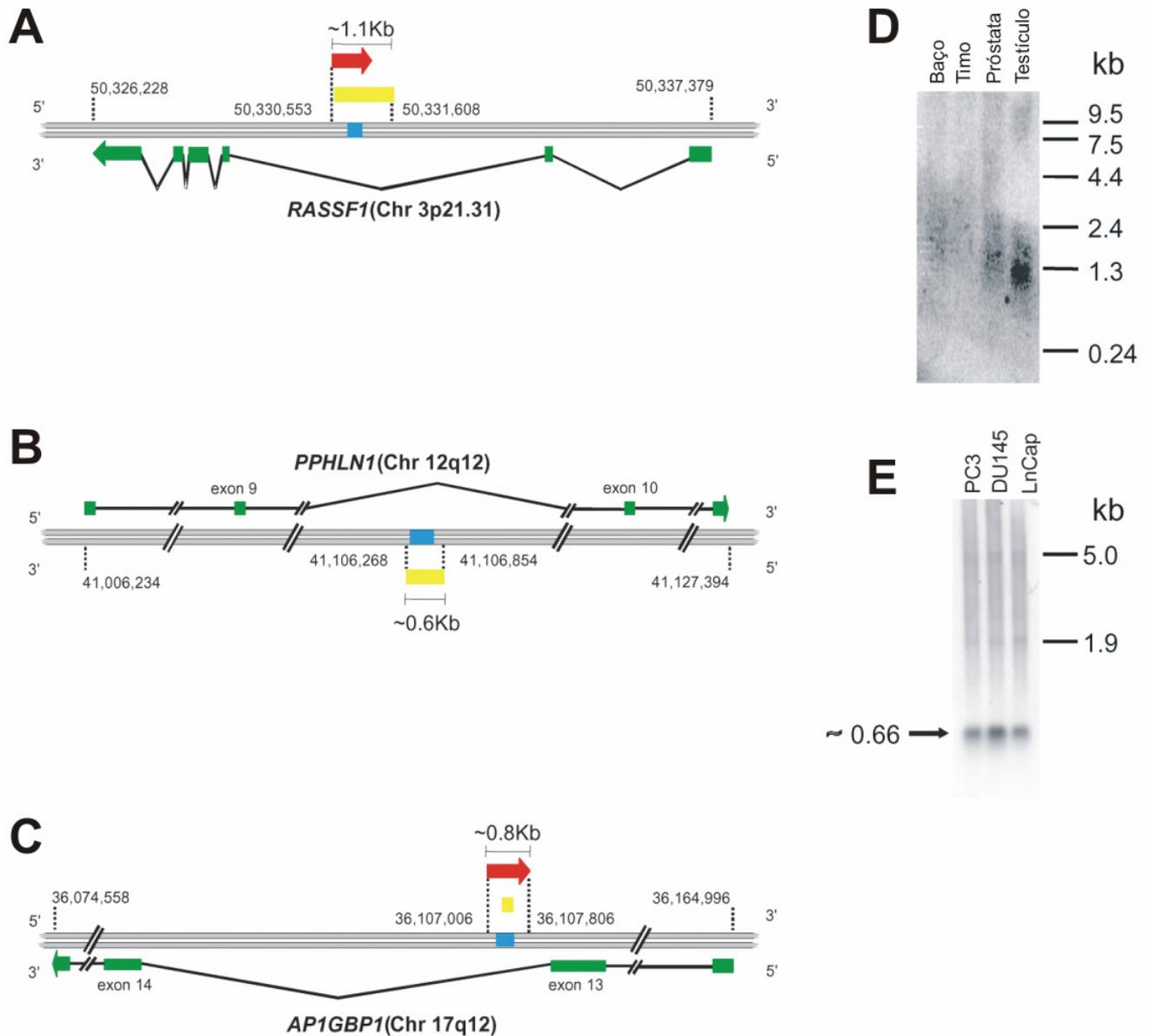


Figura 30 - Longos RNAs Intrônicos antisense correlacionados com GS caracterizados por RACE-PCR e Northern Blot. O mapeamento genômico dos fragmentos de cDNA dupla-fita presentes no *microarray* (em azul) e as coordenadas cromossômicas estão indicadas. (A–C) Mapeamento dos longos transcritos intrônicos de *asRASSF1*, *asPPHLN1* e *asAP1GBP1* estendidos por RACE-PCR. Flechas vermelhas mostram transcritos obtidos por RACE-PCR utilizando bibliotecas de cDNA humano com orientação definida (côlon). Blocos amarelos mostram transcritos obtidos por RACE-PCR com biblioteca de pulmão. (D–E) Longos transcritos intrônicos de *asRASSF1* e *asPPHLN1* detectados por Northern Blot utilizando sondas complementares à fita antisense das respectivas regiões intrônicas.

A abordagem de OSRT-PCR (ver item 3.7.1) mostrou, e também confirmou, que das sete mensagens intrônicas analisadas, seis eram transcritas com orientação antisenso em relação ao intron correspondente (Figura 28C). Surpreendentemente, foram também detectadas mensagens intrônicas senso sobrepostas às antisenso nas amostras de mRNA testadas (Figura 28C).

#### **4.2.4 Validação por PCR quantitativo em tempo real das mensagens intrônicas de *RASSF1***

Utilizando *microarrays* de cDNA fita dupla não é possível determinar separadamente os níveis de expressão provenientes de mensagens senso e antisenso sobrepostas. Assim os resultados de expressão diferencial observados para essas mensagens intrônicas poderiam refletir tanto alterações em uma, quanto em ambas mensagens. Para discriminar entre essas possibilidades foi utilizada uma abordagem de PCR quantitativo em tempo real, após a realização de transcrições reversas com orientação específica (OSRT-qPCR) (para detalhes ver item 3.7.2) para confirmar a expressão diferencial dos transcritos intrônicos de *RASSF1*, o mais correlacionado com o grau de diferenciação tumoral.

A figura 31A mostra que ambos transcritos intrônicos, senso e antisenso, estão significativamente diminuídos em amostras de alto GS, em comparação com amostras de baixo GS ( $p < 0,024$  e  $p < 0,014$ ; respectivamente). Como mostra a figura 31B, os níveis dos transcritos exônicos codificantes de proteína de *RASSF1* também são significativamente diminuídos em amostras de alto GS ( $p < 0,009$ ). Na figura 31C vê-se que os níveis de *RASSF1* codificante de proteína são menores ( $p < 0,002$ ) em amostras de tumores de próstata GS 7 tumors cujos perfis de expressão

são negativamente correlacionados com o padrão de baixo GS (pacientes 17, 19, 47 e 3), o que confirma os dados do experimento de *microarray* mostrado na figura 27B. Já havia sido demonstrado na literatura que a hipermetilação da região promotora de *RASSF1* correlacionava-se com o GS de tumores de próstata (Liu *et al.*, 2002). Nossos dados acrescentam o fato de que a expressão do transcrito codificante da proteína *RASSF1* também está correlacionada com o GS de tumores de próstata.

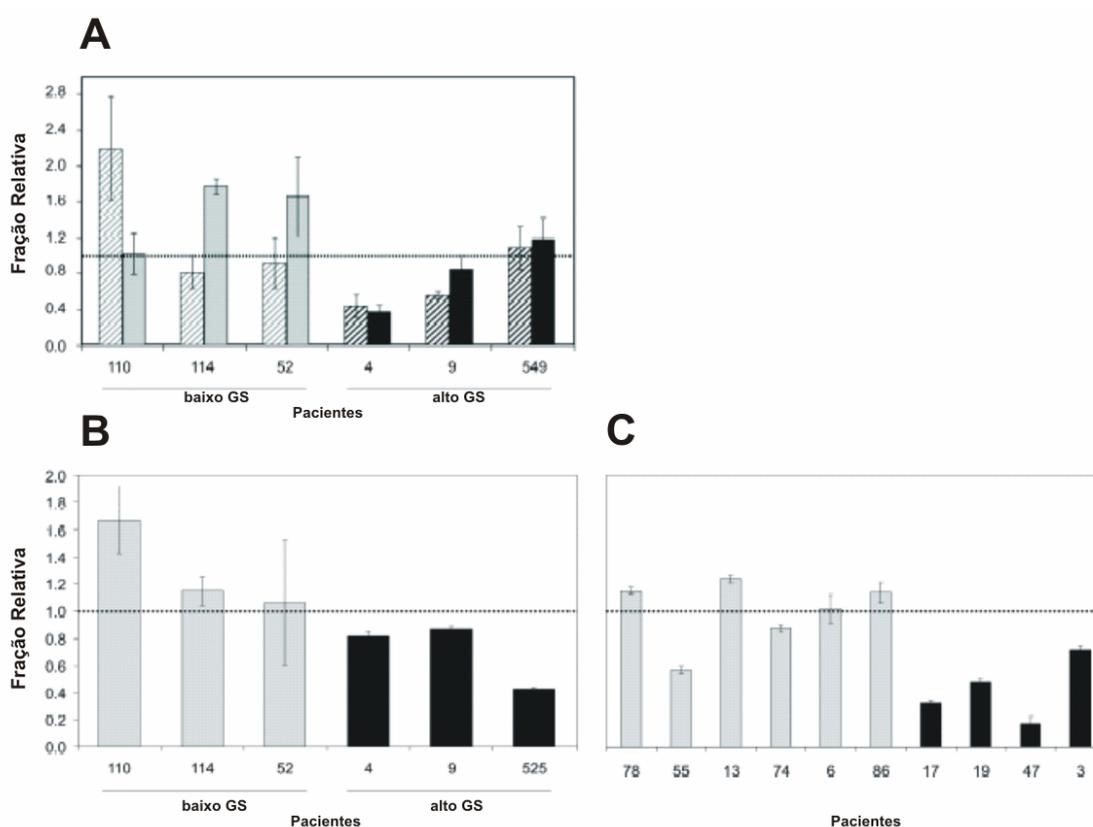


Figura 31 - **Expressão diminuída de *asRASSF1* está em tumores com alto GS.** OSRT-qPCR em tempo real foi usada para medir os níveis de mRNA. Frações relativas foram calculadas em relação à abundância média de mRNA entre todas as amostras testadas. **(A)** Os níveis de mensagens intrônicas antisense (barras hachuradas) e senso (barras cheias) transcritas a partir do *locus* de *RASSF1* foram medidos em 3 amostras de tumor com grau de baixo GS (barras cinzas) e alto GS (barras pretas). **(B)** Transcritos codificantes para a proteína *Rassf1*, medidos nas mesmas amostras. **(C)** Transcritos codificantes para a proteína *Rassf1* medidos em amostras de tumor com grau de GS 7 que apresentaram correlação negativa (barras pretas) ou positiva (barras cinzas) com o perfil de expressão médio das amostras de baixo GS obtido nos experimentos de *microarray*. Diferenças de expressão estatisticamente significativas com  $p < 0,05$ .

### **4.3 Transcritos intrônicos regulados pelo estímulo de andrógeno.**

O trabalho que identificou os Longos RNAs Intrônicos não-codificantes (ncRNAs), e os relacionou com a progressão do câncer de próstata (Reis *et al.*, 2004) descrito anteriormente (ver item 4.2), além de ilustrar a enorme complexidade do programa transcricional humano, permitiu a elaboração de hipóteses sobre as possíveis funções biológicas destes transcritos intrônicos relacionadas ao controle da expressão gênica em eucariotos (Reis *et al.*, 2005). Assim, foram priorizados os estudos relacionados à caracterização dessas mensagens, visando principalmente agregar evidências sobre essas funções regulatórias dos ncRNAs intrônicos. Para alcançar esse objetivo foram desenhados experimentos para avaliar a existência de um mecanismo comum de controle da transcrição dos RNAs intrônicos, que fosse compartilhado com o de mRNAs codificantes para proteína. Essa modulação da expressão em condições fisiológicas normais envolveria, por exemplo, um estímulo hormonal, que acarreta sensíveis alterações nos perfis transcricionais.

#### **4.3.1 Identificação de transcritos intrônicos regulados por andrógeno**

Utilizando a plataforma de *microarray* de cDNA enriquecida em transcritos intrônicos (detalhes da plataforma no item 3.5.1) foi realizado um estudo de expressão gênica para identificar mensagens cujos níveis de RNA fossem significativamente alterados em células da linhagem tumoral de próstata em resposta a um tratamento com andrógeno sintético R1881 em diferentes tempos de incubação (detalhes no item 3.2.3). Precedente às hibridizações, a eficiência desses tratamentos foi avaliada por PCR quantitativo em tempo real para determinação dos níveis de expressão de *KLK3* (*PSA, Prostate Specific Antigen*) (Nelson *et al.*, 2002) e

*TMEPAI* (Xu *et al.*, 2000), dois transcritos que já haviam sido demonstrados como regulados por andrógeno no epitélio da próstata (Figura 32).

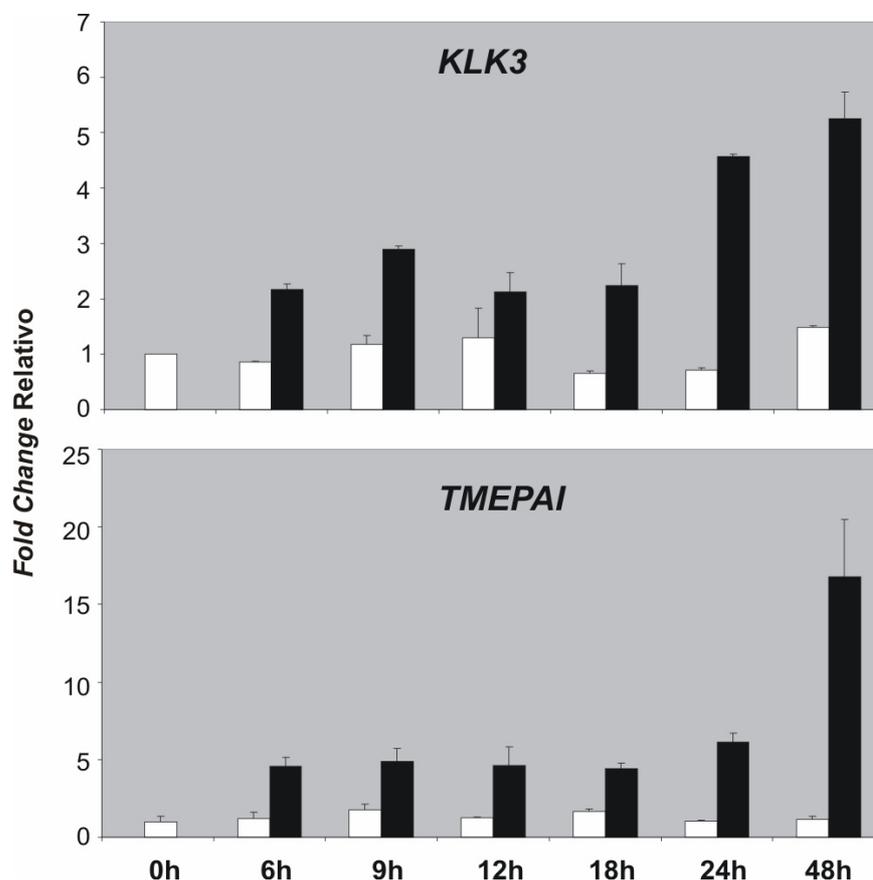


Figura 32 – **Indução de genes regulados por andrógeno em células de próstata.** Níveis de expressão de *KLK3* (*PSA, Prostate Specific Antigen*) e *TMEPAI* determinados por PCR quantitativo em tempo real para confirmar que o tratamento foi efetivo para induzir a resposta de genes regulados em células LNCaP suplementadas com andrógeno sintético de 6 a 48h. *Fold-change* Relativo indica os níveis de RNA de *KLK3* (**superior**) e *TMEPAI* (**inferior**) para cada ponto com relação ao tempo 0, e são mostrados para células tratadas com andrógeno (barras pretas) e para as células controle sem tratamento (barras brancas).

Com a validação do estímulo do andrógeno nas células LNCaP, um total de 26 hibridizações, correspondentes aos tempos de tratamento e controles, foi realizada. Os valores de intensidade normalizados provenientes desses experimentos foram analisados com a abordagem de SAM (Tusher *et al.*, 2001) para a identificação dos transcritos com alterações de abundância significativas após

exposição ao andrógeno de 6 a 48 horas (detalhes no item 3.1.3). Como resultado, foram detectadas 168 mensagens exônicas codificantes para proteína (Figura 33A) e 39 transcritos intrônicos (Figura 33B) com padrão de expressão temporal estatisticamente diferente em células tratadas com andrógeno em comparação com os controles sem tratamento. A lista completa dos transcritos exônicos codificantes para proteína regulados por andrógeno é apresentada na Tabela Suplementar 1 (ANEXO 5).

Os RNAs intrônicos regulados pelo andrógeno puderam ser agrupados em três diferentes conjuntos baseados em distintos padrões de expressão temporal após o tratamento (Figura 33C). O Grupo I contém RNAs cujos níveis foram diminuídos de 2 a 3 vezes após 24 horas de exposição ao andrógeno. Os Grupos II e III incluem RNAs cujos níveis foram aumentados com 24 horas de tratamento: enquanto o Grupo II apresenta RNAs aumentados de 2 a 3 vezes, o grupo III apresenta mensagens com aumentos de 3 a 6 vezes. Em todos os casos, os níveis de expressão dos transcritos intrônicos retorna a um valor próximo do inicial (tempo 0) após 48 hs de introdução do andrógeno na cultura (Figura 33C).

Todos os 39 RNAs intrônicos identificados são mensagens sem *splicing*, e aparentemente não são utilizados para sintetizar proteínas, pois não apresentam potencial codificante significativo em análises por *ESTScan* (Iseli *et al.*, 1999). Assim, é pouco provável que esses transcritos intrônicos representem novos exons das mensagens codificantes de proteína correspondentes. A lista completa dos RNAs intrônicos regulados por andrógeno é apresentada na Tabela Suplementar 2 (ANEXO 6).

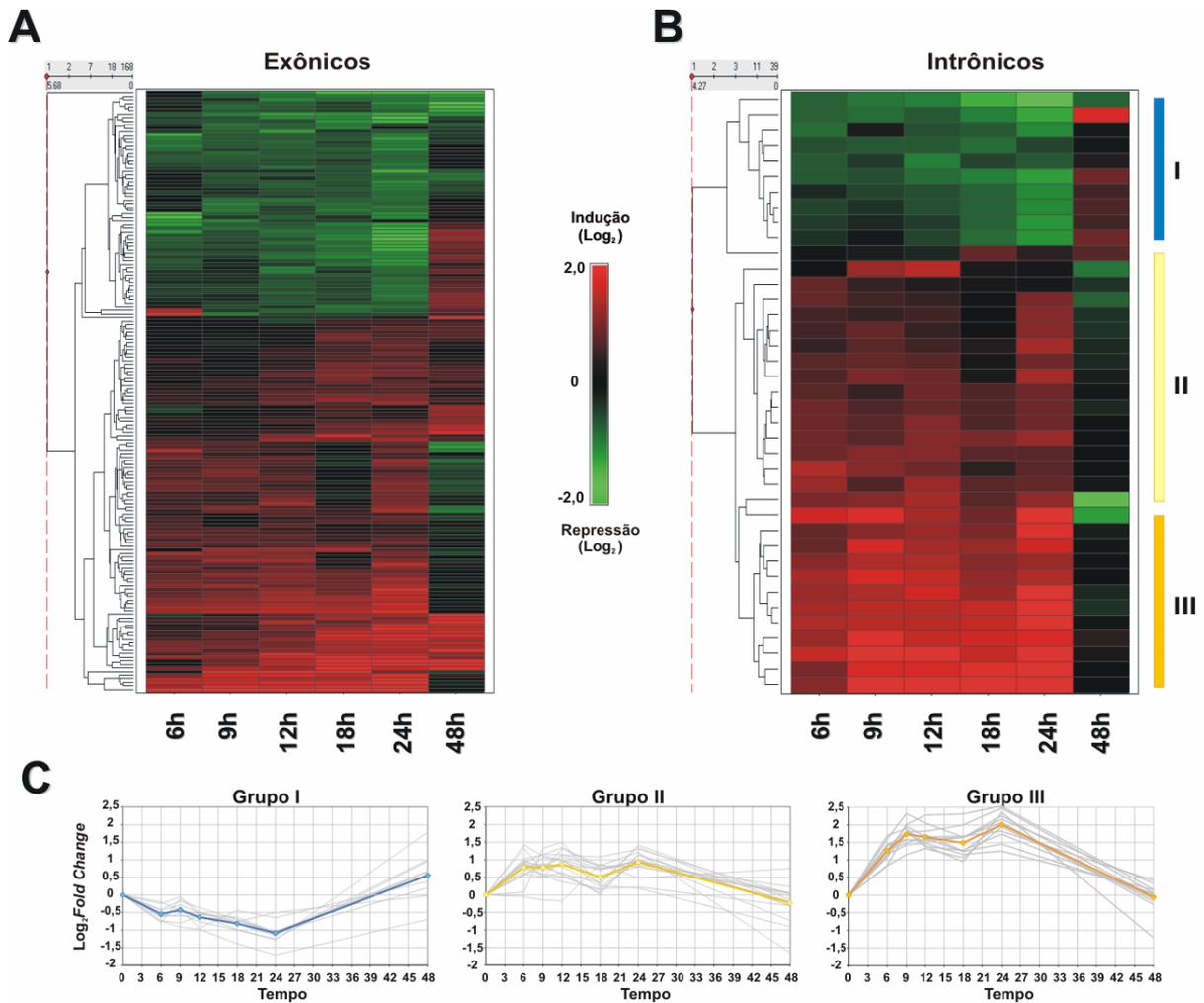


Figura 33 – Perfil temporal de expressão de transcritos que respondem ao tratamento com **andrógeno**. Análise por *Hierarchical clustering* (Sanchez-Carbayo *et al.*, 2002) de transcritos codificantes para proteínas (exônicos) (**A**) e intrônicos (**B**) significativamente ativados ou reprimidos após a exposição das células ao andrógeno é mostrada. Transcritos intrônicos que respondem ao tratamento com andrógeno e que apresentam perfis temporais de expressão semelhantes foram agrupados usando *K-means* (**C**). Grupos de mensagens intrônicas com padrões similares de expressão são indicados por barras verticais coloridas (I–III) no painel **B**. *Fold-change* médio de cada grupo (Grupos I–III) é indicado em **C** por uma linha colorida.

### 4.3.2 RNAs intrônicos regulados por andrógeno mapeiam em genes envolvidos em transdução de sinais

A ferramenta de comparação de ontologias gênicas (GO) denominada eGOn (<http://www.genetools.microarray.ntnu.no/egon/>) foi utilizada para investigar o enriquecimento de categorias específicas de GO entre os transcritos regulados por andrógeno (detalhes no item 3.1.5).

Para os transcritos intrônicos, foi postulado que representam uma nova classe de RNAs que podem atuar em *cis* como fatores regulatórios (Reis *et al.*, 2005). Dessa forma, cada mensagem intrônica foi anotada com as categorias de GO do mRNA codificante de proteína transcrito a partir do mesmo *locus* gênico. A análise dos *loci* que contém transcritos intrônicos regulados por andrógeno mostrou enriquecimento em categorias de GO normalmente envolvidas em respostas celulares a hormônios: *cell organization and biogenesis* (GO:0016043), *protein binding* (GO:0005515) and *signal transducer activity* (GO:0004871).

A análise com as mensagens exônicas reguladas por andrógeno mostrou várias categorias de GO significativamente enriquecidas ( $p < 0,05$ ), entre elas algumas diretamente relacionadas ao estímulo por andrógeno, como *regulation of metabolism* (GO:0019222), *specific DNA-binding* (GO:0043566) e *intracellular receptor-mediated signaling pathway* (GO:0030522).

### 4.3.3 Caracterização dos RNAs intrônicos regulados por andrógeno

Um sub-conjunto de RNAs intrônicos (17 mensagens) regulados por andrógeno foi selecionado para caracterização. Transcrições reversas com orientação específica (OSRT) seguida de PCR (detalhes no item 3.7.1) foram utilizadas para determinação da orientação transcricional desse grupo de

mensagens (Figura 34) em relação aos mRNAs codificantes de proteína de mesmo *loci*. Para a grande maioria (13 de 17) o PCR forneceu resultados. Para 10 *loci* dos 13, um transcrito intrônico antisense foi detectado (Figura 34B), sendo que em seis desses (*DNAJC3*, *SAP18*, *KDELR2*, *FLJ10154*, *MYO5A* e *ATF2*) foi também detectada uma mensagem senso sobreposta; nos quatro *loci* restantes (*GAS6*, *RAB25*, *ITGA6* e *ADD3*) apenas o RNA antisense foi detectado. Ainda, para três outras mensagens intrônicas entre as 13 (*loci* *ACTN4*, *STARD13* e *ERG*) foram detectadas apenas transcritos com orientação senso (Figura 34B).

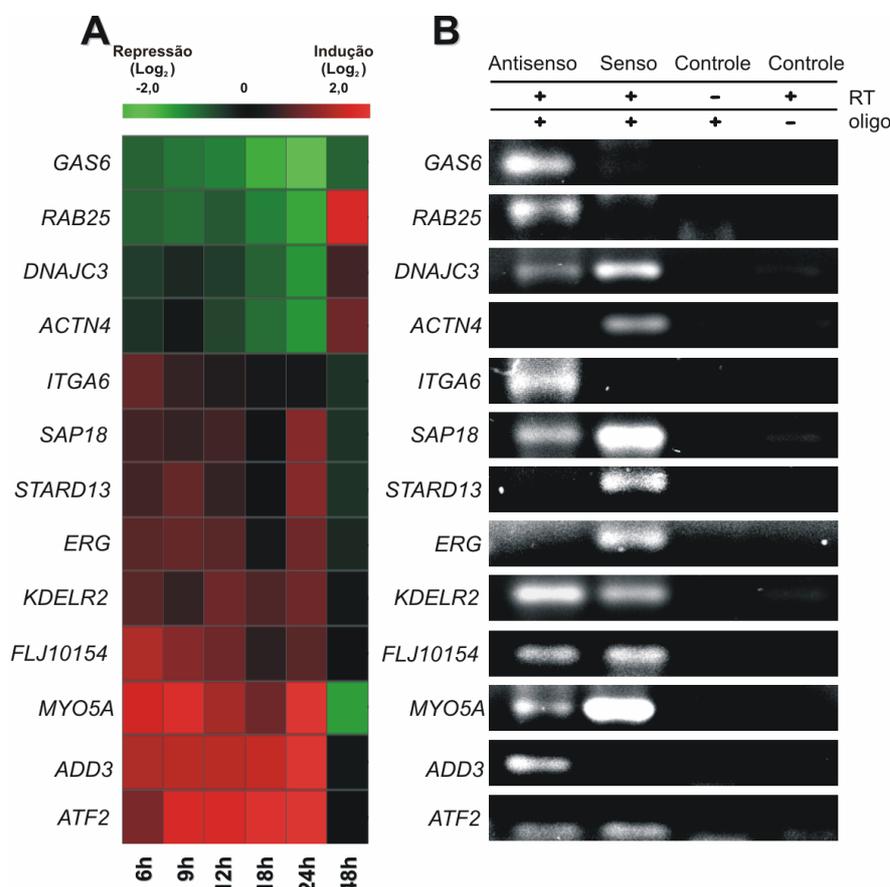


Figura 34 – **Orientação transcricional de mensagens intrônicas que respondem ao tratamento por andrógeno.** Transcrição reversa orientação-específica seguida de PCR (OSRT-PCR) foi realizada em um conjunto selecionado de 13 mensagens intrônicas reguladas por andrógeno identificadas pelas análises de *microarray* (A) para determinação da orientação transcricional da mensagem intrônica relativa à mensagem codificante para proteína de mesmo *locus*. Nomes a esquerda se referem aos mRNAs RefSeq codificantes para proteína que mapeiam nos mesmos *loci* e definem as regiões intrônicas onde atividade transcricional foi identificada. (B) Produtos de PCR separados em gel de agarose 3% detectaram transcrição antisense (*GAS6*, *RAB25*, *ITGA6* e *ADD3*), senso (*ACTN4*, *STARD3* e *ERG*) ou antisense/senso simultânea (*DNAJC3*, *SAP18*, *KDELR2*, *FLJ10154*, *MYO5A* e *ATF2*). Os controles negativos, para garantir ausência de contaminação com DNA genômico (-RT) e de auto-anelamento do RNA (- oligo) durante a transcrição, estão indicados.

Como todos transcritos intrônicos identificados não apresentaram potencial codificante pela análise por *ESTScan*, eles realmente devem representar RNAs não-codificantes (ncRNAs) originados de unidades transcricionais independentes. Provavelmente isso também é verdade para as três mensagens com orientação senso, pois apesar de expressos a partir da mesma fita, eles aparentemente não são exons novos das mensagens codificantes para proteína dos *loci*. Transcrição intrônica senso em unidades independentes têm sido observada em outros trabalhos (Kapranov *et al.*, 2002; Bertone *et al.*, 2004).

Para determinação do tamanho completo de algumas mensagens intrônicas, foram realizados experimentos de RACE-PCR com uma biblioteca feita a partir de RNA de adenocarcinoma de próstata (detalhes no 3.4.1). Entre os RNAs intrônicos selecionados para determinação do transcrito completo estavam mensagens expressas com orientação antisenso (*GAS6* e *ITGA6*), senso (*STARD3*) e mensagens antisenso/senso sobrepostas (*MYO5A* e *DNAJC3*).

Os resultados dessa abordagem revelaram que os RNAs antisenso dos *loci* *DNAJC3*, *MYO5A* e *ITGA6* são longos (0,5 a 0,8kb), e mapeiam sem *splicing* totalmente dentro de regiões intrônicas das mensagens codificantes de proteína correspondentes (Figura 35A-C). Um transcrito intrônico do *locus* *STARD3*, também sem *splicing*, com aproximadamente 1,4kb foi estendido a partir do fragmento com orientação senso determinada por OSRT-PCR (Figura 35D).

O RNA intrônico antisenso que mapeia no *locus* *GAS6* foi estendido utilizando uma abordagem de PCR com oligonucleotídeos iniciadores específicos (detalhes no item 3.4.2). A presença de uma única mensagem composta pelo mRNA com cerca de 3,2kb obtido de um biblioteca de cDNA de cérebro, cuja seqüência encontra-se disponível no GenBank (AK123478), e o fragmento intrônico de *GAS6* depositado no

*microarray* foi confirmada pela obtenção de um produto com cerca de 4,2kb em cDNA de adenocarcinoma de próstata (Figura 35E – caixa amarela).

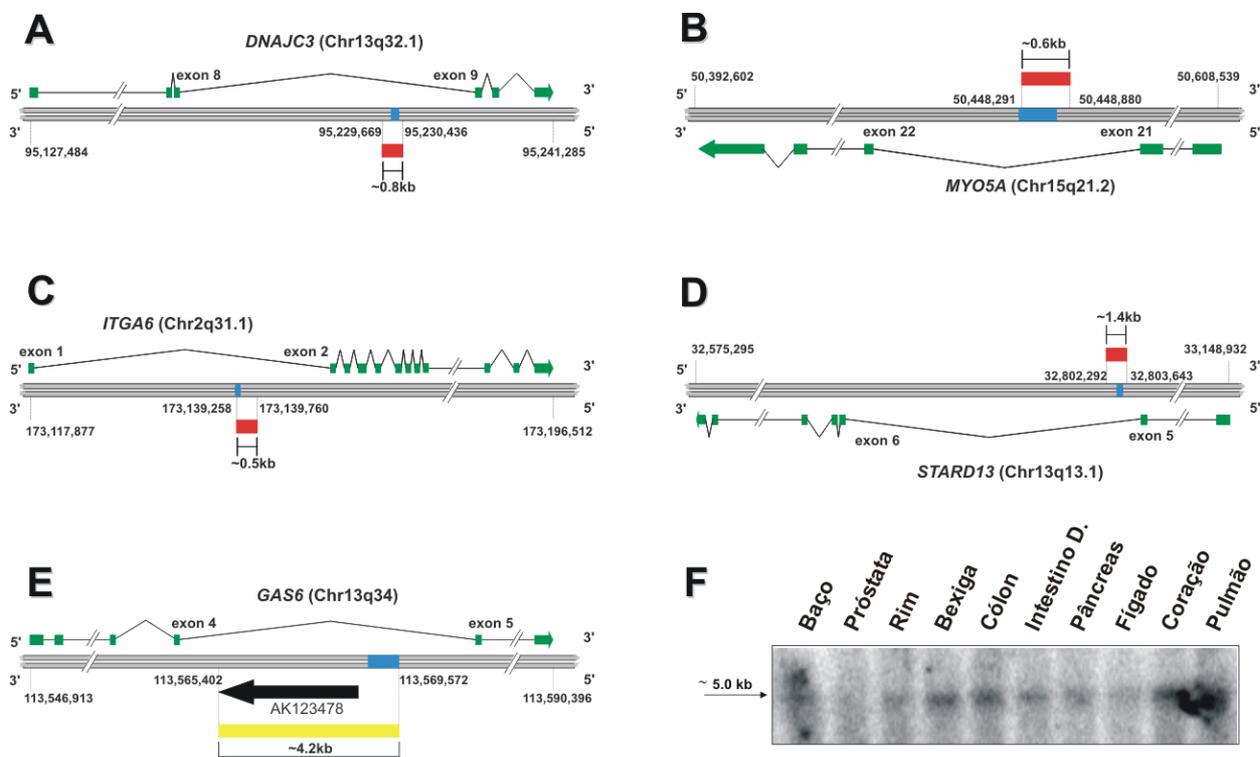


Figura 35 – **Caracterização dos transcritos intrônicos não codificantes que respondem ao tratamento com andrógeno.** Transcritos intrônicos selecionados foram estendidos por RACE-PCR (A–D, caixas vermelhas) ou PCR com oligonucleotídeos específicos (E, caixa amarela). Os longos transcritos intrônicos antisense sem *splicing* estendidos estão mapeados na sequência genômica (A – E, linhas cinza) com orientação relativa ao transcrito codificante para proteína expresso a partir da fita oposta (A – E, caixas verdes). Fragmentos de cDNA dupla-fita depositados nos *microarrays* estão representados como caixas azuis (A – E). Um transcrito intrônico antisense sem *splicing* que mapeia no *locus* de *GAS6* é mostrado (E, flecha preta). Coordenadas cromossômicas e o tamanho dos fragmentos intrônicos estendidos estão indicadas. *Northern blot* de múltiplos tecidos usando sondas complementares ao transcrito antisense do *locus* de *GAS6* confirma a existência de uma mensagem com aproximadamente 5,0kb.

Mais uma vez, a análise *in silico* utilizando *ESTScan* (Iseli *et al.*, 1999) apontou que não existem fases abertas de leitura (ORFs – *Open Reading Frames*) nas mensagens intrônicas estendidas dos *loci* *DNAJC3*, *ITGA6*, *MYO5A* e *GAS6*. E apesar de um peptídeo com 74 resíduos ter sido predito para o *STARD13* intrônico, buscas utilizando essa sequência protéica com *BLASTp* (Altschul *et al.*, 1997) não resultaram em identificação de proteínas similares no GenBank, nem de domínios

conservados presentes no CDD (Marchler-Bauer *et al.*, 2005). Dessa maneira, é definitivamente improvável que os RNAs intrônicos regulados por andrógeno, especialmente os estendidos por PCR, sejam utilizados para codificar proteínas.

Ainda para o transcrito intrônico de *GAS6*, o tamanho e a orientação transcricional foram confirmados, e a expressão em outros tecidos foi determinada por *Northern blot* utilizando sondas de fita simples geradas por transcrição *in vitro* (detalhes no item 3.6.2). Esse resultado mostra a expressão de uma mensagem longa, sem *splicing* e transcrita antisenso a partir do *locus* de *GAS6* com aproximadamente 5,0kb em diferentes tecidos humanos (Figura 35F). A observação que RNAs intrônicos regulados por andrógeno puderam ser detectados em tecidos que normalmente não sofrem esse tipo de controle hormonal (Figura 35F), sugere que a regulação dos níveis dessas mensagens pode também ocorrer por vias independentes do andrógeno.

#### **4.3.4 Efeitos do andrógeno nos níveis dos RNAs intrônicos e dos transcritos codificantes de proteína correspondentes**

Transcrições reversas com orientação específica (OSRT) seguidas de PCR quantitativo em tempo real (detalhes no item 3.7.2), usando RNA de células tratadas e células controle, foram utilizadas para confirmar as alterações na abundância das mensagens intrônicas após a exposição ao andrógeno. Foram testadas duas mensagens intrônicas detectadas apenas com orientação antisenso (*loci GAS6* e *ITGA6*) e duas onde transcritos antisenso e senso sobrepostos foram observados (*loci DNAJC3* e *KDELR2*) (Figuras 36A – D, Painéis “Intrônico”). Para cada *locus*, a abundância de mensagens exônicas codificantes para proteína correspondentes foi também quantificada (Figuras 36A – D, Painéis “Exônico”).

Foi confirmada a diminuição significativa ( $p < 0,005$ ) dos níveis do transcrito intrônico antisense de *GAS6* após o tratamento com andrógeno (Figura 36A, Painel “Intrônico”). Porém as mensagens codificantes de proteína do *locus GAS6*, que contêm os exons constitutivos 4 e 5 estão significativamente aumentadas ( $p < 0,025$ ) nas mesmas condições experimentais (Figura 36A, Painel “Exônico”). Para o *locus DNAJC3*, foi observado que as abundâncias dos RNAs intrônicos senso e antisense são moduladas em direções opostas com o tratamento com andrógeno (Figura 36B, Painel “Intrônico”). O aumento observado para o transcrito intrônico senso de *DNAJC3* corrobora o perfil observado nos experimentos de *microarray* (Figura 36B, Painel “Intrônico”), sendo observado comportamento similar para o transcrito codificante de proteína de *DNAJC3* (Figura 36B, Painel “Exônico”).

A alteração na abundância dos RNAs intrônicos de *KDELR2* e *ITGA6* também foi confirmada (Figura 36C e D, Painéis “Intrônico”). Para *KDELR2*, tanto os níveis do RNA antisense, quanto do senso intrônico foram aumentados pelo andrógeno ( $p < 0,0035$  e  $p < 0,02$ , respectivamente) (Figura 36C, Painel “Intrônico”). De maneira similar, o RNA intrônico antisense do *locus ITGA6* foi significativamente diminuído ( $p < 0,0012$ ) pelo tratamento (Figura 36D, Painel “Intrônico”). Entretanto, para esses últimos *loci* não foram detectadas alterações significativas na abundância dos respectivos transcritos codificantes de proteína, quando oligonucleotídeos iniciadores desenhados para medir exons constitutivos foram utilizados (Figura 36C e D, Painéis “Exônico”).

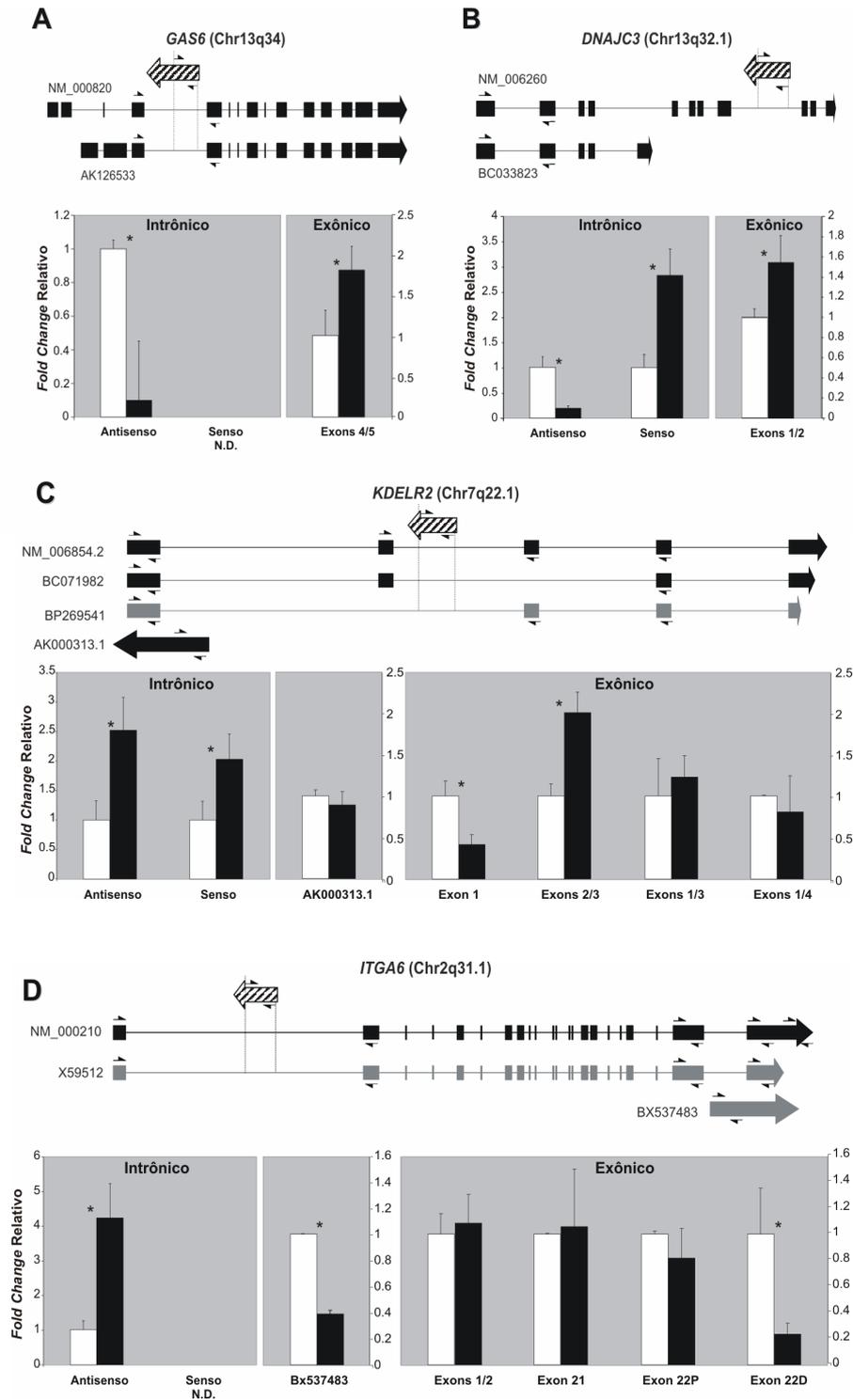


Figura 36 – Transcrição intrônica antisense pode regular a abundância e o processamento pós-transcricional de mRNAs codificantes de proteína do mesmo *locus*. (A - D) Níveis de expressão de transcritos intrônicos senso e antisense (Painéis “Intrônico”) ou de exons específicos de mensagens codificantes de proteína (Painéis “Exônico”) dos *loci* de *GAS6* (A), *DNAJC3* (B), *KDELR2* (C) ou *ITGA6* (D) medidos a partir de células tratadas com (barras pretas) ou células controle (barras brancas). mRNAs curados completos que mapeiam em cada *loci* são mostrados como caixas e setas pretas. mRNAs não curados e ESTs são mostrados como caixas e setas cinza. Transcritos intrônicos sem *splicing*, regulados por andrógeno, são mostrados como setas hachuradas. *Fold-change* Relativo (eixo Y) indica aumento ou diminuição dos níveis do transcrito após o tratamento com andrógeno em relação aos seus controles. Diferenças de expressão com significância estatística ( $p < 0,05$ ) estão indicadas com asterisco.

Em seguida, experimentos similares aos descritos acima foram desenhados, porém visando testar o possível envolvimento dos longos transcritos intrônicos antisense não-codificantes regulados por andrógeno no mecanismo de *splicing*, assim como já demonstrado anteriormente para um transcrito antisense super-expresso artificialmente (Yan *et al.*, 2005). Foi investigada a existência de correlação entre as abundâncias das mensagens intrônicas e os níveis de alguns mRNAs codificantes de proteína expressos a partir do mesmo *locus*, usando oligonucleotídeos iniciadores desenhados para exons regulados (ausentes em pelo menos uma das isoformas documentadas). Para o *locus KDELR2* foi observado que os exons 2/3 foram significativamente ( $p < 0,015$ ) aumentados pelo tratamento com andrógeno (Figura 36C, Painel “Exônico”). De fato, o exon 2 está ausente no transcrito de *KDELR2* representado pelo mRNA BP269541, sendo possível que o aumento observado nos níveis do RNA intrônico antisense do *locus KDELR2* após o tratamento induza a retenção do exon 2. Não foi observada nenhuma mudança significativa na abundância dos transcritos que contêm os exons 1/3 ou 1/4 (Figura 36C, Painel “Exônico”). Mas surpreendentemente, foi observada diminuição na abundância do exon 1 de *KDELR2* ( $p < 0,026$ ), quando a quantificação foi feita com oligonucleotídeos iniciadores desenhados apenas para esse exon (Figura 36C, Painel “Exônico”). Esse último resultado só pode ser explicado se ocorrer acúmulo após o tratamento com andrógeno de um transcrito ainda não caracterizado de *KDELR2* com a região 3' do exon 1 mais curta.

Para o *locus ITGA6*, a abundância das isoformas codificantes de proteína que contêm um exon terminal mais longo (exon 22D) foi diminuída ( $p < 0,04$ ) pelo tratamento (Figura 36D, Painel “Exônico”), sendo inversamente correlacionada com o aumento observado ( $p < 0,0012$ ) para o RNA intrônico antisense nas mesmas

condições (Figura 36C, Painel “Intrônico”). Entretanto, não é possível determinar se o resultado observado para o transcrito intrônico definitivamente causou a diminuição das isoformas de mRNA de *ITGA6* que contêm o exon 22D, pois foi também detectada diminuição de outra mensagen intrônica, naturalmente transcrita a partir do *locus ITGA6* (BX537483) que ainda possui uma sobreposição com a região 3' do exon terminal da mensagen codificante de proteína (Figura 36D).

#### **4.3.5 Aumento da ligação de AR a um motivo ARE a 5' do *asMYO5A***

A regulação da atividade transcricional pelo andrógeno envolve sua ligação direta com o receptor de andrógeno citoplasmático e a translocação do complexo formado para o núcleo, onde interage especificamente com regiões da cromatina, que possuem o motivo de ligação de andrógeno (ARE - *Androgen response elements*), localizados próximos das seqüências alvo (Nelson *et al.*, 2002). Visando agregar evidências de um mecanismo de controle transcricional direto atuando nos RNAs intrônicos regulados pelo andrógeno, foi utilizada uma abordagem bioinformática baseada na literatura (Nelson *et al.*, 2002), de busca de possíveis motivos AREs localizados na região intrônica, a 5' da região de mapeamento das mensagens intrônicas antisense (detalhes no item 3.1.4). Com essa abordagem foram identificados motivos ARE putativos para 24 dos 39 transcritos intrônicos identificados como regulados por andrógeno (Figura 37).

Para demonstrar experimentalmente a ligação do complexo andrógeno/receptor de andrógeno a um dos motivos ARE identificados para os RNAs intrônicos, foi utilizado o ensaio de imunoprecipitação da cromatina (*Chromatin Immunoprecipitation* – ChIP) com anticorpos reativos contra receptor de andrógeno (detalhes no item 3.8.1). Com essa metodologia foi possível constatar

que o aumento de abundância verificado por OSRT-qPCR da mensagem intrônica antisenso do *locus MYO5A (asMYO5A)* após o tratamento com andrógeno (Figura 38A), correlaciona-se diretamente com o aumento na detecção da ligação do receptor de andrógeno ao DNA (Figura 38B) em sua possível região promotora com motivo ARE.

GenBank Accession (EST)	Nome do Locus	Seqüência ARE	Posição	Localização Genômica	% Match
CK327134	NA	A G G A C A G A A T G T T T T	-2,056	2p23.2	83
BF333281	NIBP	A G A A G G T G C T G T T C A	-621	8q24.3	75
BF879099	GAS6	G G A A A A T C A T G T T C A	-1,528	13q34	75
AW860958	DST	A G A A C A T C T T T T T G C	-799	6p12.1	75
BF881469	ALMS1	G G A C C A G T T T T T T C T	-1,091	2p13.1	75
CK327189	P2RY14	A G A A C A G C A G G C T C T	-1,837	3q25.1	83
BF368584	DNAJC3	A G A A T A C C A T G T T G C	-1,608	13q32.1	75
BF768459	ACTN4	G G A A C A T G A A T T T C T	-785	19q13.2	75
BF768642	NA	T G A A C T G G G T T T T C T	+170	Xq27.1	75
BF882783	ACTN4	G G A A C C A C T G G T T C T	-1,839	19q13.2	75
CK327094	ATF2	T G A A C A G A C A C T T C T	-920	2q31.1	75
BE087892	STARD13	A G A T C A T G T G G T T C T	-2,239	13q13.1	83
CK327184	ITGA6	A G A A C A G G T T T T T G A	-2,788	2q31.1	75
BF926454	ADD3	T G A C C A T G C T G G T C T	-2,285	10q25.1	75
BF848956	MYO5A	A G T A C A T A A T T T A C T	-812	15q21.2	75
BF805741	PMF1	A G A A A T T C T T G C T C T	-2,104	1q22	75
AW805635	ERG	A G A A C T G C T T T G T C T	+246	21q22.2	75
BF350758	KDEL2	A G A A C A G C T G C T T C T	+219	7p22.1	83
BE156190	NA	A G A C C A G T T T T A T C T	-181	4q32.3	75
BE062809	UBE2V1	A G A A C A A F A T G T T A A	-2,786	20q13.13	83
AW819863	PPP2R2A	A G A G C A G G A A G C T C T	-881	17q23.3	75
BF332494	SAP18	A G A C C A G C C T G G T C A	-414	13q12.11	75
BF364131	NA	T G A C C A G G C T G G T C T	-2,481	16q23.2	75
CK327190	ZNF644	A G A A T C T A C T G T T C C	-283	1p22.2	75

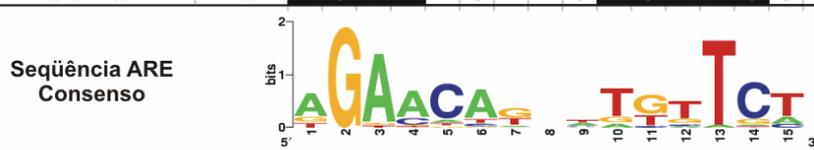


Figura 37 – Identificação *in silico* de motivos ARE em regiões a 5' de RNAs intrônicos regulados por andrógeno. Seqüências de DNA genômico (3 kb) localizado a 5' dos sitios postulados de início de transcrição de cada um dos 39 RNAs intrônicos regulados por andrógeno identificados foram utilizadas em uma busca pelo consenso AGAACAnnnTGTTCT, característico de elementos de resposta a andrógeno (ARE). GenBank Accession esta indicado para os 24 transcritos onde os motivos ARE foram encontrados. Alinhamentos dos 24 AREs identificados foi realizado utilizando CLUSTALW, e os resíduos conservados estão mostrados em preto (**superior**). Uma representação gráfica da seqüência consenso dos 24 AREs, com a freqüência de cada base sendo proporcional ao tamanho do caractere foi gerada utilizando WEBLOGO (Crooks *et al.*, 2004).

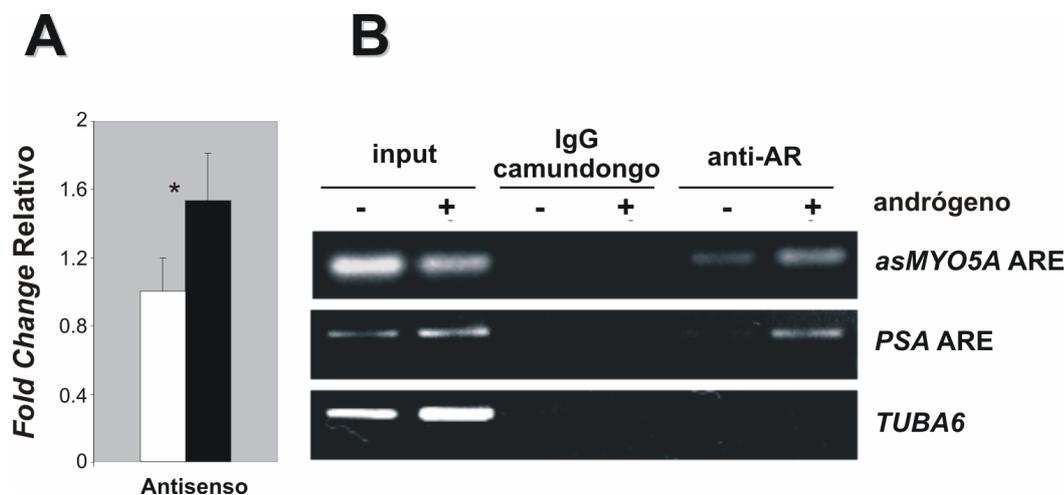


Figura 38 – **Aumento na abundância do transcrito intrônico antisense *MYO5A* está correlacionada com a ligação do complexo andrógeno/receptor de andrógeno ao ARE postulado.** (A) Níveis de expressão do transcrito intrônico antisense *MYO5A* foram medidos em células tratadas com andrógeno (barras pretas) ou células controle (barras brancas). *Fold-change* Relativo (eixo Y) indica o aumento ou diminuição dos níveis do transcrito após o tratamento com andrógeno em relação aos seus controles. Diferença de expressão apresentou significância estatística ( $p < 0,05$ ) assim como indicado pelo asterisco. (B) Extratos celulares de culturas tratadas com andrógeno (+) ou controles (-) foram utilizadas em ensaios de ChIP na presença de anticorpos contra a proteína do receptor do andrógeno (anti-AR). Reações de PCR foram utilizadas para detecção do DNA imunoprecipitado contendo o ARE para transcrito intrônico antisense *MYO5A* (*asMYO5A* ARE). O ARE do *Prostate Specific Antigen* ARE (*PSA* ARE), e uma região do gene da tubulina alfa 6 não relacionada com o motivo ARE (*TUBA6*) foram utilizados como controles positivos e negativos respectivamente. Um anticorpo não-relacionado (IgG camundongo) foi utilizado como controle para imunoprecipitação inespecífica. A quantidade inicial de DNA utilizada relativa a cada ensaio de ChIP é mostrada (input).

#### 4.4 Transcritos intrônicos sintênicos com baixa conservação expressos diferencialmente em tecidos humanos e de camundongos.

A observação de RNAs intrônicos com envolvimento em câncer (Reis *et al.*, 2004), assim como de mensagens dessa classe sendo reguladas por estímulos hormonais (Louro *et al.*, 2006)<sup>5</sup>, aumentou ainda mais o interesse no estudo dessas moléculas. Entre as estratégias utilizadas para inferência de funcionalidade para novas seqüências, as mais utilizadas são a avaliação da especificidade de expressão em diferentes tecidos, e as análises de genômica comparativa em

<sup>5</sup> Louro *et al.* 2006 (Artigo submetido). ANEXO 7 - incluso no CD-ROM suplementar a esta Tese.

diferentes organismos. Assim, utilizamos a análise comparativa entre homem e camundongo como uma abordagem alternativa. Foram obtidos os resultados descritos abaixo, que identificam um conjunto de mensagens intrônicas com perfil de expressão espacialmente conservado em humanos e camundongos, o que sugere funções tecido-específicas desempenhas por esses RNAs; contudo, essas funções ainda necessitam ser determinadas.

#### **4.4.1 Detecção de transcritos intrônicos em tecidos humanos e de camundongos.**

A plataforma de *microarray* de cDNA enriquecida em transcritos intrônicos humanos (detalhes da plataforma no item 3.5.1) foi utilizada na identificação de transcritos intrônicos não-codificantes funcionalmente conservados entre humanos e camundongos. A observação que mensagens não-codificantes para proteína, mesmo sem altos níveis de conservação de identidade entre diferentes espécies, podem desempenhar importantes funções celulares (Numata *et al.*, 2003; Ravasi *et al.*, 2006) embasaram o desenho experimental desse estudo (descrito no item 3.5.4). Amostras de RNA de três tecidos diferentes (fígado, rim e próstata) de espécimes humanos e camundongos foram hibridizados independentemente contra a plataforma de *microarray* de cDNA humana.

De todas as sondas intrônicas depositadas na plataforma, foram identificadas 63 (7,7 %) e 21 (2,6 %) expressas em pelo menos um dos tecidos de humanos ou camundongos, respectivamente. A grande maioria (90%) dos transcritos detectados em camundongos foram também detectados como expressos em humanos. De maneira similar, das 2.292 sondas exônicas foram observadas 212 (9.2 %) e 65 (2.8 %) expressas em pelo menos um tecido humano ou de camundongo, com uma

sobreposição de 94 % dos transcritos detectados em camundongo com os detectados em humano.

#### **4.4.2 Expressão diferencial de transcritos intrônicos em tecidos humanos e de camundongos.**

Os valores de intensidade normalizados, provenientes dos experimentos realizados independentemente com amostras de tecido de humanos e camundongos foram analisados estatisticamente para identificação de mensagens diferencialmente expressas (detalhes no item 3.1.3). Dos 275 transcritos detectados nos tecidos humanos 227 (82%) foram identificados como diferencialmente expressos por ANOVA, compreendendo 51 transcritos intrônicos e 176 exônicos. Proporções similares foram encontradas nas análises com os tecidos de camundongo, onde 87% (75/86) dos transcritos expressos foram identificados por ANOVA como diferencialmente expressos. Entre esses, um total de 17 provenientes de sondas intrônicas e 58 de exônicas. Importante notar que 100 % das mensagens intrônicas, assim como 93% das exônicas diferencialmente expressas em tecidos de camundongo também foram detectadas como diferencialmente expressas nos tecidos humanos.

Para deixar a abordagem mais estrita foi também realizada uma análise utilizando SAM (Tusher *et al.*, 2001). Com as mensagens intrônicas expressas, a análise por SAM revelou 17 diferencialmente expressas compartilhadas por humanos e camundongos, sendo 14 mais expressas em fígado, duas mais expressas em rim e apenas uma considerada mais expressa em próstata humana e em fígado de camundongo. Similarmente, com as mensagens exônicas, 54 transcritos compartilhados por humanos e camundongos foram detectados como

diferencialmente expressos, sendo 42 mais expressos em fígado, oito mais expressos em rim, um em próstata, e ainda três mais expressos em rim humano e em fígado de camundongo. Mais uma vez, foi possível notar a alta proporção de mensagens (~94%) com padrão de expressão similar nas duas espécies analisadas.

Uma análise de *hierarchical clustering* combinada foi realizada com as listas de transcritos com diferença de expressão, obtidas de maneira independente para cada espécie. Assim, foi possível identificar uma assinatura de expressão tecido-específica independente da espécie analisada, como pode ser observado pelo agrupamento dos perfis de expressão pelos tecidos e não por espécie (Figura 39). A Tabela Suplementar 3 (ANEXO 8) apresenta os transcritos identificados nessa análise.

#### **4.4.3 Validação da existência e orientação transcricional de mensagens intrônicas sintênicas diferencialmente expressas em tecidos humanos e de camundongos.**

Para confirmar a existência e determinar a orientação transcricional das mensagens intrônicas cuja expressão preferencial em tecidos parece ser conservada entre as espécies estudadas, foram selecionados 10 transcritos intrônicos diferencialmente expressos em ambas espécies. A seqüência conhecida de cDNA humano foi alinhada contra o genoma humano utilizando BLAT (*UCSC Genome Browser*), para obtenção da região sintênica intrônica no genoma de camundongo. Esse conjunto de 10 regiões intrônicas sintênicas foi caracterizado por transcrição reversa com orientação específica seguida de PCR (OSRT-PCR) com RNA extraído do fígado de cinco camundongos. Como resultados preliminares, foram identificados dois transcritos antisense, quatro senso, e quatro pares de transcritos antisense e senso expressos nas regiões intrônicas sintênicas dos 10 *loci* testados.

Com a mesma abordagem, foi determinada a orientação transcricional de 5 dessas mensagens intrônicas humanas, assim como apresentado figura 40, ao lado do resultado confirmado para as mensagens sintênicas de camundongo. A determinação da orientação transcricional foi realizada com RNA extraído de células da linhagem HepG2, de tumor de fígado humano (Figura 40B).

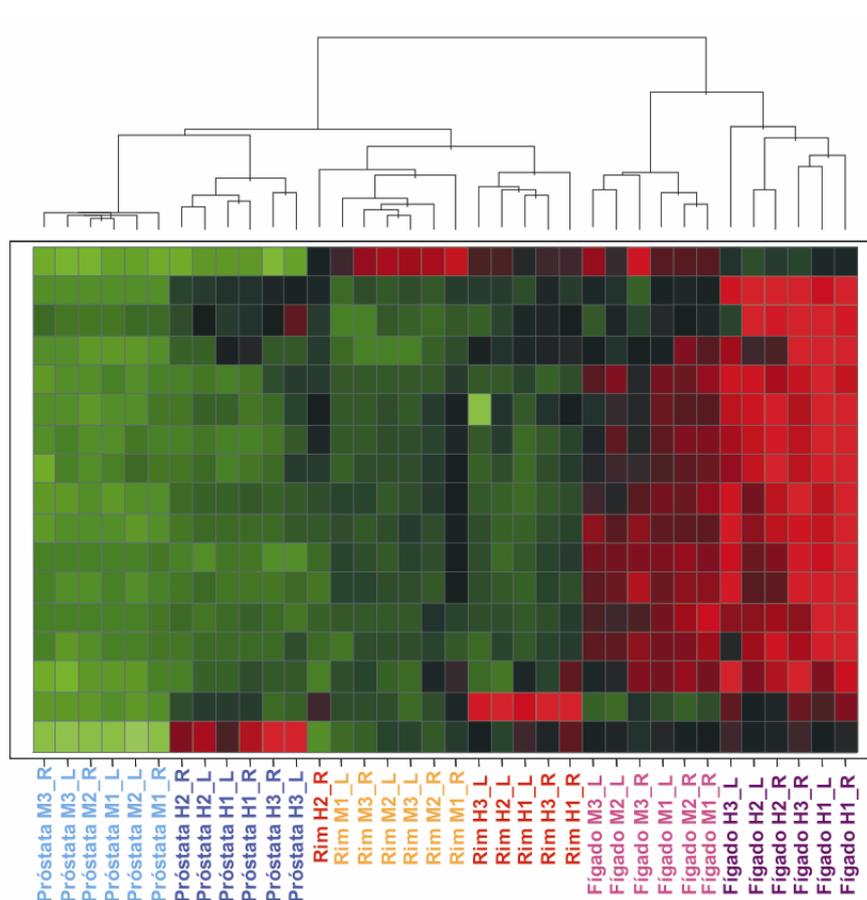


Figura 39 – **Assinatura de expressão intrônica específica de tecidos e comum entre humanos e camundongos.** *Hierarchical clustering* (Sanchez-Carbayo *et al.*, 2002) dos transcritos intrônicos independentemente identificados como diferencialmente expressos em tecidos humanos e de camundongos. Rótulos coloridos abaixo do painel indicam a espécie (H – humano, M – camundongo), o tecido (próstata, rim ou fígado) e a réplicas (1, 2 ou 3; L – lado esquerdo da lâmina, R – lado direito da lâmina) das hibridizações realizadas. O padrão de expressão de cada mensagem está representado pelo número de desvios-padrão acima (vermelho) ou abaixo (verde) do valor médio da expressão da mensagem em todas as amostras.

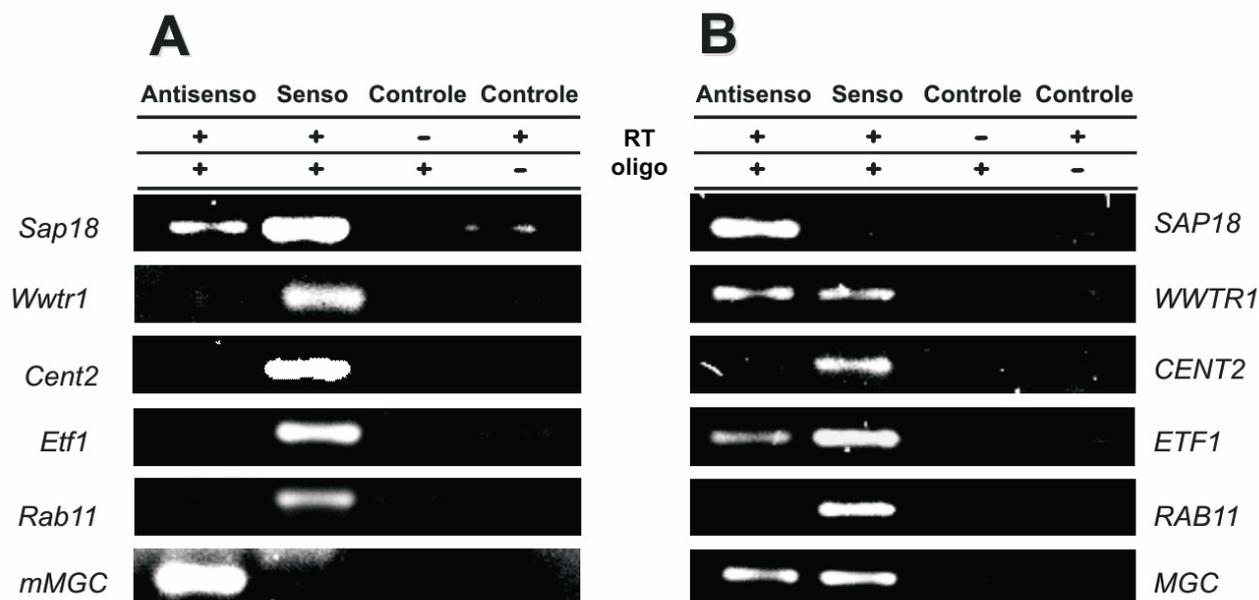


Figura 40 – **Determinação da orientação transcripcional das mensagens intrônicas expressas diferencialmente em tecidos humanos e de camundongos.** (A e B) Transcrição reversa orientação-específica seguida de PCR (OSRT-PCR) foi realizada em um conjunto selecionado de 5 mensagens intrônicas diferencialmente expressas em tecidos humanos e de camundongos, identificadas pelas análises de *microarray*, para determinação da orientação da mensagem intrônica relativa a mensagens codificante para proteína de mesmo *locus*. Nomes a esquerda se referem aos mRNAs RefSeq codificante para proteína que mapeiam nos mesmos *loci* e definem as regiões intrônicas onde atividade transcripcional foi identificada. (A) Produtos de PCR gerados a partir de amostras de fígado de camundongo e separados em gel de agarose 3% detectaram transcrição antisenso (*mMGC*), senso (*Wwtr1*, *Cent2*, *Etf1* e *Rab11*) ou antisenso/senso simultânea (*Sap18*). (B) Produtos de PCR gerados a partir de células da linhagem HepG2 de fígado humano e separados em gel de agarose 3% mostram transcrição antisenso (*SAP18*), senso (*CENT2* e *RAB11*) ou antisenso/senso simultânea (*WWTR1*, *ETF1* e *MGC4707*). Os controles negativos, que avaliam a ausência de contaminação com DNA genômico (–RT) e de auto-anelamento do RNA (– oligo) durante a transcrição, estão indicados.

Surpreendentemente, para as mensagens provenientes dos *loci* humanos foi observada correspondência total na orientação com a observada para os transcritos de camundongo apenas para as mensagens intrônicas senso de *CENT2* e *RAB11* (Figura 40A e B). Apesar de também terem sido detectados transcritos antisenso para os *loci* de *SAP18* e *MGC4707*, um senso correspondente foi apenas observado em humano para o *locus MGC*, e em camundongo apenas para o *Sap18* (Figura 40A e B). E ainda, os pares de mensagens antisenso/senso para *WWTR1* e *ETF1*

observados são exclusivos de fígado humano, pois em camundongo *Wwtr1* e *Etf1* apresentaram apenas mensagens intrônicas senso (Figura 40A e B).

#### **4.5 Interação de transcritos intrônicos com proteínas ligadoras de RNA.**

Tendo como hipótese que essa nova classe de mensagens não-codificantes, os Longos RNAs Intrônicos, identificada e extensivamente estudada nesse projeto seja biologicamente relevante e não apenas um tipo de ruído transcricional, foi investigada a existência de associação entre transcritos intrônicos e proteínas ligadoras de RNA (RBPs – *RNA Binding Proteins*). Essas proteínas estão amplamente descritas na literatura como essenciais na regulação pós-transcricional de RNAs, através de processamento do RNA mensageiro imaturo (*splicing*), transporte e estabilidade do mRNA, assim como de sua tradução (Siomi & Dreyfuss, 1997). Assim, os resultados descritos a seguir apesar de ainda necessitarem de experimentos confirmatórios, já apresentam evidências que sustentam o envolvimento dessas mensagens em atividades regulatórias no programa transcricional humano.

##### **4.5.1 Expressão das proteínas ligadoras de RNA em células 293T**

Após as transfecções com as construções para expressão episossomal de RBPs (ver item 3.2.3), as células 293 T foram coletadas e lisadas para permitir a extração de RNA associado às proteínas ligadoras de RNA em estudo (ver item 3.3.3). Uma alíquota com cerca de 100µl dos lisados foi utilizada para quantificação protéica através da metodologia de Bradford (Harlow & Lane, 1998), assim como para a verificação da expressão das RBPs específicas por *Western Blot*.

Os lisados celulares de 293T produzidos apresentaram tipicamente concentrações de proteína total bastante similares (8-10mg/ml), independente da RBP expressa, não sendo necessário qualquer tipo de normalização precedente as imunoprecipitações. As quantidades similares de proteína podem ser confirmadas na figura 41. Todo o conteúdo das amostras de RNA derivadas das imunoprecipitações (ver item 3.3.3) foi utilizado na geração de alvos fluorescentes marcados (ver detalhes no item 3.5.2), assim como toda a quantidade de alvo fluorescente obtida foi utilizada nas hibridizações (ver item 3.5.4).

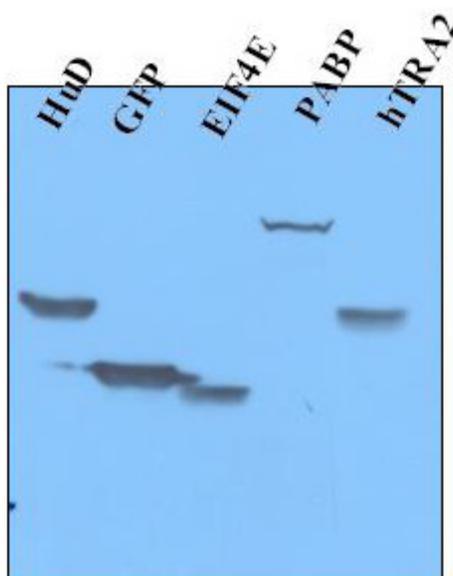


Figura 41 – **Expressão transitória das proteínas ligadoras de RNA (RBPs)**. Western Blot para detecção da expressão das RBP em 293T. Um volume de lisado celular referente à cerca de 50µg de proteína total de cada proteína expressa (RBPs e o controle GFP) separado por SDS-PAGE, 10% de acrilamida foi transferido para uma membrana de nitrocelulose, para detecção da proteína de fusão com anti-proteína A. **Poço 1:** HuD (40kD); **Poço 2:** GFP (27kD); **Poço 3:** EIF4E (24kD); **Poço 4:** PABP (72kD); **Poço 5:** h-TRA2 (40kD). Massas moleculares indicadas referentes às proteínas com o *tag* de proteína A.

#### 4.5.2 Identificação, utilizando *microarrays* de cDNA, de RNAs longos intrônicos não-codificantes associados a proteínas ligadoras de RNA

Alvos fluorescentes imunoprecipitados e marcados foram hibridizados com *microarrays* intrônicos (ver item 3.5.4). Após as análises para determinação das

mensagens especificamente co-imunoprecipitadas (ver item 3.5.5) foi possível detectar 296, 201, 175 e 122 transcritos que foram imunoprecipitados por estarem presentes nas células superexpressando PABP, HUD, hTRA2 e EIF4E, respectivamente.

De modo a tentar caracterizar o conjunto de transcritos intrônicos candidatos a ligação com RBP, foi feita uma análise de enriquecimento de genes pertencentes a uma determinada categoria de ontologia gênica (GO), utilizando o programa eGOn v1.0 (Explore Gene Ontology - <http://www.genetools.no/>). Algumas categorias apresentaram-se como significativamente enriquecidas em comparação com todos os transcritos intrônicos presentes em nossa plataforma de *microarrays* (Tabela Suplementar 4 – ANEXO 9).

#### **4.6 Análise de expressão de RNAs intrônicos em células LNCaP tratadas com $\alpha$ -amanitina**

Para ampliar o estudo da biogênese dos transcritos intrônicos iniciado com a identificação de RNAs regulados por andrógeno (Louro *et al.*, 2006), foram realizados experimentos de análise de expressão gênica com células de próstata tumoral LNCaP tratadas com  $\alpha$ -amanitina. Essa molécula é um octapeptídeo bicíclico extraído do cogumelo *Amanita phalloides* que inibe fortemente a atividade da RNA polimerase II (RNAP II) impedindo sua translocação sobre o DNA durante a elongação da transcrição (Bushnell *et al.*, 2002). Como tem sido apenas postulado o envolvimento da RNAP II na transcrição de muitos RNAs não codificantes, esse estudo visou testar se essa é a polimerase responsável pela síntese da maioria dos longos RNAs intrônicos que vêm sendo identificados pelas análises descritas.

Um conjunto de experimentos piloto (não apresentados aqui) utilizou a plataforma de *microarrays* de cDNA, tendo sido hibridizadas duas amostras tratadas e duas controle; uma parcela considerável dos transcritos intrônicos cuja expressão foi detectada em LNCaP, apresentaram sua expressão significativamente aumentada na células tratadas com o inibidor da RNAP II. Esse inesperado aumento na abundância dos transcritos intrônicos está em concordância com resultado observado na literatura para algumas mensagens codificadoras de proteínas humanas (Kravchenko *et al.*, 2005), e apesar de sugerir várias hipóteses, ainda necessita de experimentação para uma melhor compreensão do fenômeno. Entretanto, deixa claro que a RNAP II não está envolvida na transcrição das mensagens intrônicas onde esse fenômeno foi observado.

#### **4.6.1 Uma grande fração de RNAs intrônicos aparentemente não é transcrita pela RNAP II**

Para ampliar e validar as observações iniciais obtidas com a plataforma de cDNA onde estão representados 822 mensagens intrônicas, foram realizadas novas hibridizações com RNA de células LNCaP tratadas com  $\alpha$ -amanitina, agora usando uma plataforma de *microarrays* de oligonucleotídeos desenhada pelo grupo (descrita no item 3.11.1) que possui 13.004 sondas intrônicas.

Como pode ser observado na figura 42, cerca de 39 % (3.909) das mensagens exônicas expressas foram significativamente afetadas pela inibição da RNAP II enquanto o restante, provavelmente mRNAs mais estáveis, não tiveram sua abundância significativamente alterada. Como esperado, a grande maioria das mensagens codificantes para proteínas (exônicas) alteradas pelo tratamento (96 %) tiveram seus níveis diminuídos, enquanto apenas 4 % das mensagens codificantes

para proteínas foram aumentadas (Figura 42A), resultado similar ao já observado em células HeLa (Kravchenko *et al.*, 2005).

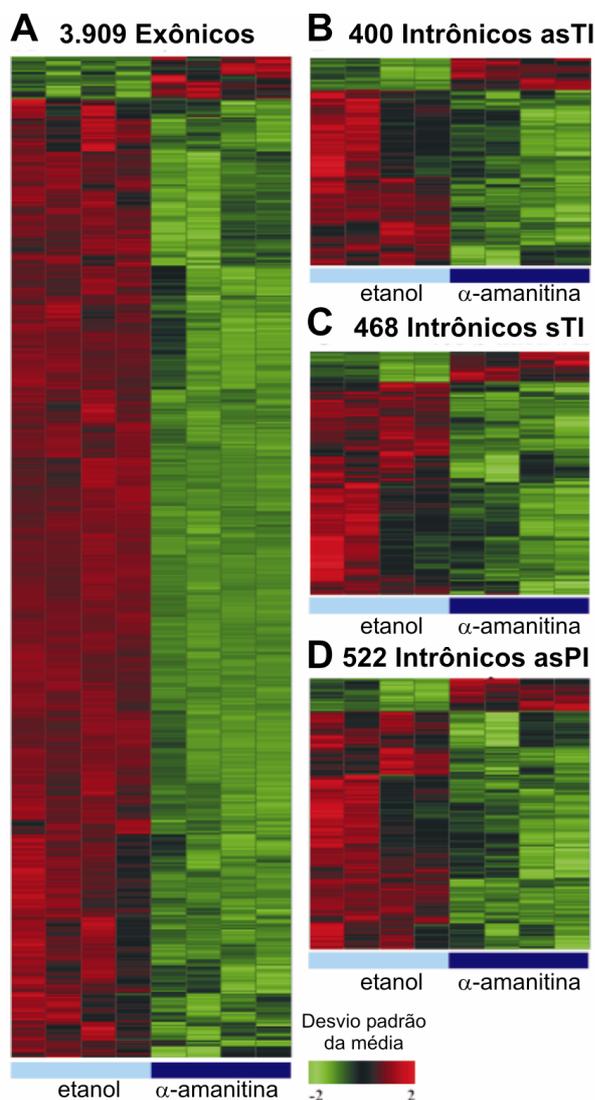


Figura 42 – Efeito do tratamento com  $\alpha$ -amanitina sobre a expressão de transcritos intrônicos e exônicos. Alteração na expressão após tratamento com  $\alpha$ -amanitina (azul escuro) de 3.909 transcritos codificadores de proteínas (exônicos) (A), 400 transcritos totalmente intrônicos antisense (asTI) (B), 468 transcritos totalmente intrônicos senso (sTI) (C), e 522 transcritos parcialmente intrônicos antisense (asPI) (D). Todos os transcritos cujos padrões de expressão estão representados foram identificados como diferencialmente expressos pelas análises estatísticas combinadas, descritas no texto, em amostras tratadas com  $\alpha$ -amanitina (azul escuro) em relação ao controle (veículo etanol, azul claro). Cada linha representa um transcrito, e dentro de cada amostra (separadas pelas barras coloridas indicadas) as colunas indicam as réplicas.

Frações bastante menores de transcritos expressos totalmente intrônicos antisenso (asTI, 16 %) e senso (sTI, 18 %), e transcritos parcialmente intrônicos antisenso (asPI, 19%) foram afetados pela  $\alpha$ -amanitina (Figura 42B-D). Surpreendentemente, entre os transcritos cujos níveis foram alterados nas células tratadas, 16 % (62/400), 13 % (59/468) e 12 % (64/522) de asTI, sTI e asPI foram, respectivamente, aumentados. A Tabela Suplementar 5 (ANEXO 10) apresenta todos os transcritos cuja expressão foi afetada pelo tratamento com  $\alpha$ -amanitina.

#### **4.6 Sistema para modulação da expressão de *asRASSF1* e *asPPHLN1* em células humanas**

A determinação das funções celulares específicas desempenhadas pelos Longos RNAs intrônicos não-codificantes identificados pelas diferentes abordagens descritas até aqui é o próximo passo na caracterização dessa classe de ncRNAs. Utilizando a estratégia descrita no item 3.9.6 já foi construído um sistema de expressão para os Longos RNAs intrônicos não-codificantes correlacionados com o grau de diferenciação tumoral de próstata *asRASSF1* e *asPPHLN1* (Reis *et al.*, 2004). Esse sistema, esquematizado na figura 14, será utilizado para superexpressar os transcritos intrônicos em estudo, e observar as alterações celulares desencadeadas.

## 5 DISCUSSÃO

### 5.1 Identificação e caracterização parcial de *RASL11A* e *ARL-9*, membros representativos de novas famílias de pequenas GTPases monoméricas.

Analisando os resultados obtidos com os alinhamentos das proteínas Rasl11 e Arl-9/Arl10 com alguns membros representantes da superfamília de pequenas GTPases monoméricas humanas e com seus ortólogos, foi possível notar que se tratam de novas subfamílias que apesar de apresentarem as principais características da superfamília, possuem certas peculiaridades. Essas características diferenciais que permitem sua separação em ramos distintos estão mantidas em muitos eucariotos, sugerindo que essas proteínas são membros de famílias gênicas evolutivamente conservadas.

A análise da seqüência do produto do gene *ARL-9*, descrito aqui, mostra que o domínio G, incluindo o *Interswitch Toggle* típico das proteínas Arf, está apenas representado de forma completa se for utilizado como iniciador de tradução um códon alternativo não-AUG. O uso de códons de iniciação alternativos, no caso CUG, já foi documentado para genes humanos funcionais, incluindo proto-oncogenes, fatores de transcrição e fatores de proliferação (Boeck & Kolakofsky, 1994; Tee & Jaffe, 2001). Em todos os casos estudados, esses códons aparecem em um contexto similar, o qual é caracterizado pela presença de uma purina na posição -3 e uma guanosina (mais raramente uma citosina) na posição +4 (Boeck & Kolakofsky, 1994). Essas características de seqüência estão também presentes na ORF que se inicia com o códon alternativo CUG no transcrito de *ARL-9*. Contudo, a confirmação de seu uso, assim como sua significância biológica só pode ser verificada através de experimentação bioquímica mais aprofundada.

A proteína Rasl11a, identificada, clonada e parcialmente caracterizada durante esse projeto apresenta na região do *switch* 1 do domínio G um resíduo de Asn na posição 59 de sua seqüência primária, ao invés de um resíduo de Thr. Essa modificação é observada também na proteína Rasl11b, em todos os ortólogos de Rasl11, mas não é restrito a essa subfamília, visto que Ris, uma família recentemente descrita de proteínas relacionadas à Ras, apresenta a mesma substituição (Wistow *et al.*, 2002). Na proteína H-Ras, membro prototípico da superfamília, o resíduo Thr35 no *switch* 1, em conjunto com a Gly60 no *switch* 2, contribuem para a estabilização da forma ligada a GTP através de pontes de hidrogênio entre os NH da cadeia principal e o oxigênio do fosfato  $\gamma$  do GTP (Vetter & Wittinghofer, 2001a). É possível que essa substituição de Thr para Asn nas proteínas Rasl11, preservando o caráter polar não carregado da cadeia lateral, não interfira significativamente com a estabilização do GTP e nem com a atividade GTPásica das proteínas. Estudos com mutantes de proteínas Ras revelaram que uma troca do resíduo Thr35 por um resíduo de Ser, nas proteínas Ras, reduz a atividade GTPásica intrínseca, mas mantém a capacidade de ligação com alta afinidade aos seus efetores (Spoerner *et al.*, 2001). Contudo, apenas uma caracterização bioquímica detalhada, inclusive dos efetores ainda desconhecidos das proteínas Rasl11, pode determinar as conseqüências funcionais da substituição de Thr para Asn observada.

Pequenas GTPases relacionadas a Ras geralmente apresentam na região carboxi-terminal seqüências que sinalizam para modificações que permitem sua associação com a membrana plasmática. Essa associação é freqüentemente indispensável para sua funcionalidade biológica. As proteínas Rasl11 não apresentam nenhum sinal já descrito de reconhecimento para adição de lipídeos na

região C-terminal, o que é encontrado nas proteínas Ras (Reuther & Der, 2000a), tais como sinais de farnesilação ou palmitoilação (Dai *et al.*, 1998; Sinensky, 2000). Entretanto, uma assinatura conservada de resíduos polibásicos foi encontrada, nas análises aqui descritas, em todos os ortólogos de vertebrados de Rasl11, sugerindo que essas proteínas apresentam um mecanismo de ligação à membrana plasmática baseado em resíduos carregados semelhante ao postulado para Rin e Rit, membros da família Ras humana (Lee *et al.*, 1996; Sinensky, 2000).

O tamanho de um dos transcritos observados por *Northern Blot* (~1,6kb) para *RASL11A* apresenta concordância com a seqüência completa obtida por RACE-PCR (1.543bp). A outra forma de menor tamanho (~1,2kb) poderia apenas ser explicada por iniciação ou termino alternativo de transcrição, o que ainda carece de confirmação experimental.

As análises quantitativas de PCR em tempo real revelaram que o transcrito de *RASL11A* está em quantidade significativamente menor em tumores de próstata, quando comparado com sua abundância em amostras de tecido prostático normal adjacente ao tumor. Esses resultados sugerem para esse novo membro humano da superfamília de pequenas GTPases monoméricas um papel de supressor tumoral. Existem diversos trabalhos na literatura que mostram que pequenas GTPases podem agir como supressoras tumorais. Perda de heterozigoze no gene *NOEY/AHR1* foi detectada em 41% dos cânceres de ovário e mama (Yu *et al.*, 1999b). Ainda nesse mesmo trabalho foi verificado que a re-expressão de epissomal Noey inibe a proliferação das células tumorais de ovário e mama, indicando que esse membro da família Ras tem atividade supressora tumoral. Superexpressão de Rig é capaz de inibir a transformação celular mediada por Ras em células NIH 3T3 (Ellis *et al.*, 2002b). A co-expressão de Rap1a é capaz de reverter a instabilidade

genômica induzida por H-Ras oncogênica (Wani *et al.*, 1997a). E recentemente foi demonstrada que a superexpressão de Rerg em células de carcinoma de mama inibe sua proliferação *in vitro* e também a formação de tumor em camundongos *nude* (Finlin *et al.*, 2001). O fato de Rasl11a ser bastante similar a Rerg aumenta as evidências do papel supressor tumoral do gene *RASL11A* em câncer de próstata.

Finalmente, o conjunto destes resultados aponta para uma nova subfamília de pequenas GTPases monoméricas, Rasl11, que pode apresentar um importante papel na supressão de tumores. Dessa forma a geração de anticorpos contra as proteínas dessa subfamília, assim como uma caracterização bioquímica de seu mecanismo de ativação, iniciadas durante esse projeto com a obtenção de proteínas heterólogas em um sistema em bactérias, devem contribuir no entendimento das funções dessas proteínas durante o surgimento de tumores, assim como na fisiologia de células normais.

## **5.2 Transcritos intrônicos marcadores de prognóstico em câncer de próstata.**

Os resultados obtidos no estudo de expressão gênica em câncer de próstata por *microarrays* de cDNA sugerem que a maioria das regiões intrônicas de genes codificantes de proteína onde estão mapeadas ESTs, ao longo de todo o genoma humano, são de fato transcricionalmente ativas, assim como as observadas em estudos específicos com os cromossomos 21 e 22 (Kapranov *et al.*, 2002; Rinn *et al.*, 2003; Kampa *et al.*, 2004). A observação de longos RNAs antisense sem *splicing* nessa parte de nosso projeto é inteiramente concordante com resultados prévios de outros grupos, que mostram mensagens de 0,4–1,3 kb antisense sem *splicing* e totalmente intrônicas transcritas a partir do cromossomo 22 (Rinn *et al.*, 2003).

Nas amostras de RNA poli-(A)+ foram detectados simultaneamente pares de transcritos intrônicos antisenso e senso sobrepostos que são correlacionados com o grau de diferenciação tumoral em seis diferentes *loci*. Essas mensagens antisenso totalmente intrônicas provavelmente tem como alvo a seqüência senso intrônica complementar presente no pré-mRNA. Dessa forma podem desempenhar suas funções regulatórias de maneira seqüência-específica no processamento das mensagens codificantes de proteína correspondentes.

Pares de transcritos senso/antisenso, assim como mensagens antisenso totalmente intrônicas, têm sido detectados também em camundongos através de análises de mapeamento em escala genômica (Kiyosawa *et al.*, 2003). Esses transcritos têm sido implicados nos fenômenos de *imprinting* e inativação do cromossomo X em mamíferos (Lee *et al.*, 1999; Sleutels *et al.*, 2000). RNAs antisenso também têm sido demonstrados como mediadores do silenciamento gênico e metilação na doença autossômica alfa-talassemia (Tufarelli *et al.*, 2003). Ainda mais importante, trabalhos antigos da literatura mostram que alguns transcritos antisenso intrônicos agem como efetores negativos de genes relacionados com câncer, tais como *cMyc* (Nepveu & Marcu, 1986) e *p53* (Khochbin & Lawrence, 1989), e como efetores positivos de *RPS14* (Tasheva & Roufa, 1995). Assim, esses pares de transcritos antisenso e senso devem fazer parte de um mecanismo geral de regulação da expressão gênica em eucariotos ainda não completamente compreendido.

De fato, através de uma abordagem de mapeamento de sítios para ligação de fatores de transcrição (TFBS), sem viés para regiões codificantes para proteína, nos cromossomos 21 e 22 foi revelado que 36% dos TFBS para Sp1, *cMyc* e *p53* são

significativamente correlacionados com RNAs não-codificantes, e os pares de mensagens geralmente são co-regulados (Cawley *et al.*, 2004).

Tanto os níveis do transcrito intrônico antisense, quanto os do senso de *RASSF1* estão significativamente diminuídos em tumores com alto GS. Ainda, foi observado que a fração de mensagens correlacionadas com o grau de diferenciação tumoral com relação a todas expressas em próstata, é consideravelmente maior de transcritos intrônicos (6.6%, [23/346]), quando comparada à fração de transcritos que mapeiam em exons de *RefSeq* (2%, [27/1,369]). Assim, da mesma forma que evidências apontando para uma maior complexidade do programa transcricional humano continuam sendo acumuladas, também se acumulam as perguntas a respeito da relevância biológica dos RNAs não-codificantes, que aparentemente respondem pela grande maioria dos transcritos. A correlação com o estágio de progressão tumoral observada nessa parte do projeto é claramente um forte argumento contra a hipótese de simples ruído transcricional. Esses resultados apresentados também apontam fortemente para a necessidade de inclusão de sondas para RNAs não-codificantes, principalmente os expressos a partir de regiões intrônicas, na busca de metodologias alternativas para diagnóstico e prognóstico molecular de câncer, até então exclusivamente focadas em produtos de origem protéica.

Finalmente, um paradigma corrente em oncogênese aponta diretamente para o acúmulo de mutações em fases abertas de leitura de mensagens codificantes para proteínas, que possuem tanto comportamento de oncoproteínas quanto de supressoras tumorais, como causas do surgimento e desenvolvimento de câncer (Mcmanus, 2003). O envolvimento de vários tipos de RNAs não-codificantes, como os microRNAs (Mcmanus, 2003; Calin *et al.*, 2004), em câncer vem se tornando

mais evidente. Assim a identificação de transcritos intrônicos correlacionados com o surgimento ou desenvolvimento de neoplasias, apresentado aqui pela primeira vez na literatura, pode e deve aumentar definitivamente a complexidade desse paradigma.

### **5.3 Transcritos intrônicos regulados pelo estímulo de andrógeno.**

O resultado obtido a partir das análises de GO das mensagens que responderam ao tratamento é mais uma observação altamente sugestiva que os RNAs intrônicos não são um mero ruído transcricional. Na verdade, aponta para possíveis funções desempenhas por essas mensagens durante a expressão gênica desencadeada com a exposição celular ao andrógeno. Essas mensagens provavelmente agem na regulação pós-transcricional de genes envolvidos na transdução de sinal em células de próstata. E de fato, alguns desses *loci* já haviam sido demonstrados como envolvidos em eventos celulares relacionados ao crescimento e diferenciação de células da próstata. Por exemplo, a proteína codificada pelo mRNA *GAS6* é um ligante do receptor tirosina quinase *Axl*, e o complexo *Gas6/Axl* apresenta atividade mitogênica em linhagens celulares humanas de câncer de próstata, modulando a via de transdução de sinal de *PI3K/AKT* e *MEK* (Sainaghi *et al.*, 2005). De maneira similar, o fator de transcrição codificado por *ERG*, que é membro da família *ETS* de genes responsáveis pela integração dos sinais que regulam o crescimento e a diferenciação celular, respostas ao estresse, e tumorigênese, foi anteriormente identificado como o proto-oncogene cujos níveis encontram-se aumentados com maior frequência em células epiteliais malignas de próstata (Petrovics *et al.*, 2005).

O envolvimento de RNAs não-codificantes em eventos de processamento pós-transcricional em moléculas de mRNA codificantes para proteína já tem sido demonstrado. *HBII-52*, um pequeno RNA nucleolar (snoRNA) foi descrito como um elemento que atua em *trans* na regulação do processamento do transcrito *5-HT2CRA*, codificante para o receptor de serotonina (Kishore & Stamm, 2006). Alteração no processamento ocorre devido à ligação do *HBII-52* a um elemento de silenciamento do exon Vb presente no pré-mRNA de *5-HT2CRA*, gerando uma mensagem defeituosa induzida pelo snoRNA, causando a síndrome congênita de Prader-Willi (Kishore & Stamm, 2006).

Em outro trabalho, Yan e seus colaboradores identificaram no *locus FAS* um longo RNA antisenso parcialmente intrônico que foi denominado *SAF*, naturalmente transcrito a partir da região genômica relativa ao primeiro intron do mRNA senso codificante para a proteína Fas (Yan *et al.*, 2005). Nesse estudo, os autores mostraram que a superexpressão de *SAF* pode induzir grandes alterações no processamento dos exons 3' do mRNA de *FAS*, que passa a codificar para uma proteína sem a propriedade de ficar ancorada à membrana plasmática, onde normalmente realiza suas funções (Yan *et al.*, 2005). Assim, células com expressão de *SAF* aumentada deixam de induzir apoptose mediada por Fas (Yan *et al.*, 2005). Os resultados obtidos nessa parte do projeto apontaram para novos RNAs antisenso intrônicos, que de maneira análoga a *SAF*, podem estar envolvidos em mecanismos regulatórios, ainda não totalmente elucidados. Aparentemente as mensagens identificadas atuam na modulação do processamento de pré-mRNAs de mensagens codificantes que apresentam uma complexa utilização de exons, influenciando na função desempenhada pelo produto proteico.

Os resultados apresentados aqui fornecem evidências originais que RNAs não-codificantes transcritos a partir de regiões intrônicas devem ser regulados por sinais fisiológicos que normalmente atuam sobre os transcritos codificantes de proteína, como os hormônios. A existência de um número considerável de sítios para ligação de fatores de transcrição (TFBS – *Transcription Factor Binding Site*) como Sp1, c-Myc, p53 e Creb, posteriores a regiões transcricionalmente ativas, e portanto fora da região tradicionalmente considerada como região promotora, foi demonstrada (Cawley *et al.*, 2004; Euskirchen *et al.*, 2004). Alguns desses TFBS estão orientados de maneira a possibilitar a geração de transcritos ainda não conhecidos e com orientação antisense em relação às mensagens já estudadas (Cawley *et al.*, 2004). Cawley e colaboradores ainda observaram que frações similares de mensagens codificantes de proteína e RNAs não-codificantes, com TFBS experimentalmente determinados, são reguladas após tratamento com ácido retinóico (Cawley *et al.*, 2004). Ainda, essas observações são reforçadas pela identificação de longos RNAs não-codificantes de camundongos que respondem ao estímulo com lipopolissacarídeos (Ravasi *et al.*, 2006).

O mRNA de *MYO5A* codifica para uma proteína envolvida no tráfego intracelular de vesículas e exocitose (Watanabe *et al.*, 2005). Interessante notar nesse contexto, que isoformas de *Myo5a* são tecido-específicas e originadas a partir de uso alternativo de exons, conferindo a especificidade de cargas transportadas por essa proteína no meio intracelular (Au & Huang, 2002). Com isso, é perfeitamente concebível especular que o aumento nos níveis de *asMYO5A* em células de próstata em resposta direta ao estímulo do andrógeno, afete a abundância relativa das diferentes formas do mRNA que codifica para *Myo5a*, modulando finalmente a atividade e especificidade dessa proteína.

Apesar de não existirem evidências conclusivas a respeito dos mecanismos moleculares, é postulado que os transcritos intrônicos realizem suas funções atuando em *cis* através de pareamento de bases do tipo Watson-Crick ao pré-mRNA senso complementar (Reis *et al.*, 2005). Os resultados obtidos nessa etapa do projeto estão de acordo com conceito de que pelo menos uma fração dos transcritos intrônicos deve afetar diretamente a abundância ou o processamento de mensagens codificantes de proteína originadas da fita oposta, assim como já observado por outros autores (Lee *et al.*, 1999; Katayama *et al.*, 2005). Entretanto, não pode ser excluída a possibilidade que alguns desses longos RNAs não-codificantes possam modular processos transcricionais, pós-transcricionais e até nos produtos protéicos de outros *loci*. Para isso poderiam tanto assumir estruturas secundárias complexas, ser parcialmente complementares às seqüências alvo, quanto ser processados em seqüências funcionais menores, similares aos microRNAs (Bartel, 2004). Nesse caso, a identificação dos genes afetados seria a maior limitação na caracterização funcional de novos RNAs não-codificantes biologicamente relevantes.

Finalmente, esse conjunto completo de resultados sugere fortemente que ao menos uma parte dos RNAs intrônicos de células de mamíferos são transcricionalmente controlados e devem exercer funções regulatórias, corroborando a visão da existência de um sistema complexo, mas ainda pouco compreendido, de regulação baseado em RNAs não-codificantes em organismos superiores (Johnson *et al.*, 2005; Mattick & Makunin, 2006; Willingham & Gingeras, 2006).

#### **5.4 Transcritos intrônicos diferencialmente expressos em tecidos humanos e de camundongos.**

Hibridizações cruzadas entre espécies, similares às realizadas nessa etapa do projeto, também têm sido utilizadas em análises de expressão com diversas motivações, que vão desde ausência de plataformas de *microarrays* específicas para a espécie em estudo (Moody *et al.*, 2002; Tsoi *et al.*, 2003); para possibilitar comparações dos padrões de expressão de um conjunto de transcritos em múltiplas espécies (Enard *et al.*, 2002; Adjaye *et al.*, 2004); e até para possibilitar a identificação de genes conservados com a observação de sua expressão em diferentes estágios de desenvolvimento (Vallee *et al.*, 2006). Entretanto esse tipo de abordagem necessita de cautela, pois mesmo pequenas divergências na seqüência de nucleotídeos entre as mensagens afetam diretamente seus níveis de intensidade e a especificidade da hibridização em comparações entre espécies na mesma plataforma (Gilad *et al.*, 2005).

Assim, no estudo mostrado aqui para evitar esse tipo de variação não controlável, as hibridizações de cada espécie foram realizadas em lâminas separadas. Da mesma forma, as comparações apresentadas foram apenas realizadas depois das análises estatísticas serem finalizadas independentemente em cada espécie. Ainda, a observação que todas as mensagens intrônicas expressas em camundongo foram também detectadas nos tecidos humanos correspondentes, sugere fortemente que com nosso desenho experimental e com a abordagem bastante estrita de análise estamos apontando apenas para mensagens hibridizadas com uma relativa especificidade.

Os 17 transcritos intrônicos compartilhados entre humanos e camundongos, detectados como diferencialmente expressos em fígado, rim e próstata, têm seus

perfis de expressão agrupados por tecido, revelando uma assinatura tecidual comum de transcrição intrônica entre as duas espécies. Esse resultado sugere que a expressão intrônica não é randômica, similar à expressão tecido-específica já observada para outros RNAs não-codificantes, os microRNAs (Lu *et al.*, 2005). Do conjunto de mensagens intrônicas aparentemente sintênicas identificados aqui, 10 foram confirmadas quanto a sua existência na população de RNA, e tiveram sua orientação transcricional determinada em amostras provenientes de fígado de camundongo. A determinação da orientação transcricional em paralelo de 5 dessas 10 mensagens em amostras humanas mostrou uma correspondência apenas parcial, sendo observada uma maior porcentagem (apesar do número baixo de *loci* testados) de transcrição antisense.

Assim, especula-se que talvez tanto os transcritos intrônicos senso quanto os antisense poderiam representar papéis distintos na regulação do programa transcricional de mamíferos. Sendo, provavelmente as mensagens antisense as determinantes da maior complexidade transcricional humana, através de um mecanismo de regulação ainda mais fino, que modularia os efeitos desempenhados pelas mensagens senso no controle da expressão gênica. Esse mecanismo foi postulado pela observação de um outro estudo atual onde é mostrado que a superexpressão de certos introns (região intrônica com mesma orientação transcricional dos exons de mesmo *locus*) em linhagens celulares onde a expressão da mensagem codificante de proteína correspondente não era detectável causa alteração na expressão de alguns mRNAs de maneira altamente reprodutível e específica para cada intron (Hill *et al.*, 2006). Esse resultado indica claramente uma funcionalidade para seqüências intrônicas senso não totalmente relacionada às mensagens codificantes correspondentes, talvez como proposto pelos autores,

atuando como sinais de notificação para a transcrição de outros mRNAs que participam das mesmas vias metabólicas, ou de vias relacionadas (Hill *et al.*, 2006).

Por outro lado, até o momento apenas uma clássica exceção (*Xist*) têm sido utilizada para indicar que uma baixa conservação de identidade não implica necessariamente em ausência de conservação de funcionalidade (Numata *et al.*, 2003; Pang *et al.*, 2006; Ravasi *et al.*, 2006). Casos como esse geralmente são apenas identificados acidentalmente, pois metodologias tradicionais de análise *in silico* por definição não apontam para seqüências com baixa identidade que sejam funcionalmente preservadas durante a evolução. Os resultados obtidos nessa etapa do projeto mostram uma conservação, entre duas espécies, de expressão tecido-específica para um número de RNAs intrônicos com baixa conservação de identidade, sugerindo que a expressão funcional desse tipo de mensagem não-codificante é mais comum do que o inicialmente esperado. Pistas para explicar essa observação podem ser obtidas do trabalho de Torarinsson e colaboradores onde foi mostrado que milhares de regiões genômicas sintênicas de humanos e camundongos, mas não alinháveis quanto a estrutura primária, podem conter RNAs que possuem conservação de estrutura secundária (Torarinsson *et al.*, 2006). Isso aponta fortemente para a existência de um fenômeno já observado em proteínas, a conservação de estrutura para realização de funções similares, também em moléculas de RNA.

Finalmente, sob a luz dessas observações, especula-se sobre o papel dos transcritos intrônicos antisense, identificados aqui em maior proporção em humanos em relação a camundongos, adicionando ainda mais complexidade ao modelo de regulação intrônica proposto (Hill *et al.*, 2006). Enquanto a sinalização intrônica senso ativaria as redes de genes relacionadas, os transcritos antisense poderiam

exercer um papel de modulação negativa das funções desempenhadas por cada intron (senso) de uma maneira espacialmente (tecido), e talvez temporalmente, bastante controlada. Assim, alguns transcritos antisenso poderiam ter um papel indireto na regulação da expressão gênica, tanto capturando a sinalização intrônica e a redirecionando para as diferentes vias, quanto através de impedimento para que essa seqüência adquirira a estrutura secundária correta que permite a realização de suas funções celulares.

### **5.5 Interação de transcritos intrônicos com proteínas ligadoras de RNA.**

Os resultados obtidos nos ensaios de interação de mensagens intrônicas com proteínas ligadoras de RNA (RBPs) envolvidas em reconhecidos processos regulação fortalecem nossas hipóteses de funcionalidade para essa classe de transcritos. A existência de transcritos intrônicos candidatos a ligação funcional com PABP, EIF4E, hTRA2 e HUD, aumentam a confiabilidade do conjunto de resultados com RNAs intrônicos obtidos nesse projeto como um todo, e ainda encorajam a caracterização funcional de alguns membros dessa peculiar classe de RNAs.

Ainda, o enriquecimento significativo ( $p < 0,05$ ) de transcritos intrônicos imunoprecipitados com PABP, EIF4E e HUD na categoria *regulation of progression through cell cycle*, assim como na categoria *signal transducer activity* com PABP e HUD são novas evidências providas pela análise de ontologia gênica (GO) que também se somam às já discutidas sobre as funções regulatórias dos Longos RNAs Intrônicos não-codificantes.

## 5.6 Análise de expressão de RNAs intrônicos em células LNCaP tratadas com $\alpha$ -amanitina

Apesar da enorme quantidade de resultados acumulados que demonstram que a maioria dos transcritos humanos são RNAs não-codificantes de proteína expressos a partir de regiões intrônicas e intergênicas do genoma (Willingham & Gingeras, 2006), até o momento não existem trabalhos que exploraram sistematicamente aos mecanismos de biogênese dessas mensagens. Na verdade, tem se assumido que os longos RNAs não-codificantes são em sua grande maioria transcritos pela RNAP II pois esses compartilham certas características estruturais em comum com os mRNAs, como poliadenilação e *capping* 5' (Erdmann *et al.*, 2001; Numata *et al.*, 2003). Ainda, a observação da ligação específica do receptor de andrógeno a regiões promotoras intrônicas (Louro *et al.*, 2006), assim como a identificação de sítios de ligação para conhecidos fatores de transcrição por RNAP II localizados em regiões intrônicas e intergências ativas (Martone *et al.*, 2003; Cawley *et al.*, 2004; Euskirchen *et al.*, 2004; Kim *et al.*, 2005), podem ser utilizadas como indícios para expressão pela RNAP II dessa classe de mensagens.

Até pouco tempo apenas três RNA polimerases nucleares eram conhecidas, sendo bem diferenciadas por suas respostas ao tratamento com  $\alpha$ -amanitina, que não tem efeito na atividade da RNAP I, inibe fortemente a RNAP II, e levemente a RNAP III (Jacob *et al.*, 1970; Kedinger *et al.*, 1970; Lindell *et al.*, 1970). Entretanto uma quarta RNAP nuclear foi recentemente descrita (spRNAP IV), e surpreendentemente foi observado que tratamento com  $\alpha$ -amanitina aumenta a expressão de mRNAs codificantes de proteína transcritos por ela (Kravchenko *et al.*, 2005).

Nessa etapa do projeto foi avaliada pela primeira vez a contribuição da RNAP II na expressão dos RNAs não codificantes intrônicos em células humanas, bloqueando a sua atividade com a  $\alpha$ -amanitina e medindo os efeitos na expressão gênica com a plataforma de *microarray* de oligonucleotídeos com cerca de 13.000 sondas intrônicas. Constatamos que além de uma menor porcentagem de transcritos intrônicos (16-19%) serem sensíveis ao tratamento com  $\alpha$ -amanitina, quando comparada com a dos exônicos (39%), uma fração significativamente maior dos transcritos intrônicos (12–16 %), em relação aos exônicos afetados pelo tratamento (4%), tiveram seus níveis aumentados. Esse comportamento observado para os transcritos intrônicos nas células tratadas com  $\alpha$ -amanitina levanta a possibilidade de que um outro sistema de síntese esteja atuando na transcrição de muitos RNAs não codificantes. Apesar da spRNAP IV, que aparentemente é ativada pela  $\alpha$ -amanitina (Kravchenko *et al.*, 2005), ser uma ótima candidata, a participação direta da spRNAP IV na ativação da transcrição intrônica aqui observada, ainda necessita de confirmação experimental.

## 6 CONCLUSÕES E CONSIDERAÇÕES FINAIS

O conjunto de resultados originais obtidos durante esse doutoramento ilustram a complexidade e contribuem para desvendar o repertório transcricional humano completo. Novos transcritos foram identificados e parcialmente caracterizados tanto em um contexto celular normal, quanto no surgimento e desenvolvimento de neoplasias de próstata.

Com a abordagem de busca de novos membros parálogos de famílias gênicas envolvidas com câncer foram identificadas duas novas subfamílias, Rasl11 e Arl9/10, da superfamília de pequenas GTPases monoméricas. As proteínas aparentemente codificadas por essas mensagens são conservadas entre os eucariotos e apresentam certas peculiaridades na seqüência primária. A observação que os níveis do mRNA de *RASL11A* encontram-se significativamente diminuídos em amostras tumorais de próstata aponta para possíveis papéis supressores de tumor também para esse novo membro da superfamília. Futuros estudos funcionais, principalmente com as proteínas Rasl11, tornam-se muito interessantes para o entendimento do processo tumoral em próstata, e para validação da abordagem de busca por proteínas parálogas como estratégia para reconhecimento das possíveis vias bioquímicas alternativas, que eventualmente estejam dificultando o tratamento efetivo do câncer.

A estratégia de observar o perfil de expressão de amostras de próstata normais e tumorais, utilizando uma plataforma de *microarray* enriquecida com mensagens ainda não anotadas, acabou revelando a existência de uma nova classe de RNAs não-codificantes. Esses transcritos são longos, não sofrem *splicing*, são expressos a partir dos introns e freqüentemente com orientação oposta (antisense) ao RNA mensageiro (mRNA) de genes codificantes para proteína. O padrão de

transcrição observado para estes Longos RNAs Intrônicos não-codificantes em amostras de câncer de próstata correlaciona-se com o grau de diferenciação tumoral, sendo o primeiro relato da implicação de transcritos dessa natureza em uma doença humana. Essa descoberta motiva a elaboração de hipóteses sobre possíveis funções biológicas destes transcritos no controle fino da expressão gênica. Ainda, apontou para a possibilidade de inclusão dessas mensagens no arsenal molecular para ajudar aos médicos na determinação do prognóstico de pacientes com câncer.

A caracterização da biogênese desses ncRNAs foi iniciada nesse projeto, revelando que assim como os mRNAs codificantes para proteínas, eles respondem a estímulos fisiológicos hormonais. Tomando como um exemplo o transcrito antisense *asMYO5A*, foi demonstrado que a ativação da expressão em resposta ao andrógeno ocorre de forma independente do mRNA, transcrito a partir da fita oposta; a ativação ocorre através da ligação do complexo andrógeno/receptor de andrógeno diretamente a um elemento de resposta intrônico, posicionado no genoma na região a 5' do local de início de transcrição intrônica. Outros resultados obtidos ainda revelam que possivelmente existe o envolvimento de diferentes RNA polimerases (RNAPs) na transcrição de uma fração significativa dessas mensagens. O aumento observado na abundância de uma porcentagem considerável dos Longos RNAs Intrônicos não-codificantes após a inibição da atividade da RNAP II é um forte indicativo do envolvimento de outra RNAP, sendo a spRNAP IV uma candidata bastante provável.

Papéis funcionais para esses transcritos intrônicos em eventos celulares normais ganharam novas evidências com os experimentos realizados durante esse projeto. A conservação de expressão específica em tecidos, observada entre humanos e camundongos, fortalece a visão de uma transcrição controlada que deve

estar envolvida em eventos característicos de cada tipo celular. A provável interação desses RNAs com proteínas ligadoras de RNA (RBPs) envolvidas em mecanismos de regulação pós-transcricional, assim como a observação de transcrição intrônica correlacionada tanto com a abundância, quanto com o padrão de utilização de exons de mensagens codificantes para proteína de mesmo *locus* em células tratadas com andrógeno, reforçam a atuação dos RNAs intrônicos no controle do programa transcricional humano que foi postulada.

De fato os resultados aqui apresentados abrem muitas possibilidades de estudos com os Longos RNAs Intrônicos não-codificantes. A futura utilização do sistema de expressão para células de mamíferos, já construído durante o projeto, para modular os níveis desses transcritos trará mais pistas, ainda mais conclusivas a respeito da funcionalidade dessa classe de mensagens, contribuindo com um considerável avanço no entendimento do controle do complexo programa transcricional humano.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

Abate-Shen, C. & M. M. Shen. (2000). Molecular genetics of prostate cancer. *Genes Dev*, **14**, 19, 2410-34.

Adjaye, J., R. Herwig, D. Herrmann, W. Wruck, A. Benkahla, *et al.* (2004). Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays. *BMC Genomics*, **5**, 1, 83.

Adjei, A. A. (2001). Blocking oncogenic Ras signaling for cancer therapy. *J Natl Cancer Inst*, **93**, 14, 1062-74.

Adjei, A. A., G. A. Croghan, C. Erlichman, R. S. Marks, J. M. Reid, *et al.* (2003). A Phase I trial of the farnesyl protein transferase inhibitor R115777 in combination with gemcitabine and cisplatin in patients with advanced cancer. *Clin Cancer Res*, **9**, 7, 2520-6.

Adjei, A. A., A. Mauer, L. Bruzek, R. S. Marks, S. Hillman, *et al.* (2003). Phase II study of the farnesyl transferase inhibitor R115777 in patients with advanced non-small-cell lung cancer. *J Clin Oncol*, **21**, 9, 1760-6.

Akhtar, A., D. Zink & P. B. Becker. (2000). Chromodomains are protein-RNA interaction modules. *Nature*, **407**, 6802, 405-9.

Akker, S. A., P. J. Smith & S. L. Chew. (2001). Nuclear post-transcriptional control of gene expression. *J Mol Endocrinol*, **27**, 2, 123-31.

Albertsen, P. C., J. A. Hanley, D. F. Gleason & M. J. Barry. (1998). Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. *Jama*, **280**, 11, 975-80.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 17, 3389-402.

Au, J. S. & J. D. Huang. (2002). A tissue-specific exon of myosin Va is responsible for selective cargo binding in melanocytes. *Cell Motil Cytoskeleton*, **53**, 2, 89-102.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 2, 281-97.

Bertone, P., V. Stolc, T. E. Royce, J. S. Rozowsky, A. E. Urban, *et al.* (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science*, **306**, 5705, 2242-2246.

Best, C. J., I. M. Leiva, R. F. Chuaqui, J. W. Gillespie, P. H. Duray, *et al.* (2003). Molecular differentiation of high- and moderate-grade human prostate cancer by cDNA microarray analysis. *Diagn. Mol. Pathol.*, **12**, 2, 63-70.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, **72**, 291-336.

Boeck, R. & D. Kolakofsky. (1994). Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. *EMBO J.*, **13**, 15, 3608-3617.

Boettner, B. & L. Van Aelst. (2002). The role of Rho GTPases in disease development. *Gene*, **286**, 2, 155-74.

Bos, J. L. (1997). Ras-like GTPases. *Biochim Biophys Acta*, **1333**, 2, M19-31.

Bruno, I. G., W. Jin & G. J. Cote. (2004). Correction of aberrant FGFR1 alternative RNA splicing through targeting of intronic regulatory elements. *Hum Mol Genet*, **13**, 20, 2409-20.

Brunschwig, E. B., K. Wilson, D. Mack, D. Dawson, E. Lawrence, *et al.* (2003). PMEPA1, a transforming growth factor-beta-induced marker of terminal colonocyte differentiation whose expression is maintained in primary and metastatic colon cancer. *Cancer Res.*, **63**, 7, 1568-1575.

Bushnell, D. A., P. Cramer & R. D. Kornberg. (2002). Structural basis of transcription: alpha-amanitin-RNA polymerase II cocystal at 2.8 A resolution. *Proc Natl Acad Sci U S A*, **99**, 3, 1218-22.

Calin, G. A., C. G. Liu, C. Sevignani, M. Ferracin, N. Felli, *et al.* (2004). MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A*, **101**, 32, 11755-60.

Calvo, A., O. Gonzalez-Moreno, C. Y. Yoon, J. I. Huh, K. Desai, *et al.* (2005). Prostate cancer and the genomic revolution: Advances using microarray analyses. *Mutat Res*, **576**, 1-2, 66-79.

Camargo, A. A., H. P. Samaia, E. Dias-Neto, D. F. Simao, I. A. Migotto, *et al.* (2001). The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc Natl Acad Sci U S A*, **98**, 21, 12103-8.

Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, *et al.* (2005). The transcriptional landscape of the mammalian genome. *Science*, **309**, 5740, 1559-63.

Caron, P. R., M. D. Mullican, R. D. Mashal, K. P. Wilson, M. S. Su, *et al.* (2001). Chemogenomic approaches to drug discovery. *Curr Opin Chem Biol*, **5**, 4, 464-70.

Carroll, P. (2001). Rising PSA after a radical treatment. *Eur Urol*, **40 Suppl 2**, 9-16.

Cavaille, J., K. Buiting, M. Kieffmann, M. Lalande, C. I. Brannan, *et al.* (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A*, **97**, 26, 14311-6.

- Cavaille, J., P. Vitali, E. Basyuk, A. Huttenhofer & J. P. Bachellerie. (2001). A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J Biol Chem*, **276**, 28, 26374-83.
- Cawley, S., S. Bekiranov, H. H. Ng, P. Kapranov, E. A. Sekinger, *et al.* (2004). Unbiased Mapping of Transcription Factor Binding Sites along Human Chromosomes 21 and 22 Points to Widespread Regulation of Noncoding RNAs. *Cell*, **116**, 4, 499-509.
- Chen, G. & A. J. Courey. (2000). Groucho/TLE family proteins and transcriptional repression. *Gene*, **249**, 1-2, 1-16.
- Chen, J., M. Sun, W. J. Kent, X. Huang, H. Xie, *et al.* (2004). Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res*, **32**, 16, 4812-20.
- Chirgwin, J. M., A. E. Przybyla, R. J. Macdonald & W. J. Rutter. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*, **18**, 24, 5294-9.
- Chomczynski, P. & N. Sacchi. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*, **162**, 1, 156-9.
- Claessens, F. & D. T. Gewirth. (2004). DNA recognition by nuclear receptors. *Essays Biochem*, **40**, 59-72.
- Cohen, P. (1999). The development and therapeutic potential of protein kinase inhibitors. *Curr Opin Chem Biol*, **3**, 4, 459-65.
- Cox, A. D. & C. J. Der. (2002). Ras family signaling: therapeutic targeting. *Cancer Biol Ther*, **1**, 6, 599-606.
- Cox, D. N., A. Chao, J. Baker, L. Chang, D. Qiao, *et al.* (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev*, **12**, 23, 3715-27.
- Crooks, G. E., G. Hon, J. M. Chandonia & S. E. Brenner. (2004). WebLogo: a sequence logo generator. *Genome Res*, **14**, 6, 1188-90.
- Dai, Q., E. Choy, V. Chiu, J. Romano, S. R. Slivka, *et al.* (1998). Mammalian prenylcysteine carboxyl methyltransferase is in the endoplasmic reticulum. *J. Biol. Chem.*, **273**, 24, 15030-15034.
- Davis, E., F. Caiment, X. Tordoir, J. Cavaille, A. Ferguson-Smith, *et al.* (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol*, **15**, 8, 743-9.
- Dennis, G., Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, *et al.* (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*, **4**, 5, P3.

Dermitzakis, E. T., A. Reymond & S. E. Antonarakis. (2005). Conserved non-genic sequences - an unexpected feature of mammalian genomes. *Nat Rev Genet*, **6**, 2, 151-7.

Desilva, U., L. Elnitski, J. R. Idol, J. L. Doyle, W. Gan, *et al.* (2002). Generation and comparative analysis of approximately 3.3 Mb of mouse genomic sequence orthologous to the region of human chromosome 7q11.23 implicated in Williams syndrome. *Genome Res*, **12**, 1, 3-15.

Dhanasekaran, S. M., T. R. Barrette, D. Ghosh, R. Shah, S. Varambally, *et al.* (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature*, **412**, 6849, 822-6.

Dias-Neto, E., R. G. Correa, S. Verjovski-Almeida, M. R. Briones, M. A. Nagai, *et al.* (2000). Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. USA*, **97**, 7, 3491-3496.

Du, T. & P. D. Zamore. (2005). microPrimer: the biogenesis and function of microRNA. *Development*, **132**, 21, 4645-52.

Duret, L., F. Dorkeld & C. Gautier. (1993). Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression. *Nucleic Acids Res*, **21**, 10, 2315-22.

Dvir, A. (2002). Promoter escape by RNA polymerase II. *Biochim Biophys Acta*, **1577**, 2, 208-223.

Eckley, D. M. & T. A. Schroer. (2003). Interactions between the evolutionarily conserved, actin-related protein, Arp11, actin, and Arp1. *Mol. Biol. Cell*, **14**, 7, 2645-2654.

Ellis, C. A., M. D. Vos, H. Howell, T. Vallecorsa, D. W. Fults, *et al.* (2002a). Rig is a novel Ras-related protein and potential neural tumor suppressor. *Proc Natl Acad Sci U S A*, **99**, 15, 9876-81.

Ellis, C. A., M. D. Vos, H. Howell, T. Vallecorsa, D. W. Fults, *et al.* (2002b). Rig is a novel Ras-related protein and potential neural tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 15, 9876-9881.

Enard, W., P. Khaitovich, J. Klose, S. Zollner, F. Heissig, *et al.* (2002). Intra- and interspecific variation in primate gene expression patterns. *Science*, **296**, 5566, 340-3.

Epstein, J. I., A. W. Partin, J. Sauvageot & P. C. Walsh. (1996). Prediction of progression following radical prostatectomy. A multivariate analysis of 721 men with long-term follow-up. *Am. J. Surg. Pathol.*, **20**, 3, 286-292.

Erdmann, V. A., M. Z. Barciszewska, M. Szymanski, A. Hochberg, N. De Groot, *et al.* (2001). The non-coding RNAs as riboregulators. *Nucleic Acids Res*, **29**, 1, 189-93.

Esquela-Kerscher, A. & F. J. Slack. (2006). Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*, **6**, 4, 259-69.

Euskirchen, G., T. E. Royce, P. Bertone, R. Martone, J. L. Rinn, *et al.* (2004). CREB binds to multiple loci on human chromosome 22. *Mol Cell Biol*, **24**, 9, 3804-14.

Ewing, B. & P. Green. (2000). Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet*, **25**, 2, 232-4.

Farrell, C. M. & L. N. Lukens. (1995). Naturally occurring antisense transcripts are present in chick embryo chondrocytes simultaneously with the down-regulation of the alpha 1 (I) collagen gene. *J Biol Chem*, **270**, 7, 3400-8.

Finlin, B. S., C. L. Gau, G. A. Murphy, H. Shao, T. Kimel, *et al.* (2001). RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. *J Biol Chem*, **276**, 45, 42259-67.

Frazer, K. A., L. Elnitski, D. M. Church, I. Dubchak & R. C. Hardison. (2003). Cross-species sequence comparisons: a review of methods and available resources. *Genome Res*, **13**, 1, 1-12.

Futreal, P. A., A. Kasprzyk, E. Birney, J. C. Mullikin, R. Wooster, *et al.* (2001). Cancer and genomics. *Nature*, **409**, 6822, 850-2.

Garcia-Echeverria, C., P. Traxler & D. B. Evans. (2000). ATP site-directed competitive and irreversible inhibitors of protein kinases. *Med Res Rev*, **20**, 1, 28-57.

Gebuhr, T. C., S. J. Bultman & T. Magnuson. (2000). Pc-G/trx-G and the SWI/SNF connection: developmental gene regulation through chromatin remodeling. *Genesis*, **26**, 3, 189-97.

Geck, P., J. Szelei, J. Jimenez, T. M. Lin, C. Sonnenschein, *et al.* (1997). Expression of novel genes linked to the androgen-induced, proliferative shutoff in prostate cancer cells. *J Steroid Biochem Mol Biol*, **63**, 4-6, 211-8.

Gemmill, R. M., J. D. West, F. Boldog, N. Tanaka, L. J. Robinson, *et al.* (1998). The hereditary renal cell carcinoma 3;8 translocation fuses FHIT to a patched-related gene, TRC8. *Proc. Natl. Acad. Sci. U S A*, **95**, 16, 9572-9577.

Gilad, Y., S. A. Rifkin, P. Bertone, M. Gerstein & K. P. White. (2005). Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res*, **15**, 5, 674-80.

Girard, A., R. Sachidanandam, G. J. Hannon & M. A. Carmell. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, **442**, 7099, 199-202.

Gleason, D. Histologic grading and clinical staging of carcinoma of the prostate. In: M. Tannenbaum (Ed.). Urologic pathology. Philadelphia: Lea & Febiger, 1977. Histologic grading and clinical staging of carcinoma of the prostate., p.171-197

- Glinsky, G. V., A. B. Glinskii, A. J. Stephenson, R. M. Hoffman & W. L. Gerald. (2004). Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest*, **113**, 6, 913-23.
- Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, *et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 5439, 531-7.
- Griffiths-Jones, S., R. J. Grocock, S. Van Dongen, A. Bateman & A. J. Enright. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*, **34**, Database issue, D140-4.
- Hardison, R. C., J. Oeltjen & W. Miller. (1997). Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res*, **7**, 10, 959-66.
- Harlow, E. & D. Lane. Antibodies: A Laboratory Manual. 1998
- Harris, C. J. & A. P. Stevens. (2006). Chemogenomics: structuring the drug discovery process to gene families. *Drug Discov Today*, **11**, 19-20, 880-8.
- Hill, A. E., J. S. Hong, H. Wen, L. Teng, D. T. Mcpherson, *et al.* (2006). Micro-RNA-like effects of complete intronic sequences. *Front Biosci*, **11**, 1998-2006.
- Hogenesch, J. B., K. A. Ching, S. Batalov, A. I. Su, J. R. Walker, *et al.* (2001). A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell*, **106**, 4, 413-5.
- Howe, K. J. (2002). RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta*, **1577**, 2, 308-24.
- Huang, C., E. A. Sloan & C. F. Boerkoel. (2003). Chromatin remodeling and human disease. *Curr Opin Genet Dev*, **13**, 3, 246-52.
- Hughes, T. R., M. Mao, A. R. Jones, J. Burchard, M. J. Marton, *et al.* (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.*, **19**, 4, 342-347.
- Imamura, T., S. Yamamoto, J. Ohgane, N. Hattori, S. Tanaka, *et al.* (2004). Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochem Biophys Res Commun*, **322**, 2, 593-600.
- Isaacs, J. T., P. I. Lundmo, R. Berges, P. Martikainen, N. Kyprianou, *et al.* (1992). Androgen regulation of programmed death of normal and malignant prostatic cells. *J Androl*, **13**, 6, 457-64.
- Iseli, C., C. V. Jongeneel & P. Bucher. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 138-148.

Jacob, S. T., E. M. Sajdel & H. N. Munro. (1970). Specific action of alpha-amanitin on mammalian RNA polymerase protein. *Nature*, **225**, 5227, 60-2.

Jensen, R. A. (2001). Orthologs and paralogs - we need to get it right. *Genome Biol*, **2**, 8, INTERACTIONS1002.

Ji, P., S. Diederichs, W. Wang, S. Boing, R. Metzger, *et al.* (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, **22**, 39, 8031-41.

Johnson, J. M., S. Edwards, D. Shoemaker & E. E. Schadt. (2005). Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet*, **21**, 2, 93-102.

Jones, D. O., I. G. Cowell & P. B. Singh. (2000). Mammalian chromodomain proteins: their role in genome organisation and expression. *Bioessays*, **22**, 2, 124-37.

Jongeneel, C. V., M. Delorenzi, C. Iseli, D. Zhou, C. D. Haudenschild, *et al.* (2005). An atlas of human gene expression from massively parallel signature sequencing (MPSS). *Genome Res*, **15**, 7, 1007-14.

Kampa, D., J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, *et al.* (2004). Novel RNAs Identified From an In-Depth Analysis of the Transcriptome of Human Chromosomes 21 and 22. *Genome Res.*, **14**, 3, 331-342.

Kapranov, P., S. E. Cawley, J. Drenkow, S. Bekiranov, R. L. Strausberg, *et al.* (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, **296**, 5569, 916-919.

Kapranov, P., J. Drenkow, J. Cheng, J. Long, G. Helt, *et al.* (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res*, **15**, 7, 987-97.

Katayama, S., Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, *et al.* (2005). Antisense transcription in the mammalian transcriptome. *Science*, **309**, 5740, 1564-6.

Kedinger, C., M. Gniazdowski, J. L. Mandel, Jr., F. Gissinger & P. Chambon. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. *Biochem Biophys Res Commun*, **38**, 1, 165-71.

Khochbin, S. & J. J. Lawrence. (1989). An antisense RNA involved in p53 mRNA maturation in murine erythroleukemia cells induced to differentiate. *EMBO J.*, **8**, 13, 4107-4114.

Kim, T. H., L. O. Barrera, C. Qu, S. Van Calcar, N. D. Trinklein, *et al.* (2005). Direct isolation and identification of promoters in the human genome. *Genome Res*, **15**, 6, 830-9.

Kishore, S. & S. Stamm. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 5758, 230-2.

Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 2, 145-8.

Kiyosawa, H., I. Yamanaka, N. Osato, S. Kondo & Y. Hayashizaki. (2003). Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res.*, **13**, 6B, 1324-1334.

Kondo, S., N. Yamamoto, T. Murakami, M. Okumura, A. Mayeda, *et al.* (2004). Tra2 beta, SF2/ASF and SRp30c modulate the function of an exonic splicing enhancer in exon 10 of tau pre-mRNA. *Genes Cells*, **9**, 2, 121-30.

Kravchenko, J. E., I. B. Rogozin, E. V. Koonin & P. M. Chumakov. (2005). Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin. *Nature*, **436**, 7051, 735-9.

Kuge, O., C. Dascher, L. Orci, T. Rowe, M. Amherdt, *et al.* (1994). Sar1 promotes vesicle budding from the endoplasmic reticulum but not Golgi compartments. *J. Cell Biol.*, **125**, 1, 51-65.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 6822, 860-921.

Lapointe, J., C. Li, J. P. Higgins, M. Van De Rijn, E. Bair, *et al.* (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A*, **101**, 3, 811-6.

Lee, C. H., N. G. Della, C. E. Chew & D. J. Zack. (1996). Rin, a neuron-specific and calmodulin-binding small G-protein, and Rit define a novel subfamily of ras proteins. *J. Neurosci.*, **16**, 21, 6784-6794.

Lee, J. T., L. S. Davidow & D. Warshawsky. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.*, **21**, 4, 400-404.

Lee, T. I. & R. A. Young. (2000). Transcription of eukaryotic protein-coding genes. *Annu Rev Genet*, **34**, 77-137.

Lewis, B. P., C. B. Burge & D. P. Bartel. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 1, 15-20.

Liang, F., I. Holt, G. Pertea, S. Karamycheva, S. L. Salzberg, *et al.* (2000). Gene index analysis of the human genome estimates approximately 120,000 genes. *Nat Genet*, **25**, 2, 239-40.

Lindell, T. J., F. Weinberg, P. W. Morris, R. G. Roeder & W. J. Rutter. (1970). Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science*, **170**, 956, 447-9.

Liu, L., J. H. Yoon, R. Dammann & G. P. Pfeifer. (2002). Frequent hypermethylation of the RASSF1A gene in prostate cancer. *Oncogene*, **21**, 44, 6835-6840.

Liu, Y., G. Dehni, K. J. Purcell, J. Sokolow, M. L. Carcangiu, *et al.* (1996). Epithelial expression and chromosomal location of human TLE genes: implications for notch signaling and neoplasia. *Genomics*, **31**, 1, 58-64.

Lonard, D. M. & B. W. O'malley. (2005). Expanding functional diversity of the coactivators. *Trends Biochem Sci*, **30**, 3, 126-32.

Loots, G. G., R. M. Locksley, C. M. Blankespoor, Z. E. Wang, W. Miller, *et al.* (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, **288**, 5463, 136-40.

Louro, R., H. I. Nakaya, P. P. Amaral, F. Festa, M. C. Sogayar, *et al.* (2006). Androgen responsive intronic non-coding RNAs. *BMC Genomics*, (**Submitted**), ANEXO 7.

Louro, R., H. I. Nakaya, A. C. Paquola, E. A. Martins, A. M. Da Silva, *et al.* (2004). RASL11A, member of a novel small monomeric GTPase gene family, is down-regulated in prostate tumors. *Biochem Biophys Res Commun*, **316**, 3, 618-27. ANEXO 4

Lu, J., G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, *et al.* (2005). MicroRNA expression profiles classify human cancers. *Nature*, **435**, 7043, 834-8.

Luo, J., S. Zha, W. R. Gage, T. A. Dunn, J. L. Hicks, *et al.* (2002). Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. *Cancer Res*, **62**, 8, 2220-6.

Luo, R. Z., X. Fang, R. Marquez, S. Y. Liu, G. B. Mills, *et al.* (2003). ARHI is a Ras-related small G-protein with a novel N-terminal extension that inhibits growth of ovarian and breast cancers. *Oncogene*, **22**, 19, 2897-909.

Maffini, M. V., P. Geck, C. E. Powell, C. Sonnenschein & A. M. Soto. (2002). Mechanism of androgen action on cell proliferation: AS3 protein as a mediator of proliferative arrest in the rat prostate. *Endocrinology*, **143**, 7, 2708-14.

Makalowski, W., J. Zhang & M. S. Boguski. (1996). Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res*, **6**, 9, 846-57.

Mangus, D. A., M. C. Evans & A. Jacobson. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol*, **4**, 7, 223.

Mansfield, J. H., B. D. Harfe, R. Nissen, J. Obenaus, J. Srineel, *et al.* (2004). MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other

developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet*, **36**, 10, 1079-83.

Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri, C. Deweese-Scott, L. Y. Geer, *et al.* (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res*, **33**, Database issue, D192-6.

Margulies, E. H., M. Blanchette, D. Haussler & E. D. Green. (2003). Identification and characterization of multi-species conserved sequences. *Genome Res*, **13**, 12, 2507-18.

Martone, R., G. Euskirchen, P. Bertone, S. Hartman, T. E. Royce, *et al.* (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc Natl Acad Sci U S A*, **100**, 21, 12247-52.

Maston, G. A., S. K. Evans & M. R. Green. (2006). Transcriptional Regulatory Elements in the Human Genome. *Annu Rev Genomics Hum Genet*.

Mata, J., S. Marguerat & J. Bahler. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci*, **30**, 9, 506-14.

Mattick, J. S. (2004). RNA regulation: a new genetics? *Nat Rev Genet*, **5**, 4, 316-323.

Mattick, J. S. & I. V. Makunin. (2006). Non-coding RNA. *Hum Mol Genet*, **15 Suppl 1**, R17-29.

Matys, V., E. Fricke, R. Geffers, E. Gossling, M. Haubrock, *et al.* (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, **31**, 1, 374-8.

Mcmanus, M. T. (2003). MicroRNAs and cancer. *Semin. Cancer Biol.*, **13**, 4, 253-258.

Miyaso, H., M. Okumura, S. Kondo, S. Higashide, H. Miyajima, *et al.* (2003). An intronic splicing enhancer element in survival motor neuron (SMN) pre-mRNA. *J Biol Chem*, **278**, 18, 15825-31.

Moody, D. E., Z. Zou & L. McIntyre. (2002). Cross-species hybridisation of pig RNA to human nylon microarrays. *BMC Genomics*, **3**, 1, 27.

Nadon, R., P. Shi, A. Skandalis, E. Woody, H. Hubschle, *et al.* (2001). Statistical inference methods for gene expression arrays. Microarrays: Optical Technologies and Informatics. *Proceedings of SPIE*, **4266**, 46-55.

Nakaya, H. I., P. P. Amaral, R. Louro, A. Lopes, A. A. Fachel, *et al.* (2006). Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription *Genome Biol*, **(Submitted)**, ANEXO 2.

- Nelson, P. S., N. Clegg, H. Arnold, C. Ferguson, M. Bonham, *et al.* (2002). The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad Sci U S A*, **99**, 18, 11890-5.
- Nepveu, A. & K. B. Marcu. (1986). Intragenic pausing and anti-sense transcription within the murine c-myc locus. *EMBO J.*, **5**, 11, 2859-2865.
- Noguchi, M., S. Miyamoto, T. A. Silverman & B. Safer. (1994). Characterization of an antisense Inr element in the eIF-2 alpha gene. *J Biol Chem*, **269**, 46, 29161-7.
- Numata, K., A. Kanai, R. Saito, S. Kondo, J. Adachi, *et al.* (2003). Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res*, **13**, 6B, 1301-6.
- Oesterling, J. E., J. Roy, A. Agha, T. Shown, T. Krarup, *et al.* (1997). Biologic variability of prostate-specific antigen and its usefulness as a marker for prostate cancer: effects of finasteride. The Finasteride PSA Study Group. *Urology*, **50**, 1, 13-8.
- Okazaki, Y., M. Furuno, T. Kasukawa, J. Adachi, H. Bono, *et al.* (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, **420**, 6915, 563-573.
- Ota, T., Y. Suzuki, T. Nishikawa, T. Otsuki, T. Sugiyama, *et al.* (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet*, **36**, 1, 40-5.
- Oxford, G. & D. Theodorescu. (2003). Ras superfamily monomeric G proteins in carcinoma cell motility. *Cancer Lett*, **189**, 2, 117-28.
- Pabo, C. O. & R. T. Sauer. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem*, **61**, 1053-95.
- Paduch, M., F. Jelen & J. Otlewski. (2001). Structure of small G proteins and their regulators. *Acta Biochim Pol*, **48**, 4, 829-50.
- Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci*, **12**, 4, 357-8.
- Pang, K. C., M. C. Frith & J. S. Mattick. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet*, **22**, 1, 1-5.
- Pasqualato, S., L. Renault & J. Cherfils. (2002a). Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication. *EMBO Rep*, **3**, 11, 1035-41.
- Pasqualato, S., L. Renault & J. Cherfils. (2002b). Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication. *EMBO Rep.*, **3**, 11, 1035-1041.

Paule, M. R. & R. J. White. (2000). Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res*, **28**, 6, 1283-98.

Penalva, L. O. & J. D. Keene. (2004). Biotinylated tags for recovery and characterization of ribonucleoprotein complexes. *Biotechniques*, **37**, 4, 604, 606, 608-10.

Penalva, L. O., S. A. Tenenbaum & J. D. Keene. (2004). Gene expression analysis of messenger RNP complexes. *Methods Mol Biol*, **257**, 125-34.

Penny, G. D., G. F. Kay, S. A. Sheardown, S. Rastan & N. Brockdorff. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, **379**, 6561, 131-7.

Petrovics, G., A. Liu, S. Shaheduzzaman, B. Furasato, C. Sun, *et al.* (2005). Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. *Oncogene*, **24**, 23, 3847-52.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, 9, e45.

Potter, S. R., W. Horniger, M. Tinzl, G. Bartsch & A. W. Partin. (2001). Age, prostate-specific antigen, and digital rectal examination as determinants of the probability of having prostate cancer. *Urology*, **57**, 6, 1100-4.

Prescott, E. M. & N. J. Proudfoot. (2002). Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A*, **99**, 13, 8796-801.

Ptashne, M. & A. Gann. (1997). Transcriptional activation by recruitment. *Nature*, **386**, 6625, 569-77.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat. Genet.*, **32 Suppl**, 496-501.

Ramaswamy, S., K. N. Ross, E. S. Lander & T. R. Golub. (2003). A molecular signature of metastasis in primary solid tumors. *Nat. Genet.*, **33**, 1, 49-54.

Rao, S., D. Cunningham, A. De Gramont, W. Scheithauer, M. Smakal, *et al.* (2004). Phase III double-blind placebo-controlled study of farnesyl transferase inhibitor R115777 in patients with refractory advanced colorectal cancer. *J Clin Oncol*, **22**, 19, 3950-7.

Ravasi, T., H. Suzuki, K. C. Pang, S. Katayama, M. Furuno, *et al.* (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res*, **16**, 1, 11-9.

Reis, E. M., R. Louro, H. I. Nakaya & S. Verjovski-Almeida. (2005). As antisense RNA gets intronic. *Omic*s, **9**, 1, 2-12. ANEXO 1

- Reis, E. M., H. I. Nakaya, R. Louro, F. C. Canavez, A. V. Flatschart, *et al.* (2004). Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene*, **23**, 39, 6684-6692. ANEXO 3
- Reuther, G. W. & C. J. Der. (2000a). The Ras branch of small GTPases: Ras family members don't fall far from the tree. *Curr. Opin. Cell Biol.*, **12**, 2, 157-165.
- Reuther, G. W. & C. J. Der. (2000b). The Ras branch of small GTPases: Ras family members don't fall far from the tree. *Curr Opin Cell Biol*, **12**, 2, 157-65.
- Rhodes, D. R., T. R. Barrette, M. A. Rubin, D. Ghosh & A. M. Chinnaiyan. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res*, **62**, 15, 4427-33.
- Richter, J. D. & N. Sonenberg. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*, **433**, 7025, 477-80.
- Rinn, J. L., G. Euskirchen, P. Bertone, R. Martone, N. M. Luscombe, *et al.* (2003). The transcriptional activity of human Chromosome 22. *Genes Dev.*, **17**, 4, 529-540.
- Rogelj, B. & K. P. Giese. (2004). Expression and function of brain specific small RNAs. *Rev Neurosci*, **15**, 3, 185-98.
- Saha, S., A. B. Sparks, C. Rago, V. Akmaev, C. J. Wang, *et al.* (2002). Using the transcriptome to annotate the genome. *Nat Biotechnol*, **20**, 5, 508-12.
- Sainaghi, P. P., L. Castello, L. Bergamasco, M. Galletti, P. Bellosta, *et al.* (2005). Gas6 induces proliferation in prostate carcinoma cell lines expressing the Axl receptor. *J Cell Physiol*, **204**, 1, 36-44.
- Saito, K., T. Fujiwara, J. Katahira, K. Inoue & H. Sakamoto. (2004). TAP/NXF1, the primary mRNA export receptor, specifically interacts with a neuronal RNA-binding protein HuD. *Biochem Biophys Res Commun*, **321**, 2, 291-7.
- Sanchez-Carbayo, M., N. D. Socci, E. Charytonowicz, M. Lu, M. Prystowsky, *et al.* (2002). Molecular profiling of bladder cancer using cDNA microarrays: defining histogenesis and biological phenotypes. *Cancer Res.*, **62**, 23, 6973-6980.
- Sanchez-Diaz, P. & L. O. Penalva. Post-transcriptional meets post-genomic. The saga of RNA binding proteins in a new era. 2006.
- Sandelin, A., P. Bailey, S. Bruce, P. G. Engstrom, J. M. Klos, *et al.* (2004). Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics*, **5**, 1, 99.
- Shaffer, P. L., A. Jivan, D. E. Dollins, F. Claessens & D. T. Gewirth. (2004). Structural basis of androgen receptor binding to selective androgen response elements. *Proc Natl Acad Sci U S A*, **101**, 14, 4758-63.

- Shatkin, A. J. & J. L. Manley. (2000). The ends of the affair: capping and polyadenylation. *Nat Struct Biol*, **7**, 10, 838-42.
- She, X., Z. Jiang, R. A. Clark, G. Liu, Z. Cheng, *et al.* (2004). Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature*, **431**, 7011, 927-30.
- Shendure, J. & G. M. Church. (2002). Computational discovery of sense-antisense transcription in the human and mouse genomes. *Genome Biol.*, **3**, 9, research0044.1-research0044.14.
- Shoemaker, D. D., E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engle, *et al.* (2001). Experimental annotation of the human genome using microarray technology. *Nature*, **409**, 6822, 922-927.
- Silverman, T. A., M. Noguchi & B. Safer. (1992). Role of sequences within the first intron in the regulation of expression of eukaryotic initiation factor 2 alpha. *J Biol Chem*, **267**, 14, 9738-42.
- Sinensky, M. (2000). Recent advances in the study of prenylated proteins. *Biochim. Biophys. Acta*, **1484**, 2-3, 93-106.
- Singh, D., P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, *et al.* (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, **1**, 2, 203-209.
- Siomi, H. & G. Dreyfuss. (1997). RNA-binding proteins as regulators of gene expression. *Curr Opin Genet Dev*, **7**, 3, 345-53.
- Sleutels, F., D. P. Barlow & R. Lyle. (2000). The uniqueness of the imprinting mechanism. *Curr. Opin. Genet. Dev.*, **10**, 2, 229-233.
- Sleutels, F., R. Zwart & D. P. Barlow. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, **415**, 6873, 810-3.
- Smith, C. W. & J. Valcarcel. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, **25**, 8, 381-8.
- Sorek, R. & H. M. Safer. (2003). A novel algorithm for computational identification of contaminated EST libraries. *Nucleic Acids Res.*, **31**, 3, 1067-1074.
- Spoerner, M., C. Herrmann, I. R. Vetter, H. R. Kalbitzer & A. Wittinghofer. (2001). Dynamic properties of the Ras switch I region and its importance for binding to effectors. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 9, 4944-4949.
- Strausberg, R. L., K. H. Buetow, M. R. Emmert-Buck & R. D. Klausner. (2000). The cancer genome anatomy project: building an annotated gene index. *Trends Genet*, **16**, 3, 103-6.

Swinnen, J. V. & G. Verhoeven. (1998). Androgens and the control of lipid metabolism in human prostate cancer cells. *J Steroid Biochem Mol Biol*, **65**, 1-6, 191-8.

Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, *et al.* (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA*, **96**, 6, 2907-2912.

Tasheva, E. S. & D. J. Roufa. (1995). Regulation of human RPS14 transcription by intronic antisense RNAs and ribosomal protein S14. *Genes Dev.*, **9**, 3, 304-316.

Tee, M. K. & R. B. Jaffe. (2001). A precursor form of vascular endothelial growth factor arises by initiation from an upstream in-frame CUG codon. *Biochem. J.*, **359**, Pt 1, 219-226.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin & D. G. Higgins. (1997). The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, **25**, 24, 4876-82.

Thompson, J. D., D. G. Higgins & T. J. Gibson. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 22, 4673-80.

Torarinsson, E., M. Sawera, J. H. Havgaard, M. Fredholm & J. Gorodkin. (2006). Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res*, **16**, 7, 885-9.

Tsoi, S. C., J. M. Cale, I. M. Bird, V. Ewart, L. L. Brown, *et al.* (2003). Use of human cDNA microarrays for identification of differentially expressed genes in Atlantic salmon liver during *Aeromonas salmonicida* infection. *Mar Biotechnol (NY)*, **5**, 6, 545-54.

Tufarelli, C., J. A. Stanley, D. Garrick, J. A. Sharpe, H. Ayyub, *et al.* (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.*, **34**, 2, 157-165.

Tusher, V. G., R. Tibshirani & G. Chu. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA*, **98**, 9, 5116-5121.

Tycowski, K. T., A. Aab & J. A. Steitz. (2004). Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr Biol*, **14**, 22, 1985-95.

Vallee, M., C. Robert, S. Methot, M. F. Palin & M. A. Sirard. (2006). Cross-species hybridizations on a multi-species cDNA microarray to identify evolutionarily conserved genes expressed in oocytes. *BMC Genomics*, **7**, 113.

Van 'T Veer, L. J., H. Dai, M. J. Van De Vijver, Y. D. He, A. A. Hart, *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 6871, 530-536.

Van Der Vlag, J. & A. P. Otte. (1999). Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet*, **23**, 4, 474-8.

Varambally, S., S. M. Dhanasekaran, M. Zhou, T. R. Barrette, C. Kumar-Sinha, *et al.* (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*, **419**, 6907, 624-9.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, *et al.* (2001). The sequence of the human genome. *Science*, **291**, 5507, 1304-1351.

Vetter, I. R. & A. Wittinghofer. (2001a). The guanine nucleotide-binding switch in three dimensions. *Science*, **294**, 5545, 1299-1304.

Vetter, I. R. & A. Wittinghofer. (2001b). The guanine nucleotide-binding switch in three dimensions. *Science*, **294**, 5545, 1299-304.

Wani, M. A., N. C. Denko & P. J. Stambrook. (1997a). Expression of Rap 1 suppresses genomic instability of H-ras transformed mouse fibroblasts. *Somat. Cell Mol. Genet.*, **23**, 2, 123-133.

Wani, M. A., N. C. Denko & P. J. Stambrook. (1997b). Expression of Rap 1 suppresses genomic instability of H-ras transformed mouse fibroblasts. *Somat Cell Mol Genet*, **23**, 2, 123-33.

Watanabe, M., K. Nomura, A. Ohyama, R. Ishikawa, Y. Komiya, *et al.* (2005). Myosin-Va regulates exocytosis through the submicromolar Ca<sup>2+</sup>-dependent binding of syntaxin-1A. *Mol Biol Cell*, **16**, 10, 4519-30.

Williamson, A. R. (1999). The Merck Gene Index project. *Drug Discov Today*, **4**, 3, 115-122.

Willingham, A. T. & T. R. Gingeras. (2006). TUF love for "junk" DNA. *Cell*, **125**, 7, 1215-20.

Wistow, G., S. L. Bernstein, S. Ray, M. K. Wyatt, A. Behal, *et al.* (2002). Expressed sequence tag analysis of adult human iris for the NEIBank Project: steroid-response factors and similarities with retinal pigment epithelium. *Mol. Vis.*, **8**, 185-195.

Woolfe, A., M. Goodson, D. K. Goode, P. Snell, G. K. McEwen, *et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, **3**, 1, e7.

Wutz, A., O. W. Smrzka, N. Schweifer, K. Schellander, E. F. Wagner, *et al.* (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature*, **389**, 6652, 745-9.

- Xu, L. L., N. Shanmugam, T. Segawa, I. A. Sesterhenn, D. G. Mcleod, *et al.* (2000). A novel androgen-regulated gene, PMEPA1, located on chromosome 20q13 exhibits high level expression in prostate. *Genomics*, **66**, 3, 257-63.
- Yan, M. D., C. C. Hong, G. M. Lai, A. L. Cheng, Y. W. Lin, *et al.* (2005). Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas. *Hum. Mol. Genet.*, **14**, 11, 1465-1474.
- Yekta, S., I. H. Shih & D. P. Bartel. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 5670, 594-6.
- Yelin, R., D. Dahary, R. Sorek, E. Y. Levanon, O. Goldstein, *et al.* (2003). Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.*, **21**, 4, 379-386.
- Yu, Y., F. Xu, H. Peng, X. Fang, S. Zhao, *et al.* (1999a). NOEY2 (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. *Proc Natl Acad Sci U S A*, **96**, 1, 214-9.
- Yu, Y., F. Xu, H. Peng, X. Fang, S. Zhao, *et al.* (1999b). NOEY2 (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 1, 214-219.
- Zohn, I. M., S. L. Campbell, R. Khosravi-Far, K. L. Rossman & C. J. Der. (1998). Rho family proteins and Ras transformation: the RHOad less traveled gets congested. *Oncogene*, **17**, 11 Reviews, 1415-38.

## 8 LISTA DE ANEXOS

### Anexos inclusos no CD-ROM complementar a esta Tese

**ANEXO 1:** Artigo publicado - Reis, E. M., R. Louro, H. I. Nakaya & S. Verjovski-Almeida. (2005). As antisense RNA gets intronic. *Omics*, **9**, 1, 2-12.

**ANEXO 2:** Artigo submetido - Nakaya, H. I., P. P. Amaral, R. Louro, A. Lopes, A. A. Fachel, *et al.* (2006). Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.*

**ANEXO 3:** Artigo publicado - Reis, E. M., H. I. Nakaya, R. Louro, F. C. Canavez, A. V. Flatschart, *et al.* (2004). Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene*, **23**, 39, 6684-6692.

**ANEXO 4:** Artigo publicado - Louro, R., H. I. Nakaya, A. C. Paquola, E. A. Martins, A. M. Da Silva, *et al.* (2004). RASL11A, member of a novel small monomeric GTPase gene family, is down-regulated in prostate tumors. *Biochem Biophys Res Commun*, **316**, 3, 618-27.

**ANEXO 5:** Tabela Suplementar 1 - Transcritos exônicos regulados por andrógeno.

**ANEXO 6:** Tabela Suplementar 2 - Transcritos intrônicos regulados por andrógeno.

**ANEXO 7:** Artigo Submetido - Louro, R., H. I. Nakaya, P. P. Amaral, F. Festa, M. C. Sogayar, *et al.* (2006). Androgen responsive intronic non-coding RNAs. *BMC Genomics*.

**ANEXO 8:** Tabela Suplementar 3 - Assinatura de expressão de transcritos intrônicos conservada em tecidos humanos e de camundongos.

**ANEXO 9:** Tabela Suplementar 4 - Categorias de GO enriquecidas de transcritos intrônicos imunoprecipitados com RBPs.

**ANEXO 10:** Tabela Suplementar 5 - Transcritos exônicos e intrônicos alterados pelo tratamento com  $\alpha$ -amanitina.

**ANEXO 11:** Súmula Curricular.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)