

**IDENTIFICAÇÃO DE MARCADORES MOLECULARES  
DE PROGNÓSTICO PARA ADENOCARCINOMA DE  
PRÓSTATA UTILIZANDO ABORDAGEM  
QUANTITATIVA BASEADA EM SEQUÊNCIAS**

**GUSTAVO DE CAMPOS MOLINA**

**Tese apresentada à Fundação Antônio  
Prudente para obtenção do título de Doutor  
em Ciências**

**Área de Concentração: Oncologia**

**Orientadora: Dra. Dirce Maria Carraro**

**São Paulo**

**2012**

## FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Molina, Gustavo de Campos.

**Identificação de marcadores moleculares de prognóstico para adenocarcinoma de próstata utilizando abordagem quantitativa baseada em sequências** / Gustavo de Campos Molina. São Paulo; 2012. 116p.

Tese (Doutorado)-Fundação Antônio Prudente.

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

Orientadora: Dirce Maria Carraro

Descritores: 1. MARCADORES BIOLÓGICOS. 2. NEOPLASIAS DA PRÓSTATA 3. SEQUENCIAMENTO. 4. EXPRESSÃO GÊNICA. 5. PROGNÓSTICO

*"Porque, agora, vemos como em espelho, obscuramente; então, veremos face a face. Agora, conheço em parte; então, conhecerei como também sou conhecido."*

*1ª Coríntios 13: 12*

## DEDICATÓRIA

Dedico esse trabalho a Deus, por me conceder o dom da vida e me permitir completar mais um importante passo na minha jornada.

A minha amável esposa Amanda, pela compreensão da minha ausência em determinados momentos, mas acima de tudo pelo carinho, amor e encorajamento nos momentos difíceis.

Aos meus pais, Léia e Carlos, por toda dedicação e apoio incondicional que me deram durante esses quatro anos.

Aos meus irmãos Thais e Bruno por estarem sempre do meu lado mesmo à distância, e aos meus novos familiares, Meyres e André pelo incentivo e carinho.

Sem vocês, não faria sentido chegar até aqui, muito obrigado a todos.

## AGRADECIMENTOS

Meus sinceros agradecimentos...

... em especial à minha orientadora, Dra. Dirce Maria Carraro, pela amizade, por tantas oportunidades e por todo o apoio e dedicação. Por todos os momentos que passamos elaborando minuciosamente cada etapa deste trabalho. Muito obrigado pelo aprendizado e por apoiar minha decisão profissional.

...à Dra. Elisa Ferreira e Napolitano, pela contribuição constante e fundamental no decorrer desses quatro anos, mas acima de tudo pela convivência e exemplo de dedicação.

... à Giovana Torrezan, pelo enorme suporte com os experimentos neste último um ano e meio.

... ao pessoal do Laboratório de Genômica e Biologia Molecular, fundamentais de tantas formas para o desenvolvimento do trabalho, Bruna D. F. Barros, Eloisa Oliviere, Tatiana I. Ricca, Bianca C. G. Lisboa, Carolina Sens, Louise D. Mota, Alex Fiorini, Márcia, Roberto Ferreira e Cristiane Gonçalves. E também a Letícia A. Martins, Fábio S. Piccole, Mev D. Valentin, Vera C. Prescinoti, Elen P. Bastos, Maria Cristina Rangel, Nádia P. Castro, Thiago Saraiva e Aderbal R. Silva, que já não estão no laboratório.

... a Dra. Helena Brentani, ao Renato Puga e Leandro Lima pelas análises de bioinformática desse trabalho.

... ao Dr. Francisco Fonseca, pela contribuição dos dados clínicos das amostras utilizadas neste trabalho

... ao Dr. Antônio Hugo Campos pela participação da avaliação da amostras e microdissecção a *laser* e na padronização dos anticorpos juntamente com Dr. Rafael Malagoli, André Lavorato e Sueli.

... ao, Severino Silva e Carlos Nascimento, do departamento de Anatomia Patológica que auxiliarem com os experimentos desse trabalho.

... aos responsáveis pela Pós-Graduação, especialmente Ana Maria Kurinari, e Luciane Pitombeira.

... aos responsáveis pela Biblioteca: Renata e Sueli Francisco pela formatação da tese.

... aos meus supervisores do Cristália: Adriana Yokomizo, Marcos Alegria e Milena Oliveira, pela compreensão dos momentos que precisei me ausentar; e aos meus colegas de trabalho Renata Carvalho, Aline Leal, Marcelo Aguiar, André Barros e Nadja Andrade que me ajudaram em muito durante minha ausência.

E a todas às pessoas que, direta ou indiretamente, me auxiliaram e contribuíram na realização deste estudo. Este trabalho é o resultado de todos vocês!

*O presente trabalho contou com a participação e apoio de um grande pesquisador da ciência brasileira: Professor Dr. Ricardo Renzo Brentani. Esse grande personagem não somente nos ensinou a fazer ciência, mas também a amar a pesquisa e retribuir para a sociedade o investimento que ela nos proporciona.*

*A vida neste mundo é passageira e nosso querido professor já descansou, mas que sua obra seja promulgada com muito empenho e carinho através dos muitos alunos que ele formou.*

## RESUMO

Molina GC. **Identificação de marcadores moleculares de prognóstico para adenocarcinoma de próstata utilizando abordagem quantitativa baseada em sequências.** São Paulo; 2012. [Tese de Doutorado-Fundação Antônio Prudente].

Câncer de próstata (CaP) é um dos cânceres mais comuns na população masculina sendo a segunda causa de morte por neoplasias no ocidente. O adenocarcinoma de próstata é classificado de acordo com a escala de Gleason em graus de malignidade que variam de 2 a 10. Os tumores de escore de Gleason 7 apresentam respostas variáveis ao tratamento, e são sub classificados em 4+3 ou 3+4 tendo, geralmente o primeiro apresenta um pior prognóstico quanto a recidiva bioquímica. A falta de correspondência da classificação da biopsia e da peça cirúrgica, e a inespecificidade do teste de PSA sérico têm incentivado estudos de identificação de novos marcadores moleculares, principalmente em relação aos tumores Gleason 7 nos quais a evolução da doença é de difícil predição. Assim o objetivo desse estudo foi identificar putativos marcadores de prognóstico de pacientes com CaP de escore de Gleason 7. Neste estudo foi elaborado um protocolo de construção de biblioteca de cDNA que combinou microdissecção à laser e a utilização da plataforma de sequenciamento Roche-454 para acessar o transcrito de células tumorais de próstata. Foram construídas bibliotecas de cDNA 3' de 21 amostras de adenocarcinoma de próstata, 11 sem recidiva e 10 com recidiva bioquímica, e sequenciadas em paralelo no 454. Foram geradas um total 930.102 sequências, com um tamanho médio de 149,29pb, as quais identificaram 19.933 transcritos e 14.102 genes distintos. Foram aplicados dois métodos matemáticos para identificar diferenças transcricionais entre os grupos de amostras. A análise de diferença de expressão pelo método de proporção identificou 246 transcritos entre as amostras de CaP com e sem recidiva bioquímica. O padrão de expressão



desses 246 genes permitiu discriminar 100% das amostras dos grupos com e sem recidiva através da clusterização hierárquica. Um segundo método, baseado na probabilidade de um gene, ou conjunto de genes, ser mais expresso em um determinado grupo identificou 314 transcritos. Entre os dois métodos houve uma correspondência de 150 transcritos em comum. Dos 76 genes escolhidos para validação por RT-qPCR, 57 foram identificados pelos dois métodos e de forma geral, apresentaram níveis de validação semelhantes. Um total de 29 de 76 (38,2%) e 22 de 57 (38,6%) tiveram a diferença de expressão confirmada pelo método de proporção e probabilidade, respectivamente. Interessantemente, foi observado um aumento no percentual de validação entre os genes com valores extremos de probabilidade. O método de probabilidade permitiu identificar um conjunto de 6 genes (*RPL35*, *RPS28*, *C12orf57*, *PODXL*, *SRSF5*, *LOC100293090*) que discriminaram os grupos com e sem recidiva com uma probabilidade de 98,73%. Dois genes, *BRP44* e *NFKBIZ*, identificados como mais expressos em amostras que não recidivaram, foram selecionados para avaliar o potencial prognóstico em amostras independentes pelo padrão de expressão de suas respectivas proteínas, BRP44 e IκB-ζ. A expressão da proteína IκB-ζ foi associada com menor incidência de vesícula seminal comprometida (P=0,03) e com menor escore de Gleason (P=0,01), sugerindo que a positividade de IκB-ζ é um fator de prognóstico favorável em adenocarcinoma de próstata. Não foram observadas associações estatisticamente significativas entre a expressão da proteína BRP44 e as características clínico patológicas. Apesar da necessidade de avaliar nossos resultados em um grupo independente de amostras, nossos dados mostram que o método de construção de biblioteca cDNA a partir de células tumorais capturadas a laser combinado com o método matemático baseado em probabilidade apresentou um grande potencial para identificar marcadores moleculares aptos para discriminar amostras de adenocarcinoma de próstata Gleason 7 com diferentes evoluções.

## SUMMARY

Molina GC. **[Identification of prognostic molecular markers for prostate adenocarcinoma exploiting a quantitative approach based on sequencing]**. São Paulo; 2012. [Tese de Doutorado-Fundação Antônio Prudente].

Prostate cancer is one of the most common forms of cancer within male population, being the second major cause of death due to neoplasia in the Western-World. The prostate adenocarcinoma malignancy is classified according to the Gleason scale, varying between 2 and 10. Tumors having a Gleason score of 7 present variable responses to the treatment, and are sub-classified in 4+3 or 3+4, where usually the first presents a worse prognosis when considering biochemical relapse. The lack of correlation between the different systematization of the biopsy and surgical specimen itself and specificity of the PSA test have encouraged studies which wish to identify new molecular markers, especially those related to tumors with Gleason score 7 whose development portrays a hard prediction. Therefore, the aim of this study was to identify putative prognostic markers of patients with Gleason 7 prostate cancer. This work combined the elaboration of a protocol for building a cDNA library from laser microdissected samples and a methodology which uses 454 platform for analyzing the transcriptome in a quantitative manner. 3'-end cDNA libraries were built from 21 samples of prostate adenocarcinoma, 11 not presenting relapse and 10 with biochemical relapse, all of which were sequenced in parallel on the 454. 930,102 sequences were generated, having an average size of 149.29 bp, of which 19,933 transcripts and 14,102 genes were identified. The differentially expressed genes between relapse and non relapse samples were assessed by two mathematical methods. The proportion method identified 246 transcripts differentially expressed when analyzing prostate cancer with and without biochemical relapse. The expression pattern of these 246 genes

allowed us to distinguish 100% of the samples from both groups, with and without relapse, through hierarchical clustering. The second mathematical method, based in the odds of a gene, or a gene set, to be more expressed in one of the samples group, identified 314 differentially expressed transcripts. Between both methods there was an overlap of 150 transcripts. Of the 76 chosen genes for confirmation by RT-qPCR, 57 were pointed out by both methods. A total of 29 from 76 (38.16%) and 22 from 57 (38,6%) had their difference in expression confirmed through RT-qPCR. Notably, we found a major ratio of validation between the genes with extreme values of probability. The odds method allowed indentifying a set of 6 genes (RPL35, RPS28, C12orf57, PODXL, SRSF5, LOC100293090) which discriminate the relapse and non relapse group with 98.73% of probability. Two genes, *BRP44* e *NFKBIZ*, identified as more expressed between non relapse samples, were chosen in order to analyze the potential prognosis in independent samples based on the expression pattern of their respective protein, BRP44 e I $\kappa$ B- $\zeta$ . The expression of protein I $\kappa$ B- $\zeta$  was associated with the lowest incidence of compromised seminal vesicle ( $P=0.03$ ) and with the lowest Gleason score ( $P=0.01$ ), indicating indicates that the presence of I $\kappa$ B- $\zeta$  may be a favorable factor in prostate adenocarcinoma prognosis. It was not possible to observe any significant correlation between the expression protein BRP44 and the pathological characteristics of the samples. Although there is the need to evaluate our results in an independent sample set, this work shows that combining the cDNA library method construction from laser microdissected cells with the odds mathematical method may reveal a powerful tool to identify molecular markers to discriminate Gleason 7 prostate cancer samples with distinct prognosis.

## LISTA DE FIGURAS

<b>Figura 1</b>	Representação do adaptador 454 <sup>a</sup> .....	28
<b>Figura 2</b>	Esquema da metodologia padronizada para a construção da biblioteca de cDNA para sequenciamento no 454.....	30
<b>Figura 3</b>	Eletroforese capilar da amplificação do RNA total da linhagem celular Hb4.....	47
<b>Figura 4</b>	Eletroforese em gel de agarose 1% da amostra amplificadas por 15 ciclos de PCR após ligação dos adaptadores.....	49
<b>Figura 5</b>	Representação da microdissecção a laser das células epiteliais das glândulas prostáticas.....	50
<b>Figura 6</b>	Resultados da transcrição <i>in vitro</i> e amplificação por PCR de uma amostra microdissecada a laser.....	53
<b>Figura 7</b>	Resultados de purificação das amostras com o kit Agencourt <sup>®</sup> AMPure <sup>®</sup> XP.....	56
<b>Figura 8</b>	Eletroforese capilar no DNA 1000pb Chip do Bionalyser.....	57
<b>Figura 9</b>	Classificação das categorias biológicas dos genes diferencialmente expressos de acordo com o KEGG.....	67
<b>Figura 10</b>	Clusterização hierárquica baseada nos 246 genes diferencialmente expressos entre os grupos de amostras com recidiva e sem recidiva.....	69

<b>Figura 11</b>	Gráfico da média dos valores de estabilidade de expressão dos genes <i>18S</i> , <i>ALAS1</i> , <i>B2M</i> , <i>HPRT1</i> , <i>SDH</i> , <i>C2orf28</i> e <i>TSEN34</i> .....	75
<b>Figura 12</b>	Gráfico da expressão relativa dos genes <i>BRP44</i> e <i>NFKBIZ</i> com os dados da validação técnica por RT-qPCR.....	78
<b>Figura 13</b>	Discriminação das amostras com se sem recidiva com os genes classificadores.....	81
<b>Figura 14</b>	Expressão proteica de <i>BRP44</i> em adenocarcinoma de próstata.....	83
<b>Figura 15</b>	Expressão proteica de $\text{I}\kappa\text{B-}\zeta$ em adenocarcinoma de próstata.....	84
<b>Figura 16</b>	Curvas de Kaplan-Meier para cálculo de sobrevida livre de doença para $\text{I}\kappa\text{B-}\zeta$ e <i>BRP44</i> .....	86

## LISTA DE TABELAS

<b>Tabela 1</b>	Características clínico patológicas das amostras utilizadas na imunohistoquímica.....	40
<b>Tabela 2</b>	Distribuição das amostras de adenocarcinoma de próstata selecionadas para captura a <i>laser</i> e extração de RNA para confecção das bibliotecas a serem sequenciadas no 454.....	45
<b>Tabela 3</b>	Análise <i>in silico</i> das enzimas de corte frequente.....	48
<b>Tabela 4</b>	Resultado das etapas da construção das bibliotecas das 21 amostras sequenciadas no 454.....	52
<b>Tabela 5</b>	Resultado do sequenciamento do <i>pool</i> das 21 bibliotecas pelo método de Sanger.....	55
<b>Tabela 6</b>	Resultados do sequenciamento na plataforma 454 das 21 bibliotecas de cDNA 3' .....	59
<b>Tabela 7</b>	Genes diferencialmente expressos do subgrupo 1 entre as amostras de adenocarcinoma de próstata de pacientes com e sem recidiva bioquímica.....	61
<b>Tabela 8</b>	Genes diferencialmente expressos do subgrupo 2 entre as amostras de adenocarcinoma de próstata de pacientes com e sem recidiva bioquímica.....	64
<b>Tabela 9</b>	Representação dos transcritos com probabilidades maiores que 70% de serem mais expressos nos grupos com e sem recidiva..	71

<b>Tabela 10</b>	Comparação entre os resultados de diferença de expressão dos dados do 454 (RNA-Seq) com a validação técnica (RT-qPCR)...	77
<b>Tabela 11</b>	Distribuição dos 57 genes testados por RT-qPCR de acordo com a faixa de probabilidade.....	79
<b>Tabela 12</b>	Valores de probabilidade dos 12 genes escolhidos como putativos classificadores entre as amostras sem recidiva e com recidiva.....	80
<b>Tabela 13</b>	Associação entre expressão proteica de BRP44 e I $\kappa$ B- $\zeta$ com as características clinicopatológicas das amostras de adenocarcinoma de próstata usada no imunoistoquímica.....	85
<b>Quadro 1</b>	Processos biológicos com relevância estatística no grupo de genes diferencialmente expressos entre as amostras com e sem recidiva bioquímica de acordo com o <i>Gene Ontology</i> .....	66
<b>Quadro 2</b>	Descrição das categorias biológicos enriquecidas e dos respectivos genes mais expressos representados nas amostras com e sem recidiva bioquímica de acordo com o KEGG.....	68

## LISTA DE ABREVIACOES

<b>BRP44</b>	Nomenclatura do gene e de seu produto proteico Proteína de Cérebro 44, do inglês <i>Brain Protein 44</i>
<b>CaP</b>	Câncer de Próstata
<b>cDNA</b>	DNA complementar a fita de RNA
<b>cm</b>	centímetro
<b>DEPC</b>	Dietil Pirocarbonato
<b>DNA</b>	Ácido Desoxirribonucléico
<b>dNTPs</b>	Nucleotídeos de DNA
<b>EST</b>	Do inglês, <i>Expressed Sequence Tag</i>
<b>ETOH</b>	Etanol
<b>HE</b>	Hematoxilina-eosina
<b>HPB</b>	Hiperplasia Prostática Benigna
<b>IκB-ζ</b>	Proteína Inibidora Zeta do Fator Nuclear NF Kappa Beta, do inglês NF-kappa-B inhibitor zeta, codificada pelo gene NFKBIZ,
<b>LCM</b>	Do inglês, <i>Laser Capture Microdissection</i>
<b>M</b>	Molar
<b>mL</b>	Mililitro
<b>mM</b>	Milimolar
<b>mm</b>	milímetro
<b>MPSS</b>	Do inglês, <i>Massive Parallel Signature Sequencing</i> ,
<b>NaOH</b>	Hidróxido de Sódio
<b>NFKBIZ</b>	Nomenclatura do gene Inibidor Zeta da Proteína NF Kappa Beta, do inglês <i>NF-kappa-B inhibitor zeta</i>
<b>PCR</b>	Do inglês, <i>Polimerase Chain Reaction</i>
<b>PB</b>	par de base
<b>PFDR</b>	Do inglês, <i>positive False Discovery Ratio</i>
<b>PIN</b>	Do inglês <i>prostatic intraepithelial neoplasia</i>
<b>RNA</b>	Ácido Ribonucléico
<b>RNAa</b>	RNA amplificado



<b>RNA<sub>m</sub></b>	RNA mensageiro
<b>RNTP<sub>s</sub></b>	Nucleotídeos de RNA
<b>RPM</b>	Rotações por minuto
<b>RT-qPCR</b>	Do inglês, <i>Reverse Transcriptase quantitative PCR</i>
<b>SAGE</b>	Do inglês, <i>Serial Analysis of Gene Expression</i>
<b>U</b>	Unidade
<b>μg</b>	microgramas
<b>μg/μL</b>	microgramas por microlitro
<b>μL</b>	microlitro
<b>μm</b>	micrometro
<b>μM</b>	micromolar
<b>&gt;</b>	Maior
<b>≥</b>	Maior ou igual
<b>&lt;</b>	Menor

# ÍNDICE

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
1.1	Câncer de próstata: epidemiologia e aspectos clínicos .....	1
1.2	Genética e evolução do câncer de próstata .....	6
1.3	Biomarcadores em câncer de próstata .....	9
1.4	Estudos transcricionais em câncer de próstata .....	12
1.5	Análises transcricionais com sequenciadores de nova geração.....	16
<b>2</b>	<b>OBJETIVOS</b> .....	<b>22</b>
2.1	Objetivo geral .....	22
2.2	Objetivos específicos.....	22
<b>3</b>	<b>MATERIAL E MÉTODOS</b> .....	<b>23</b>
3.1	Seleção das amostras de adenocarcinoma de próstata para construção das bibliotecas de cDNA 3' .....	23
3.2	Estabelecimento do protocolo de construção de biblioteca de cDNA 3' .....	25
3.2.1	Amplificação do RNA total e síntese de cDNA .....	25
3.2.2	Digestão com enzima de restrição .....	26
3.2.3	Montagem dos adaptadores 454A.....	26
3.2.4	Seleção dos fragmentos 3' biotinilados e ligação do adaptador 454A...	28
3.2.5	Amplificação por PCR .....	29
3.3	Sequenciamento e validação das bibliotecas pelo método de Sanger..	30
3.3.1	Sequenciamento individual das bibliotecas de cDNA 3' .....	30
3.3.2	Análise das sequências geradas pelo método de Sanger .....	31
3.3.3	Montagem do pool e validação no ABI .....	31
3.4	Sequenciamento na plataforma GS FLX Titanium System 454 e análise dos dados.....	32
3.4.1	Mapeamento dos dados do sequenciamento 454 .....	32
3.4.2	Identificação dos genes diferencialmente expressos .....	33

3.4.2.1	Baseado em <i>Fold change</i> e/ ou no método de proporções.....	33
3.4.2.2	Baseado no método de cálculo de probabilidades .....	34
3.4.3	Anotação de vias e processos biológicos e clusterização hierárquica ..	35
3.5	Validação dos genes diferencialmente expressos.....	36
3.5.1	Amplificação do RNA e síntese do cDNA.....	36
3.5.2	Reação de RT-qPCR com o sistema Taqman Low Density Array (TLDA).....	37
3.5.3	Seleção do gene normalizador .....	38
3.5.4	Cálculo da expressão relativa dos ensaios de RT-qPCR .....	38
3.6	Imunoistoquímica (IHQ).....	39
3.6.1	Análise da expressão proteica de candidatos previamente validados por RT-qPCR.....	39
3.6.2	Reações de imunoistoquímica.....	41
3.6.3	Análise dos resultados de Imunoistoquímica.....	43
<b>4</b>	<b>RESULTADOS .....</b>	<b>44</b>
4.1	Seleção das amostras de adenocarcinoma de próstata .....	44
4.2	Estabelecimento do protocolo de construção de biblioteca de cDNA....	46
4.2.1	Amplificação do RNA total .....	46
4.2.2	Digestão com enzima de restrição .....	47
4.2.3	Amplificação por PCR .....	49
4.3	Construção das bibliotecas de cDNA 3' .....	50
4.4	Sequenciamento e validação das bibliotecas pelo método de Sanger..	53
4.4.1	Sequenciamento individual das bibliotecas de cDNA 3' .....	53
4.4.2	Montagem do <i>pool</i> e validação no ABI .....	54
4.5	Sequenciamento na plataforma GS FLX Titanium 454 e análise dos dados.....	58
4.5.1	Mapeamento das sequências geradas na plataforma GS FLX Titanium 454 .....	58
4.5.2	Genes diferencialmente expressos entre as amostras com recidiva e sem recidiva bioquímica .....	60
4.5.2.1	Baseado em <i>Fold change</i> e/ ou <i>p</i> valor.....	60

4.5.2.2	Baseado no cálculo de probabilidades.....	70
4.6	Validação técnica dos genes diferencialmente expressos .....	74
4.6.1	Reação de RT-qPCR com o sistema Taqman Low Density Array (TLDA).....	74
4.6.2	Análise dos resultados das reações de RT-qPCR dos candidatos selecionados pelo método de proporção .....	76
4.6.3	Análise dos resultados das reações de RT-qPCR dos candidatos selecionados pelo método de probabilidades.....	78
4.7	Identificação de genes classificadores baseado no método de probabilidades .....	79
4.8	Validação biológica dos genes diferencialmente expressos por imunistoquímica.....	82
4.8.1	Reações de imunistoquímica e classificação das amostras .....	82
4.8.2	Avaliação do potencial prognóstico de BRP44 e NFKBIZ .....	84
<b>5</b>	<b>DISCUSSÃO .....</b>	<b>87</b>
<b>6</b>	<b>CONCLUSÕES .....</b>	<b>96</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>97</b>

## **ANEXOS**

**Anexo 1** Lista dos 122 Processo biológicos do Gene Ontology anotados nas amostras de CaP com  $p < 0,01$

**Anexo 2** Lista das 51 vias bioquímicas do KEGG enriquecidas nas amostras de CaP com  $p < 0,01$

**Anexo 3** Lista dos 896 genes diferencialmente expressos identificados pelo método de probabilidades

# 1 INTRODUÇÃO

## 1.1 CÂNCER DE PRÓSTATA: EPIDEMIOLOGIA E ASPECTOS CLÍNICOS

O câncer de próstata (CaP) é a neoplasia mais comum no gênero masculino após as neoplasias cutâneas, sendo a segunda em número de óbitos diretamente relacionado com a doença nos Estado Unidos (JEMAL et al. 2011). No Brasil, segundo as estimativas do Instituto Nacional do Câncer (INCA), serão cerca de 60 mil novos casos para o ano de 2012. Os fatores de risco do CaP incluem idade avançada, etnia afro americana, história familiar e dieta, sendo que a história familiar é o fator mais relevante (MACKINNON et al. 2009).

O CaP é comumente descrito como um câncer multifocal, uma vez que tumores primários, via de regra, aparecem em regiões distintas e com características histológicas e genéticas heterogêneas (AIHARA et al. 1994; MACINTOSH et al. 1998; BOSTWICK et al. 1998). Enquanto que em determinados casos o tumor pode se manter confinado à glândula prostática sem afetar a qualidade de vida do paciente, o oposto pode ocorrer nos casos que progridem rapidamente para metástase podendo levar os pacientes a óbito (MACKINNON et al. 2009).

Os principais métodos de investigação de alterações na próstata baseiam-se no exame digital retal, ou toque retal (TR), e na dosagem sérica

dos níveis de Antígeno Prostático Específico (PSA – do inglês *Prostate Specific Antigen*).

O TR é um exame rápido, de baixo custo e fácil realização. Por este procedimento é possível reconhecer as dimensões, o formato, os limites do órgão e alterações de consistência e mobilidade, tornando-o assim essencial na decisão da conduta clínica. Apesar da facilidade, este exame tem certa deficiência em detectar tumores na extremidade oposta da glândula que passam despercebidos pelo toque e, além disso, a maioria dos tumores detectados por TR são tumores com evolução mais adiantada (MAHON 2005).

A molécula do PSA é uma serina protease secretada pelas células epiteliais da próstata e tem importante função na liquefação dos fluidos seminais (BICKERS e AUKIM-HASTIE 2009). A introdução na rotina do exame de diagnóstico do PSA permitiu um aumento na detecção do CaP em estágios menos avançados (ILYIN et al. 2004) e localmente confinados (MILLER et al. 2003).

A concentração do PSA total varia conforme a idade, raça e peso, e a sua medida pode ser diretamente influenciada pelo modo como é armazenado o anticorpo e pela maneira como é realizada a calibração durante sua detecção (LILJA et al. 2008).

O teste do PSA tem uma baixa especificidade e muitas vezes a distinção entre hiperplasia prostática benigna (HPB), câncer de próstata, ou mesmo prostatite, se torna difícil apenas com os resultados do exame. Dos pacientes compreendidos na faixa de concentração sanguínea do PSA entre

3 a 15 ng/mL, cerca de 70 a 80% das biopsias são negativas (HUGOSSON et al. 2004; MAATTANEN et al. 2007), e 30% dos tumores extraídos de prostatectomia radical são classificados como pouco agressivos (SARDANA et al. 2008). Em contra partida, alguns tipos de câncer de próstata são extremamente agressivos e não produzem PSA acima do considerado normal (PRYOR e SCHELLHAMMER 2002).

Até o momento, não se chegou a um valor consenso da concentração de PSA total que possa discriminar com segurança cânceres de alto risco daqueles de baixo risco (BARRY 2009). A concentração de PSA total de 4 ng/mL é normalmente empregada como valor de corte no diagnóstico do câncer de próstata, no entanto, há evidências de que esse número é arbitrário, refletindo a sua baixa especificidade (VICKERS e LILJA 2009). Em pacientes de idade mais avançada a acurácia do PSA total é ainda mais reduzida uma vez que a glândula prostática aumenta de tamanho com o passar dos anos e, conseqüentemente, aumenta a produção do PSA (VICKERS et al. 2007).

Apesar de sua especificidade limitada no diagnóstico da CaP, o teste de PSA é de grande valia no acompanhamento dos pacientes submetidos a prostatectomia radical. Após a remoção da glândula prostática juntamente com as vesículas seminais, o paciente é acompanhado com o exame de PSA a cada três ou quatro meses. Uma concentração sérica de PSA acima de 0,2 ng/mL é considerada como recidiva bioquímica e pode significar a presença de tumor residual ou metástase fora da loja prostática. Uma vez

confirmada a recidiva do tumor, o paciente é submetido a radio ou quimioterapia dependendo do quadro clínico (PAPAREL et al. 2009).

Após ser detectada a alteração da glândula prostática pelo TR ou ocorrer um aumento na concentração sérica do PSA de forma inesperada, o paciente é submetido à biópsia para confirmação do tumor. Em casos positivos sua classificação histológica é fornecida pelo escore de Gleason (GLEASON 1966).

O princípio dessa classificação consiste na análise do grau de diferenciação celular do tumor, sendo ranqueada em cinco padrões histológicos distintos. Números de 1 (mais diferenciado) a 5 (menos diferenciado) são atribuídos a lesão mais frequente na biópsia, e outro número, também de 1 a 5, é atribuído ao segundo padrão mais prevalente, sendo que a soma dos dois valores resulta no escore de Gleason (GLEASON 1966; CHAN et al. 2000). No entanto, mesmo sendo o fator de prognóstico mais utilizado, o escore de Gleason apresenta uma correspondência da classificação da biópsia ligeiramente inferior em relação à peça após a prostatectomia radical (STARK et al. 2009).

Em vias de regra, os tumores de Gleason 3 a 6 apresentam boa resposta terapêutica, enquanto que os tumores 8 a 10 tem mau prognóstico com conseqüente diminuição da sobrevida do paciente (D'AMICO et al. 1998). No entanto, os tumores de escore de Gleason 7 apresentam respostas variáveis ao tratamento, o que resultam em uma evolução incerta da doença quanto a sua recidiva.



Os tumores de próstata de escore de Gleason 7 são subclassificados de acordo com o grau de diferenciação das lesões mais frequentes. Sendo que tumores classificados como 3+4 possuem a lesão principal moderadamente diferenciada e apresentam um melhor prognóstico quando comparados aos tumores 4+3, com a lesão principal pouco diferenciada (BURDICK et al. 2009). A diferença entre os dois subtipos de tumores Gleason 7 é muito discreta e dependente de análise subjetiva, o que dificulta a decisão do melhor tipo de tratamento com consequências diretas na recidiva bioquímica (CHAN et al. 2000; RASIAH et al. 2003).

O tratamento mais empregado contra o CaP é a remoção completa da glândula prostática chamada de prostatectomia radical. Entretanto, 25 a 40% dos pacientes submetidos a esse procedimento sofrem recidiva bioquímica em 10 anos (HAN et al. 2001; AMLING 2006).

Após a prostatectomia radical é feito o levantamento das características anatomopatológicas como: extensão do tumor, invasão perineural, extensão extraprostática, acometimento da vesícula seminal, margem cirúrgica, número de linfonodos acometidos e escore de Gleason da peça. Essas informações, mais o histórico do PSA são avaliadas com o auxílio do nomograma.

O nomograma é um painel multivariável para o prognóstico do paciente que permite calcular a probabilidade de tempo livre de doença de 5 a 7 anos (KATTAN et al. 1999). Apesar de ser a ferramenta de prognóstico mais empregada, nem sempre é possível determinar corretamente o tempo livre de doença com o nomograma. Desta forma, a ausência de um

marcador prognóstico independente, que seja específico e sensível na triagem de subgrupos de pacientes com câncer de próstata, têm incentivado estudos de identificação de marcadores moleculares, principalmente em tumores de escore de Gleason 7, nos quais, a evolução da doença é de difícil predição.

## 1.2 GENÉTICA E EVOLUÇÃO DO CÂNCER DE PRÓSTATA

Apesar de carecer de provas contundentes, o evento mais aceito como precursor do CaP é a neoplasia prostática intraepitelial (PIN – do inglês *prostatic intraepithelial neoplasia*) (DEMARZO et al. 2003). A PIN é caracterizada por ter a aparência do epitélio luminal hiperplásico, redução no número de células basais, aumento do núcleo e nucléolo, hiper Cromasia citoplasmática e atipia nuclear; além disso, lesões de PIN de alto grau geralmente possuem marcação positiva para marcadores de proliferação celular (SHAPPELL et al. 2004).

Uma importante diferença histológica entre o CaP e a PIN é a presença de células basais no segundo tipo histológico, mas ausentes no CaP. Marcadores de células basais como p63 e citoceratina podem ser empregados para distinção desses dois tipos histológicos (HUMPHREY 2007; GRISANZIO e SIGNORETTI 2008).

A redução da expressão do gene homeobox *NKX3.1* demonstra ser frequentemente um evento crítico no processo inicial do desenvolvimento do câncer de próstata (ABATE-SHEN et al. 2008). Este gene está localizado no

cromossomo 8p21.2, região com perda de heterozigose em até 85% da lesões de neoplasia intraepitelial prostática de alto grau (EMMERT-BUCK et al 1995; VOCKE et al. 1996; HÄGGMAN et al. 1997; SWALWELL et al. 2002; BETHEL et al. 2006). Em casos onde não ocorre a perda de heterozigose, é comum observar o silenciamento do gene *NKX3.1* por meio de alterações epigenéticas como hipermetilação da região promotora (ASATIANI et al. 2005).

A inativação do gene *NKX3.1* em camundongos adultos resulta em uma resposta deficiente ao dano por estresse oxidativo, enquanto que a expressão desse gene em linhagens celulares de câncer de próstata humano protege contra o dano ao DNA de forma regulada por resposta inflamatória (OUYANG et al. 2005; MARKOWSKI et al. 2008; BOWEN e GELMANN 2010). Esses e outros achados têm sugerido que o gene *NKX3.1* representa um gene supressor de tumor haploinsuficiente que age como um guardião dos eventos iniciais do desenvolvimento do câncer de próstata (KIM et al. 2002a; GELMANN 2003).

Recentes estudos revelam que o aumento de expressão do gene *MYC* estaria envolvido com o processo inicial do câncer de próstata. Em muitas neoplasias intraepiteliais prostáticas, a proteína nuclear MYC está com a expressão aumentada mesmo sem ocorrer amplificação gênica (GUREL et al. 2008). Em modelos com camundongos transgênicos, o aumento da expressão do gene *MYC* desencadeia a rápida formação de PIN seguida por uma progressão para adenocarcinoma invasivo e, em certas ocasiões, para metástase (ELLWOOD-YEN et al. 2003).

Outro evento molecular frequentemente encontrado no câncer de próstata são rearranjos cromossômicos que ativam a expressão dos fatores de transcrição da família ETS, como *ERG*, *ETV1* e *ETV4* (ILJIN et al. 2006; MEHRA et al. 2007; HU et al. 2008; MOSQUERA et al. 2008; SARAMÄKI et al. 2008)

O rearranjo mais comum entre esses genes origina a fusão gênica *TMPRSS2-ERG*, resultando na expressão do gene *ERG* truncado e controlada pelo promotor responsivo a andrógeno do gene *TMPRSS2* (TOMLINS et al. 2005; PERNER et al. 2006; WANG et al. 2006; CLARK et al. 2007). A fusão *TMPRSS2-ERG* é encontrada com uma frequência de 15% em lesões de neoplasia intraepitelial prostática de alto grau e em 50% dos casos de CaP localizado (CLARK et al. 2008; MOSQUERA et al. 2008; ALBADINE et al. 2009), sugerindo que esse rearranjo ocorre depois do processo inicial de formação do câncer de próstata mas que predispõe a progressão clínica da doença.

Outro gene frequentemente encontrado com expressão reduzida em CaP é o *PTEN* (SALMENA et al. 2008). Como resultado em comum de trabalhos independentes, ficou comprovado que o gene *PTEN* sofre uma perda do número de cópias como um evento inicial do CaP, perda esta correlacionada com a progressão para um câncer mais agressivo e hormônio refratário. Em modelos animais, a inativação do gene *PTEN* demonstra-se contribuir com a perda de função da proteína codificada pelo gene homeobox *NKX3.1*, com o aumento na expressão do proto-oncogene

*MYC* e com a formação da fusão *TMPRSS-ERG* (KIM et al. 2002b; CARVER et al 2009; KIM et al. 2009; KING et al. 2009).

A perda de expressão de *PTEN* resulta no aumento da sinalização da via Akt/mTOR por meio da ativação da proteína AKT1 (THOMAS et al. 2004; CHEN et al. 2006; MULHOLLAND et al. 2006; SHEN e ABATE-SHEN 2007). Juntamente com a via da MAPK/Erk, a ativação dessas vias promovem a progressão do tumor e resistência à castração em modelos com linhagens celulares e camundongos (UZGARE e ISAACS 2004; GAO et al. 2006). Enquanto que a inativação conjunta dessas vias inibe o desenvolvimento de tumores hormônio refratários em camundongos (KINKADE et al. 2008).

### **1.3 BIOMARCADORES EM CÂNCER DE PRÓSTATA**

Com a inclusão do teste do PSA, o diagnóstico do CaP ocorre cada vez mais, em estágios menos avançados da doença quando comparado a realidade do final da década de 90. No entanto, a determinação da melhor e mais acurada conduta clínica é prejudicada pela carência de marcadores de prognóstico específicos. Este fato tem incentivado a busca de novos biomarcadores.

A enzima AMACR, localizada dentro do peroxissomo e com papel na  $\beta$ -oxidação de ácidos graxos ramificados, foi identificada como promotora de crescimento em câncer de próstata de forma independente de andrógeno (KUEFER et al. 2002; ZHA et al. 2003). Em 88% dos casos, a expressão do RNAm da AMACR no tumor está aumentado em 9 vezes quando comparado

com o tecido normal (ROGERS et al. 2004). A detecção da AMACR no soro mostrou ter maior sensibilidade e especificidade do que o PSA para o diagnóstico de câncer de próstata (SREEKUMAR et al. 2004).

Um fator negativo da AMACR é que em outros tipos de câncer de pacientes com doença autoimune sua expressão também pode ser detectada no soro (TRICOLI et al. 2004). No entanto, seu emprego em conjunto com o PSA aumenta a capacidade de diagnosticar o CaP confinado à glândula ou metastático quando comparado ao teste do PSA isolado (CARDILLO et al. 2005).

O gene supressor de tumor, Caveolina 1, é frequentemente encontrado com expressão reduzida em CaP, e em tumores mais agressivos sua expressão é ainda menor. A concentração de Caveolina 1 plasmática pré operatória pode ser uma ferramenta no prognóstico do CaP (DI VIZIO et al. 2009; THOMPSON et al. 2010).

Além dos marcadores proteicos mencionados acima, o RNA não codificante (ncRNA) do gene *PCA3* tem expressão aumentada nos tecidos neoplásicos cerca de 66 vezes em relação ao tecido normal e vem sendo cada mais empregado na clínica como método auxiliar na classificação dos tumores de câncer de próstata (BUSSEMAKERS et al. 1999; DE KOK et al. 2002), principalmente por ser de fácil detecção por RT-qPCR em amostras de urina após o toque retal (KIRBY et al. 2009). Apesar do *PCA3* ter uma menor sensibilidade no diagnóstico do CaP ele se mostrou mais específico em relação ao PSA e quando utilizado de forma combinada esses dois

marcadores tiveram maior precisão na predição do CaP (MARKS et al. 2008).

Poucas associações de polimorfismos com câncer de próstata têm sido estabelecidas, devido à alta prevalência desse câncer na população masculina de mais idade e também na heterogeneidade da doença (REYNOLDS et al. 2007). As mutações correlacionadas com a susceptibilidade ao câncer de próstata são encontradas nos genes *HPC1*, *HPC2*, *HPCX*, *HPC20*, *CAPB*, *PCAP* (HSING e CHOKKALINGAM 2006), além de mutações no gene *BRCA1* (GILLANDERS et al. 2004) e *BRCA2* (Anonymous 1999). Um estudo recente associou a presença de um polimorfismo no gene da esteroide-5- $\alpha$ -redutase (*SRD5A2*), diminuindo a atividade da enzima com maior risco de desenvolver um câncer agressivo (CUSSENOT et al. 2007).

A hipermetilação de ilhas CpGs localizadas nas regiões promotoras de genes supressores de tumor é um importante mecanismo de inativação gênica e está associada a vários tipos de câncer, alterando vias importantes de regulação levando à transformação e progressão do tumor (JONES e BAYLIN 2007). A primeira alteração epigenética descrita em câncer de próstata é a hipermetilação no promotor do gene pirofosfato S-transferase 1, *GSTP1*, envolvida da desintoxicação protegendo as células de radicais livres (LEE et al. 1994; HAYES e PULFORD 1995).

Essa alteração é a mais comum em câncer de próstata, sendo encontrada em até 90% dos casos (SINGAL et al. 2004). Foi proposto um método de detecção não invasivo, utilizando amostras de urina após o toque

retal, mas o emprego do bissulfito em um DNA de baixa qualidade limitou a implementação deste método (FRADET 2009). JERÓNIMO et al. (2004) relataram que a dosagem combinada dos níveis de metilação de GSTP1 e APC, aumentou a sensibilidade para detectar o câncer de próstata em 98,3%.

Outras regiões promotoras hipermetiladas foram correlacionadas com o desenvolvimento do câncer, como a do membro da subfamília A do citocromo P450, P4501A1, (OKINO et al. 2006) e do inibidor tecidual de metaloproteinase, TIMP-2, (PULUKURI et al. 2007). No entanto, são necessários mais estudos de correlação para empregar essas metilações como marcadores moleculares na rotina (FEBBO 2009).

#### **1.4 ESTUDOS TRANSCRICIONAIS EM CÂNCER DE PRÓSTATA**

O transcrito de uma célula é o conjunto total dos transcritos expressos em uma determinada condição fisiológica e a taxa de expressão individual de cada transcrito. Decodificar o transcrito é essencial para entender os elementos funcionais do genoma revelando os constituintes moleculares de células e tecidos, assim como compreender o desenvolvimento de doenças.

O estudo do transcrito está focado em catalogar todas as espécies de transcritos contendo cauda Poli A, incluindo RNA mensageiro (RNAm), RNA não codificante (ncRNA), como aqueles sem a cauda, como os micro RNA; determinar a unidade transcricional do gene em relação ao sítio inicial,



regiões 5' e 3' não traduzidas (UTRs), padrões de *splicing* e modificações pós transcricionais; e por fim, quantificar a taxa de expressão de cada transcrito.

Metodologias baseadas em análise de sequências, tais como: EST - *Expressed Sequence Tag* (OKUBO et al. 1992), SAGE - *Serial Analysis Gene Expression* (VELCULESCU et al. 1995), e MPSS - *Massive Parallel Signature Sequencing* (BRENNER et al. 2000), permitem comparar quantitativamente o padrão de expressão gênica entre amostras ou tecidos distintos. Estas técnicas possibilitam a avaliação da expressão de um número muito grande de genes sem a necessidade do conhecimento prévio de suas sequências, possuindo grande potencial analítico e sensibilidade. No entanto, pelo fato de serem, até pouco tempo atrás, proibitivas para grande número de amostras, devido ao enorme trabalho e custo na geração dos dados, eram ineficientes para acessar diferenças genéticas individuais, prejudicando assim sua aplicação na identificação de diferenças transcricionais para doenças heterogêneas e complexas (CARRARO et al. 2007).

Um método de quantificação do transcrito em larga escala muito utilizado é o cDNA *microarray* (SCHENA et al. 1995). Esta metodologia consiste na hibridização das sequências imobilizadas com moléculas de cDNA ou cRNA provenientes de duas situações celulares distintas marcadas com diferentes fluoróforos e hibridizadas de forma competitiva. A intensidade de fluorescência emitida das moléculas hibridizadas por complementaridade

à sequência imobilizada na lâmina é proporcional ao nível de expressão do gene.

Apesar de suas limitações como falta de sensibilidade em analisar genes diferencialmente expressos com baixa expressão, dependência de um conhecimento prévio da sequência genômica para o desenho das sondas, um alto ruído durante a aquisição do sinal, a possibilidade dessa técnica avaliar milhares de genes dentro de um universo maior de amostras a um custo relativamente reduzido permitiu acessar a variabilidade biológica em doenças complexas como o câncer.

A metodologia do *microarray* tem sido usada em oncologia para associar os dados de expressão com características clínico patológica (BRENTANI et al. 2005) com a finalidade de identificar alterações moleculares envolvidas com o aparecimento (MASCHIETTO et al. 2008) e progressão de tumores em geral (CASTRO et al. 2008), além de marcadores moleculares potenciais para resposta a tratamento (FOLGUEIRA et al. 2005).

Estudos de diferença de expressão gênica empregando o *microarray* identificaram um vasto repertório de produtos gênicos diferencialmente expressos entre tecido normal e tumor de próstata (LUO et al. 2001; MAGEE et al. 2001; SINGH et al. 2002; BEST et al. 2003; KRISTIANSEN et al. 2005; HALVORSEN et al. 2005; TRUE et al. 2006) ou mesmo comparando neoplasia intraepitelial prostática (PIN) com câncer de próstata (ASHIDA et al. 2004), tumor localizado contra câncer metastático (LATULIPPE et al. 2002; VARAMBALLY et al. 2002) tumor hormônio sensível ou hormônio

refratário (TRUE et al. 2006), e por fim, câncer de alto grau contra de baixo grau (WELSH et al. 2001).

Com o objetivo de refinar a classificação clínico - morfológica do CaP, estudos de diferença de expressão gênica tem sido desenvolvidos para auxiliar sua subclassificação. Essa abordagem permitiu correlacionar a expressão de um painel multigênico de 62 amostras de tumor de próstata de acordo com o grau do tumor, estágio da doença e tempo de recidiva bioquímica (LAPOINTE et al. 2004). Neste mesmo trabalho, foi identificado que a expressão aumentada do gene *MUC1* foi associada de forma significativa com tumores avançados e com alto risco de recorrência. Por outro lado, a expressão do gene *AZGP1* se mostrou como um fator de bom prognóstico.

SINGH et al. (2002) utilizaram o *microarray* com o objetivo de buscar marcadores prognósticos para o câncer de próstata. Em um grupo de 52 amostras, eles chegaram a um painel de cinco genes: cromogranina A,  $\beta$  receptor do fator de crescimento derivado de plaqueta (PDGFR  $\beta$ ), HOXC6, receptor 3 de inositol trifosfato (IPTR3) e sialitransferase 1, capazes de predizer a recorrência do PSA com 90% de acurácia por validação cruzada.

VARAMBALLY et al. (2005) compararam o padrão de expressão gênico, tanto ao nível transcricional como proteico, de amostras de câncer de próstata localizado contra metastático e levantaram um painel de 50 genes que foi capaz de predizer a recorrência bioquímica em um grupo independente de amostras de câncer de próstata. Enquanto que YU et al. (2004), chegaram a um painel de 70 genes capaz de predizer com 78% de

acurácia, em um grupo de amostras independentes, presença de invasão loco regional, metástase e recorrência bioquímica.

A utilização do nomograma no cotidiano da clínica é uma importante ferramenta na conduta clínica de cada caso. Informações provindas de estudos de *microarray* podem contribuir com um nomograma mais completo e aumentar o poder de predição do mesmo. O painel de genes escolhidos não necessariamente é um forte marcador prognóstico independente, mas quando associado pode aumentar o poder preditivo do nomograma (GLINSKY et al. 2004; STEPHENSON et al. 2005).

## **1.5 ANÁLISES TRANSCRICIONAIS COM SEQUENCIADORES DE NOVA GERAÇÃO**

Apesar de todo o projeto Genoma Humano ter sido feito com o método de sequenciamento Sanger, a realização de projetos dessa conjectura é de custo extremamente elevado e a construção da biblioteca para sua validação é muito laboriosa. Da mesma forma, estudos de genômica de diferentes amostras ou organismos em um mesmo trabalho realizado por um único laboratório são financeira e temporalmente inviáveis.

O desafio de reduzir o custo do sequenciamento para projetos que demandam *giga* bases de informação impulsionou o surgimento dos chamados Sequenciadores de Nova Geração (NGS, do inglês *next generation sequencing*).

Atualmente, dos principais NGS disponíveis comercialmente, três são as marcas que contribuíram com maior número de publicações: 454 GS FLX (Roche), *Genome Analyser IIx* (Illumina) e SOLiD™ Systems (Applied Biosystems). Apesar de se basearem em métodos químicos distintos, tamanho da sequência e diferente quantidade de bases geradas, para os três foi necessário contornar duas limitações do método convencional por Sanger.

Uma limitação do método de Sanger que foi superada nos NGS foi a amplificação do molde por PCR ou a clonagem. Essa é uma etapa que exige tempo, uma vez que cada fragmento de DNA a ser sequenciado deve ser processado individualmente. Além disso, obter o molde de milhares de colônias de *E. coli* é extremamente laborioso e certas sequências não são clonáveis em bactéria, gerando um viés na representação dos dados (GODISKA et al. 2010).

A segunda limitação é a velocidade de geração dos dados. Um sequenciamento do genoma humano por um único equipamento de 96 capilares levaria, por exemplo, cerca de 10 anos. Desta forma, a eletroforese capilar foi substituída por métodos de sequenciamento em paralelo nas três metodologias, onde milhares ou mesmo milhões de clones são sequenciados ao mesmo tempo. Além disso, a maior prerrogativa do NGS é a possibilidade do sequenciamento de diversas amostras, no qual as bibliotecas preparadas separadamente podem ser misturadas e sequenciadas em uma mesma corrida. Para isso, cada biblioteca é marcada com uma etiqueta de bases específicas (do inglês *barcode*), e por meio de

ferramentas de informática, cada sequência gerada é relacionada a sua biblioteca de origem (MEYER et al. 2007; WANG et al. 2008).

O conjunto de dados oriundos do transcrito por NGS tem sido denominado RNA-Seq (NAGALAKSHMI et al. 2008). Por se tratar de um monitoramento base a base, o RNA-Seq possibilita a interrogação de todas as sequências únicas do genoma, incluindo aquelas que são muito similares entre si e seriam falsamente tratadas como a mesma sequência como em um experimento de *microarray*. Aliado a isso, um grande número de sequências pode ser rapidamente obtido o que resulta em um método sensível e de aplicação variável.

Desta forma, quando comparado com *microarray*, o RNA-Seq permite detectar e quantificar níveis de expressão de RNA de transcritos raros (MARIONI et al. 2008), além de determinar precisamente os limites dos exons, mapear a região 3' UTR e identificar novas regiões transcritas assim como novas variantes de *splicing* (TRAPNELL et al. 2010). As novas implicações levantadas por esse método resultaram em um grande número de publicações em um espaço de tempo inferior a dois anos. Com as vantagens citadas acima e com o contínuo aprimoramento tecnológico dos sequenciadores de nova geração, cada vez mais, diferentes aplicações migram das plataformas de *microarray* para as plataformas de NGS (COPPÉE 2008).

Os dados gerados pelo RNA-seq não são usados apenas para análise de expressão gênica diferencial, mas também para identificar novos transcritos, fusões gênicas, desbalanço alélico, polimorfismos (SNPs) e

*splicing* alternativo, levando a uma melhor compreensão do transcrito do câncer (MORTAZAVI et al. 2008; MAHER et al 2009; BERGER et al. 2010; TUCH et al. 2010; CARRARO et al. 2011). A versatilidade desse método também pode contribuir na compreensão da complexa atuação de micro-RNAs (miRNAs) na biologia do tumor (HE et al. 2007; WYMAN et al. 2009; NYGAARD et al. 2009; UZIEL et al. 2009).

Apesar do grande potencial dos NGS para abordar o transcrito tanto de forma qualitativa como quantitativa, até o momento há um escasso número de publicações usando sequenciadores de nova geração para identificar genes diferencialmente expressos em câncer de próstata (NACU et al. 2011; ORR et al. 2012). Isso ressalta a importância do desenvolvimento de trabalhos de sequenciamento em larga escala a fim de contribuir com um maior esclarecimento do transcrito dessa neoplasia.

É marcadamente observada a carência de marcadores moleculares que possam auxiliar na discriminação entre tumores de próstata agressivos com potencial de invadir e se espalhar fora da glândula prostática daqueles mais indolentes que não apresentam essa capacidade, influenciando diretamente na melhor ou pior evolução da doença. Esse tipo de marcador é especialmente procurado para o grupo de tumores Gleason 7, os quais apresentam evolução mais variável, sendo portanto a classificação mais problemática do ponto de vista de decisão da conduta clínica.

A utilização de marcadores moleculares com esse potencial indicaria os pacientes que se beneficiariam de tratamentos radicais mais eficientes e

evitaria esse tipo de tratamento aqueles que apresentam um tipo de tumor pouco agressivo.

A identificação de genes diferencialmente expressos, através de metodologias com alto potencial de avaliação quantitativo, poderia auxiliar na identificação de marcadores moleculares prognósticos, isto é, marcadores preditivos confiáveis de má ou boa evolução de tumores de escore de Gleason 7.

Nesse sentido, considerando os avanços na tecnologia de geração de sequências e relativa diminuição de seu valor, as metodologias baseadas em sequências voltaram a ser extremamente atrativas do ponto de vista de investigação transcricional, além de possuírem menores vieses considerando o fato de ser uma medida direta de expressão. Dessa forma, nosso grupo vem trabalhando no estabelecimento de estratégias baseadas no uso de enzima de restrição e ligação de adaptadores com *barcodes* para acessar o transcrito de diferentes amostras tumorais em uma mesma corrida utilizando a plataforma 454-ROCHE.

No primeiro estudo foi estabelecida a abordagem para acessar o transcrito completo (*full length*) considerando os aspectos quantitativos e qualitativos de cada transcrito (CARRARO et al. 2011). No segundo a metodologia foi adaptada para utilização de quantidades mínimas de RNA total (na ordem de picogramas) incluindo a etapa de amplificação linear de RNA mensageiro (MOLINA et al. em preparação). A avaliação dos dados gerados mostrou que os aspectos quantitativos dos transcritos foram



preservados após o procedimento de amplificação permitindo sua utilização para análise de expressão gênica comparativa.

Com os avanços e aprendizados adquiridos no desenvolvimento dos dois trabalhos anteriores, propiciou a proposta desse estudo que combina avaliação quantitativa baseada em sequência com microdissecção a laser das células de interesse.

Assim, o objetivo desse trabalho é comparar o transcrito de células epiteliais tumorais capturadas de adenocarcinoma de próstata de escore de Gleason 7, cujo paciente tenha ou não apresentado recidiva bioquímica, a fim de identificar diferenças no nível de expressão que possam ser usadas como marcadores de prognóstico para este tipo de tumor.

## **2 OBJETIVOS**

### **2.1 OBJETIVO GERAL**

Avaliar a diferença de expressão gênica entre células tumorais de adenocarcinoma de próstata de escore de Gleason 7 de pacientes que tenham ou não apresentado recidiva bioquímica para identificar possíveis marcadores de prognóstico.

### **2.2 OBJETIVOS ESPECÍFICOS**

- Estabelecer um protocolo para construção de bibliotecas de cDNA para avaliação de múltiplas amostras;
- Sequenciar as bibliotecas na plataforma Genome Sequencer FLX Titanium 454 (Roche-Life Science);
- Comparar o padrão de expressão gênica entre as amostras com e sem recidiva bioquímica e identificar genes diferencialmente expressos;
- Testar o potencial de alguns genes candidatos identificados para marcador molecular por associação da expressão com características clínico patológicas.

### **3 MATERIAL E MÉTODOS**

#### **3.1 SELEÇÃO DAS AMOSTRAS DE ADENOCARCINOMA DE PRÓSTATA PARA CONSTRUÇÃO DAS BIBLIOTECAS DE cDNA 3´**

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antônio Prudente do Hospital do Câncer-Hospital A.C. Camargo-SP, no processo de número 955/07 e apresenta um termo de consentimento informado devidamente assinado por todos os participantes.

Foram selecionadas 26 amostras de tecido congelado de adenocarcinoma de próstata do banco de tumores do Hospital A.C. Camargo. A fim de selecionar apenas os casos com escore de Gleason 7, foi feita a revisão de lâminas dos casos com o patologista Dr. Antonio Hugo J. F. M. Campos.

Em relação às amostras sem recidiva, foram selecionadas somente aquelas com seguimento maior que 36 meses, enquanto que os casos com recidiva foram definidos como pacientes com concentração de PSA sérico maior que 0,2 ng/mL após a prostatectomia radical.

Cortes histológicos de 0,4 µm foram realizados em condições de baixa temperatura (de 20 a 25°C negativos) para fins de preservação da integridade do RNA total. Foram coletadas 10 lâminas de cada amostra.

Instantes antes da microdissecção, cada lâmina foi corada seguindo as seguintes etapas: 1) Imersão da lâmina em etanol 75% por 30 segundos

para fixação do tecido, 2) Imersão da lâmina em água nuclease-free (Sigma) por até 30 segundos para hidratação, 3) Hematoxilina por 30 segundos, 4) Reagente *blue* (aproximadamente 1ml de hidróxido de amônio em 50ml de água nuclease-free (SIGMA) por 30 segundos, 5) Eosina por 30 segundos para coloração da lâmina, 6) Etanol 75% por 30 segundos, 7) Etanol 95% por 30 segundos e 8) Etanol 100% por 2 minutos para desidratação, 9) Imersão da lâmina em xilol por 5 minutos para diafanização e 10) Secagem a temperatura ambiente.

Em torno de 2.000 células epiteliais foram capturadas de cada amostra por microdissecção a laser, usando o instrumento Pixcell Laser Capture Microdissection System (Arcturus Systems for Microgenomics), pertencente ao Departamento de Anatomia Patológica do Hospital do Câncer A.C. Camargo. O RNA total das amostras microdissecadas a laser foi extraído utilizando o kit PicoPure™ RNA Isolation Kit (Catalog # KIT0202 – Arcturus).

A amplificação do RNA e a construção das bibliotecas de cDNA 3' foram feitas segundo o protocolo estabelecido com RNA total da linhagem celular Hb4, descrito no item 3.2.

## 3.2 ESTABELECIMENTO DO PROTOCOLO DE CONSTRUÇÃO DE BIBLIOTECA DE cDNA 3'

### 3.2.1 Amplificação do RNA total e síntese de cDNA

Inicialmente, foi realizada uma etapa de amplificação de RNA partindo de aproximadamente 5ng de RNA total da linhagem Hb4a. A primeira etapa do protocolo consiste na síntese de cDNA e consequente a transcrição *in vitro*, ambas feitas com o kit RiboAmp® HSPlus (Arcturus) conforme as instruções do fabricante.

Depois da transcrição *in vitro* do RNA, o RNA amplificado (aRNA) foi incubado com 1µg de oligonucleotídeos degenerados de 15 bases (dN15) a 65°C por 5 minutos e resfriado em banho de gelo. Em seguida foi acrescentado 1X do tampão *First Strand Buffer*, 0,01M DTT; 1mM dNTP; 1 unidade de RNaseIN e 4 unidades de SuperscriptIII (Invitrogen). A reação foi mantida a 42°C por 2 horas.

Para síntese da segunda fita de DNA, foi utilizado um oligonucleotídeo dT(24) que contém a sequência correspondente ao *primer* de sequenciamento 454B e uma molécula de biotina na porção 5'. Para esta reação foram adicionados 1X do tampão da enzima Advantage DNA Polimerase (Clontech), 0,5µg do oligonucleotídeo dT<sub>(24)</sub>454B-biot (5'BIOT-GCCTTGCCAGCCCGCTCAGAGGCGGT<sub>(24)</sub>-3'), 0,2mM dNTP, 1 unidade de RNase H (Invitrogen) e 2 unidades da enzima Advantage DNA Polimerase (Clontech). A reação foi incubada por 37°C por 10 min, 94°C por 3 min, 65°C

por 5 min e 75°C por 30 min. Em seguida a reação foi inativada a 65°C com 50mM NaOH e 0,1mM EDTA.

A dupla fita de DNA foi purificada pelo método do fenol: clorofórmio:álcool isoamílico (25:24:1) (Sigma-Aldrich) e precipitada com etanol e acrilamida, lavado com etanol 70% 3 vezes, e por fim, ressuspendido em água deionizada.

### **3.2.2 Digestão com enzima de restrição**

Inicialmente foi feita uma análise *in silico* com o banco de dados RefSeq para determinar qual enzima de restrição de corte frequente, com sítio de reconhecimento de quatro nucleotídeos, que poderia gerar o maior número de fragmentos dentro da faixa de 70 a 600pb.

Após a purificação do cDNA dupla fita, este foi digerido com a enzima de restrição *DpnII*, que reconhece o sítio GATC. A reação foi feita em volume final de 50µl, em 1X de tampão NEB Buffer 2, 25 unidades de *DpnII* (New Englan Biolabs) e incubada a 37°C por 3 horas. Em seguida a enzima foi inativada a 65°C por 20 min.

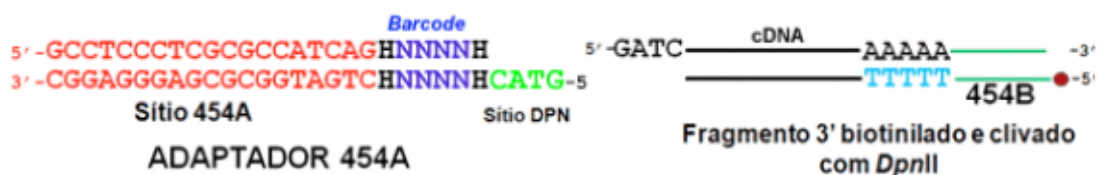
### **3.2.3 Montagem dos adaptadores 454A**

Para cada amostra foi montado um adaptador de dupla fita de DNA contendo a sequência correspondente ao primer de sequenciamento 454A (sítio 454A) com uma sequência específica de 6 nucleotídeos chamada etiqueta, ou, do inglês, *barcode*. Na extremidade 3' do adaptador há o sítio

coesivo de 4 bases GATC para que seja ligado ao fragmento de DNA digerido com a enzima *DpnII* (Figura 1).

Como cada amostra deve conter um *barcode* único, um cuidado especial foi empregado na escolha de sua sequência. Pelo fato do sequenciamento de regiões de homopolímero (acima de 6 nucleotídeos) pelo 454 não ser preciso, o desenho de cada *barcode* seguiu os seguintes parâmetros: (1) não apresentar homopolímeros de qualquer tamanho na sequência do *barcode*; (2) a primeira e a última base do *barcode* não podem ser uma guanina, para evitar a formação de homopolímeros de 2 bases G devido às sequências dos sítios 454A e do sítio da *DpnII* (Figura 1); (3) mesmo que haja um erro de sequenciamento e uma consequente troca de 1 base na sequência do *barcode*, todos *barcodes* ainda serão distintos entre si.

Com esses critérios, foi possível desenhar 72 *barcodes* de 6nt e seus respectivos adaptadores foram montados a partir da hibridização de dois oligonucleotídeos (Figura 1). Para esse estudo somente 23 *barcodes* foram utilizados.



**Figura 1** - Representação do adaptador 454A: O adaptador 454A é formado pela hibridação de 2 oligonucleotídeos. Em vermelho está representado o sítio 454A, necessário na PCR e no sequenciamento na plataforma 454; em azul é a sequência de 6 bases correspondente ao *barcode* específico para cada amostra; e em verde é a extremidade coesiva complementar à extremidade 5' do fragmento clivado com *DpnII*. À direita representação de um fragmento 3' biotilado e digerido com *DpnII* que será ligado ao adaptador. N: representa as 4 bases A T C e G, e H representa as bases A T e C (IUPAC)

### 3.2.4 Seleção dos fragmentos 3' biotilados e ligação do adaptador 454A

Partículas magnéticas recobertas com estreptavidina (F. Hoffmann-La Roche Ltd.) foram equilibradas em tampão de ligação (TEN 1000: 10mM Tris-HCl; 1mM EDTA e 1M NaCl, pH 7,5) e incubadas por 1 hora com o produto da digestão. Após duas lavagens com o tampão TEN 1000 seguidas de duas lavagens com água deionizada, as partículas foram equilibradas no tampão 1X da Ligase (New England Biolabs) em um volume final de 20 $\mu$ l. Foram adicionados 300ng dos adaptadores 454A e aquecidos a 50°C por 2 min. A solução foi resfriada à temperatura ambiente e colocada no gelo. Em seguida foram adicionadas 1000 unidades da enzima T4 DNA Ligase (New England Biolabs). A reação de ligação foi incubada a 16°C por duas horas e a cada 30min foi homogeneizada.

Em seguida, os adaptadores em excesso, não incorporados na ligação, foram lavados e as partículas magnéticas contendo os fragmentos 3' e seus adaptadores foram ressuspensos em 25 $\mu$ l de água deionizada.

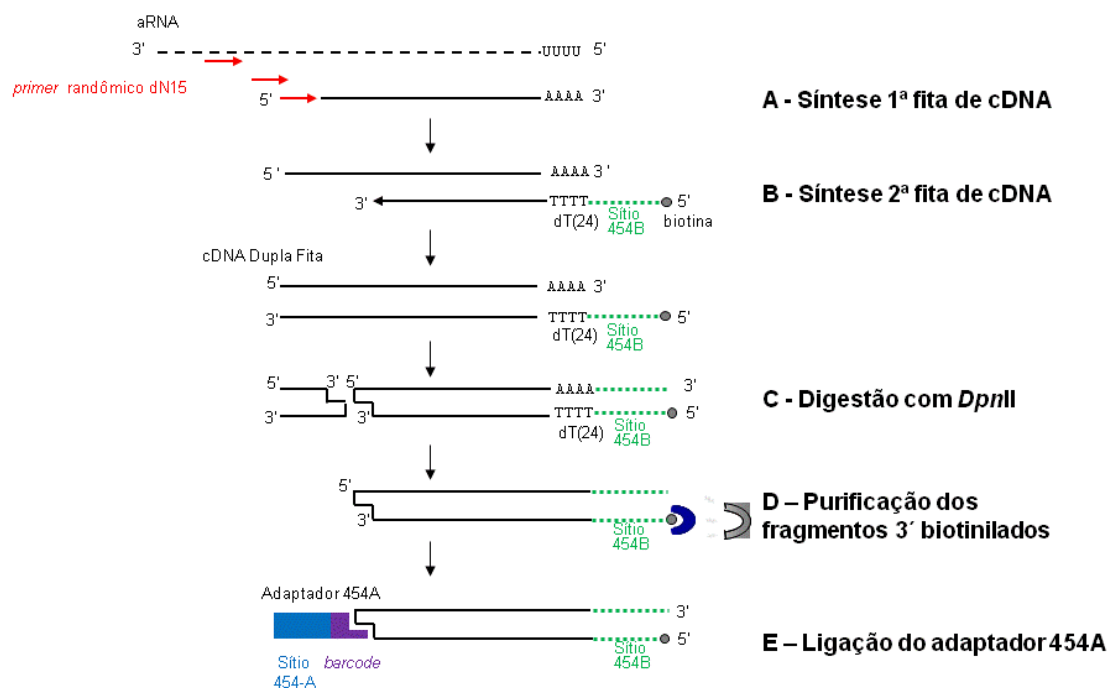


### 3.2.5 Amplificação por PCR

A reação da PCR foi feita com 2 unidades de Platinum Taq DNA Polimerase High Fidelity (Invitrogen), 1,5mM MgSO<sub>4</sub>, 200pmol dos *primers* 454A (5'-GCCTCCCTCGCGCCATCAG-3') e 454B (5'-GCCTTGCCAGCCCGCTCAG-3'), 0,2mM dNTP em um volume final de 20µl, segundo as seguintes condições: 94°C por 3 min de desnaturação inicial; e 15 ciclos a 94°C por 30s, 58°C por 30s e 68°C por 1 min e 30s.

O produto de PCR foi purificado pelo kit Agencourt® AMPure® XP (Beckman Coulter, Beverly MA) para eliminação de fragmentos menores que 150pb, seguida da etapa de quantificação feita por fluorometria pelo kit Quant-iT™ dsDNA HS Assay Kit (Invitrogen) no fluorímetro Qubit (Invitrogen), ou por espectrofotometria no equipamento NanoDrop 1000 (Thermo Scientific).

As principais etapas da construção da biblioteca de cDNA 3' descritas acima estão representadas na Figura 2.



**Figura 2** - Esquema da metodologia padronizada para a construção da biblioteca de cDNA para sequenciamento no 454. **A:** a partir do RNA amplificado antissenso (aRNA) é feita a síntese da primeira fita cDNA com *primers* randômicos dN15. **B:** O cDNA é convertido em dupla fita, com o oligonucleotídeo dT<sub>(24)</sub>454B-biot, o qual contém o sítio 454B (linha tracejada verde) e é biotilado na extremidade 5'. **C:** O cDNA dupla fita é clivado com a enzima *DpnII*. **D** Apenas os fragmentos biotinizados 3' são selecionados com as *beads* magnéticas cobertas com estreptavidina. **E:** Aos fragmentos purificados é ligado o adaptador 454A contendo o sítio 454A (retângulo azul) e o *barcode* paciente específico (hexágono roxo)

### 3.3 SEQUENCIAMENTO E VALIDAÇÃO DAS BIBLIOTECAS PELO MÉTODO DE SANGER

#### 3.3.1 Sequenciamento individual das bibliotecas de cDNA 3'

Para monitorar a estrutura das bibliotecas de cDNA 3', cada uma foi clonada individualmente com o InstAclone™ PCR Cloning Kit (Fermentas), segundo instruções do fabricante, em bactérias eletrocompetentes da linhagem *Escherichia coli* DH10B.

Para cada transformação, cerca de 50 colônias de bactéria tiveram seus insertos amplificados por PCR com os *primers* M13Rv (5' CAGGAAACAGCTATGAC3'), que anela no vetor pTZ57R/T, e com o *primer* 454B (5'-GCCTTGCCAGCCCGCTCAG-3'). Os produtos de PCR foram sequenciados pelo método de Sanger com o kit BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) e com o *primer* M13Rv, em seguida analisados no sequenciador ABI 3130 xl (Applied Biosystems).

### **3.3.2 Análise das sequências geradas pelo método de Sanger**

Um *pipeline* de bioinformática foi desenvolvido para analisar as sequências geradas pelo método Sanger e está descrito a seguir.

Inicialmente este *pipeline* utiliza o programa PHRED (EWING et al. 1998) para eliminar regiões de baixa qualidade nas sequências obtidas e o programa CrossMatch para mascarar sequências derivadas dos vetores de clonagem. O programa CrossMatch é novamente utilizado para identificar de forma sequencial os adaptadores, a sequência de 6 bases correspondente ao *barcode* e o sítio de restrição (*DpnII*). Somente as sequências nas quais todas as regiões descritas acima foram identificadas foram consideradas válidas.

### **3.3.3 Montagem do *pool* de bibliotecas e validação no ABI**

Após purificar e quantificar as bibliotecas de cDNA 3', um *pool* das 23 bibliotecas foi feito com 10ng provenientes de cada amostra somando um total de 230ng. Desse *pool*, 30ng foram clonados e sequenciados conforme

descrito anteriormente, 50ng foram avaliados por eletroforese capilar no Bionalyser DNA 1000 (Agilent) e o restante enviado para o sequenciamento na plataforma *Genome Sequencer FLX Titanium System 454* (Roche-Life Sciences).

### **3.4 SEQUENCIAMENTO NA PLATAFORMA GS FLX TITANIUM SYSTEM 454 E ANÁLISE DOS DADOS**

#### **3.4.1 Mapeamento dos dados do sequenciamento do 454**

O sequenciamento na plataforma *GS FLX Titanium System 454 Roche-Life Sciences* (Roche-Life Sciences) foi terceirizado na *Core Facility* do Hospital M.D. Anderson Cancer Center (Houston, Texas – EUA).

Os dados recebidos foram analisados preliminarmente pelo *pipeline* de bioinformática desenvolvido com a colaboração da Dra. Helena Brentani e Renato Puga. Inicialmente as sequências geradas no 454 foram submetidas à identificação dos adaptadores, e sequências contendo adaptadores truncados ou internamente posicionado foram removidas. Em seguida, sequências com alta similaridade ( $E\text{-value} \leq 1 \times 10^{-20}$ ,  $\text{identity} \geq 85\%$  and  $\text{coverage} \geq 85\%$ ) ao RNA ribossomal humano ou DNA mitocondrial foram filtradas pelo MEGABLAST. As sequências restantes foram alinhadas contra o genoma humano (release hg19, Fevereiro de 2009) através da ferramenta BLAT (KENT 2002) e os melhores alinhamentos foram selecionados com a ferramenta psiReps usando os seguintes parâmetros:

minCover = 0.70, minAli = 0.96, nearTop = 0.005. Sequências com alinhamento múltiplo em locais distintos do genoma foram eliminadas.

O banco de dados KnownGene (HSU et al. 2006), cujas coordenadas estão disponíveis pela Universidade California Santa Cruz foi usado como referência para mapear as sequências em relação aos transcritos e exons descritos. As sequências com sobreposição em pelo menos uma base da unidade transcritiva (UT), definida como a região genômica compreendida entre a primeira e última base do transcrito, foram classificadas como intragênica.

### **3.4.2 Identificação dos genes diferencialmente expressos**

#### **3.4.2.1 Baseado em *Fold change* ou no método de proporções**

A comparação do perfil de expressão gênica entre os grupos com e sem recidiva bioquímica foi realizada com a ferramenta RNA-Seq do pacote CLC Genomics Workbench 4.5 (CLC bio).

Os dados foram subdivididos em dois subconjuntos. O subconjunto 1 compreende os transcritos que foram representados em pelo menos 50% das amostras de cada um dos grupos (com e sem recidiva bioquímica). Para este subconjunto foi empregado o teste estatístico de proporção (BAGGERLY et al. 2003), que considera o número total de sequências de cada biblioteca em relação ao total de sequências geradas nos dois grupos (com e sem recidiva bioquímica), e atribui pesos assumindo uma distribuição

Beta binomial. Foram considerados como diferencialmente expressos os genes com *fold change*  $\geq |2|$  e  $p \leq 0.05$ .

Já o subconjunto 2 corresponde aos transcritos que foram reportados por pelo menos 50% das amostras em um dos grupos, enquanto no outro grupo o mesmo transcrito foi reportado por no máximo uma amostra. Para os transcritos que não foram reportados por nenhuma das amostras em um dos grupos, o valor zero, referente ao total de sequências deste grupo, foi substituído por um, a fim de realizar o cálculo do *fold change*. Para esse segundo subconjunto foram considerados diferencialmente expressos os genes com *fold change*  $\geq |2,5|$ .

#### **3.4.2.2 Baseado no método de cálculo de probabilidades**

Foi desenvolvido um modelo matemático pelo Dr. Carlos AB Pereira para identificar genes diferencialmente expressos a partir de dados de sequenciamento.

Nesta segunda análise foi usada a distribuição Normal Logística. Esta é uma distribuição a posteriori de uma transformação da frequência relativa que representa a expressão de um determinado gene (VARUZZA e PEREIRA 2008). A frequência é relativa ao tamanho da biblioteca. Esse método é baseado em logitos, logaritmo da razão de probabilidades a posteriori. De fato como tratamos de todos os genes de interesse a construção produz uma distribuição normal multivariada dos logitos.

Para definir o grupo de genes com diferença de expressão entre as amostras com e sem recidiva bioquímica, foi calculado a densidade

multivariada dos genes dentro das bibliotecas de cDNA com os dados de sequenciamento do 454. Em seguida foi calculado a densidade do grupo marginal considerando todas as bibliotecas de cDNA para cada um dos dois grupos (com e sem recidiva bioquímica).

As duas densidades de grupo permitiram calcular a probabilidade de cada gene, ou de combinação de genes, ser mais expresso no grupo sem recidiva em relação ao grupo com recidiva. Os valores de probabilidade próximos de 1 correspondem aos genes mais expressos no grupo sem recidiva, por outro lado os valores próximos a 0 correspondem aos genes mais expressos no grupo com recidiva.

### **3.4.3 Anotação de vias e processos biológicos e clusterização hierárquica**

Os dados do sequenciamento das 21 amostras de adenocarcinoma de próstata foram classificados de acordo com os processos biológicos através do Gene Ontology (GO) e com as vias metabólicas representativas usando o KEGG (Kyoto Encyclopedia of Genes and Genomes - [www.kegg.com](http://www.kegg.com)). Para avaliação da representatividade significativa dos processos biológicos e vias metabólicas no grupo de genes diferencialmente expressos foi usada a distribuição qui-quadrado do *Onto-Tools* desenvolvida pelo *The Intelligent Systems and Bioinformatics Laboratory (ISBL)* (<http://vortex.cs.wayne.edu>). Processos biológicos com  $pFDR < 0,05$  e vias com  $p < 0,05$  foram considerados significantes, porém se representados por

um único gene, não foram levados em consideração, na tentativa de se evitar resultados artefatuais.

A mesma avaliação de vias e processos biológicos foi feita com os genes diferencialmente expressos entre as amostras com recidiva e sem recidiva. Além disso, foi feita uma clusterização hierárquica baseada na distância Euclidiana com *average linkage* que discrimina as amostras baseada no perfil de expressão. A confiabilidade dessa discriminação foi verificada pela técnica de Bootstrap. Essas análises foram feitas no programa TMEV versão 4.1 (SAEED et al. 2003).

### **3.5 VALIDAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS**

#### **3.5.1 Amplificação do RNA e síntese do cDNA**

Para a validação utilizando o grupo inicial de amostras foi utilizado o aRNA por dois *rounds* subsequentes com o kit RiboAmp® HSPlus (Arcturus), que foi convertido na fita complementar de DNA (cDNA) conforme o protocolo a seguir:

Quatro microgramas do aRNA, amplificado a partir do RNA total das células capturadas a laser, foram incubados com 9ug de oligonucleotídeos degenerados de 6 bases (dN6) a 65°C por 5 minutos e resfriado por 1 minuto a 4°C. Em seguida foi acrescentado 1x tampão *First Strand Synthesis*; 0,01M DTT; 1mM dNTP; 1 unidade de RNaseIN e 400 unidades de SuperscriptIII (Invitrogen). A reação foi mantida a 42°C por 2 horas.



A qualidade do cDNA foi aferida através da reação de PCR com a enzima Platinum® Taq DNA Polimerase (Invitrogen) com um par de oligonucleotídeos do gene endógeno *GAPDH* (*primer forward* 5'-GAAGGTGAAGGTCGGA-3' e *primer reverse* 5'-GGGTCATTGATGGCAAC-3') com 10ng de cDNA segundo o seguinte ciclo: desnaturação inicial de 95°C por 2 minutos, e 35 ciclos a 95°C por 30 segundos, 60°C por 30 segundos e 72°C por 30 segundos. O produto amplificado de 102pb foi analisado por eletroforese em gel de agarose 1,5%.

### **3.5.2 Reação de RT-qPCR com o sistema *Taqman Low Density Array* (TLDA)**

Dos genes diferencialmente expressos pela análise de bioinformática foram escolhidos aqueles com  $p \leq 0.01$  para o subconjunto 1 (53 genes), e aqueles com *fold change*  $\geq |2,5|$  e reportados em pelo menos 60% das amostras de um dos dois grupos do subconjunto 2 (34 genes).

Para todos os 87 genes diferencialmente expressos foram selecionados ensaios padronizados com a ferramenta disponível online TaqMan® Gene Expression Assays.

Os critérios para a escolha do ensaio foram: i) o ensaio deve se localizar na posição 3' do gene quando possível; ii) os oligonucleotídeos devem se anelar em dois exons distintos; iii) o produto amplificado não deve ser superior a 150pb; iv) o ensaio deve compreender o maior número de isoformas dos RNAs mensageiros de um mesmo gene.

Juntamente com mais cinco genes endógenos, *ALAS1*, *B2M*, *HPRT1*, *SDH*, *TBP*, escolhidos com base na literatura (OHL et al. 2005), esses ensaios foram entregues prontos em placas de 384 poços. Para cada reação foram usados 40ng de cDNA em um volume total de 2 $\mu$ l. O preparo do cartão foi feito segundo as instruções do fabricante (Applied Biosystems), e analisado no 7900HT Fast Real-Time PCR System (Applied Biosystems).

### **3.5.3 Seleção do gene normalizador**

Os genes mais estáveis nas amostras foram selecionados utilizando a ferramenta geNorm (VANDESOMPELE et al. 2002). O programa geNorm calcula a medida de estabilidade gênica (M) para os genes endógenos em um grupo de amostras indicando os mais estáveis para serem utilizados como normalizadores. O valor M sugerido pelo programa é de 1.5, abaixo do qual se tem os melhores genes endógenos. A estabilidade é determinada partindo-se do princípio de que dois genes normalizadores ideais possuem razões de expressão idênticas em todas as amostras de cDNA analisadas, independente das condições experimentais ou de tipos celulares.

### **3.5.4 Cálculo da expressão relativa dos ensaios de RT-qPCR**

A expressão diferencial dos genes selecionados foi determinada pelo método de quantificação relativa  $2^{\Delta Ct}$ , que leva em conta a expressão do gene alvo em relação ao gene normalizador.

Para determinar as diferenças de expressão entre amostras do grupo recidiva contra o grupo não recidiva, foi aplicado o teste *t* de Student que

leva em conta a média e desvio padrão da medida relativa de cada grupo, sendo considerados como significativos valores com *fold change*  $\geq |2|$  e  $p < 0,05$ .

### **3.6 IMUNOISTOQUÍMICA (IHQ)**

#### **3.6.1 Análise da expressão proteica de candidatos previamente validados por RT-qPCR**

Os candidatos validados por RT-qPCR foram avaliados por IHQ em um grupo independente de amostras de tecidos tumorais fixados por formaldeído e emblocados em parafina (FFPE).

Para o marcador BRP44 foram avaliadas um total de 92 amostras, enquanto que para o marcador NFKBIZ um total de 50 casos. As amostras são provenientes do arquivo de anatomia patológica do Hospital A.C. Camargo e as características clínico patológicas estão descritas na Tabela 1:

**Tabela 1** - Características clínico patológicas das amostras utilizadas na imunohistoquímica.

Variável	Categoria	Marcador	
		NFKBIZ	BRP44
Idade	Média	66,28	66,62
	Mínimo	46	46
	Máximo	83	85
Volume Tumor	Média	24,9	22,97
	Mínimo	0,5	0,5
	Máximo	100	100
Estadiamento	pT2a	3 (6,00)	5 (5,43)
	pT2b	11 (22,00)	16 (17,39)
	pT3a	26 (52,00)	47 (51,08)
	pT3b	8 (16,00)	19 (20,65)
	pT4a	2 (4,00)	5 (5,43)
	6	20 (40,00)	39 (42,39)
	7	18 (36,00)	33 (35,86)
Gleason	8	6 (12,00)	9 (9,78)
	9	6 (12,00)	11 (11,95)
	< 10	26 (52,00)	44 (47,82)
PSA Pré cirúrgico (ng/mL)	> 10	24 (48,00)	47 (52,18)
	Positivo	7 (14,00)	16 (17,39)
Vesícula Seminal	Negativo	43 (86,00)	76 (82,61)
	Positivo	35 (70,00)	70 (76,09)
Margem	Negativo	15 (30,00)	22 (23,91)
	Positivo	6 (12,00)	7 (7,61)
Linfonodo	Negativo	44 (88,00)	85 (92,39)
	Positivo	13 (26,00)	22 (23,91)
Invasão Agiolinfática	Negativo	37 (74,00)	70 (76,09)
	Positivo	37 (74,00)	69 (75,00)
Recidiva Bioquímica	Negativo	13 (26,00)	23 (25,00)
	Positivo	10 (20,00)	17 (18,48)
Recidiva Clínica	Negativo	40 (80,00)	75 (81,52)
	<b>Total</b>	<b>50</b>	<b>92</b>

### 3.6.2 Reações de imunistoquímica

Para as reações de imunistoquímica foi utilizado o Kit EnVision™ FLEX Mini Kit, High pH (DAKO, Dinamarca). A desparafinização das lâminas foi realizada em 3 banhos de Xilol consecutivos de 5 minutos e em seguida 4 banhos de álcool absoluto para retirar o Xilol. As lâminas foram lavadas por 5 minutos em água corrente e em seguida colocadas em solução de Tris-EDTA (pH 9.0) em panela de pressão por 25 minutos para a recuperação antigênica.

Após esse período as lâminas foram retiradas da panela e mantidas na solução de Tris-EDTA por 20 minutos para resfriar. Em seguida foram lavadas 2 vezes em tampão de lavagem (EnVision™ FLEX Wash Buffer) por 5 minutos e depois em água corrente por 5 minutos. O bloqueio da peroxidase endógena foi realizado em dois banhos de imersão de 5 minutos consecutivos em Peróxido de Hidrogênio 10 volumes.

Após esse período as lâminas foram lavadas duas vezes em tampão de lavagem (EnVision™ FLEX Wash Buffer) por 5 minutos. Em seguida as lâminas foram acondicionadas em câmaras úmidas e foi adicionado aproximadamente 50µL de anticorpo primário (Anti-BPR44, marca Abnova, com uma diluição 1:50; e Anti-NFKBIZ, marca Sigma, com uma diluição 1:25). A incubação do anticorpo primário foi feita por 20 minutos e após esse período as lâminas foram lavadas duas vezes em tampão de lavagem (EnVision™ FLEX Wash Buffer) por 5 minutos cada.

Para o Anti-NFKBIZ as lâminas foram incubadas por mais 20 minutos com 100µL de EnVision™ FLEX+ Rabbit (LINKER) para amplificar o sinal do

anticorpo primário. Essa incubação foi seguida de duas lavagens com tampão de lavagem (EnVision™ FLEX Wash Buffer) por 5 minutos cada.

Após as lavagens as lâminas foram incubadas por 20 minutos com 100µL de EnVision™ FLEX /HRP, um reagente de detecção conjugado à moléculas de peroxidase e anticorpos secundários. Em seguida foram realizadas duas lavagens em tampão de lavagem (EnVision™ FLEX Wash Buffer) por 5 minutos cada e as lâminas seguiram para a etapa de revelação. Nessa etapa as lâminas foram incubadas com 100µL de DAB (Substrate working solution mix) cada, por 5 minutos e em seguida foram lavadas em água corrente por 5 minutos.

A contracoloração foi feita imergindo as lâminas em Hematoxilina de Harris por 1 minuto e retirando o excesso de hematoxilina em duas imersões rápidas em solução diferenciador (álcool absoluto + ácido clorídrico fumante). As lâminas foram lavadas em água corrente por 5 minutos e seguiram para a desidratação e montagem de forma automatizada no equipamento Tissue-Tek® Prisma®/Film® (SAKURA FINETEK USA).

Os resultados foram analisados em microscópio óptico pelo Dr. Rafael Malagoli Rocha e classificados em relação à intensidade da reação, à porcentagem de células positivas e ao padrão de marcação.

A marcação nuclear foi classificada segundo a presença e ausência da marcação. As amostras foram categorizadas como negativas (score 0) e positivas (score 1).

A marcação citoplasmática foi classificada segundo a intensidade em: ausente, fraca, moderado-fraco, moderado-forte e forte. Para avaliação do

potencial prognóstico dos marcadores estudados, as amostras foram categorizadas em negativas (ausência de marcação e marcação fraca) e positivas (marcação moderada fraca, moderada forte e forte).

### **3.6.3 Análise dos resultados de Imunoistoquímica**

O teste de frequências do qui-quadrado foi utilizado para verificar a associação entre as variáveis categóricas. O teste exato de *Fisher* foi adotado quando pelo menos uma frequência esperada foi menor do que 5. A sobrevida livre de doença foi calculada empregando o método de Kaplan-Meier, utilizando o teste de Log-rank, para a comparação entre as curvas de sobrevida. O intervalo livre de doença foi mensurado a partir da data da cirurgia de prostatectomia radical até a data em que recidiva bioquímica foi diagnosticada. Todos os testes estatísticos foram feitos usando o programa GraphPad Prism versão 5.0.

## **4 RESULTADOS**

### **4.1 SELEÇÃO DAS AMOSTRAS DE ADENOCARCINOMA DE PRÓSTATA**

Um total de 26 amostras congeladas foram retiradas do biobanco do Hospital AC Camargo. Os espécimes foram reavaliados pelo patologista Dr. Antônio Hugo Campos para confirmação das lesões tumorais e do respectivo grau de Gleason representado na peça avaliada. Três casos foram excluídos por não conterem células tumorais no corte de tecido congelado presente na lâmina e outros dois por terem graus de Gleason diferente de 3 ou 4. As 21 amostras de tumor selecionadas estão distribuídas conforme a Tabela 2, sendo 11 de pacientes sem recidiva e 10 com recidiva.



**Tabela 2** - Distribuição das amostras de adenocarcinoma de próstata selecionadas para captura a *laser* e extração de RNA para confecção das bibliotecas a serem sequenciadas no 454.

Ama	Idade (anos)	Gleason Primário da Peça	Gleason Secundário da Peça	Escore de Gleason	PSA pré cirúrgico (ng/mL)	Volume Tunmor (%)	Estadio Patológico	Margem Cirúrgica	Recidiva Bioquímica	Recidiva Clínica	Risco D'Amico	Gleason da lâmina microdissecada	Cirurgião (n) Experiência (Anos)
2-Rec	59	3	4	7	30,25	12	T3b	0	1	1	Alto	3	3 - (20 anos)
3-Rec	64	4	3	7	16,9	22,5	T3b	1	1	0	Inter.	3+4	5 - (12 anos)
4-Rec	50	3	4	7	8,15	11	T2b	0	1	1	Inter.	3	2- (20 anos)
5-Rec	52	3	4	7	4,31	1,75	T2a	0	1	0	Inter.	3	5 - (12 anos)
6-Rec	63	3	4	7	8,93	7	T2b	0	1	0	Inter.	4 cribriforme	5 - (12 anos)
30-Rec	74	3	4	7	5,4	3,25	T3a	0	1	0	Inter.	3	5 - (12 anos)
40-Rec	56	4	3	7	6,1	15	T3a	1	1	1	Inter.	3	2- (20 anos)
50-Rec	57	4	3	7	3,61	13	T3a	1	1	0	Inter.	3	2- (20 anos)
60-Rec	61	3	5	8	5,59	15,25	T3a	1	1	0	Inter.	3+4	5 - (12 anos)
70-Rec	47	3	4	7	16,17	7	T2b	0	1	0	Inter.	4	5 - (12 anos)
1-Nrec	72	3	4	7	6,8	9,25	T3a	0	0	0	Inter.	4 cribriforme	1 - (25 anos)
10-Nrec	56	4	3	7	8,06	3,25	T2b	0	0	0	Inter.	3+4	1 - (25 anos)
17-Nrec	58	4	3	7	6,5	15	T2b	0	0	0	Inter.	4	4 - (19 anos)
20-Nrec	71	3	4	7	5,66	4,6	T2b	0	0	0	Inter.	3	3 - (20 anos)
25-Nrec	62	3	4	7	8,12	15,5	T2b	0	0	0	Inter.	3	3 - (20 anos)
27-Nrec	59	3	4	7	5,84	28	T2b	0	0	0	Inter.	3	2- (20 anos)
35-Nrec	43	3	4	7	10,06	13	T2b	0	0	0	Inter.	3	5 - (12 anos)
37-Nrec	65	3	4	7	3,85	12	T3a	0	0	0	Inter.	4	5 - (12 anos)
45-Nrec	55	4	3	7	22,1	13	T3a	0	0	0	Alto	3+4	1 - (25 anos)
55-Nrec	47	3	4	7	4,67	1,5	T2b	0	0	0	Inter.	3	4 - (19 anos)
65-Nrec	53	3	4	7	4,3	9,5	T2b	1	0	0	Inter.	3	1 - (25 anos)

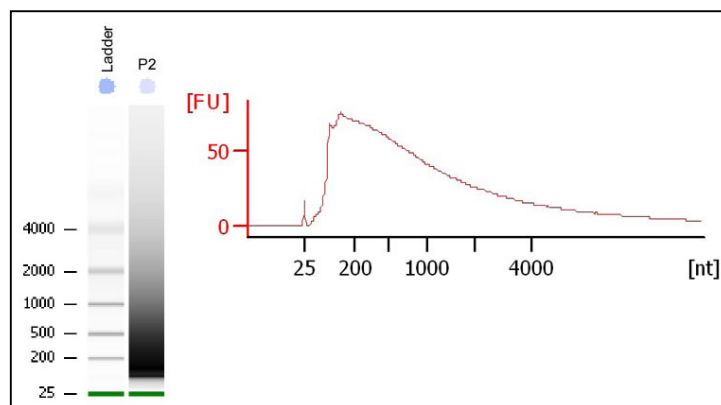
Na coluna Cirurgião (n) representa qual cirurgião realizou a prostatetomia e entre parênteses o tempo de experiência do profissional desde a conclusão da residência

## 4.2 ESTABELECIMENTO DO PROTOCOLO DE CONSTRUÇÃO DE BIBLIOTECA DE cDNA

### 4.2.1 Amplificação do RNA total

Antes de iniciarmos a construção das bibliotecas das amostras microdissecadas a laser, testamos o método de amplificação de RNA partindo de 5ng de RNA total de HB4a, com kit RiboAmp® HSPlus (Arcturus). Foi possível obter um rendimento de 355,20ng de aRNA e para avaliarmos a eficiência da amplificação conferimos o tamanho e a faixa da concentração do *smear* de aRNA por eletroforese capilar no Bionalyser (Agilent) (Figura 3).

Como mostrado na Figura 3 amplificação do RNA total apresentou uma concentração de aRNA de 100 a 1000nt, o que é indicativo de uma amplificação de boa qualidade, com pouca concentração de dímeros de *primer* que podem gerar artefatos durante o procedimento de amplificação (SARAIVA et al. 2006). Assim, decidimos prosseguir a construção da biblioteca para essa duas amostra de aRNA até a etapa de amplificação por PCR a fim de verificar se o método proposto poderia ser empregado para as amostras microdissecadas a *laser*.



**Figura 3** - Eletroforese capilar da amplificação do RNA total da linhagem celular Hb4. 100ng de aRNA foram aplicados no PicoChip do Bionalyser e avaliados pela eletroforese digital e pelo eletroferograma.

#### 4.2.2 Digestão com enzima de restrição

Como o tamanho médio das sequências geradas pelo sequenciador 454 (Roche) varia de uma faixa de aproximadamente 500nt, com variação de 100 a 600, e durante a construção da biblioteca nós eliminamos os fragmentos menores que 70pb correspondentes aos dímeros de adaptadores, foi realizada uma análise computacional com diferentes enzimas de restrição de corte frequente (reconhecimento de 4 bases) com as sequências do banco de dados RefSeq, que compreende um total de 38.774 transcritos.

O objetivo era selecionar a enzima capaz de gerar o maior número de fragmentos na faixa de tamanho que seriam potencialmente sequenciados, levando em consideração as características da plataforma 454-ROCHE e da biblioteca. Na Tabela 3 estão apresentados os números de transcritos que quando digeridos com as respectivas enzimas de restrição apresentariam fragmentos dentro das diferentes faixas de tamanho (70 a 600pb; <70pb e

>600pb), sendo que a faixa ideal para o transcrito ser sequenciado, e consequentemente identificado na biblioteca, seria de 70pb a 600pb. A análise computacional mostrou que as enzimas *NlaIII*, *AluI* e *DpnII* são as que apresentam o maior número de transcritos na faixa ideal de tamanho (27.137, 26.775 e 25.157, respectivamente).

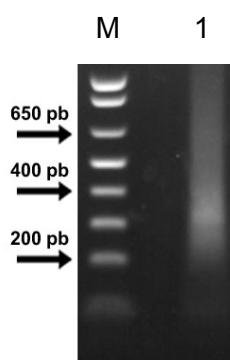
**Tabela 3** - Análise *in silico* das enzimas de corte frequente. Em vermelho são as enzimas de corte frequente com o potencial em gerar o maior número de fragmentos dentro da faixa de tamanho ótima para o sequenciamento no 454 (70 a 600pb)

Enzima	Sítio	< 70pb		70 a 600pb		> 600pb	
		Total	%	Total	%	Total	%
Acil	CCGC	3.030	08	14.550	38	21.194	55
Afal	GTAC	4.307	11	20.121	52	14.346	37
<b>AluI</b>	<b>AGCT</b>	<b>8.131</b>	<b>21</b>	<b>26.775</b>	<b>69</b>	<b>3.868</b>	<b>10</b>
Bfal	CTAG	4.386	11	20.427	53	13.961	36
BstUI	CGCG	1.665	04	8.339	22	28.770	74
CviRI	TGCA	9.805	25	25.344	65	3.625	09
<b>DpnII</b>	<b>GATC</b>	<b>4.357</b>	<b>11</b>	<b>25.157</b>	<b>65</b>	<b>9.260</b>	<b>24</b>
HaeIII	GGCC	6.410	17	23.408	60	8.956	23
HhaI	GCGC	2.311	06	12.644	33	23.819	61
HpaII	CCGG	3.004	08	15.613	40	20.157	52
Maell	ACGT	2.239	06	14.837	38	21.698	56
MseI	TTAA	1.7127	44	13.982	36	7.665	20
<b>NlaIII</b>	<b>CATG</b>	<b>8.140</b>	<b>21</b>	<b>27.137</b>	<b>70</b>	<b>3.497</b>	<b>09</b>
TaqI	TCGA	1.547	04	12.432	32	24795	64
Tsp509I	AATT	13.603	35	17.139	44	8032	21

Apesar de as enzimas *NlaIII* e *AluI* gerarem um número maior de fragmentos dentro da faixa de 70 a 600pb (70%), nós optamos pela utilização da enzima *DpnII* uma vez que *NlaIII* e *AluI* geram o um número muito maior de sequências menores que 70pb, que seriam descartados durante a construção da biblioteca. Enquanto que fragmentos maiores que 600pb, apesar de estarem acima da faixa, podem ser sequenciadas mesmo que parcialmente. Dessa forma usando a enzima *DpnII* nossa estratégia tem o potencial de identificar cerca de 34.417 (25.157 de 70 a 600pb + 9.260 >600pb) representando 89% dos transcritos do RefSeq.

#### 4.2.3 Amplificação por PCR

Na amostra digerida com *DpnII* foi ligado o adaptador 454A, e após purificação do produto da ligação as amostras foram submetidas à amplificação por PCR de 15 ciclos. A Figura 4 mostra que a eficiência da amplificação por PCR foi satisfatória, sendo que a amostra apresentou tamanho do *smear* com maior concentração entre 200 a 400pb.

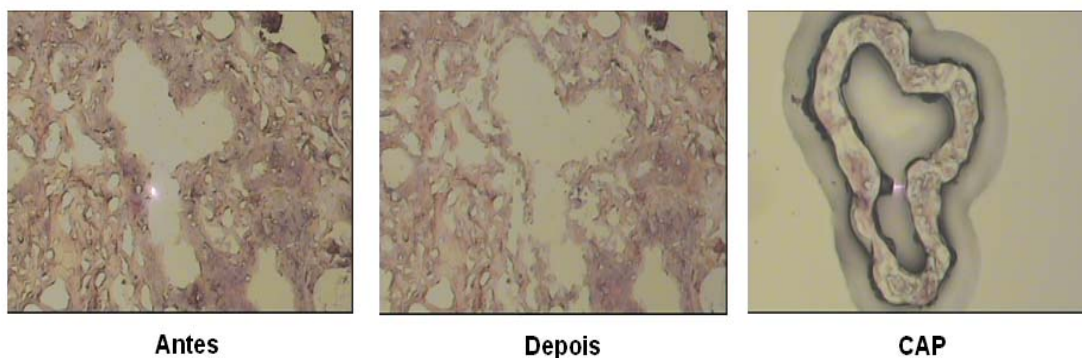


**Figura 4** - Eletroforese em gel de agarose 1% da amostra amplificadas por 15 ciclos de PCR após ligação dos adaptadores 400ng da amostra da linhagem Hb4 amplificadas por 15 ciclos PCR. M: marcador de peso molecular 1 Kb Plus DNA Ladder (Invitrogen); 1: PCR 15 ciclos da amostra do protocolo P2

Assim, após o estabelecimento da amplificação e da ligação de adaptadores, nós seguimos para a construção das bibliotecas a partir de RNA total purificado de células tumorais capturadas por microdissecação a *laser* de adenocarcinoma de próstata.

#### 4.3 CONSTRUÇÃO DAS BIBLIOTECAS DE cDNA 3'

Para cada amostra foram realizados de 800 a 2000 tiros de *laser* com diâmetro de 7 $\mu$ m, o que representa, aproximadamente, uma faixa de 2400 a 6000 células capturadas. Uma representação da microdissecação a *laser* das células epiteliais das glândulas prostáticas pode ser visualizada na Figura 5.



**Figura 5** - Representação da microdissecação a laser das células epiteliais das glândulas prostáticas. Coloração por Hematoxilina-Eosina; antes – representa o tecido antes da captura das células, depois – após a captura e CAP – as células de interesse isoladas e aderidas na membrana.

A fim de evitar a degradação do RNA após descongelar o tecido da lâmina de LCM, o tempo de captura foi no máximo de 10 minutos para cada caso. A variação do número de células capturadas está diretamente relacionada com a densidade das células tumorais, como por exemplo,

amostras com padrão de Gleason 4 cribriforme, apresentam uma densidade de células neoplásicas maior que tecidos apenas com o padrão 3, possibilitando assim um número maior de tiros em um mesmo intervalo de tempo.

Como a quantidade de RNA total extraída com o kit PicoPure™ RNA Isolation Kit foi extremamente baixa (na ordem de picograma), mesmo com a utilização do PicoChip não foi possível avaliar a qualidade do RNA extraído no Bionalyser (Agilent). Por isso, o ponto de checagem para iniciar a biblioteca ficou estabelecido como sendo o resultado do aRNA, e a decisão de sequenciá-la ficou dependente do tamanho do *smear* do produto amplificado por PCR.

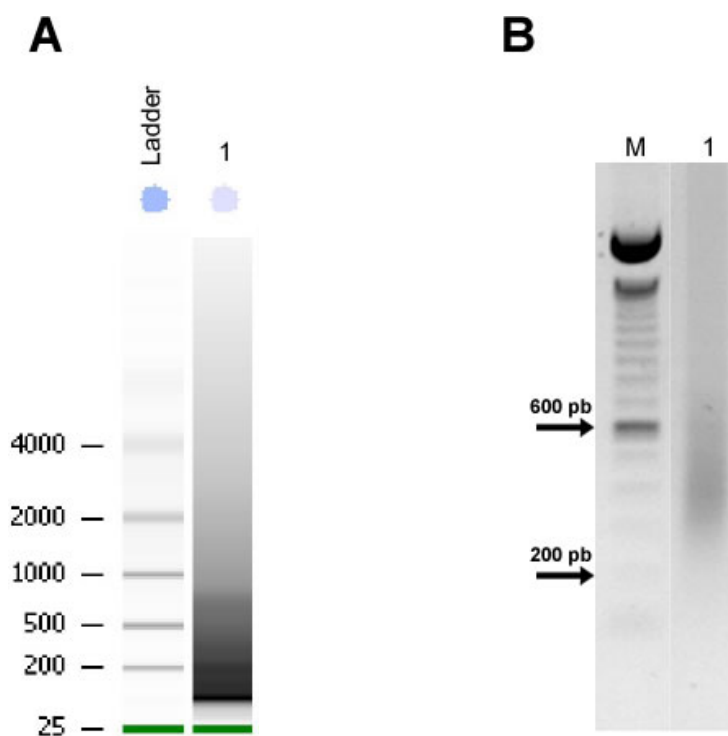
Na Tabela 4 estão apresentados os resultados para as 23 amostras de todas as etapas de checagem da construção da biblioteca. Como mostrado na Tabela 4, foi calculado pelo programa 2100expert (Agilent) a distribuição da massa do aRNA em relação ao tamanho, e todas as amostras obtiveram entre 60-80% da massa total distribuídas entre 200 a 800nt. Sendo assim, seguimos para a próxima etapa com todas as amostras.

Na mesma Tabela 4 é mostrado que a distribuição do DNA amplificado por PCR de todas as amostras ficou dentro da faixa desejável para sequenciamento no 454 (70 a 600pb), (MARGULIES et al. 2005). Como exemplificação, na Figura 6, está a eletroforese em gel de agarose 1% do material amplificado pela PCR de 15 ciclos após a etapa final da construção da biblioteca.

**Tabela 4** - Resultado das etapas da construção das bibliotecas das 21 amostras sequenciadas no 454. A porcentagem do smear do RNA é a relação da massa de aRNA compreendida entre 200 e 800nt em relação a massa total do amplificado. Os rendimentos e massa de RNA e DNA foram calculados pelo Nanodrop (Thermo). O tamanho do *smear* de DNA foi definido por eletroforese em gel de agarose 1% de cada amostra. Rec – amostras de adenocarcinoma com recidiva bioquímica; Nrec – amostras de câncer de próstata sem recidiva bioquímica

Amostra	Número de	<i>Smear</i> aRNA	Rendimento total	<i>Smear</i> DNA	Rendimento
	Tiros	(%200-800nt)	RNA (ng)	(pb)	total DNA (ng)
2-Rec	863	76	199	200-450	175
3-Rec	1200	78	270	200-450	98
4-Rec	1245	75	213	200-700	135
5-Rec	1679	62	24	150-400	1154
6-Rec	1796	68	54	150-600	489
30-Rec	1645	79	330	150-500	686
40-Rec	907	71	232	150-400	861
50-Rec	1200	79	333	150-500	365
60-Rec	1064	68	152	100-400	328
70-Rec	2000	80	763	250-600	394
1-Nrec	1094	77	277	200-600	148
10-Nrec	2001	85	147	150-700	381
17-Nrec	1500	71	124	150-600	109
20-Nrec	1000	84	130	150-500	441
25-Nrec	1205	81	140	150-400	141
27-Nrec	1000	75	91	150-400	54
35-Nrec	1556	81	129	100-500	1570
37-Nrec	1500	80	36	100-400	56
45-Nrec	1281	73	89	150-400	391
55-Nrec	1682	69	174	100-500	691
65-Nrec	2009	73	35	200-500	362





**Figura 6** – Resultados da transcrição *in vitro* e amplificação por PCR de uma amostra microdissecada a *laser*. A: Eletroforese do aRNA no Picochip do Bionalyser. Ladder: padrão de peso molecular de RNA. B: Eletroforese da PCR de 15 ciclos da mesma amostra M: padrão de peso molecular 100bp DNA Ladder (Invitrogen)

#### 4.4 SEQUENCIAMENTO E VALIDAÇÃO DAS BIBLIOTECAS PELO MÉTODO DE SANGER

##### 4.4.1 Sequenciamento individual das bibliotecas de cDNA 3'

Inicialmente, cada biblioteca foi clonada individualmente e cerca de 50 sequências foram geradas para cada uma delas. Neste primeiro passo, o objetivo foi apenas confirmar se os adaptadores estavam corretos e completos, com o *barcode* e o sítio da enzima *DpnII*. Foi observado que das sequências com boa qualidade e com adaptador 454A, 92,0% continham a sequência do adaptador com *barcode* e o sítio da *DpnII* completos, e não

apresentaram artefatos como concatâmeros ou adaptadores com *barcode* inespecíficos.

#### **4.4.2 Montagem do *pool* e validação no ABI**

Após confirmar que a estrutura de cada biblioteca estava correta, 30ng do *pool* de todas as bibliotecas foi clonado em bactérias eletrocompetentes da linhagem *Escherichia coli* DH10B. Cerca de 1000 colônias de bactéria tiveram seus insertos amplificados por PCR com os *primers* M13Rv e 454B. A partir dos *amplicons* gerados, foi feita a reação de sequenciamento com o kit BigDye TerminatorV3.1 e com o *primer* M13Rv no sequenciador ABI 3130 xl (Applied Biosystems). Cerca de 930 sequências geradas foram analisadas.

Do total de sequências geradas, 828 (89%) continham o adaptador 454A completo, das quais, 767 (83%) continham um *barcode* de 6 bases e o sítio *DpnII*. Esse grande aproveitamento de sequências demonstrou a ausência de artefatos nas diferentes etapas de construção das bibliotecas.

Para que fosse possível obter uma representação significativa do transcrito entre as amostras, foi necessário que a quantidade de sequências para cada uma delas não tivesse muita variação, de forma que o número total de sequências de cada amostra fosse semelhante. Como mostrado na Tabela 5, com exceção da amostra 5-Rec, todas as bibliotecas foram representadas no *pool*. No entanto, as 767 sequências não estão distribuídas da forma esperada, sugerindo que a quantificação por

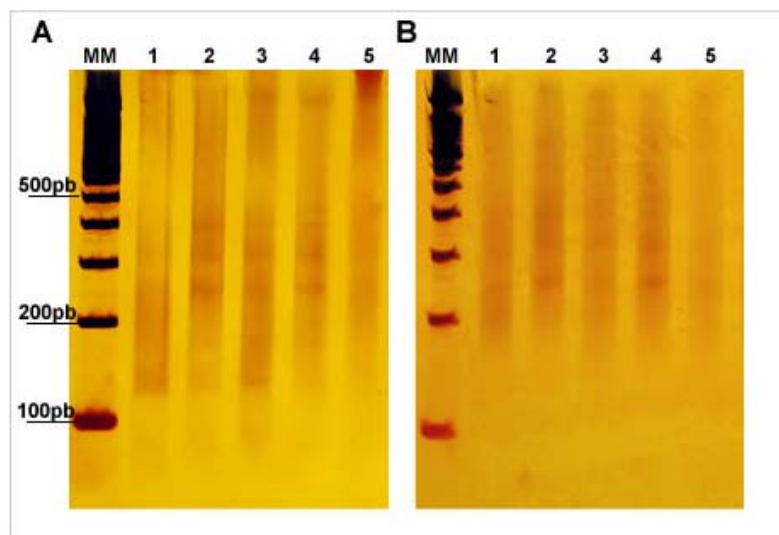
espectrofotometria não é acurada o suficiente para ser usada como calibração da concentração de cada população de DNA.

**Tabela 5** - Resultado do sequenciamento do *pool* das 21 bibliotecas pelo método de Sanger. *Barcode*: são as sequências de seis nucleotídeos utilizadas para identificar cada amostra após o sequenciamento. Rec – amostras de adenocarcinoma com recidiva bioquímica; Nrec – amostras de câncer de próstata sem recidiva bioquímica e HPB – amostras de Hiperplasia Prostática Benigna.

<b>Amostra</b>	<b>Barcode</b>	<b>Número total de sequências</b>	<b>Frequência (%)</b>
2-Rec	ATCTGT	73	9,52
3-Rec	TACTGT	16	2,09
4-Rec	TCATGT	78	10,17
5-Rec	AGTCGT	0	0,00
6-Rec	TATCGT	11	1,43
30-Rec	ACTAGA	71	9,26
40-Rec	AGCATA	15	1,96
50-Rec	AGTACA	2	0,26
60-Rec	ATCGCT	6	0,78
70-Rec	ATCTGT	88	11,47
1-Nrec	ACGTGT	31	4,04
10-Nrec	ATGCTA	127	16,56
17-Nrec	CTATGT	12	1,56
20-Nrec	CACTGA	13	1,69
25-Nrec	CAGATA	20	2,61
27-Nrec	TCGCTA	11	1,43
35-Nrec	CAGTGT	19	2,48
37-Nrec	TATGCT	9	1,17
45-Nrec	CGAGAT	16	2,09
55-Nrec	CGCTCA	102	13,30
65-Nrec	TACGCA	19	2,48
<b>Total</b>		<b>739</b>	<b>100</b>
<b>Média</b>		<b>33,34</b>	<b>4,34</b>

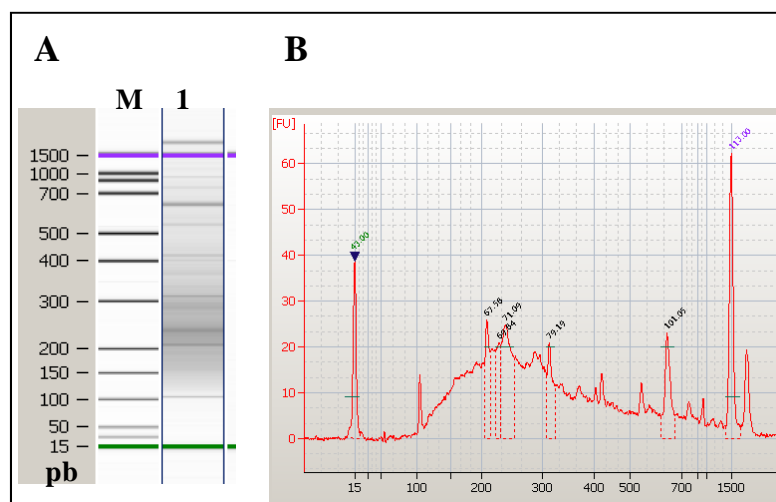
Assim, para obter uma quantificação mais precisa, as amostras foram submetidas ao processo de purificação utilizando as partículas magnéticas do kit Agencourt® AMPure® XP (Beckman Coulter, Beverly MA) para eliminação de fragmentos menores que 150pb seguida da etapa de quantificação feita por fluorometria pelo kit Quant-iT™ dsDNA HS Assay Kit (Invitrogen, Eugene, OR), a qual é mais específica pois apenas o DNA dupla fita incorpora o fluoróforo e é quantificado.

A eletroforese de cinco amostras em gel de poliacrilamida está mostrada na Figura 7 mostrando grande homogeneidade tanto na quantidade de material como na distribuição de tamanho entre as amostras. Como esperado, para todos os casos, os fragmentos menores que 150pb, que incluem excesso de *primers* e de adaptadores, foram eliminados na purificação.



**Figura 7** - Resultados de purificação das amostras com o kit Agencourt® AMPure® XP. Eletroforese em gel de poliacrilamida 8% revelado com nitrato de prata 4,5%. MM: Marcador de peso molecular 100bp (Invitrogen) A: 10ng das amostras de 1 a 5 antes da purificação; B: 10ng das amostras de 1 a 5 após a purificação.

Após quantificar por fluorometria, um segundo *pool* com 10ng de cada amostra foi formado e uma alíquota de 50ng foi analisada por eletroforese capilar no Bionalyser (Agilent). Como mostrado na Figura 8, a distribuição do *smear* está entre a faixa de 100pb e 700pb considerada aceitável para sequenciamento na plataforma Titanium 454 (Roche-Life Sciences).



**Figura 8** - Eletroforese capilar no DNA 1000pb Chip do Bionalyser. 50ng do *pool* de DNA foram aplicados no DNA 1000pb Chip do Bionalyser. A: eletroforese digital, M: marcador de peso molecular, 1: *pool*; B: eteroferograma do *pool*.

Uma fração de 20ng do *pool* foi clonado em *E. coli* DH10B e foram geradas 100 sequências pelo método de Sanger. Nesta etapa foi verificado apenas se a estrutura das bibliotecas foi mantida após a inserção de um novo método de purificação. Uma vez que estrutura se manteve inalterada seguiu-se com o sequenciamento na plataforma Titanium 454 (Roche).

## **4.5 SEQUENCIAMENTO NA PLATAFORMA GS FLX TITANIUM 454 E ANÁLISE DOS DADOS**

### **4.5.1 Mapeamento das sequências geradas na plataforma GS FLX Titanium 454**

Após duas corridas na plataforma Genome Sequencer FLX Titanium 454, foram geradas 930.102 sequências, com um tamanho médio de 149.29pb e todas elas continham a estrutura barcode e sítio da enzima DpnII corretos. Este fato mostra a eficácia da metodologia desenvolvida quanto à ausência de artefatos provenientes de etapas como ligação de adaptadores, purificação e PCR.

As sequências com alta similaridade com o DNA mitocondrial e RNA ribossomal foram filtradas pelo MEGABLAST e o restante foi mapeado contra o banco RefSeq. Um total de 634.285 sequências (68,29%) alinharam com cobertura  $\geq 80\%$  e identidade  $\geq 90\%$  no banco de dados do transcrito humano RefSeq.

As sequências mapeadas no RefSeq reportaram um total de 19.933 transcritos distintos representando 47,45% do total de sequências anotadas neste mesmo banco. Foram representados 14.102 genes distintos, 50,36% do número de genes do RefSeq.

Quando comparado os dois grupos de pacientes de CaP separadamente, foram mapeadas 262.511 sequências para o grupo das amostras com recidiva (41,39%) e 331.885 sequências para o grupo das

amostras sem recidiva (52,32%). Para o grupo com recidiva foram anotados 17.094 transcritos distintos enquanto que para o grupo sem recidiva 14.933.

**Tabela 6** - Resultados do sequenciamento na plataforma 454 das 21 bibliotecas de cDNA 3'. Seqs – número total de sequências geradas pelo 454; RefSeq – número de sequências mapeadas no banco de dados do RefSeq; Tamanho – tamanho médio das sequências mapeadas no RefSeq; RNAm – número de transcritos identificados por pelo menos uma sequência; Rec – amostras de adenocarcinoma de próstata de pacientes que sofreram recidiva bioquímica. Nrec – amostras de adenocarcinoma de pacientes que não sofreram recidiva bioquímica.

Amostra	Grupo com Recidiva				Grupo sem Recidiva				
	Seqs	RefSeq	Tamanho	RNAm	Amostra	Seqs	RefSeq	Tamanho	RNAm
2-Rec	51.855	35.649	148,9	5.542	1-Nrec	41.850	27.538	136,36	5.604
3-Rec	39.066	25.789	152,39	5.240	10-Nrec	60.360	40.799	156,71	7.539
4-Rec	60.562	43.431	155,05	5.859	17-Nrec	48.014	34.572	147,29	5.687
5-Rec	31.727	22.380	136,61	4.441	20-Nrec	60.508	45.827	168,55	7.600
6-Rec	35.398	24.699	134,29	5.372	25-Nrec	48.326	33.329	157,45	6.630
30-Rec	62.563	43.666	169,43	7.739	27-Nrec	23.272	15.019	128,73	4.015
40-Rec	47.053	27.998	153,71	6.342	35-Nrec	38.179	30.228	141,74	5.233
50-Rec	4.039	2.913	144,1	1.170	37-Nrec	46.349	26.300	124,52	5.208
60-Rec	31.126	16.303	149,89	3.596	45-Nrec	20.294	12.641	110,71	4.524
70-Rec	25.798	19.683	168,07	4.146	55-Nrec	58.420	41.760	142,17	6.734
					65-Nrec	33.795	23.872	137,97	5.402
<b>Média</b>	<b>38.918</b>	<b>26.251</b>	<b>151,24</b>	<b>4.945</b>		<b>43.579</b>	<b>30.171</b>	<b>141,11</b>	<b>5.834</b>
<b>Total</b>	<b>389.187</b>	<b>262.511</b>		<b>17.094</b>		<b>479.367</b>	<b>331.885</b>		<b>14.933</b>

Os 13.691 genes resultantes de todas as amostras de células tumorais puderam ser anotados em 122 classes funcionais segundo o *Gene Ontology Consortium Annotation* com  $p < 0,01$ . Sendo que as classes mais enriquecidas foram regulação da transcrição (648 genes), redução oxidativa (359 genes), ciclo celular (312 genes); transporte de proteínas (311 genes) e

apoptose (281 genes); essas classes e as demais estão apresentadas no Anexo 1.

Segundo o banco de dados KEGG; os 13.691 genes puderam ser classificados em 51 vias bioquímicas enriquecidas com  $p < 0,01$ ; das quais se pode ressaltar Vias em Câncer, com 235 genes e a Via específica de Câncer de Próstata com 71 genes (Anexo 2).

#### **4.5.2 Genes diferencialmente expressos entre as amostras com recidiva e sem recidiva bioquímica**

##### **4.5.2.1 Baseado em *Fold change* e/ ou *p* valor**

O cálculo de diferença de expressão gênica foi realizado com as 594.396 sequências, mapeadas no RefSeq, com o programa CLC Genomics Workbench 4.5. Os dados foram subdivididos em dois subconjuntos. O subconjunto um foi composto pelos genes que foram reportados em pelo menos 50% das amostras de cada grupo (recidiva e não recidiva). Para este subconjunto foi usado o teste estatístico de proporção *Baggerley et al.'s test*, e foram encontrados 128 genes diferencialmente expressos com *fold change*  $\geq |2|$  e  $p \leq 0.05$ . Deste total, 23 genes tiveram maior expressão no grupo com recidiva e 105 genes com maior expressão no grupo sem recidiva (Tabela 7).



**Tabela 7** - Genes diferencialmente expressos do subgrupo 1 entre as amostras de adenocarcinoma de próstata de pacientes com e sem recidiva bioquímica. Sinal negativo e positivo corresponde aos genes mais expressos nos grupos sem recidiva e com recidiva, respectivamente.

Gene	Fold change	P	Gene	Fold change	P	Gene	Fold change	P
<i>PODXL</i>	3,39	1,68E-03	<i>CYHR1</i>	-2,32	2,60E-02	<i>REPS2</i>	-3,12	1,10E-02
<i>EPB41L4A</i>	3,18	4,90E-02	<i>PRSS8</i>	-2,33	2,24E-03	<i>FAM82B</i>	-3,12	2,10E-02
<i>PNN</i>	3,07	3,20E-02	<i>CBS</i>	-2,35	8,56E-03	<i>MRPL21</i>	-3,12	2,60E-02
<i>DNAJB12</i>	3,06	1,10E-02	<i>MPZL1</i>	-2,36	6,40E-03	<i>AK2</i>	-3,13	7,05E-03
<i>C1orf63</i>	2,93	2,90E-02	<i>ATXN7L3B</i>	-2,36	4,90E-02	<i>ELF3</i>	-3,14	7,20E-03
<i>LEMD3</i>	2,84	1,90E-02	<i>DULLARD</i>	-2,39	3,30E-02	<i>HSPE1</i>	-3,15	4,40E-02
<i>LOC100293090</i>	2,81	4,48E-03	<i>PSMA4</i>	-2,40	1,80E-02	<i>NDUFB4</i>	-3,16	1,60E-02
<i>LOC100288998</i>	2,68	2,70E-02	<i>GNG5</i>	-2,46	4,96E-03	<i>C12orf57</i>	-3,17	3,31E-05
<i>SLC22A3</i>	2,67	5,01E-11	<i>ARPC2</i>	-2,50	4,48E-04	<i>LOC100291837</i>	-3,18	1,40E-02
<i>LOC100287932</i>	2,67	4,30E-02	<i>CNPY2</i>	-2,50	1,80E-02	<i>DEPDC6</i>	-3,18	1,50E-02
<i>GALNT7</i>	2,64	3,80E-02	<i>C6orf108</i>	-2,52	3,17E-03	<i>NAMPT</i>	-3,22	1,30E-02
<i>MTRNR2L10</i>	2,54	1,10E-02	<i>RPLP1</i>	-2,54	3,40E-02	<i>ESRP2</i>	-3,25	2,46E-03
<i>ITPR3</i>	2,44	2,70E-02	<i>PXMP4</i>	-2,60	4,60E-02	<i>RPS28</i>	-3,25	5,02E-06
<i>ADCY6</i>	2,38	1,20E-02	<i>PIH1D1</i>	-2,61	4,92E-04	<i>RPS12</i>	-3,31	1,30E-02
<i>GPR107</i>	2,28	3,65E-03	<i>C22orf13</i>	-2,63	2,80E-02	<i>C7orf50</i>	-3,33	4,50E-02
<i>CKAP5</i>	2,26	1,70E-02	<i>NFKBIZ</i>	-2,63	3,55E-03	<i>AHCYL1</i>	-3,36	4,80E-02
<i>EGFR</i>	2,20	1,57E-03	<i>CHD3</i>	-2,65	1,10E-02	<i>UBA52</i>	-3,45	7,72E-04
<i>CEP350</i>	2,20	4,20E-02	<i>RPS5</i>	-2,67	3,47E-03	<i>PEX6</i>	-3,51	4,30E-02
<i>MGEA5</i>	2,11	1,90E-02	<i>TMSB10</i>	-2,68	1,73E-03	<i>MYL6</i>	-3,52	2,30E-02
<i>SRSF5</i>	2,05	2,50E-02	<i>AKR7A2</i>	-2,68	3,60E-02	<i>CTSH</i>	-3,71	4,40E-02
<i>SKAP2</i>	2,04	1,31E-03	<i>POLD4</i>	-2,70	2,70E-02	<i>COX6A1</i>	-3,82	2,48E-03
<i>ZNF226</i>	2,03	6,49E-03	<i>RPS20</i>	-2,71	6,13E-03	<i>OSTC</i>	-3,83	3,70E-02
<i>NDUFAF1</i>	2,02	3,80E-02	<i>NDUFV1</i>	-2,73	2,00E-02	<i>C2orf24</i>	-3,84	2,30E-02
<i>RPL32</i>	-2,03	2,10E-02	<i>SAP18</i>	-2,73	4,50E-02	<i>REEP5</i>	-3,91	2,10E-02
<i>PYGO2</i>	-2,06	7,25E-03	<i>COX7A2</i>	-2,73	4,70E-02	<i>CHCHD2</i>	-3,97	1,25E-03

Cont./ Tabela 7

<i>EDC4</i>	-2,06	4,90E-02	<i>FUNDC2</i>	-2,77	2,00E-02	<i>ANKRD37</i>	-4,05	4,90E-02
<i>CAMTA1</i>	-2,07	1,40E-02	<i>EMG1</i>	-2,82	5,94E-03	<i>PARK7</i>	-4,08	4,80E-02
<i>PLOD1</i>	-2,08	1,20E-02	<i>MRPL30</i>	-2,84	1,20E-02	<i>S100A6</i>	-4,09	4,16E-03
<i>RPL34</i>	-2,09	4,28E-03	<i>RNPEP</i>	-2,84	1,20E-02	<i>ETFB</i>	-4,12	4,21E-04
<i>TMC4</i>	-2,09	4,90E-02	<i>ESYT1</i>	-2,86	4,94E-04	<i>ATPIF1</i>	-4,14	3,20E-02
<i>ZNF148</i>	-2,10	4,10E-02	<i>LOC100291443</i>	-2,86	4,00E-02	<i>FTL</i>	-4,26	2,67E-03
<i>RPL24</i>	-2,11	2,20E-02	<i>RPL38</i>	-2,89	3,90E-02	<i>BRP44</i>	-4,29	5,16E-03
<i>RPS15</i>	-2,15	2,14E-03	<i>POLR2L</i>	-2,93	1,70E-02	<i>C2orf28</i>	-4,63	6,11E-03
<i>EIF3K</i>	-2,16	1,20E-02	<i>RPL8</i>	-2,93	3,90E-02	<i>ACTR3</i>	-4,68	3,10E-07
<i>FAM134A</i>	-2,17	1,84E-03	<i>TSPAN1</i>	-2,93	4,30E-02	<i>RPS29</i>	-4,75	2,90E-02
<i>RPL29</i>	-2,17	2,60E-02	<i>TRPM4</i>	-2,95	1,70E-03	<i>GADD45G</i>	-4,98	3,40E-02
<i>RPL39</i>	-2,17	1,80E-02	<i>TMEM50B</i>	-2,96	2,70E-02	<i>RPS27</i>	-5,04	4,90E-03
<i>C3orf33</i>	-2,18	2,80E-02	<i>ATOX1</i>	-2,98	1,10E-02	<i>UQCRQ</i>	-5,08	3,60E-02
<i>RAB7A</i>	-2,19	1,50E-02	<i>COMMD3</i>	-2,99	4,00E-02	<i>RPS26</i>	-5,63	1,10E-02
<i>LRRC47</i>	-2,19	2,40E-02	<i>CRK</i>	-2,99	4,80E-02	<i>DYNLL1</i>	-5,64	1,76E-03
<i>XPO7</i>	-2,24	1,70E-02	<i>ITM2C</i>	-3,01	8,47E-03	<i>KRT19</i>	-6,25	4,30E-02
<i>GSPT1</i>	-2,27	4,60E-02	<i>STUB1</i>	-3,03	2,00E-02	<i>PTMA</i>	-6,25	4,60E-02
<i>JUND</i>	-2,31	1,92E-04	<i>RPL35</i>	-3,04	7,27E-05			

O subconjunto 2 foi composto pelos genes que foram reportados em pelo menos 50% das amostras em um dos grupos e no máximo em uma amostra no outro grupo. Para o subconjunto dois, foram considerados diferencialmente expressos os genes com *fold change*  $\geq |2,5|$ , totalizando 118 genes, sendo que 36 genes mais expressos no grupo com recidiva e 82 genes mais expressos no grupo sem recidiva (Tabela 8).

**Tabela 8** - Genes diferencialmente expressos do subgrupo 2 entre as amostras de adenocarcinoma de próstata de pacientes com e sem recidiva bioquímica. Sinal negativo e positivo corresponde aos genes mais expressos nos grupos sem recidiva e com recidiva, respectivamente.

Gene	Fold change	Gene	Fold change	Gene	Fold change	Gene	Fold change
<i>DNAJC25</i>	17,60	<i>MMP11</i>	5,50	<i>ZNF844</i>	-6,36	<i>ANKZF1</i>	-9,09
<i>ZNF317</i>	12,10	<i>GPATCH8</i>	5,00	<i>HRSP12</i>	-7,00	<i>EDEM1</i>	-9,09
<i>LOXL4</i>	10,00	<i>CWC25</i>	4,40	<i>SENP2</i>	-7,00	<i>FBN1</i>	-9,09
<i>ZCCHC4</i>	10,00	<i>LOC440600</i>	3,85	<i>SYPL1</i>	-7,00	<i>LOC100294411</i>	-9,09
<i>C14orf132</i>	10,00	<i>ACBD5</i>	3,30	<i>TCF19</i>	-7,00	<i>MAGOHB</i>	-9,09
<i>HMG5</i>	10,00	<i>JARID2</i>	2,75	<i>TMEM101</i>	-7,00	<i>WDR17</i>	-9,09
<i>PSMF1</i>	10,00	<i>POLRMT</i>	-2,73	<i>TSEN34</i>	-7,00	<i>ACOX3</i>	-10,00
<i>INTS6</i>	9,90	<i>RAB1F</i>	-2,73	<i>COMMD4</i>	-7,27	<i>RPL36</i>	-10,00
<i>ZNF548</i>	8,80	<i>HNRNPUL1</i>	-2,73	<i>CTNNA1</i>	-7,27	<i>GDI2</i>	-10,91
<i>FAM76B</i>	7,70	<i>PRIC285</i>	-3,18	<i>IQCE</i>	-7,27	<i>HIF1A</i>	-10,91
<i>LOC100292047</i>	7,70	<i>BMS1</i>	-3,64	<i>LOC100129195</i>	-7,27	<i>HIST1H4J</i>	-10,91
<i>PDE4C</i>	7,70	<i>TXNL1</i>	-3,86	<i>OLFML3</i>	-7,27	<i>MFSD7</i>	-10,91
<i>SMA5</i>	7,70	<i>DNMT1</i>	-3,94	<i>SYTL4</i>	-7,27	<i>PLXNA1</i>	-10,91
<i>SNX16</i>	7,70	<i>CTNNA1</i>	-4,09	<i>PTGS2</i>	-7,73	<i>PXMP2</i>	-10,91
<i>TRIM16L</i>	7,70	<i>MARS2</i>	-4,09	<i>NDUFB5</i>	-8,00	<i>GADD45B</i>	-11,00
<i>UBE2L3</i>	7,70	<i>ADAMTS4</i>	-4,55	<i>PAQR7</i>	-8,00	<i>XRRA1</i>	-11,82
<i>MFSD6</i>	7,00	<i>C10orf10</i>	-4,55	<i>ADCY2</i>	-8,18	<i>ACLY</i>	-12,00
<i>CCDC6</i>	6,60	<i>C1orf163</i>	-4,55	<i>C12orf43</i>	-8,18	<i>DEAF1</i>	-12,73
<i>IRAK2</i>	6,60	<i>IPO4</i>	-5,00	<i>CCDC22</i>	-8,18	<i>PPCDC</i>	-12,73
<i>PDE6B</i>	6,60	<i>AXL</i>	-5,46	<i>FXRD2</i>	-8,18	<i>DPT</i>	-13,64
<i>PPP1R12A</i>	6,60	<i>BANF1</i>	-5,46	<i>KLHDC8B</i>	-8,18	<i>NR2C2AP</i>	-14,00
<i>RG9MTD3</i>	6,60	<i>ZNF337</i>	-6,00	<i>MFAP1</i>	-8,18	<i>SUMO2</i>	-15,46
<i>DEF6</i>	6,00	<i>FAM129B</i>	-6,00	<i>RFC3</i>	-8,18	<i>TMEM18</i>	-15,46
<i>LOC100132870</i>	6,00	<i>HSD3B7</i>	-6,00	<i>SUCLG2</i>	-8,18	<i>GSS</i>	-16,00
<i>RBM22</i>	6,00	<i>KIAA0408</i>	-6,00	<i>SUMO1</i>	-8,18	<i>LOC100293228</i>	-18,18
<i>CHRM3</i>	5,50	<i>PDCD5</i>	-6,36	<i>EVL</i>	-9,00	<i>TBK1</i>	-19,09
<i>APIP</i>	5,50	<i>CHD6</i>	-6,36	<i>NUP210</i>	-9,00	<i>RNF5</i>	-23,00
<i>DICER1</i>	5,50	<i>HPGD</i>	-6,36	<i>TMEM205</i>	-9,00	<i>DAD1</i>	-27,27
<i>KLHL2</i>	5,50	<i>LOC152024</i>	-6,36	<i>MRPS24</i>	-9,09		
<i>KREMEN1</i>	5,50	<i>STIM1</i>	-6,36	<i>UBE2E2</i>	-9,09		

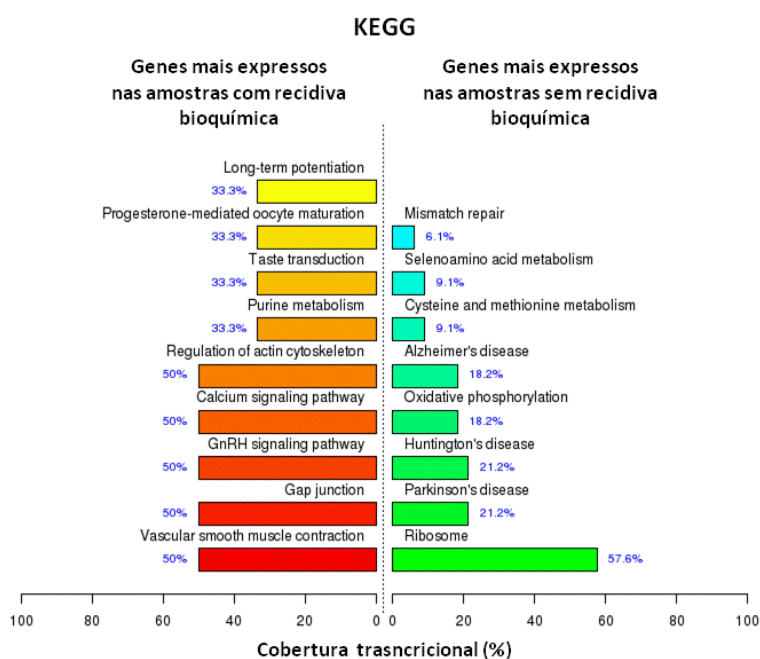
Ao todo foram definidos como diferencialmente expressos 246 genes, sendo 59 genes com maior expressão nas amostras do grupo com recidiva e 187 genes com maior expressão nas amostras sem recidiva.

A fim de investigar as funções biológicas dos genes diferencialmente expressos, foi feita a classificação de acordo com *Gene Ontology* (GO) e com o banco de dados de vias bioquímicas do KEGG. Com o uso da ferramenta FunNet28, foi possível identificar 3 categorias de processos biológicos do *Gene Ontology* enriquecidos no grupo com recidiva e outros 12 enriquecidos no grupo sem recidiva. Os genes diferencialmente expressos de cada processo biológico enriquecido estão apresentados no Quadro 1.

**Quadro 1** - Processos biológicos com relevância estatística no grupo de genes diferencialmente expressos entre as amostras com e sem recidiva bioquímica de acordo com o *Gene Ontology*.

Processo Biológico	p	Genes
<b>Amostras com Recidiva Bioquímica</b>		
Resposta a Íons Cálcio	5.38e-03	<i>EGFR; ITPR3</i>
Montagem do Complexo Proteico	2.63e-02	<i>NDUFAF1; SKAP2</i>
Comunicação Celular	1.70e-02	<i>KREMEN1; SNX16</i>
<b>Amostras sem Recidiva Bioquímica</b>		
Elongação Traducional	7.75e-20	<i>RPL8; RPL24, RPL29; RPL32;RPL34; RPL35; RPL36; RPL38; RPL39; RPLP1; RPS5; RPS12; RPS15; RPS20; RPS26; RPS27; RPS28; RPS29; UBA52</i>
Transporte	9.12e-03	<i>ETFB; MAGOHB; NDUFB4; NDUFB5; NDUFV1; SYPL1; SYTL4; TXNL1; UQCRQ</i>
Cadeia de Transporte de Elétron	4.13e-04	<i>ETFB; NDUFB4; NDUFB5; NDUFV1; TXNL1; UQCRQ</i>
Regulação da Transcrição do Promotor da RNA Polimerase II	2.87e-02	<i>SAP18; CHD3; CRK; JUND; TCF19</i>
Resposta ao Estresse Oxidativo	3.09e-03	<i>ATOX1; GSS; NDUFB4; PARK7; PTGS2</i>
Tradução	1.08e-02	<i>GSPT1; LRRC47; MRPL21; MRPL30; MRPS24</i>
Transporte de RNAm	2.26e-03	<i>MAGOHB; NUP210; SENP2; XPO7</i>
Regulação Negativa da Atividade da Proteína Sinais	6.05e-03	<i>DEPDC6; GADD45B; GADD45G</i>
Regulação do Ciclo Celular	6.52e-03	<i>GADD45B; GADD45G; PTGS2</i>
Transporte Mitocondrial de Elétron, de NADH para Ubiquinona	9.2e-03	<i>NDUFB4; NDUFB5; NDUFV1</i>
Regulação Positiva da Polimerização dos Filamentos de Actina	3.46e-03	<i>ACTR3; ARPC2</i>
Ativação da MAPKKK	9.87e-04	<i>GADD45B; GADD45G</i>

Dentre os 246 genes diferencialmente expressos, 9 categorias do banco de dados KEGG foram enriquecidas pelos genes com maior expressão no grupo com recidiva, e 7 categorias no grupo sem recidiva (Figura 9). Os genes diferencialmente expressos de cada categoria enriquecida estão apresentados no Quadro 2.



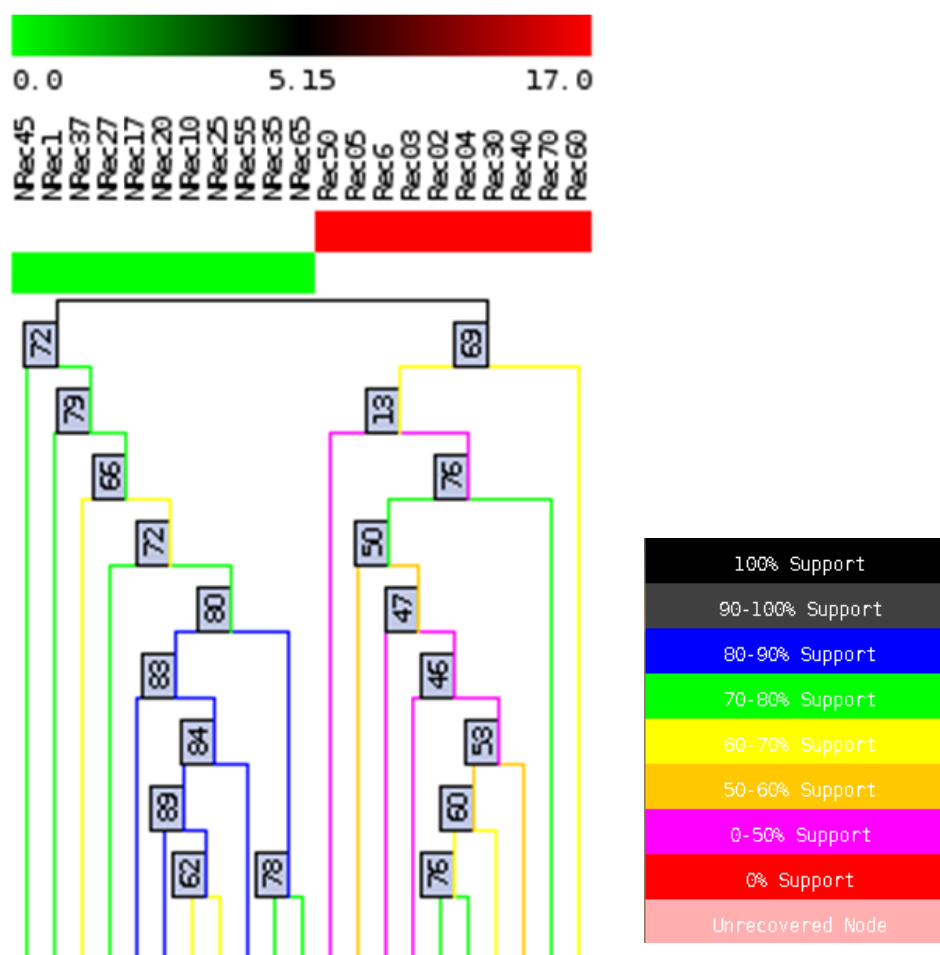
**Figura 9** - Classificação das categorias biológicas dos genes diferencialmente expressos de acordo com o KEGG. Os genes *up regulated* são os genes mais expressos no grupo com recidiva, enquanto que os genes *down regulated* são os genes mais expressos no grupo sem recidiva.

**Quadro 2** - Descrição das categorias biológicas enriquecidas e dos respectivos genes mais expressos representados nas amostras com e sem recidiva bioquímica de acordo com o KEGG.

Via Metabólica	p	Genes
<b>Amostras com Recidiva Bioquímica</b>		
Contração Músculo Liso Vascular	2.02e-03	<i>ADCY6; ITPR3; PPP1R12A</i>
Adesão Celular	9.92e-04	<i>ADCY6; EGFR; ITPR3</i>
Via de Sinalização do Hormônio Liberador de Gonadotropina (GnRH)	1.39e-03	<i>ADCY6; EGFR; ITPR3</i>
Via de Sinalização do Cálcio	6.95e-03	<i>CHRM3; EGFR; ITPR3</i>
Regulação do Citoesqueleto de Actina	1.19e-02	<i>EGFR; ITPR3; PPP1R12A</i>
Metabolismo de Purina	4.76e-02	<i>ADCY6; PDE4C</i>
Transdução de Sabor	6.1e-03	<i>ADCY6; ITPR3</i>
Maturação do Oócito Mediada por Progesterona	2.35e-02	<i>ADCY6; PDE4C</i>
Potenciação de Longa Duração	1.09e-02	<i>ADCY6; PPP1R12A</i>
<b>Amostras sem Recidiva Bioquímica</b>		
Ribossomos	3.82e-19	<i>RPL8; RPL24, RPL29; RPL32;RPL34; RPL35; RPL36; RPL38; RPL39; RPLP1; RPS5; RPS12; RPS15; RPS20; RPS26; RPS27; RPS28; RPS29; UBA52</i>
Doença de Parkinson	1.50e-03	<i>COX6A1; COX7A2; NDUFB4; NDUFB5; NDUFV1; PARK7; UQCRQ</i>
Doença de Huntington	9.35e-03	<i>COX6A1; COX7A2; NDUFB4; NDUFB5; NDUFV1; POLR2L; UQCRQ</i>
Fosforilação Oxidativa	7.55e-03	<i>COX6A1; COX7A2; NDUFB4; NDUFB5; NDUFV1; UQCRQ</i>
Doença de Alzheimer	2.23e-02	<i>COX6A1; COX7A2; NDUFB4; NDUFB5; NDUFV1; UQCRQ</i>
Metabolismo de Cisteína e Metionina	9.26e-03	<i>AHCYL1; CBS; DNMT1</i>
Metabolismo de Selenioamina	4.33e-03	<i>AHCYL1; CBS; MARS2</i>
Reparo de Alteração no DNA	3.49e-02	<i>POLD4; RFC3</i>



Para verificar se o padrão de expressão baseado nos 246 genes diferencialmente expressos poderia discriminar as amostras segundo a recidiva bioquímica, foi realizada uma clusterização hierárquica das 21 amostras de adenocarcinoma de próstata (Figura 10). Como mostrado na Figura 10, 100% das amostras com recidiva foram corretamente discriminadas de 100% das amostras sem recidiva, revelando um padrão de expressão associado a recidiva em células tumorais de próstata, e sugerindo que esses genes devem desempenhar importante papel neste processo.



**Figura 10** - Clusterização hierárquica baseada nos 246 genes diferencialmente expressos entre os grupos de amostras com recidiva e sem recidiva. Em vermelho são as amostras com recidiva e em verde amostras sem recidiva bioquímica. No detalhe ao lado está a legenda do suporte do cluster avaliado pela técnica do Bootstrap.

#### **4.5.2.2 Baseado no cálculo de probabilidades**

Em colaboração com o Dr. Carlos AB Pereira foi desenvolvido um modelo matemático para identificar transcritos diferencialmente expressos baseado em probabilidades. Inicialmente foi realizado um teste de qui-quadrado que identificou 896 transcritos com potencial para serem diferencialmente expressos entre os grupos com e sem recidiva.

Usando o modelo matemático descrito no item 3.4.2.2, foi calculada a probabilidade de cada gene ser mais expresso no grupo sem recidiva em relação ao grupo com recidiva. Os transcritos foram ordenados de acordo com o padrão de expressão e aqueles com probabilidades próximo a 1 são os genes mais expresso no grupo sem recidiva em relação ao grupo com recidiva, enquanto que os transcritos com probabilidade próximo a zero são os genes mais expressos no grupo com recidiva. Na Tabela 9 estão representados os transcritos com probabilidades maiores que 70% de serem mais expressos nos grupos com recidiva bioquímica (66 genes) e sem recidiva bioquímica (248 genes), respectivamente, totalizando 314 genes. Os valores de probabilidade para os 896 transcritos estão apresentados no Anexo 3.

**Tabela 9** - Representação dos transcritos com probabilidades maiores que 70% de serem mais expressos nos grupos com e sem recidiva.

Gene	Prob	Gene	Prob	Gene	Prob	Gene	Prob	Gene	Prob
<i>LOC100293090</i>	12,77	<i>PAQR7</i>	69,53	<i>ACLY</i>	71,33	<i>LRRC47</i>	73,25	<i>WBSCR22</i>	77,75
<i>SRSF5</i>	16,48	<i>DEAF1</i>	69,58	<i>TMEM205</i>	71,42	<i>AKR7A2</i>	73,37	<i>RPL34</i>	77,77
<i>PODXL</i>	16,60	<i>KRT19</i>	69,58	<i>PPCDC</i>	71,46	<i>RPL32</i>	73,48	<i>S100A6</i>	77,77
<i>SLC25A4</i>	18,35	<i>RHOC</i>	69,59	<i>CSDE1</i>	71,50	<i>TBK1</i>	73,51	<i>GSS</i>	77,87
<i>MTRNR2L10</i>	18,74	<i>MFSD7</i>	69,60	<i>NUPR1</i>	71,51	<i>COX6C</i>	73,77	<i>C2orf24</i>	77,88
<i>PNN</i>	19,07	<i>RPL31</i>	69,60	<i>C1R</i>	71,52	<i>RPL26L1</i>	73,80	<i>C6orf108</i>	77,89
<i>LOC100288871</i>	21,60	<i>WDR17</i>	69,60	<i>RPS14</i>	71,52	<i>KDEL1</i>	73,83	<i>RPL41</i>	78,07
<i>LOC100288998</i>	21,86	<i>HIF1A</i>	69,62	<i>RPL21P28</i>	71,58	<i>NR2C2AP</i>	73,86	<i>KLF4</i>	78,08
<i>RBM25</i>	22,07	<i>MRPS24</i>	69,63	<i>POMP</i>	71,59	<i>RBM47</i>	73,91	<i>RPL38</i>	78,09
<i>SLC25A3</i>	22,26	<i>NAMPT</i>	69,64	<i>ABCD3</i>	71,60	<i>PYGO2</i>	73,96	<i>RPL13</i>	78,16
<i>LOC653881</i>	22,43	<i>LOC100291837</i>	69,70	<i>PTRH1</i>	71,61	<i>ATPIF1</i>	74,04	<i>UQCR11</i>	78,20
<i>SRSF5</i>	23,41	<i>C19orf62</i>	69,71	<i>ITM2C</i>	71,62	<i>LOC100293228</i>	74,05	<i>RPL39</i>	78,39
<i>ACTG2</i>	23,96	<i>EGR1</i>	69,71	<i>HSD17B4</i>	71,64	<i>OSTC</i>	74,05	<i>JUND</i>	78,53
<i>GPR107</i>	24,27	<i>GPKOW</i>	69,71	<i>LOC728606</i>	71,69	<i>AHNAK</i>	74,06	<i>RPL8</i>	78,77
<i>MTRNR2L3</i>	24,31	<i>C11orf75</i>	69,75	<i>UBE2J1</i>	71,69	<i>RAB7A</i>	74,10	<i>RPL32</i>	79,23
<i>C22orf29</i>	24,70	<i>YIF1A</i>	69,76	<i>RPL26</i>	71,71	<i>EMG1</i>	74,13	<i>APRT</i>	79,59
<i>RPL6</i>	25,48	<i>NAA50</i>	69,85	<i>C14orf132</i>	71,73	<i>RPL30</i>	74,15	<i>ITM2C</i>	79,63
<i>LDHB</i>	25,51	<i>XBP1</i>	69,87	<i>CNOT6</i>	71,78	<i>RPL36</i>	74,18	<i>GNG5</i>	79,67
<i>BTBD6</i>	25,55	<i>NIT2</i>	69,89	<i>TOMM7</i>	71,82	<i>SRSF3</i>	74,18	<i>RPL23</i>	79,94
<i>ARF6</i>	26,16	<i>GTF3A</i>	69,90	<i>RNPEP</i>	71,83	<i>RNF5</i>	74,29	<i>PTMA</i>	79,95
<i>CIAO1</i>	26,21	<i>ZBTB4</i>	69,90	<i>GDI2</i>	71,84	<i>SNHG6</i>	74,32	<i>EIF3K</i>	80,05
<i>SFRS18</i>	26,30	<i>GNPAT</i>	69,91	<i>FAM134A</i>	71,87	<i>SUMO2</i>	74,34	<i>COX7A2</i>	80,13
<i>FBXL12</i>	26,44	<i>TMEM42</i>	69,91	<i>MRPS18A</i>	71,87	<i>MRPL27</i>	74,38	<i>AK2</i>	80,23
<i>SLC30A9</i>	26,52	<i>PCA3</i>	69,92	<i>ARPC2</i>	71,90	<i>RHOT2</i>	74,44	<i>C2orf28</i>	80,45
<i>DNAJB12</i>	26,60	<i>CDC37</i>	69,99	<i>RPL24</i>	71,91	<i>ABHD12</i>	74,47	<i>PRSS8</i>	80,80

TAF8	26,67	CYHR1	70,00	CDK5RAP2	71,94	LOC400590	74,56	BRP44	80,88
FBXW2	26,75	SLC33A1	70,01	PIH1D1	71,98	HAX1	74,66	CHCHD2	81,24
C11orf92	26,78	TMEM93	70,06	RPS18	72,01	NDUFB2	74,67	TMSB10	81,30
MLXIP	27,22	ATP5G1	70,07	MRPL15	72,02	C17orf49	74,72	GADD45G	81,41
ACOT9	27,29	ZCCHC6	70,12	LOC647979	72,04	ENO1	74,78	BRP44	81,99
LEMD3	27,77	DDT	70,17	ST14	72,05	C6orf108	74,85	RPS27	82,09
ALAD	27,81	PFDN1	70,21	MAGOHB	72,06	RPL37	74,94	RPS20	82,21
CCL28	27,85	EVL	70,24	CST3	72,14	RPS5	74,94	RPS26	82,63
SKAP2	27,85	VPS35	70,24	RPL35A	72,14	LRPAP1	74,99	ETFB	82,66
LYRM7	27,86	PSMF1	70,29	EPS8L2	72,19	MCM3AP	74,99	RPL32	82,81
EPB41L4A	28,02	SPECC1L	70,29	TRIM27	72,21	SNRPG	75,16	RPS10	82,89
EEF1B2	28,07	C8orf30B	70,33	MRPL21	72,24	ESRP2	75,18	NDUFB4	83,01
NSF	28,09	AMACR	70,39	ATP5L2	72,25	ROMO1	75,23	FTL	83,20
OPA3	28,23	C7orf50	70,39	MYL6	72,27	TPD52	75,23	RPL34	83,37
CPNE2	28,25	GLTSCR2	70,39	BEX4	72,31	CAMTA1	75,45	TRPM4	83,38
ACOT9	28,28	DHX9	70,43	TSPAN1	72,35	LOC400590	75,45	BRP44	83,41
DNAJC25	28,33	ARF4	70,44	ETFB	72,45	RPL13	75,50	RPL41	83,96
DCP1A	28,35	COX8A	70,46	CADM1	72,47	RPS12	75,58	FABP5	84,20
ITPR3	28,37	ENDOD1	70,48	DEPDC6	72,50	TUBB2A	75,60	TRPM4	85,33
MTRNR2L1	28,40	BCR	70,50	SPG21	72,51	UQCRQ	75,62	UBA52	85,47
EIF3I	28,42	HSPB1	70,51	LGALS1	72,55	DYNLL1	75,94	RPS15	86,01
SLC25A3	28,53	HOXB13	70,54	COX5B	72,57	ITGAE	76,13	NFKBIZ	87,53
C10orf32	28,87	COMMD3	70,61	HSPE1	72,59	UFM1	76,16	C12orf57	89,17
ZNF317	28,87	MRPL24	70,72	RPS27L	72,62	FUNDC2	76,26	RPS28	89,36
LOC100130474	29,06	CYHR1	70,76	CHMP4A	72,63	RPS17	76,26	RPL35	90,41
EEF1B2	29,18	MGST2	70,78	CLDN7	72,64	PARK7	76,30		
AFF1	29,26	ULK1	70,85	STUB1	72,66	UBE2E2	76,34		
TTC39C	29,26	RPL37A	70,88	GADD45B	72,68	TFF3	76,39		
SYAP1	29,32	PSMB6	70,93	RPL10A	72,74	RPLP1	76,53		

---

<i>RPL6</i>	29,41	<i>NDUFB5</i>	70,96	<i>TOP1MT</i>	72,75	<i>RPL29</i>	76,67
<i>FOXK1</i>	29,56	<i>NISCH</i>	70,96	<i>COX6A1</i>	72,77	<i>APRT</i>	76,77
<i>GOLGA3</i>	29,64	<i>CTSH</i>	70,97	<i>LOC400590</i>	72,77	<i>JUN</i>	76,88
<i>HSP90B1</i>	29,69	<i>ATP5F1</i>	71,02	<i>PDCD5</i>	72,83	<i>NFKBIZ</i>	76,90
<i>CNPY2</i>	29,71	<i>COMMD4</i>	71,04	<i>TMEM50B</i>	72,86	<i>DAD1</i>	76,96
<i>ANKH</i>	29,72	<i>POLD4</i>	71,05	<i>C5orf32</i>	72,94	<i>ZFP36</i>	77,04
<i>CAT</i>	30,09	<i>PXMP4</i>	71,06	<i>ELF3</i>	73,04	<i>CCDC28A</i>	77,08
<i>SLC44A4</i>	30,21	<i>C19orf10</i>	71,14	<i>ATOX1</i>	73,10	<i>TMEM9</i>	77,17
<i>CHRM3</i>	30,26	<i>FTH1</i>	71,15	<i>CCDC56</i>	73,16	<i>REEP5</i>	77,25
<i>LOC100287552</i>	30,32	<i>NAGA</i>	71,19	<i>NUP210</i>	73,17	<i>RPS29</i>	77,38
<i>CXorf38</i>	30,34	<i>ACAA1</i>	71,31	<i>KRTCAP2</i>	73,19	<i>RPL28</i>	77,52
<i>LOC100288418</i>	30,44	<i>CNPY2</i>	71,32	<i>DYNLL1</i>	73,25	<i>ACTR3</i>	77,71

---

Prob: probabilidade do gene ser mais expresso nas amostras sem recidiva bioquímica. Uma probabilidade  $\geq 70\%$  significa que o transcrito tem 70% de chances de ser mais expresso no grupo sem recidiva bioquímica, enquanto que uma probabilidade  $\leq 30\%$  mostra que o transcrito tem 70% de chances de ser mais expresso no grupo com recidiva bioquímica.

## 4.6 VALIDAÇÃO TÉCNICA DOS GENES DIFERENCIALMENTE EXPRESSOS

Para investigar a estratégia e os métodos matemáticos utilizados para identificar as diferenças no nível de expressão gênica nos dois grupos de amostras, genes candidatos foram selecionados para a validação por RT-qPCR. Foram selecionados 87 genes para serem validados por RT-qPCR no mesmo grupo de amostras representativas de células epiteliais tumorais de adenocarcinoma de próstata. Dos 87 genes selecionados, 57 deles foram identificados nas duas metodologias, de proporção e de probabilidade, apresentando valores de probabilidade de 17 a 90%. Os detalhes dessa etapa para cada método estão apresentados a seguir.

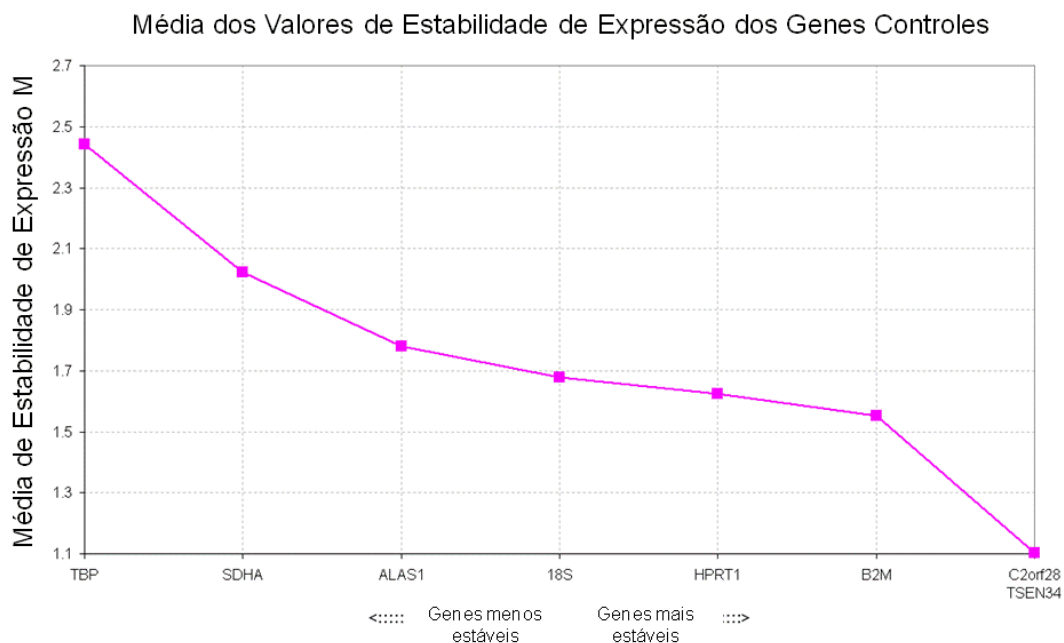
### 4.6.1 Reação de RT-qPCR com o sistema *Taqman Low Density Array* (TLDA)

Partindo do RNA amplificado primeiro *round* remanescente das 21 amostras de CaP, seguiu-se para amplificação por dois *rounds* com o kit RiboAmp® HSPlus (Arcturus) com consequente síntese de cDNA. Após realizar as reações de RT-qPCR conforme descrito no item 3.5.2, a estabilidade gênica dos 87 genes mais os 5 genes endógenos *18S*, *ALAS1*, *B2M*, *HPRT1*, *SDH*, foi avaliada através da ferramenta geNorm.

O programa geNorm calcula a medida de estabilidade gênica (M) para os genes endógenos em um grupo de amostras indicando os mais estáveis para serem utilizados como normalizadores. O valor M sugerido pelo

programa é de 1,5, abaixo do qual se tem os melhores genes endógenos. A estabilidade é determinada partindo-se do princípio de que dois genes normalizadores ideais possuem razões de expressão idênticas em todas as amostras de cDNA analisadas, independente das condições experimentais ou de tipos celulares.

Na Figura 11 está representado o valor médio da estabilidade dos 5 genes endógenos assim como dos genes *C2orf28* e *TSEN 34*. Maior estabilidade foi observada para estes dois últimos genes ( $M= 1,1$ ), sendo estes os escolhidos como normalizadores para os cálculos de diferença de expressão com os dados de RT-qPCR.



**Figura 11** - Gráfico da média dos valores de estabilidade de expressão dos genes *18S*, *ALAS1*, *B2M*, *HPRT1*, *SDH*, *C2orf28* e *TSEN34*. Os genes que estão localizados à esquerda do gráfico apresentam menor estabilidade comparada aos genes que se localizam à direita do gráfico. Assim, os genes mais estáveis determinados pelo geNorm são: *C2orf28* e *TSEN34* e o menos estável é *TBP*.

#### 4.6.2 Análise dos resultados das reações de RT-qPCR dos candidatos selecionados pelo método de proporção

Foram considerados válidos os genes que apresentaram valores de  $Ct \leq 38$ , e os genes cujas amostras obtiveram  $Ct \geq 38$  em mais de cinquenta por cento das amostras de cada grupo (recidiva e não recidiva) foram eliminadas das análises, a fim de evitar que o valor do *fold change* seja enviesado por uma parcela não representativa de amostras dos dois grupos. Sendo assim dos 87 genes iniciais seguiram para análise de diferença de expressão 76 (87,36%).

A fim de confirmar a diferença de expressão entre os grupos recidiva e não recidiva foi definido como diferencialmente expressos os genes com  $fold\ change \geq |2|$  e, validados aqueles com o mesmo sinal comparado com os dados do RNASeq.

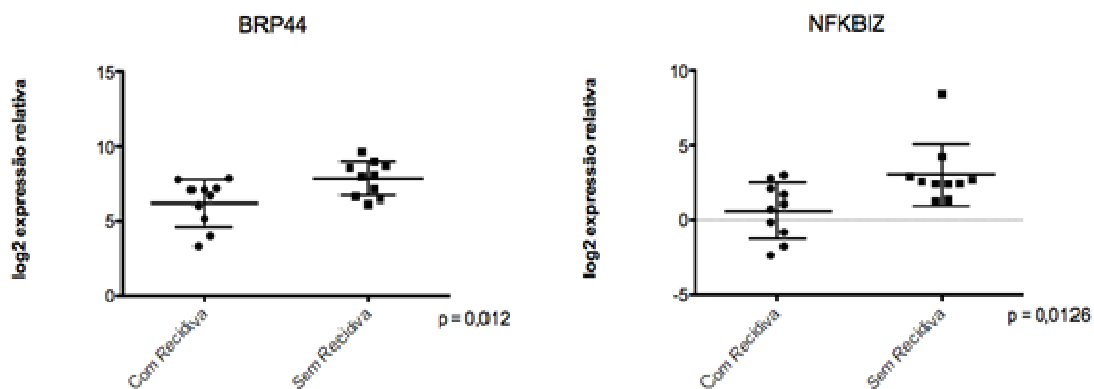
Dos 76 genes avaliados que apresentaram resultados confiáveis nos ensaios de RT-qPCR, 29 (38,16%) apresentaram dados concordantes entre as duas abordagens (RNAseq e RT-qPCR) usando o critério  $fold\ change \geq |2|$ . (Tabela 10).



**Tabela 10** - Comparação entre os resultados de diferença de expressão dos dados do 454 (RNA-Seq) com a validação técnica (RT-qPCR). Em vermelho são os genes concordantes entre os dois métodos. Sinal negativo e positivo corresponde aos genes mais expressos nos grupos de amostras sem recidiva e com recidiva, respectivamente.

Gene	<i>Fold</i> Recidiva/ Sem Recidiva		Gene	<i>Fold</i> Recidiva/ Sem Recidiva	
	RNA-Seq	qRT-PCR		RNA-Seq	qRT-PCR
ACLY	-12	1,37	MFSD7	-10,91	11,49
ACOX3	-10	3,26	MRPS24	-9,09	1,13
<b>ACTR3</b>	<b>-4,68</b>	<b>-4,83</b>	NDUFB5	-8	-1,62
AK2	-3,13	3,35	<b>NFKBIZ</b>	<b>-2,63</b>	<b>-14,42</b>
<b>ANKZF1</b>	<b>-9,09</b>	<b>-5,52</b>	NR2C2AP	-14	1,34
<b>ARPC2</b>	<b>-2,5</b>	<b>-2,29</b>	<b>NUP210</b>	<b>-9</b>	<b>-124,83</b>
<b>BRP44</b>	<b>-4,29</b>	<b>-2,76</b>	PIH1D1	-2,61	-1,04
C2orf28	-4,63	-1,01	PLXNA1	-10,91	1,3
<b>CBS</b>	<b>-2,35</b>	<b>-22,87</b>	PNN	3,07	-7,85
<b>CCDC22</b>	<b>-8,18</b>	<b>-3,25</b>	<b>PODXL</b>	<b>3,39</b>	<b>2,51</b>
<b>DAD1</b>	<b>-27,27</b>	<b>-2,2</b>	<b>PPCDC</b>	<b>-12,73</b>	<b>-1,94</b>
DEAF1	-12,73	8,12	PRSS8	-2,33	-1,13
DEF6	6	-9,57	PSMF1	-10	-1,67
DPT	-13,64	-1,6	PXMP2	-10,91	-1,71
EDEM1	-9,09	-1,4	PYGO2	-2,06	-1,72
EGFR	2,2	-372,47	RBM22	6	-1,51
<b>ELF3</b>	<b>-3,14</b>	<b>-2,7</b>	RFC3	-8,18	-1,51
<b>EMG1</b>	<b>-2,82</b>	<b>-2,52</b>	RPL34	-2,09	-1,7
<b>ESRP2</b>	<b>-3,25</b>	<b>-5,06</b>	<b>RPL35</b>	<b>-3,04</b>	<b>-1,99</b>
<b>ESYT1</b>	<b>-2,86</b>	<b>-13,2</b>	<b>S100A6</b>	<b>-4,09</b>	<b>-5,13</b>
<b>ETFB</b>	<b>-4,12</b>	<b>-2,31</b>	SENP2	-7	1,49
<b>EVL</b>	<b>-9</b>	<b>-3,19</b>	SKAP2	2,04	1,85
<b>FAM129B</b>	<b>-6</b>	<b>-4,12</b>	SLC22A3	2,67	-1,61
FAM134A	-2,17	1,78	SUCLG2	-8,18	2,46
<b>GADD45B</b>	<b>-11</b>	<b>-14,03</b>	SUMO1	-8,18	-1,19
<b>GDI2</b>	<b>-10,91</b>	<b>-9,15</b>	<b>TBK1</b>	<b>-19,09</b>	<b>-3,54</b>
GPR107	2,28	-2,47	<b>TMEM101</b>	<b>-7</b>	<b>-2,58</b>
GSS	-16	7,38	TMEM18	-15,46	-1,28
HIF1A	-10,91	-1,35	TMEM205	-9	1,44
HMGNS	-10	1,02	TMSB10	-2,68	-1,07
HRSp12	-7	-1,76	TSEN34	-7	1
<b>INTS6</b>	<b>9,9</b>	<b>2,45</b>	<b>UBE2E2</b>	<b>-9,09</b>	<b>-4,64</b>
ITM2C	-3,01	-1,29	WDR17	-9,09	-1,92
<b>KIAA0408</b>	<b>-6</b>	<b>-8,3</b>	ZCCHC4	10	-2,79
<b>KLHDC8B</b>	<b>-8,18</b>	<b>-2,01</b>	ZNF226	2,03	-1,58
LOXL4	10	1,43	ZNF317	12,1	1,55
MAGOHB	-9,09	-1,16	ZNF337	-6	1,81
<b>MFSD6</b>	<b>7</b>	<b>2,86</b>	ZNF548	8,8	-4,64

Dentre os 29 genes com dados de diferença de expressão concordantes entre o RT-qPCR e o RNA-Seq, apenas dois genes apresentaram uma diferença estatística com p valor significativo ( $p < 0,05$ ): NFKBIZ ( $p = 0,013$ ) e BRP44 ( $p = 0,012$ ), Figura 14.



**Figura 12** - Gráfico da expressão relativa dos genes *BRP44* e *NFKBIZ* com os dados da validação técnica por RT-qPCR.

#### 4.6.3 Análise dos resultados das reações de RT-qPCR dos candidatos selecionados pelo método de probabilidades

Dos 87 que foram selecionados para validação, baseado nos dados de proporção, 57 genes foram identificados como diferencialmente expressos pelo método de probabilidades. Esses 57 genes possuem uma probabilidade de serem mais expressos no grupo sem recidiva de 13 a 90%.

Os dados de RT-qPCR confirmam a diferença de expressão de 22 genes (38,60%) desses 57. Considerando que, quanto mais próximo de 1 maior a probabilidade do gene ser mais expresso nas amostras sem

recidiva, e quanto mais próximo de zero maior a probabilidade do gene ser mais expresso nas amostras com recidiva, confrontamos a validação por RT-qPCR com diferentes faixa de probabilidades desses 57 genes (Tabela 11).

**Tabela 11** - Distribuição dos 57 genes testados por RT-qPCR de acordo com a faixa de probabilidade.

Faixa de Probabilidade (%)	$0,13 \leq p \leq 0,90$	$p \geq 0,7$ $p \leq 0,3$	$p \geq 0,75$ $p \leq 0,25$	$p \geq 0,80$ $p \leq 0,20$	$p \geq 0,82$ $p \leq 0,18$	$p \geq 0,85$ $p \leq 0,15$
Número de genes candidatos	57	43	17	11	6	2
Candidatos validados	22	19	9	5	5	2
Taxa de validação (%)	38,60	44,19	52,94	45,45	83,33	100,00

Como mostrado na Tabela 12, quanto maior a probabilidade do gene ser diferencialmente expresso, maior a taxa de validação por RT-qPCR, sendo que um significativo aumento nesta taxa de validação foi observado quando comparamos a validação dos 57 genes com os genes presentes na faixa  $p \geq 0,82$  e  $p \leq 0,18\%$  (36,60% e 83,33% respectivamente).

#### **4.7 IDENTIFICAÇÃO DE GENES CLASSIFICADORES BASEADOS NO MÉTODO DE PROBABILIDADES**

Uma vez que o modelo matemático de probabilidades elaborado não se restringe a calcular a probabilidade de um único gene ser mais ou menos expresso, mas sim um conjunto de genes. Foi avaliado o potencial de discriminar as amostras de cada grupo com o conjunto de genes de genes com maior e menor valor de probabilidade que representam os genes mais

expressos nas amostras sem e com recidiva, respectivamente. Desta forma, foi calculado a probabilidade de discriminar uma amostras como pertencendo ao grupo sem ou com recidiva baseando-se no padrão de expressão dos 12 genes com as probabilidades mais extremas (Tabela 13) A discriminação das amostras foi feita com 4 agrupamentos distintos de genes: apenas com o gene com maior probabilidade de ser expresso nas amostras sem recidiva; com 2, 3 e 6 pares de genes com maior probabilidade de serem mais expressos respectivamente nas amostras sem e com recidiva (Tabela 13).

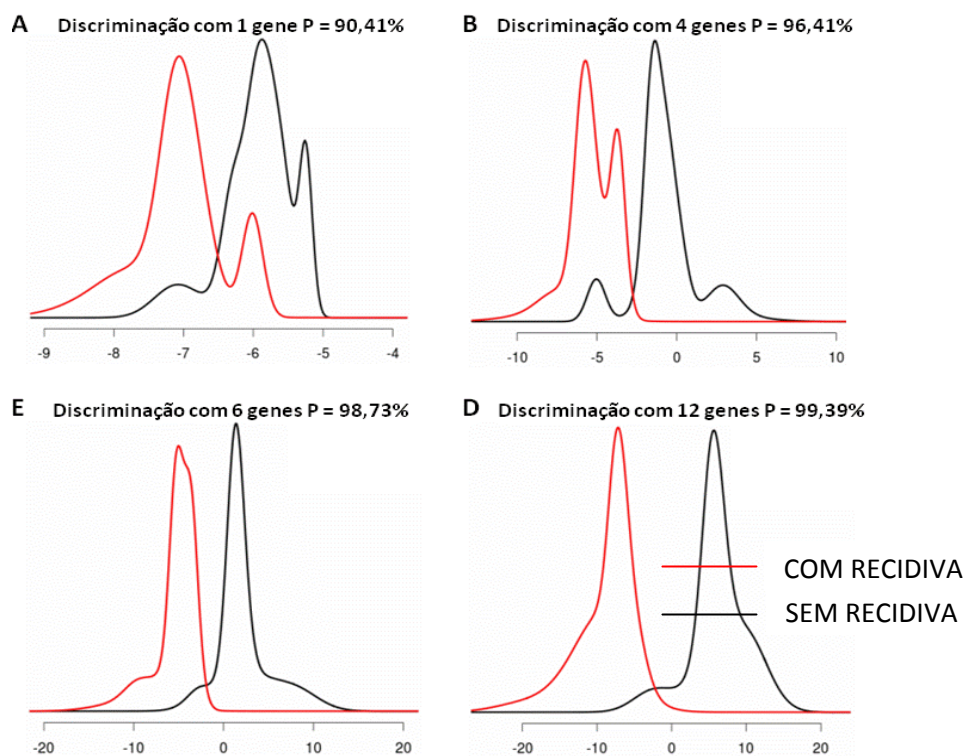
**Tabela 12** - Valores de probabilidade dos 12 genes escolhidos como putativos classificadores entre as amostras sem recidiva e com recidiva.

Gene	Prob(SR>CR) %	Gene	Prob(SR>CR) %
<i>RPL35</i>	90,41	<i>PNN</i>	19,07
<i>RPS28</i>	89,36	<i>MTRNR2L10</i>	18,74
<i>C12orf57</i>	89,17	<i>SLC25A4</i>	18,35
<i>NFKBIZ</i>	87,53	<i>PODXL</i>	16,60
<i>RPS15</i>	86,01	<i>SRSF5</i>	16,48
<i>UBA52</i>	85,47	<i>LOC100293090</i>	12,77

Prob (SR>CR): probabilidade do gene ser mais expresso nas amostras sem recidiva (SR) em relação as amostras com recidiva (CR).

A Figura 13 mostra o perfil da probabilidade de discriminação entre as amostras dos grupos com e sem recidiva de acordo com o número de genes empregado como classificadores. Quando usado apenas o gene *RPL35* existe a probabilidade de 90,41% de discriminá-la corretamente entre os dois grupos. Quando empregado o par de genes *RPL35* (p 90,41%) e *RPS28* (p 89,36%) juntamente com o par *SRSF5* (p 16,48%) e *LOC100293090* (p

12,77%) o poder de discriminação das amostras subiu para 96,41%. Com o trio de genes *RPL35* (p 90,41%), *RPS28* (p 89,36%) e *C12orf57* (p 89,17%) e trio *PODXL* *SRSF5* (p 16,48%) e *LOC100293090* (p 12,77%) a probabilidade de discriminação foi de 98,73% das amostras; e por fim com a probabilidade dos 12 genes da Tabela 13 foi de 99,39% (Figura 13).



**Figura 13** - Discriminação das amostras com e sem recidiva com os genes classificadores: (a) Com o nível de expressão de um único gene (*RPL35*). (b) Combinação do nível de expressão de 4 genes (*RPL35*, *RPS28*, *SRSF5*, *LOC100293090*), (c) 6 genes (*RPL35*, *RPS28*, *C12orf57*, *PODXL*, *SRSF5*, *LOC100293090*), e (d) 12 genes (*RPL35*, *RPS28*, *C12orf57*, *NFKBIZ*, *RPS15*, *UBA52*, *PNN*, *MTRNR2L10*, *SLC25A4*, *PODXL*, *SRSF5*, *LOC100293090*).

Vale ressaltar que os 12 genes selecionados como classificadores estão dentro da faixa de probabilidade  $p \geq 0,82$  e  $p \leq 0,18$  a qual apresentou alta taxa de validação no RT-qPCR (83,33%). Esse fato mostra que o

modelo de probabilidades é uma ferramenta valiosa na busca de conjunto de genes classificadores baseados em dados de RNASeq.

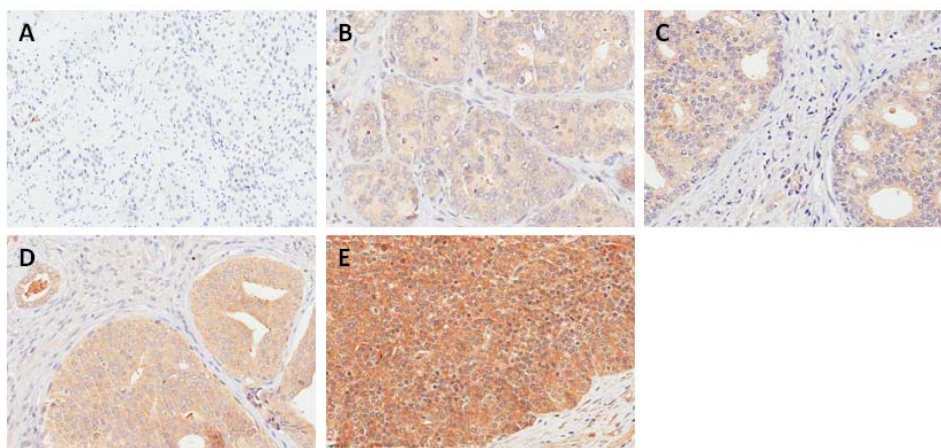
## **4.8 VALIDAÇÃO BIOLÓGICA DOS GENES DIFERENCIALMENTE EXPRESSOS POR IMUNOISTOQUÍMICA**

### **4.8.1 Reações de imunistoquímica e classificação das amostras**

Dentre os genes diferencialmente expressos apontados pelo método de proporção, os genes *NFKBIZ* e *BRP44*, mais expressos no grupo de pacientes que não apresentaram recidiva, foram validados por RT-qPCR com significância estatística ( $p = 0,013$  e  $p = 0,012$ , respectivamente). Esses mesmos genes apresentaram, respectivamente, probabilidades de 87,53 e 83,41% de serem mais expressos no grupo sem recidiva pelo método de probabilidades. Sendo assim, esses genes foram selecionados para avaliar a expressão de seus respectivos produtos proteicos.

Para a imunistoquímica da proteína BRP44, um total de 92 amostras independentes foram categorizadas como negativas (marcação ausente ou fraca) ou positivas (marcação moderada fraca, moderada forte ou forte), conforme descrito no item 3.6.2. As amostras foram distribuídas em 60 casos negativos, sendo 23 com marcação ausente e 37 com marcação fraca; e 32 casos considerados positivos, sendo 16 de marcação moderada fraca, 12 de marcação moderada-forte e 4 de marcação forte.

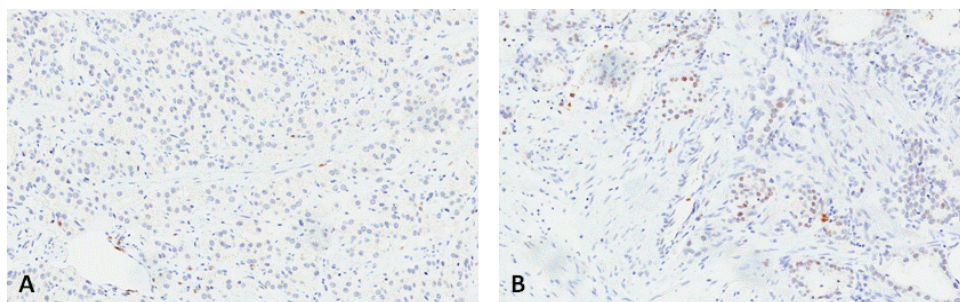
A Figura 14 mostra que a marcação da proteína BRP44 foi citoplasmática e restrita às células epiteliais do tumor.



**Figura 14** - Expressão proteica de BRP44 em adenocarcinoma de próstata. A marcação ficou restrita as células epiteliais do tumor. A: marcação negativa - escore 0, B: marcação fraca –escore 1; C marcação moderada escore 2; D marcação moderada forte – escore 3; E marcação forte – escore 4. Aumento 200X.

Para avaliar a expressão da proteína codificada pelo gene NKFBIZ (I $\kappa$ B- $\zeta$  – I $\kappa$ ppaBzeta), um total de 50 amostras independentes foram categorizadas como 30 casos negativos (marcação ausente) e 20 casos positivos (independente do número de células marcadas). Neste caso a marcação ficou restrita ao núcleo das células epiteliais do tumor. Apesar do número de células marcadas ser relativamente baixo em relação ao tecido como um todo, a marcação da proteína I $\kappa$ B- $\zeta$  ocorreu apenas em células do tumor o que permitiu prosseguir para as análises subsequentes (Figura 17).

O maior número de amostras para avalliar a expressão da proteína BRP44 se deve ao fato da quantidade limitada de anticorpo anti I $\kappa$ B- $\zeta$  que restou depois da padronização da marcação. As 50 amostras da imunohistoquímica da proteína I $\kappa$ B- $\zeta$  também foram avaliadas dentro das 92 amostras para a BRP44.



**Figura 15** - Expressão proteica de IκB-ζ em adenocarcinoma de próstata. A marcação ficou restrita as células epiteliais do tumor e com localização forma nuclear. A: marcação negativa - escore 0, B: marcação positiva –escore 1. Aumento 200X.

#### **4.8.2 Avaliação do potencial prognóstico de BRP44 e IκB-ζ**

O potencial prognóstico de BRP44 e IκB-ζ foi avaliado utilizando os resultados de imunohistoquímica das amostras de adenocarcinoma de próstata (item 3.6.1). Para as amostras foram categorizadas como negativas e positivas de acordo com o descrito no item anterior e os resultados são mostrados na Tabela 14.

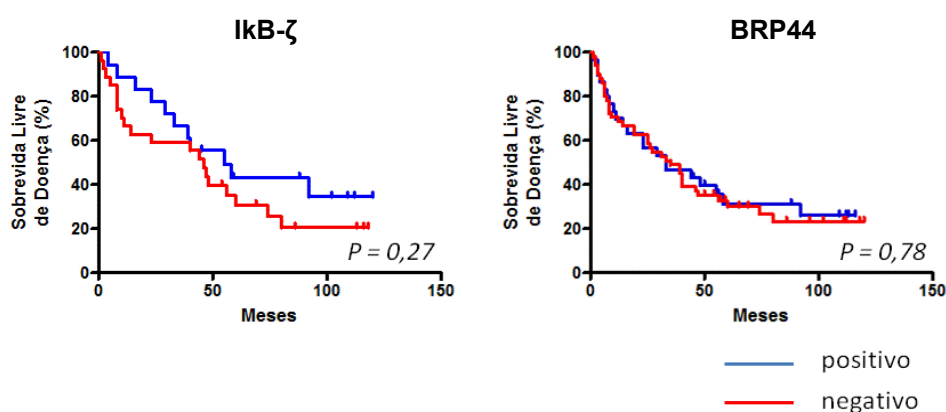


**Tabela 13** - Associação entre expressão proteica de BRP44 e I $\kappa$ B- $\zeta$  com, as características clinicopatológicas das amostras de adenocarcinoma de próstata usada no imunoistoquímica.

Variável	Categoria	BRP44, n (%)		P	I $\kappa$ B- $\zeta$ , n (%)		P
		Positivo	Negativo		Positivo	Negativo	
Vesícula Seminal	Positiva	4 (25,00)	12 (75,00)	0,06	0 (0,00)	7 (100,00)	0,03*
	Negativa	28 (36,84)	48 (63,16)		20 (46,51)	23 (53,49)	
Margem	Positiva	28 (40,00)	42 (60,00)	0,07	15 (42,85)	20 (57,15)	0,52
	Negativa	4 (18,18)	18 (81,82)		5 (33,34)	10 (66,66)	
Linfonôdo	Positivo	3 (42,85)	4 (57,15)	0,18	1 (16,66)	5 (83,33)	0,38
	Negativo	29 (34,12)	56 (65,88)		19 (43,18)	25 (56,81)	
Invasão angiolinfática	Positivo	8 (36,36)	14 (63,64)	0,85	3 (23,07)	10 (76,92)	0,14
	Negativo	24 (34,28)	46 (65,72)		17 (45,94)	20 (54,06)	
PSA antes da cirurgia	< 10ng/mL	15 (34,10)	29 (65,90)	0,84	13 (50,00)	13 (50,00)	0,133
	> 10ng/mL	17 (36,17)	30 (63,83)		7 (29,16)	17 (70,84)	
Gleason	≤6	13 (33,33)	26 (66,67)	0,65	12 (60,00)	8 (40,00)	0,01*
	7	11 (33,33)	22 (66,67)		6 (33,33)	12 (66,66)	
	≥8	8 (40,00)	12 (60,00)		2 (20,00)	10 (80,00)	
Estadiamento	pT2	4 (19,05)	17 (80,95)	0,08	4 (20,00)	10 (33,33)	0,30
	pT3	28 (39,45)	43 (60,56)		16 (80,00)	20 (66,64)	
Recidiva Bioquímica	Com Recidiva	23 (23,19)	46 (76,81)	0,61	13 (35,13)	24 (64,86)	0,23
	Sem Recidiva	9 (39,14)	14 (60,86)		7 (53,84)	6 (46,16)	
Classificação D'Ámico	Baixo	7 (30,43)	16 (69,56)	0,81	7 (30,43)	5 (69,56)	0,11
	Intermediário	14 (37,84)	23 (62,16)		10 (37,84)	13 (62,16)	
	Alto	10 (32,25)	21 (67,75)		3 (32,25)	12 (67,75)	

Foram observadas associações com significância estatística entre a positividade de I $\kappa$ B- $\zeta$  com: menor incidência de vesícula seminal comprometida ( $P=0,03$ ); com menor escore de Gleason ( $P=0,01$ ). Tanto vesícula seminal livre de doença como um escore de Gleason menor são características de bom prognóstico em CaP. Estes dados, portanto, sugerem que a presença de I $\kappa$ B- $\zeta$  pode ser um fator de prognóstico favorável em adenocarcinoma de próstata.

Em seguida, foi verificada associação entre a marcação citoplasmática de I $\kappa$ B- $\zeta$  e BRP44, sobrevida livre de doença em pacientes com câncer de próstata submetidos a prostatectomia radical. O tempo de seguimento variou de um mês até 120 meses, sendo que o seguimento mínimo para os casos sem recidiva foi de 40 meses. No entanto, não houve associação da expressão das proteínas I $\kappa$ B- $\zeta$  e BRP44 com significância estatística para sobrevida livre de recidiva bioquímica ( $P=0,27$  e  $P=0,78$ , respectivamente), Figura 16.



**Figura 16** - Curvas de Kaplan-Meier para cálculo de sobrevida livre de doença para I $\kappa$ B- $\zeta$  e BRP44

## 5 DISCUSSÃO

No presente trabalho foi estabelecida com sucesso uma metodologia para analisar quantitativamente o transcrito de células microdissecadas a *laser*. O estabelecimento de protocolos com esse grau de refinamento é de suma importância para trabalhar com populações de células homogêneas capturadas de um tecido complexo. Além disso, a microdissecção a *laser* é fundamental no estudo de tumores com histologia heterogênea e multifocal como o câncer de próstata.

As sequências geradas com o GS FLX Titanium System 454 tem um tamanho médio maior que de outros sequenciadores de nova geração, o que permite um mapeamento mais robusto com um menor número de sequências. Além disso, um fator importante da plataforma de sequenciamento escolhida é que a construção da biblioteca não depende de um kit comercial específico como no caso de outros sequenciadores. Sendo assim, por ser uma plataforma aberta, desenvolvemos um protocolo baseado no uso de enzima de restrição e sequenciamento da região 3' dos transcritos. No entanto, o protocolo desenvolvido é flexível e outras adaptações poderiam ser feitas para trabalhar com outros tipos de amostras ou mesmo de molécula.

Portanto, para aumentar a sensibilidade na identificação de transcritos, e conseqüentemente sua capacidade quantitativa, o método foi adaptado para representar somente a extremidade 3' dos transcritos. Uma

vez que a extremidade 3' do transcrito termina em uma cauda poli A, a estratégia de sequenciamento adotada foi unidirecional, iniciando na extremidade oposta da cauda poli A, ou seja, todas as bibliotecas foram sequenciadas a partir do adaptador 454A e isso permitiu que cada *barcode* possa ser identificado com sucesso.

O fato de usarmos a enzima de corte frequente *DpnII*, para fragmentação dos transcritos, confere uma menor redundância quando comparado com uma fragmentação randômica do cDNA, pelo fato de que, na estratégia com enzima de restrição, uma determinada região do transcrito é coberta por uma única sequência, enquanto que, na fragmentação randômica ocorre a sobreposição de mais de uma sequência para a mesma região. Mesmo utilizando a abordagem por restrição essa metodologia possibilita representar 89% dos transcritos do RefSeq.

O sequenciamento em paralelo das bibliotecas 21 gerou 930.102 sequências com a realização de duas corridas na plataforma GS FLX Titanium 454. No entanto, era esperado um número maior de sequências de acordo com o potencial de geração de sequências informado pelo fabricante. Na primeira corrida, foi obtido 569.318 sequências sendo que o esperado era de cerca de um milhão.

Segundo o suporte técnico da Roche, esse menor rendimento ocorreu devido a grande extensão do tamanho dos fragmentos da biblioteca, de 100 a 800pb, com uma concentração entre 150 e 300pb, sendo que o ideal deveria ser entre 300 a 800pb. Durante a PCR em emulsão (emPCR) os fragmentos menores são preferencialmente amplificados acarretando em

uma maior quantidade de massa dessas moléculas pequenas em relação às maiores. No sequenciamento, o sinal luminoso gerado pelos amplicons das moléculas pequenas é mais intenso mascarando o sinal dos poços ao lado, isso seria a razão pelo baixo aproveitamento de sequências dessa corrida.

Sendo assim, decidimos fazer mais uma corrida no 454. No entanto, a fim de obter um maior aproveitamento do número de sequências, para o próximo sequenciamento a estratégia adotada foi separar a fração de cDNA em faixas mais estreitas e sequenciar apenas a fração contendo os fragmentos maiores.

Para separar a fração do cDNA em faixas de tamanho mais estreitas, foi escolhido o método de purificação com as partículas magnéticas do kit Agencourt® AMPure® XP segundo o protocolo do próprio 454 da Roche (Paired End Library Preparation Method Manual pg 12-13). O resultado da eletroforese capilar dessa fração mostrou que houve um enriquecimento na quantidade de fragmentos maiores na fração 300-800pb, no entanto não foi possível eliminar completamente as moléculas inferiores a 300pb nesta fração. Uma vez que houve um visível enriquecimento dos fragmentos maiores que 300pb, nós decidimos em prosseguir com o sequenciamento da fração 300-800.

Apesar do resultado da eletroforese capilar mostrar um enriquecimento de moléculas acima de 300pb, essa diferença não foi identificada nos resultados do sequenciamento da fração maior que gerou um total de 360.163 sequências. Uma possível explicação para esse fenômeno seria que as poucas moléculas abaixo de 300pb ainda presentes

na fração 300-800, foram o suficiente para serem preferencialmente amplificadas na PCR em emulsão e conseqüentemente sequenciadas com maior eficiência. Uma vez que não houve diferença significativa entre as duas frações, nós decidimos juntar os dois conjuntos de dados para seguir com a análise de diferença de expressão.

Outra explicação dada pelo suporte de sequenciamento da Roche se refere a cauda poli A no final de cada fragmento. Pelo fato da sequência terminar em um homopolímero de adeninas, o sinal gerado neste ciclo é muito maior que a média dos sinais gerados ao longo da molécula de cDNA. Sendo assim, quando é feita a normalização, a intensidade do sinal do cDNA é analisado como ruído e descartado pelo programa que converte o sinal luminoso em sequência de bases.

Um forma de amenizar o efeito da cauda poli A é diminuir o número de ciclos de sequenciamento a fim de evitar que seja sequenciada essa região. Em um segundo trabalho, com bibliotecas de câncer de mama elaboradas com o mesmo protocolo apresentado nesta tese, foi realizado essa modificação para o sequenciamento. Com a modificação do número de ciclos o número de sequências obtidas foi 30% superior ao protocolo de sequenciamento convencional.

O padrão de expressão gênica entre as amostras com e sem recidiva de CaP foi determinado por dois métodos distintos. O primeiro, método de proporção, que considera o número total de sequências de cada biblioteca em relação ao total de sequências geradas nos dois grupos (com e sem recidiva bioquímica), e atribui pesos assumindo uma distribuição Beta

binomial, identificou 246 transcritos diferencialmente expressos. Sendo 56 transcritos com maior expressão nas amostras do grupo com recidiva e 190 com maior expressão nas amostras sem recidiva.

O segundo método, baseado na probabilidade de um gene ser mais expresso em dos dois grupos de amostras, permitiu identificar um total de 314 transcritos diferencialmente expressos, sendo 66 mais expressos no grupo com recidiva e 248 genes no grupo sem recidiva, entre a faixa de probabilidade  $p \geq 0,7$  e  $p \leq 0,3$ , ou seja, com 70% de probabilidade de ser mais expresso em um dos grupos. Vale ressaltar que dos 246 transcritos tidos como diferencialmente expressos pelo método de proporcionalidade, 124 (50,40%) também foram reportados pela faixa  $70\% \leq p \leq 30\%$  do método de probabilidade.

Dos 87 transcritos diferencialmente expressos dentre as amostras com e sem recidiva, determinados pelo método de proporção e escolhidos para validação, foi possível confirmar por meio da técnica do RT-qPCR, um total de 29 transcritos, representando 33,33%. Dentre esses 87 transcritos, 57 foram identificados nas duas metodologias, de proporção e de probabilidade. Sendo que os genes *BRP44* e *NFKBIZ* foram identificados com diferença de expressão estaticamente significativa ( $p = 0,013$  e  $p = 0,012$ , respectivamente).

Quanto aos 57 genes identificados como diferencialmente expressos pelo método de probabilidades, nota-se que quanto maior a probabilidade do gene ser diferencialmente expresso, maior a taxa de validação por RT-qPCR, sendo que um significativo aumento nesta taxa de validação foi

observado quando comparamos a validação dos 57 genes com os genes presentes na faixa  $p \geq 0,82$  e  $p \leq 0,18$  (36,60% e 83,33% respectivamente).

Por meio do modelo de probabilidades, foi determinado um conjunto de 12 genes (*RPL35*, *RPS28*, *C12orf57*, *NFKBIZ*, *RPS15*, *UBA52*, *PNN*, *MTRNR2L10*, *SLC25A4*, *PODXL*, *SRSF5*, *LOC100293090*) que resultaram em uma probabilidade de 99,39% de discriminar uma amostra como pertencendo ao grupo sem ou com recidiva. Um poder de discriminação muito próximo ao obtido pelo conjunto de 12 genes foi obtido quando avaliado a discriminação com 6 genes (*RPL35*, *RPS28*, *C12orf57*, *PODXL*, *SRSF5*, *LOC100293090*),  $p = 98,73\%$ .

Uma vez que um número menor de genes é menos restrigente quando comparado a 12 genes, e que os valores de discriminação entre esses dois conjuntos foram muito próximos, poderíamos propor a utilização desses 6 genes como classificadores entre pacientes com e sem recidiva com escore de Gleason 7. No entanto, ainda é necessário validar esse conjunto de 6 genes em grupo de amostras independentes.

O fato dos genes *NFKBIZ* e *BRP44* terem sido identificados como diferencialmente expressos pelo método de proporção (*fold change* -2,63 e -4,29 respectivamente) e pelo método de probabilidades ( $p = 87,53$  e  $83,41$ , respectivamente), além de terem sido validados com significância estatística pelo RT-qPCR ( $p = 0,013$  e  $p = 0,012$ ), a expressão proteica desses dois genes foi avaliada em um grupo independente de amostras por imunohistoquímica.



A marcação da proteína BRP44 foi específica para o citoplasma das células epiteliais do tumor. No entanto, não foi possível estabelecer correlações entre a marcação de BRP44 e com as características clinicopatológicas das amostras. Da mesma forma não houve diferença entre as curvas de tempo livre de doença para os grupos positivos e negativos para BRP44.

Esses dados não descartam a proteína BRP44 como um marcador de bom prognóstico específico para pacientes de CaP com escore de Gleason 7, uma vez que o RNAm foi identificado como diferencialmente expresso pelos dois métodos distintos, e validado por RT-qPCR no grupo de 21 amostras iniciais onde todas foram classificadas como Gleason 7. Estudos com maior número de amostras Gleason 7 poderiam confirmar ou não a proteína BRP44 como um marcador de bom prognóstico.

A única publicação sobre esta proteína (OTERO et al. 2010) relata sua função como ligante do anti oxidante ácido ferúlico. O ácido ferúlico é um composto fenólico disponível através de uma dieta rica em vegetais como arroz e milho. Este composto fenólico age inibindo a formação dos dímeros de timina no DNA das células e reduz a proliferação celular induzindo a apoptose em uma variedade de tumores (KHANDUJA et al. 2006). Em vista dessas informações, poderíamos supor que a maior expressão do gene *BRP44* nas amostras sem recidiva agiria sinergicamente com o ácido ferúlico.

A proteína I $\kappa$ Bzeta (I $\kappa$ B- $\zeta$ ) teve uma marcação nuclear específica com esperado com base na literatura (TRINH et al. 2008). Foram

encontradas associações com significância estatística entre a positividade de I $\kappa$ B- $\zeta$  com: menor incidência de vesícula seminal comprometida (P=0,03) e com menor escore de Gleason (P=0,01). Tanto vesícula seminal livre de doença como um escore de Gleason menor são características de bom prognóstico em CaP. Estes dados, portanto, sugerem que a presença de I $\kappa$ B- $\zeta$  pode ser um fator de prognóstico favorável em adenocarcinoma de próstata.

Apesar de ter sido possível separar as curvas de sobrevida de tempo livre de doença com a marcação de I $\kappa$ B- $\zeta$ , mostrando que os pacientes com marcação positiva para I $\kappa$ B- $\zeta$  apresentam maior tempo médio para sofrerem recidiva bioquímica, não houve diferença estatística entre as amostras com marcação positiva e negativa (P=0,27). O baixo número de amostras (n=50) para o estudo deste marcador pode ter influenciado diretamente na separação das curvas de sobrevida. Além disso, o fato de 70% das amostras avaliadas na imunistoquímica terem com margem positiva não descarta a possibilidade dessa proteína ser um bom marcador para pacientes com tumores menos agressivos.

Apesar de pertencer a família de inibidores do fator de transcrição NF- $\kappa$ B (fator nuclear kappa B), a I $\kappa$ B- $\zeta$  atua de forma distinta dos membros I $\kappa$ B- $\alpha$  e I $\kappa$ B- $\beta$  desta mesma família. Enquanto que I $\kappa$ B- $\alpha$  e I $\kappa$ B- $\beta$  sequestram a proteína NF- $\kappa$ B no citoplasma, se ligando a subunidade p65 do dímero NF- $\kappa$ B e impedem a sua translocação para o núcleo, a proteína I $\kappa$ B- $\zeta$  interage dentro do núcleo com o NF- $\kappa$ B e se liga preferencialmente a subunidade p50 do homodímero (p50/p50) ou do heterodímero (p50/p65) (TRINH et al.

2008). A associação da I $\kappa$ B- $\zeta$  com o NF- $\kappa$ B reprime a transcrição dos genes alvos de NF- $\kappa$ B (MOTOYAMA et al. 2005). A proteína I $\kappa$ B- $\zeta$  seria uma segunda forma de inibir a ativação dos genes da via do NF- $\kappa$ B constitutivamente expressos em tumores com mau prognóstico (VAN WAES 2007).

Além da proteína I $\kappa$ B- $\zeta$  inibir o fator de transcrição NF- $\kappa$ B, WU et al. publicaram em 2009 um trabalho onde relatam a inibição da proteína STAT3 pela ação da I $\kappa$ B- $\zeta$ . O fator de transcrição STAT3 é peça central da via JAK-STAT que regula proliferação celular e apoptose e sua inibição promove morte celular inibindo a proliferação celular.

Em dois outros trabalhos independentes, foi constatado que a supressão da expressão da proteína I $\kappa$ B- $\zeta$  aumenta a resistência das células à apoptose enquanto que o aumento da expressão de I $\kappa$ B- $\zeta$  promove morte celular (YAMAZAKI et al. 2001; TOTZKE et al. 2006).

Apesar de ser necessário maiores estudos sobre os efeitos do aumento da expressão em pacientes que não sofreram recidiva bioquímica, a proteína I $\kappa$ B- $\zeta$  seria um novo candidato à marcador de bom prognóstico para pacientes com câncer de próstata.

## 6 CONCLUSÕES

- 1- O presente trabalho estabeleceu um protocolo para construção de bibliotecas de cDNA de amostras microdissecadas a laser com reduzida quantidade de material inicial. A combinação entre digestão enzimática e *barcodes* paciente específicos permite análise de múltiplas amostras em uma mesma corrida permitindo investigar a variabilidade biológica.
- 2- A metodologia desenvolvida nesse estudo permite avaliar quantitativamente o transcrito de amostras microdissecadas a laser através do sequenciamento em larga escala na plataforma GS FLX Titanium 454.
- 3- Os dados gerados pelo sequenciamento na plataforma GS FLX Titanium 454 permitiram identificar genes diferencialmente expressos entre as amostras de adenocarcinoma de próstata de pacientes com e sem recidiva bioquímica.
- 4- Os genes *NFKBIZ* e *BRP44* foram identificados como mais expressos nas amostras sem recidiva e esse padrão de expressão foi confirmado por RT-qPCR. A expressão da proteína I $\kappa$ B- $\zeta$  foi associada com menor incidência de vesícula seminal comprometida e menor escore de Gleason, sugerindo que a presença de I $\kappa$ B- $\zeta$  pode ser um fator de prognóstico favorável em adenocarcinoma de próstata.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

[Anonymous]. Cancer risks in BRCA2 mutation carriers. The Breast Cancer Linkage Consortium. **J Natl Cancer Inst** 1999; 91:1310-6.

Abate-Shen C, Shen MM, Gelmann E. Integrating differentiation and cancer: the Nkx3.1 homeobox gene in prostate organogenesis and carcinogenesis. **Differentiation** 2008; 76:717-27.

Aihara M, Wheeler TM, Ohori M, Scardino PT. Heterogeneity of prostate cancer in radical prostatectomy specimens. **Urology** 1994; 43:60-6; discussion 66-7.

Albadine R, Latour M, Toubaji A, et al. TMPRSS2-ERG gene fusion status in minute (minimal) prostatic adenocarcinoma. **Mod Pathol** 2009; 22:1415-22.

Amling C. Biochemical recurrence after localized treatment. **Urol Clin North Am** 2006; 33:147-59.

Asatiani E, Huang WX, Wang A, et al. Deletion, methylation, and expression of the NKX3.1 suppressor gene in primary human prostate cancer. **Cancer Res** 2005; 65:1164-73.

Ashida S, Nakagawa H, Katagiri T, et al. Molecular features of the transition from prostatic intraepithelial neoplasia (PIN) to prostate cancer: genome-wide gene- expression profiles of prostate cancers and PINs. **Cancer Res** 2004; 64:5963-72.

Baggerly K, Deng L, Morris J, Aldaz C. Differential expression in SAGE: accounting for normal between-library variation. **Bioinformatics** 2003, 19:1477-83.

Barry MJ. Screening for prostate cancer-the controversy that refuses to die. **N Engl J Med** 2009; 360:1351-4.

Berger MF, Levin JZ, Vijayendran K, et al. Integrative analysis of the melanoma transcriptome. **Genome Res** 2010; 20:413-27.

Best CJ, Leiva IM, Chuaqui RF, et al. Molecular differentiation of high- and moderate-grade human prostate cancer by cDNA microarray analysis. **Diagn Mol Pathol** 2003; 12:63-70.

Bethel CR, Faith D, Li X, et al. Decreased NKX3.1 protein expression in focal prostatic atrophy, prostatic intraepithelial neoplasia, and adenocarcinoma: association with gleason score and chromosome 8p deletion. **Cancer Res** 2006; 66:10683-90.

Bickers B, Aukim-Hastie C New molecular biomarkers for the prognosis and management of prostate cancer – The Post PSA Era. **Anticancer Res** 2009; 29:3289-98.

Bostwick DG, Shan A, Qian J, et al. Independent origin of multiple foci of prostatic intraepithelial neoplasia: comparison with matched foci of prostate carcinoma. **Cancer** 1998; 83:1995-2002.

Bowen C, Gelmann EP. NKX3.1 activates cellular response to DNA damage. **Cancer Res** 2010; 70:3089-97.

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

Brentani RR, Carraro DM, Verjovski-Almeida S, et al. Gene expression arrays in cancer research: methods and applications. **Crit Rev Oncol Hematol** 2005; 54:95-105.

Burdick MJ, Reddy CA, Ulchaker J, et al. Comparison of biochemical relapse-free survival between primary Gleason score 3 and primary Gleason score 4 for biopsy Gleason score 7 prostate cancer. **Int J Radiat Oncol Biol Phys** 2009; 73:1439-45.

Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. **Cancer Res** 1999; 59:5975-9.

Cardillo MR, Gentile V, Ceccariello A, Giacomelli L, Messinetti S, Di Silverio F. Can p503s, p504s and p510s gene expression in peripheral-blood be useful as a marker of prostatic cancer? **BMC Cancer** 2005; 5:111.

Carraro DM, Brentani HP, Soares FA, et al. From tissue samples to tumor markers. In: Krishanarao A, editor. **From basics to diagnostics**. New Jersey: Humana Press; 2007. p.17-28.

Carraro DM, Ferreira EN, de Campos Molina G, et al. Poli (A)+ transcriptome assessment of ERBB2-induced alterations in breast cell lines. **PLoS One** 2011; 6:e21022.

Carver BS, Tran J, Gopalan A, et al. Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. **Nat Genet** 2009; 41:619-24.

Castro NP, Osório CA, Torres C, et al. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. **Breast Cancer Res** 2008; 10:R87.

Chan TY, Partin AW, Walsh PC, Epstein JI. Prognostic significance of Gleason score 3+4 versus Gleason score 4+3 tumor at radical prostatectomy. **Urology** 2000; 56:823-7.

Chen ML, Xu PZ, Peng XD, et al. The deficiency of Akt1 is sufficient to suppress tumor development in Pten+/- mice. **Genes Dev** 2006; 20:1569-74.

Clark J, Merson S, Jhavar S, et al. Diversity of TMPRSS2-ERG fusion transcripts in the human prostate. **Oncogene** 2007; 26:2667-73.

Clark J, Attard G, Jhavar S, et al. Complex patterns of ETS gene alteration arise during cancer development in the human prostate. **Oncogene** 2008; 27:1993-2003.

Coppée JY. Do DNA microarrays have their future behind them? **Microbes Infect** 2008; 10:1067-71.

Cussenot O, Azzouzi AR, Nicolaiew N, et al. Low-activity V89L variant in SRD5A2 is associated with aggressive prostate cancer risk: an explanation for the adverse effects observed in chemoprevention trials using 5-alpha-reductase inhibitors. **Eur Urol** 2007; 52:1082-7.

D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. **JAMA** 1998; 280:969-74.

DE Kok JB, Verhaegh GW, Roelofs RW, et al. DD3 (PCA3), a very sensitive and specific marker to detect prostate tumors. **Cancer Res** 2002; 62:2695-8.

DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI. Pathological and molecular aspects of prostate cancer. **Lancet** 2003; 361:955-64.



Di Vizio D, Morello M, Sotgia F, Pestell RG, Freeman MR, Lisanti MP. An absence of stromal caveolin-1 is associated with advanced prostate cancer, metastatic disease and epithelial Akt activation. **Cell Cycle** 2009; 8:2420-4.

Ellwood-Yen K, Graeber TG, Wongvipat J, et al. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. **Cancer Cell** 2003; 4:223-38.

Emmert-Buck MR, Vocke CD, Pozzatti RO, et al. Allelic loss on chromosome 8p12-21 in microdissected prostatic intraepithelial neoplasia. **Cancer Res** 1995; 55:2959-62.

Ewing B, Hillier L, Wendl MC, et al. Base-calling of automated sequencer traces using phred Accuracy assessment. **Genome Res** 1998; 8:175-85.

Febbo PG. Genomic approaches to outcome prediction in prostate cancer. **Cancer** 2009; 115:3046-57.

Folgueira MA, Carraro DM, Brentani H, et al. Gene expression profile associated with response to doxorubicin-based therapy in breast cancer. **Clin Cancer Res** 2005; 11:7434-43.

Fradet Y. Biomarkers in prostate cancer diagnosis and prognosis: beyond prostate-specific antigen. **Curr Opin Urol** 2009; 19:243-6.

Gao H, Ouyang X, Banach-Petrosky WA, Gerald WL, Shen MM, Abate-Shen C. Combinatorial activities of Akt and B-Raf/Erk signaling in a mouse model of androgen-independent prostate cancer. **Proc Natl Acad Sci U S A** 2006; 103:14477-82.

Gelmann EP. Searching for the gatekeeper oncogene of prostate cancer. **Crit Rev Oncol Hematol** 2003; 46 Suppl:S11-20.

Gillanders EM, Xu J, Chang BL, et al. Combined genome-wide scan for prostate cancer susceptibility genes. **J Natl Cancer Inst** 2004; 96:1240-7.

Gleason DF. Classification of prostatic carcinomas. **Cancer Chemother Rep** 1966; 50:125-8.

Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. **J Clin Invest** 2004; 113:913-23.

Godiska R, Mead D, Dhodda V, et al. Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. **Nucleic Acids Res** 2010; 38:e88.

Grisanzio C, Signoretti S. p63 in prostate biology and pathology. **J Cell Biochem** 2008; 103:1354-68.

Gurel B, Iwata T, Koh CM, et al. Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. **Mod Pathol** 2008; 21:1156-67.

Häggman MJ, Wojno KJ, Pearsall CP, Macoska JA. Allelic loss of 8p sequences in prostatic intraepithelial neoplasia and carcinoma. **Urology** 1997; 50:643-7.

Halvorsen OJ, Oyan AM, Bø TH, et al. Gene expression profiles in prostate cancer: association with patient subgroups and tumour differentiation. **Int J Oncol** 2005; 26:329-36.

Han M, Partin AW, Pound CR, Epstein JI, Walsh PC. Long-term biochemical disease-free and cancer-specific survival following anatomic radical retropubic prostatectomy. The 15-year Johns Hopkins experience. **Urol Clin North Am** 2001; 28:555-65.

Hayes JD, Pulford DJ. The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. **Crit Rev Biochem Mol Biol** 1995; 30:445-600.

He X, He L, Hannon GJ. The guardian's little helper: microRNAs in the p53 tumor suppressor network. **Cancer Res** 2007; 67:11099-101.

Hsing AW, Chokkalingam AP. Prostate cancer epidemiology. **Front Biosci** 2006; 11:1388-413.

Hsu F, Kent WJ, Clawson H, et al. The UCSC known genes. **Bioinformatics** 2006; 22:1036-46.

Hu Y, Dobi A, Sreenath T, et al. Delineation of TMPRSS2-ERG splice variants in prostate cancer. **Clin Cancer Res** 2008; 14:4719-25.

Hugosson J, Aus G, Lilja H, Loding P, Pihl CG. Results of a randomized, population-based study of biennial screening using serum prostate-specific antigen measurement to detect prostate carcinoma. **Cancer** 2004; 100:1397-405.

Humphrey PA. Diagnosis of adenocarcinoma in prostate needle biopsy tissue. **J Clin Pathol** 2007; 60:35-42.

Ilijin K, Wolf M, Edgren H, et al. TMPRSS2 fusions with oncogenic ETS factors in prostate cancer involve unbalanced genomic rearrangements and are associated with HDAC1 and epigenetic reprogramming. **Cancer Res** 2006; 66:10242-6.

Ilyin SE, Belkowski SM, Plata-Salamán CR. Biomarker discovery and validation: technologies and integrative approaches. **Trends Biotechnol** 2004; 22:411-6.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. **CA Cancer J Clin** 2011;61:69-90.

Jerónimo C, Henrique R, Hoque MO, et al. A quantitative promoter methylation profile of prostate cancer. **Clin Cancer Res** 2004; 10:8472-8.

Jones PA, Baylin SB. The epigenomics of cancer. **Cell** 2007; 128:683-92.

Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. **J Clin Oncol** 1999; 17:1499-507.

Kent WJ. BLAT--the BLAST-like alignment tool. **Genome Res** 2002; 12:656-64.

Khanduja KL, Avti PK, Kumar S, et al. Anti-apoptotic activity of caffeic acid, ellagic acid and ferulic acid in normal human peripheral blood mononuclear cells: a Bcl-2 independent mechanism. **Biochim et Bioph Acta** 2006; 1760:283-9.

Kim MJ, Bhatia-Gaur R, Banach-Petrosky WA, et al. Nkx3.1 mutant mice recapitulate early stages of prostate carcinogenesis. **Cancer Res** 2002a; 62:2999-3004.

Kim MJ, Cardiff RD, Desai N, et al. Cooperativity of Nkx3.1 and Pten loss of function in a mouse model of prostate carcinogenesis. **Proc Natl Acad Sci U S A** 2002b; 99:2884-9.

Kim J, Eltoum IE, Roh M, Wang J, Abdulkadir SA. Interactions between cells with distinct mutations in c-MYC and Pten in prostate cancer. **PLoS Genet** 2009; 5:e1000542.

King JC, Xu J, Wongvipat J, et al. Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. **Nat Genet** 2009; 41:524-6.

Kirby RS, Fitzpatrick JM, Irani J. Prostate cancer diagnosis in the new millennium: strengths and weaknesses of prostate-specific antigen and the discovery and clinical evaluation of prostate cancer gene 3 (PCA3). **BJU Int** 2009; 103:441-5.

Kristiansen G, Pilarsky C, Wissmann C, et al. Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. **J Pathol** 2005; 205:359-76.

Kuefer R, Varambally S, Zhou M, et al. Alpha-methylacyl-CoA racemase: expression levels of this novel cancer biomarker depend on tumor differentiation. **Am J Pathol** 2002; 161:841-8.

Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. **Proc Natl Acad Sci U S A** 2004; 101:811-6.

LaTulippe E, Satagopan J, Smith A, et al. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. **Cancer Res** 2002; 62:4499-506.

Lee WH, Morton RA, Epstein JI, et al. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. **Proc Natl Acad Sci U S A** 1994; 91:11733-7.

Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. **Nat Rev Cancer** 2008; 8:268-78.

Luo J, Duggan DJ, Chen Y, et al. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. **Cancer Res** 2001; 61:4683-8.

Maattanen L, Hakama M, Tammela TL, et al. Specificity of serum prostatespecific antigen determination in the Finnish prostate cancer screening trial. **Br J Cancer** 2007; 96:56-60.

Macintosh CA, Stower M, Reid N, Maitland NJ. Precise microdissection of human prostate cancers reveals genotypic heterogeneity. **Cancer Res** 1998; 58:23-8.

Mackinnon AC, Yan BC, Joseph LJ, et al. Molecular biology underlying the clinical. **Arch Pathol Lab Med** 2009; 133:1033-40.

Magee JA, Araki T, Patil S, et al. Expression profiling reveals hepsin overexpression in prostate cancer. **Cancer Res** 2001; 61:5692-6.

Maher CA, Kumar-Sinha C, Cao X, et al. Transcriptome sequencing to detect gene fusions in cancer. **Nature** 2009; 458:97-101.

Mahon SM. Screening for prostate cancer: informing men about their options. **Clin J Oncol Nurs** 2005; 9:625-7.

Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature** 2005; 437:376-80.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. **Genome Res** 2008; 18:1509-17.

Marks LS, Fradet Y, Deras IL, et al. PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. **Urology** 2007; 69:532-5.

Markowski MC, Bowen C, Gelmann EP. Inflammatory cytokines induce phosphorylation and ubiquitination of prostate suppressor protein NKX3.1. **Cancer Res** 2008; 68:6896-901.

Maschietto M, de Camargo B, Brentani H, et al. Molecular profiling of isolated histological components of wilms tumor implicates a common role for the Wnt signaling pathway in kidney and tumor development. **Oncology** 2008; 75:81-91.

Mehra R, Han B, Tomlins SA, et al. Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. **Mod Pathol** 2007; 20:538-44.

Meyer M, Stenzel U, Myles S, Prüfer K, Hofreiter M. Targeted high-throughput sequencing of tagged nucleic acid samples. **Nucleic Acids Res** 2007; 35:e97.

Miller DC, Hafez KS, Stewart A, Montie JE, Wei JT. Prostate carcinoma presentation, diagnosis, and staging: an update from the National Cancer Data Base. **Cancer** 2003; 98:1169-78.

Mosquera JM, Perner S, Genega EM, et al. Characterization of TMPRSS2-ERG fusion high-grade prostatic intraepithelial neoplasia and potential clinical implications. **Clin Cancer Res** 2008; 14:3380-5.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Methods** 2008; 5:621-8.

Motoyama M, Yamazaki S, Eto-Kimura A, Takeshige K, Muta T. Positive and negative regulation of nuclear factor-kappaB-mediated transcription by IkappaB-zeta, an inducible nuclear protein. **J Biol Chem** 2005; 280:7444-51.

Mulholland DJ, Dedhar S, Wu H, Nelson CC. PTEN and GSK3beta: key regulators of progression to androgen-independent prostate cancer. **Oncogene** 2006; 25:329-37.

Nacu S, Yuan W, Kan Z, et al. Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. **BMC Med Genomics** 2011; 4:11.

Nagalakshmi U, Wang Z, Waern K, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. **Science** 2008; 320:1344-9.

Nygaard S, Okayama T, Okazaki Y, et al. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. **Nat Genet** 2009; 41:553-62.



Ohl F, Jung M, Xu C, et al. Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization? **J Mol Med** 2005; 83:1014-24.

Okino ST, Pookot D, Li LC, et al. Epigenetic inactivation of the dioxinresponsive cytochrome P4501A1 gene in human prostate cancer. **Cancer Res** 2006; 66:7420-8.

Okubo K, Hori N, Matoba R, et al. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. **Nat Genet** 1992; 2:173-9.

Orr B, Riddick AC, Stewart GD, et al. Identification of stromally expressed molecules in the prostate by tag-profiling of cancer-associated fibroblasts, normal fibroblasts and fetal prostate. **Oncogene** 2012; 31:1130-42.

Otero JM, Papadakis MA, Udatha DB, Nielsen J, Panagiotou G. Yeast biological networks *unfold change* the interplay of antioxidants, genome and phenotype, and reveal a novel regulator of the oxidative stress response. **PLoS One** 2010; 5:e13606.

Ouyang X, DeWeese TL, Nelson WG, Abate-Shen C. Loss-of-function of Nkx3.1 promotes increased oxidative damage in prostate carcinogenesis. **Cancer Res** 2005; 65:6773-9.

Paparel P, Cronin AM, Savage C, Scardino PT, Eastham JA. Oncologic outcome and patterns of recurrence after salvage radical prostatectomy. **Eur Urol** 2009; 55:404-10.

Perner S, Demichelis F, Beroukhim R, et al. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. **Cancer Res** 2006; 66:8337-41.

Pryor MB, Schellhammer PF. The pursuit of prostate cancer in patients with a rising prostate-specific antigen and multiple negative transrectal ultrasound-guided prostate biopsies. **Clin Prostate Cancer** 2002; 1:172-6.

Pulukuri SM, Patibandla S, Patel J, Estes N, Rao JS. Epigenetic inactivation of the tissue inhibitor of metalloproteinase-2 (TIMP-2) gene in human prostate tumors. **Oncogene** 2007; 26:5229-37.

Rasiah KK, Stricker PD, Haynes AM, et al. Prognostic significance of Gleason pattern in patients with Gleason score 7 prostate carcinoma. **Cancer** 2003; 98:2560-5.

Reynolds MA, Kastury K, Groskopf J, Schalken JA, Rittenhouse H. Molecular markers for prostate cancer. **Cancer Lett** 2007; 249:5-13.

Rogers CG, Yan G, Zha S, et al. Prostate cancer detection on urinalysis for alpha methylacyl coenzyme a racemase protein. **J Urol** 2004; 172:1501-3.

Saeed AI, Sharov V, White J, et al. TM4: A free open-source system for microarray data management and analysis. **BioTechniques** 2003; 34:374-8.

Salmena L, Carracedo A, Pandolfi PP. Tenets of PTEN tumor suppression. **Cell** 2008; 133:403-14.

Saraiva TF. **Identificação de marcadores moleculares em câncer de mama relacionados à progressão e metástase**. São Paulo; 2008. [Dissertação de Mestrado-Fundação Antônio Prudente].

Saramäki OR, Harjula AE, Martikainen PM, Vessella RL, Tammela TL, Visakorpi T. TMPRSS2:ERG fusion identifies a subgroup of prostate cancers with a favorable prognosis. **Clin Cancer Res** 2008; 14:3395-400.

Sardana G, Dowell B, Diamandis EP. Emerging biomarkers for the diagnosis and prognosis of prostate cancer. **Clin Chem** 2008; 54:1951-60.

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

Shappell SB, Thomas GV, Roberts RL, et al. Prostate pathology of genetically engineered mice: definitions and classification. The consensus report from the Bar Harbor meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee. **Cancer Res** 2004; 64:2270-305.

Shen MM, Abate-Shen C. Pten inactivation and the emergence of androgen-independent prostate cancer. **Cancer Res** 2007; 67:6535-8.

Singal R, Ferdinand L, Reis IM, Schlesselman JJ. Methylation of multiple genes in prostate cancer and the relationship with clinicopathological features of disease. **Oncol Rep** 2004; 12:631-7.

Singh D, Febbo PG, Ross K, et al. Gene expression correlates of clinical prostate cancer behavior. **Cancer Cell** 2002; 1:203-9.

Sreekumar A, Laxman B, Rhodes DR, et al. Humoral immune response to alpha-methylacyl-CoA racemase and prostate cancer. **J Natl Cancer Inst** 2004; 96:834-43.

Stark JR, Perner S, Stampfer MJ, et al. Gleason score and lethal prostate cancer: does  $3 + 4 = 4 + 3$ ?. **J Clin Oncol** 2009; 27:3459-64.

Stephenson AJ, Smith A, Kattan MW, et al. Integration of gene expression profiling and clinical variables to predict prostate carcinoma recurrence after radical prostatectomy. **Cancer** 2005; 104:290-8.

Thomas GV, Horvath S, Smith BL, et al. Antibody-based profiling of the phosphoinositide 3-kinase pathway in clinical prostate cancer. **Clin Cancer Res** 2004; 10:8351-6.

Thompson TC, Tahir SA, Li L, et al. The role of caveolin-1 in prostate cancer: clinical implications. **Prostate Cancer Prostatic Dis** 2010; 13:6-11.

Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. **Science** 2005; 310:644-8.

Totzke G, Essmann F, Pohlmann S, et al. A novel member of the I kappa B family, human I kappa B-zeta, inhibits transactivation of p65 and its DNA binding. **J Biol Chem** 2006, 281:12645-54.

Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nat Biotechnol** 2010; 28:511-5.

Trinh DV, Zhu N, Farhang G, Kim BJ, Huxford T. The nuclear I kappa B protein I kappa B zeta specifically binds NF-kappa B p50 homodimers and forms a ternary complex on kappa B DNA. **J Mol Biol** 2008; 379:122-35.

Tricoli JV, Schoenfeldt M, Conley BA. Detection of prostate cancer and predicting progression: current and future diagnostic markers. **Clin Cancer Res** 2004; 10:3943-53.

Tuch BB, Laborde RR, Xu X, et al. Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. **PLoS One** 2010; 5:e9317.

True L, Coleman I, Hawley S, et al. A molecular correlate to the Gleason grading system for prostate adenocarcinoma. **Proc Natl Acad Sci U S A** 2006; 103:10991-6.

Uzgare AR, Isaacs JT. Enhanced redundancy in Akt and mitogen-activated protein kinase-induced survival of malignant versus normal prostate epithelial cells. **Cancer Res** 2004; 64:6190-9.

Uziel T, Karginov FV, Xie S, et al. The miR-17~92 cluster collaborates with the Sonic Hedgehog pathway in medulloblastoma. **Proc Natl Acad Sci U S A** 2009; 106:2812-7.

Varambally S, Dhanasekaran SM, Zhou M, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. **Nature** 2002; 419:624-9.

Varambally S, Yu J, Laxman B, et al. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. **Cancer Cell** 2005; 8:393-406.

Varuzza L, Pereira CAB. Comparative enumeration gene expression [serial on line] **Nature Precedings** 2008. Available from: <URL:<http://precedings.nature.com/documents/2002/version/1/files/npre20082002-1.pdf>> [2012 jul 12]

Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome Biol** 2002; 3:RESEARCH0034.

Van Waes C. Nuclear Factor- $\kappa$ B in development, prevention, and therapy of cancer. **Clin Cancer Res** 2007; 13:1076-82.

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

Vickers AJ, Ulmert D, Serio AM, et al. The predictive value of prostate cancer biomarkers depends on age and time to diagnosis: towards a biologically based screening strategy. **Int J Cancer** 2007; 121:2212-7.

Vickers AJ, Lilja H. Cutpoints in clinical chemistry: time for fundamental reassessment. **Clin Chem** 2009; 55:15-7.

Vocke CD, Pozzatti RO, Bostwick DG, et al. Analysis of 99 microdissected prostate carcinomas reveals a high frequency of allelic loss on chromosome 8p12-21. **Cancer Res** 1996; 56:2411-6.

Wang GP, Garrigue A, Ciuffi A, et al. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. **Nucleic Acids Res** 2008; 36:e49.

Wang J, Cai Y, Ren C, Ittmann M. Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer. **Cancer Res** 2006; 66:8347-51.

Welsh JB, Sapinoso LM, Su AI, et al. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. **Cancer Res** 2001; 61:5974-8.

Wu Z, Zhang X, Yang J, et al. Nuclear protein I $\kappa$ B-zeta inhibits the activity of STAT3. **Biochem Biophys Res Commun** 2009, 387:348-52.

Wyman SK, Parkin RK, Mitchell PS, et al. Repertoire of microRNAs in epithelial ovarian cancer as determined by next generation sequencing of small RNA cDNA libraries. **PLoS One** 2009; 4:e5311.

Yamazaki S, Muta T, Takeshige K. A novel IkappaB protein, IkappaB-zeta, induced by proinflammatory stimuli, negatively regulates nuclear factor-kappaB in the nuclei. **J Biol Chem** 2001; 276:27657-62.

Yu YP, Landsittel D, Jing L, et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. **J Clin Oncol** 2004; 22:2790-9.

Zha S, Ferdinandusse S, Denis S, et al. Alpha-methylacyl-CoA racemase as an androgen-independent growth modifier in prostate cancer. **Cancer Res** 2003; 63:7365-76.

**Anexo 1 - Lista dos 122 Processo biológicos do Gene Ontology anotados nas amostras de CaP com  $p < 0,01$**

RANK	ID	Term	List Hits	Population Hits	P-value
1	GO:0045449	regulation of transcription	648	950	2,61E-09
2	GO:0055114	oxidation reduction	359	524	5,53E-06
3	GO:0007049	cell cycle	312	438	8,85E-08
4	GO:0015031	protein transport	311	392	1,01E-17
5	GO:0006915	apoptosis	281	398	1,34E-06
6	GO:0019941	modification-dependent protein catabolic process	249	357	2,19E-05
7	GO:0044419	interspecies interaction between organisms	209	256	9,53E-15
8	GO:0008380	RNA splicing	188	223	3,48E-16
9	GO:0006974	response to DNA damage stimulus	179	232	5,51E-09
10	GO:0008285	negative regulation of cell proliferation	169	246	1,32E-03
11	GO:0051301	cell division	160	228	3,88E-04
12	GO:0006629	lipid metabolic process	158	233	4,23E-03
13	GO:0006886	intracellular protein transport	151	179	2,63E-13
14	GO:0005975	carbohydrate metabolic process	145	215	8,03E-03
15	GO:0016568	chromatin modification	136	179	1,75E-06
16	GO:0006397	mRNA processing	128	171	1,30E-05
17	GO:0006916	anti-apoptosis	128	174	5,40E-05
18	GO:0016192	vesicle-mediated transport	128	162	6,94E-08
19	GO:0007067	mitosis	125	177	1,09E-03
20	GO:0007264	small GTPase mediated signal transduction	125	184	9,30E-03
21	GO:0006281	DNA repair	113	151	4,20E-05
22	GO:0006457	protein <i>fold change</i> ing	110	146	3,20E-05
23	GO:0006917	induction of apoptosis	109	153	1,36E-03
24	GO:0006511	ubiquitin-dependent protein catabolic process	105	142	1,75E-04
25	GO:0000398	nuclear mRNA splicing, via spliceosome	102	111	1,02E-14
26	GO:0006464	protein modification process	97	131	2,81E-04
27	GO:0006260	DNA replication	96	126	4,68E-05
28	GO:0006414	translational elongation	90	95	2,19E-15
29	GO:0001666	response to hypoxia	89	127	7,52E-03
30	GO:0006366	transcription from RNA polimerase II promoter	85	115	7,22E-04
31	GO:0007507	heart development	85	111	9,16E-05
32	GO:0042981	regulation of apoptosis	83	111	4,45E-04
33	GO:0001701	in utero embryonic development	82	111	9,15E-04
34	GO:0006364	rRNA processing	78	84	3,61E-12
35	GO:0006928	cellular component movement	77	100	1,41E-04
36	GO:0022900	electron transport chain	76	96	2,81E-05
37	GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	76	93	3,03E-06
38	GO:0006979	response to oxidative stress	73	98	1,16E-03
39	GO:0006461	protein complex assembly	72	101	8,16E-03
40	GO:0006897	endocytosis	70	94	1,47E-03
41	GO:0043065	positive regulation of apoptosis	67	83	2,58E-05



**Cont./ Anexo 1**

42	GO:0045893	positive regulation of transcription, DNA-dependent	67	86	2,04E-04
43	GO:0046777	protein amino acid autophosphorylation	60	75	1,12E-04
44	GO:0014070	response to organic cyclic substance	58	73	2,07E-04
45	GO:0051437	positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle	55	65	9,35E-06
46	GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	53	62	7,25E-06
47	GO:0045892	negative regulation of transcription, DNA-dependent	53	70	2,93E-03
48	GO:0051436	negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle	52	62	2,76E-05
49	GO:0006631	fatty acid metabolic process	51	69	8,04E-03
50	GO:0008033	tRNA processing	49	61	3,97E-04
51	GO:0006350	transcription	48	62	2,05E-03
52	GO:0045454	cell redox homeostasis	48	63	3,67E-03
53	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	47	61	2,71E-03
54	GO:0016044	membrane organization	46	53	1,26E-05
55	GO:0016567	protein ubiquitination	46	59	1,95E-03
56	GO:0051028	mRNA transport	46	55	9,34E-05
57	GO:0015992	proton transport	45	57	1,37E-03
58	GO:0006396	RNA processing	43	54	1,26E-03
59	GO:0007265	Ras protein signal transduction	43	57	7,97E-03
60	GO:0006368	RNA elongation from RNA polymerase II promoter	39	47	4,50E-04
61	GO:0009058	biosynthetic process	39	49	2,16E-03
62	GO:0006986	response to <i>unfold changed</i> protein	38	45	2,59E-04
63	GO:0006892	post-Golgi vesicle-mediated transport	37	40	2,90E-06
64	GO:0006888	ER to Golgi vesicle-mediated transport	35	41	2,96E-04
65	GO:0051246	regulation of protein metabolic process	35	45	7,18E-03
66	GO:0006120	mitochondrial electron transport, NADH to ubiquinone	34	43	5,03E-03
67	GO:0007059	chromosome segregation	32	39	2,15E-03
68	GO:0015986	ATP synthesis coupled proton transport	32	39	2,15E-03
69	GO:0000082	G1/S transition of mitotic cell cycle	31	38	3,02E-03
70	GO:0051726	regulation of cell cycle	31	38	3,02E-03
71	GO:0030521	androgen receptor signaling pathway	30	36	1,84E-03
72	GO:0045786	negative regulation of cell cycle	30	35	7,11E-04
73	GO:0051092	positive regulation of NF-kappaB transcription factor activity	30	38	8,68E-03
74	GO:0006413	translational initiation	26	30	1,17E-03
75	GO:0000902	cell morphogenesis	25	30	4,51E-03
76	GO:0006302	double-strand break repair	24	28	2,53E-03
77	GO:0007569	cell aging	24	25	3,79E-05
78	GO:0030199	collagen fibril organization	23	26	1,26E-03
79	GO:0051289	protein homotetramerization	23	26	1,26E-03
80	GO:0000387	spliceosomal snRNP assembly	22	24	5,21E-04
81	GO:0001843	neural tube closure	22	26	5,39E-03
82	GO:0006099	tricarboxylic acid cycle	22	23	9,98E-05
83	GO:0018105	peptidyl-serine phosphorylation	22	25	1,89E-03
84	GO:0050821	protein stabilization	22	25	1,89E-03
85	GO:0051603	proteolysis involved in cellular protein catabolic process	21	24	2,84E-03

## Cont./ Anexo 1

86	GO:0006446	regulation of translational initiation	20	22	1,26E-03
87	GO:0008652	cellular amino acid biosynthetic process	20	23	4,23E-03
88	GO:0030163	protein catabolic process	20	23	4,23E-03
89	GO:0007040	lysosome organization	19	19	4,82E-05
90	GO:0030433	ER-associated protein catabolic process	19	19	4,82E-05
91	GO:0042254	ribosome biogenesis	19	22	6,28E-03
92	GO:0048538	thymus development	19	22	6,28E-03
93	GO:0006890	retrograde vesicle-mediated transport, Golgi to ER	18	21	9,26E-03
94	GO:0010332	response to gamma radiation	18	21	9,26E-03
95	GO:0000077	DNA damage checkpoint	17	18	1,09E-03
96	GO:0048008	platelet-derived growth factor receptor signaling pathway	17	18	1,09E-03
97	GO:0007030	Golgi organization	16	18	6,98E-03
98	GO:0007163	establishment or maintenance of cell polarity	16	18	6,98E-03
99	GO:0000075	cell cycle checkpoint	14	14	6,60E-04
100	GO:0001938	positive regulation of endothelial cell proliferation	14	15	4,43E-03
101	GO:0006611	protein export from nucleus	14	15	4,43E-03
102	GO:0006749	glutathione metabolic process	14	15	4,43E-03
103	GO:0008333	endosome to lysosome transport	14	14	6,60E-04
104	GO:0006261	DNA-dependent DNA replication	13	14	7,01E-03
105	GO:0006268	DNA unwinding during replication	13	14	7,01E-03
106	GO:0006401	RNA catabolic process	13	14	7,01E-03
107	GO:0006607	NLS-bearing substrate import into nucleus	13	13	1,11E-03
108	GO:0008299	isoprenoid biosynthetic process	13	14	7,01E-03
109	GO:0021987	cerebral cortex development	13	14	7,01E-03
110	GO:0048812	neuron projection morphogenesis	13	13	1,11E-03
111	GO:0001516	prostaglandin biosynthetic process	11	11	3,17E-03
112	GO:0031293	membrane protein intracellular domain proteolysis	11	11	3,17E-03
113	GO:0006878	cellular copper ion homeostasis	10	10	5,35E-03
114	GO:0018279	protein amino acid N-linked glycosylation via asparagine	10	10	5,35E-03
115	GO:0032956	regulation of actin cytoskeleton organization	10	10	5,35E-03
116	GO:0046034	ATP metabolic process	10	10	5,35E-03
117	GO:0051881	regulation of mitochondrial membrane potential	10	10	5,35E-03
118	GO:0042273	ribosomal large subunit biogenesis	9	9	9,03E-03
119	GO:0042274	ribosomal small subunit biogenesis	9	9	9,03E-03
120	GO:0048009	insulin-like growth factor receptor signaling pathway	9	9	9,03E-03
121	GO:0060325	face morphogenesis	9	9	9,03E-03

**Anexo 2 - Lista das 51 vias bioquímicas do KEGG enriquecidas nas amostras de CaP com  $p < 0,01$**

Rank	ID	Term	List Hits	Population Hits	P-value
1	1100	Metabolic pathways	776	1103	4,25E-13
2	5200	Pathways in cancer	235	332	9,29E-05
3	4510	Focal adhesion	150	201	2,70E-05
4	5016	Huntington's disease	142	185	3,24E-06
5	4144	Endocytosis	142	187	9,20E-06
6	5010	Alzheimer's disease	126	171	3,05E-04
7	3040	Spliceosome	114	128	9,52E-13
8	4310	Wnt signaling pathway	110	151	1,46E-03
9	4120	Ubiquitin mediated proteolysis	107	137	1,38E-05
10	190	Oxidative phosphorylation	104	135	5,23E-05
11	4722	Neurotrophin signaling pathway	103	126	3,65E-07
12	4910	Insulin signaling pathway	100	137	2,12E-03
13	5012	Parkinson's disease	99	133	7,38E-04
14	4142	Lysosome	97	116	9,00E-08
15	4530	Tight junction	97	134	3,77E-03
16	3010	Ribosome	82	88	4,43E-12
17	4666	Fc gamma R-mediated phagocytosis	75	97	4,76E-04
18	240	Pyrimidine metabolism	75	98	8,22E-04
19	5215	Prostate cancer	71	89	1,18E-04
20	4350	TGF-beta signaling pathway	67	87	1,15E-03
21	4210	Apoptosis	67	88	1,96E-03
22	4012	ErbB signaling pathway	66	87	2,48E-03
23	5222	Small cell lung cancer	65	86	3,13E-03
24	5210	Colorectal cancer	62	84	9,62E-03
25	5220	Chronic myeloid leukemia	59	75	8,73E-04
26	4520	Adherens junction	59	77	2,78E-03
27	5211	Renal cell carcinoma	58	70	6,58E-05
28	4370	VEGF signaling pathway	58	76	3,54E-03
29	5212	Pancreatic cancer	57	72	7,98E-04
30	5120	Epithelial cell signaling in Helicobacter pylori infection	54	68	9,54E-04
31	4115	p53 signaling pathway	54	69	1,79E-03
32	3018	RNA degradation	51	59	1,71E-05
33	4621	NOD-like receptor signaling pathway	48	62	4,74E-03
34	5110	Vibrio cholerae infection	47	56	1,78E-04
35	4150	mTOR signaling pathway	44	52	1,96E-04
36	4330	Notch signaling pathway	40	47	3,01E-04
37	3050	Proteasome	40	48	7,40E-04
38	480	Glutathione metabolism	40	50	3,41E-03
39	280	Valine, leucine and isoleucine degradation	38	44	2,25E-04
40	970	Aminoacyl-tRNA biosynthesis	37	41	3,17E-05
41	310	Lysine degradation	36	44	2,63E-03
42	5219	Bladder cancer	35	42	1,61E-03
43	520	Amino sugar and nucleotide sugar metabolism	35	44	7,24E-03
44	3420	Nucleotide excision repair	35	44	7,24E-03
45	4130	SNARE interactions in vesicular transport	32	38	1,81E-03

## Cont./ Anexo 2

46	640	Propanoate metabolism	30	33	1,38E-04
47	20	Citrate cycle (TCA cycle)	27	32	3,98E-03
48	450	Selenoamino acid metabolism	24	26	4,00E-04
49	410	beta-Alanine metabolism	19	22	9,85E-03
50	900	Terpenoid backbone biosynthesis	14	15	6,48E-03
51	903	Limonene and pinene degradation	13	14	9,97E-03

**Anexo 3** Lista dos 896 genes diferencialmente expressos identificados pelo método de probabilidades

<b>Symbol</b>	<b>Prob</b>	<b>Symbol</b>	<b>Prob</b>	<b>Symbol</b>	<b>Prob</b>	<b>Symbol</b>	<b>Prob</b>
LOC100293090	12,77	CYorf15B	43,88	LOC653889	64,48	CDC37	69,99
SRSF5	16,48	STEAP2	43,99	RPLP0	64,50	CYHR1	70,00
PODXL	16,60	PDLIM5	44,11	SEPT7	64,50	SLC33A1	70,01
SLC25A4	18,35	RPL7	44,62	AAK1	64,53	TMEM93	70,06
MTRNR2L10	18,74	RAB23	44,64	C17orf89	64,57	ATP5G1	70,07
PNN	19,07	DOCK7	44,72	TRAF4	64,57	ZCCHC6	70,12
LOC100288871	21,60	SCUBE2	44,77	COX4I1	64,59	DDT	70,17
LOC100288998	21,86	PDLIM5	44,88	NCBP1	64,61	PFDN1	70,21
RBM25	22,07	SLPI	45,10	RAB13	64,63	EVL	70,24
SLC25A3	22,26	NUMA1	45,29	ZFYVE21	64,63	VPS35	70,24
LOC653881	22,43	BTF3	45,34	ZNF765	64,64	PSMF1	70,29
SRSF5	23,41	LMX1B	45,38	ATP1B3	64,67	SPECC1L	70,29
ACTG2	23,96	CAPN2	45,44	ZDHHC2	64,71	C8orf30B	70,33
GPR107	24,27	ATP6AP2	45,55	RPLP0	64,72	AMACR	70,39
MTRNR2L3	24,31	LOC401504	45,77	H3F3B	64,80	C7orf50	70,39
C22orf29	24,70	ALDH1A1	45,84	WDR1	64,83	GLTSCR2	70,39
RPL6	25,48	DHCR24	45,99	EFS	64,90	DHX9	70,43
LDHB	25,51	KL	45,99	SPTLC2	64,90	ARF4	70,44
BTBD6	25,55	DHX29	46,07	ATPIF1	64,96	COX8A	70,46
ARF6	26,16	PSMA7	46,21	ISY1	64,97	ENDOD1	70,48
CIAO1	26,21	ENTPD4	46,85	CPSF4	65,02	BCR	70,50
SFRS18	26,30	BTF3	46,88	GARS	65,02	HSPB1	70,51
FBXL12	26,44	B2M	47,12	MARS	65,05	HOXB13	70,54
SLC30A9	26,52	NPY	47,13	STRAP	65,10	COMMD3	70,61
DNAJB12	26,60	MTRNR2L9	47,18	SSR4	65,11	MRPL24	70,72
TAF8	26,67	ATP6V1A	47,68	MRPL24	65,17	CYHR1	70,76
FBXW2	26,75	SFPQ	47,68	PPFIA1	65,18	MGST2	70,78
C11orf92	26,78	TLE1	47,87	LOC100288879	65,19	ULK1	70,85
MLXIP	27,22	RAB4A	47,96	NCRNA00201	65,19	RPL37A	70,88
ACOT9	27,29	SERPINB11	47,99	HSPC159	65,20	PSMB6	70,93
LEMD3	27,77	TMEM59	48,07	FAM84A	65,24	NDUFB5	70,96
ALAD	27,81	UNC80	48,13	ATP5A1	65,31	NISCH	70,96
CCL28	27,85	COL9A2	48,41	B3GALT6	65,32	CTSH	70,97
SKAP2	27,85	DARS	48,65	GEMIN4	65,35	ATP5F1	71,02
LYRM7	27,86	FASTKD2	48,72	PLXNA1	65,36	COMMD4	71,04
EPB41L4A	28,02	CDH1	48,88	RPL23A	65,39	POLD4	71,05
EEF1B2	28,07	LMX1B	49,03	STK11IP	65,40	PXMP4	71,06
NSF	28,09	HERC2P2	49,07	ACOX3	65,41	C19orf10	71,14
OPA3	28,23	CEACAM22P	49,08	AZGP1	65,41	FTH1	71,15
CPNE2	28,25	IFI30	49,10	RPS15A	65,44	NAGA	71,19
ACOT9	28,28	KLK2	49,38	PKP1	65,47	ACAA1	71,31
DNAJC25	28,33	PHKA2	49,88	DAZAP1	65,48	CNPY2	71,32
DCP1A	28,35	PIGH	50,16	PCOTH	65,48	ACLY	71,33
ITPR3	28,37	SRSF7	51,06	ATF4	65,51	TMEM205	71,42
MTRNR2L1	28,40	COX7C	51,30	UBE2Z	65,53	PPCDC	71,46

**Cont./ Anexo 3**

<i>EIF3I</i>	28,42	<i>ORM1</i>	51,42	<i>GRHPR</i>	65,55	<i>CSDE1</i>	71,50
<i>SLC25A3</i>	28,53	<i>RDH16</i>	51,45	<i>SCNN1A</i>	65,55	<i>NUPR1</i>	71,51
<i>C10orf32</i>	28,87	<i>OR51E2</i>	52,21	<i>DDT</i>	65,57	<i>C1R</i>	71,52
<i>ZNF317</i>	28,87	<i>ZRANB2</i>	52,34	<i>SNX17</i>	65,59	<i>RPS14</i>	71,52
<i>LOC100130474</i>	29,06	<i>RGS2</i>	52,52	<i>PDIA4</i>	65,61	<i>RPL21P28</i>	71,58
<i>EEF1B2</i>	29,18	<i>C15orf48</i>	52,98	<i>ADAMTS4</i>	65,62	<i>POMP</i>	71,59
<i>AFF1</i>	29,26	<i>LOC646347</i>	53,21	<i>HEYL</i>	65,68	<i>ABCD3</i>	71,60
<i>TTC39C</i>	29,26	<i>SMS</i>	53,46	<i>KCNS3</i>	65,74	<i>PTRH1</i>	71,61
<i>SYAP1</i>	29,32	<i>KLK2</i>	53,49	<i>STX8</i>	65,77	<i>ITM2C</i>	71,62
<i>RPL6</i>	29,41	<i>RN5-8S1</i>	53,51	<i>PEX6</i>	65,78	<i>HSD17B4</i>	71,64
<i>FOXK1</i>	29,56	<i>IFI6</i>	53,59	<i>TSC22D3</i>	65,81	<i>LOC728606</i>	71,69
<i>GOLGA3</i>	29,64	<i>CAPN2</i>	53,60	<i>TH1L</i>	65,82	<i>UBE2J1</i>	71,69
<i>HSP90B1</i>	29,69	<i>LOC646347</i>	53,71	<i>C10orf10</i>	65,85	<i>RPL26</i>	71,71
<i>CNPY2</i>	29,71	<i>FASTKD2</i>	53,83	<i>DHRS7</i>	65,85	<i>C14orf132</i>	71,73
<i>ANKH</i>	29,72	<i>IFI6</i>	54,70	<i>ANPEP</i>	65,86	<i>CNOT6</i>	71,78
<i>CAT</i>	30,09	<i>PCBP1</i>	54,87	<i>PARK7</i>	65,92	<i>TOMM7</i>	71,82
<i>SLC44A4</i>	30,21	<i>ATP5A1</i>	55,16	<i>SHROOM1</i>	65,94	<i>RNPEP</i>	71,83
<i>CHRM3</i>	30,26	<i>C15orf48</i>	55,20	<i>PTGS2</i>	65,95	<i>GDI2</i>	71,84
<i>LOC100287552</i>	30,32	<i>HINT1</i>	55,22	<i>CLDN4</i>	65,98	<i>FAM134A</i>	71,87
<i>CXorf38</i>	30,34	<i>SPON2</i>	55,24	<i>HSD17B10</i>	65,98	<i>MRPS18A</i>	71,87
<i>LOC100288418</i>	30,44	<i>CD24</i>	55,29	<i>C14orf4</i>	65,99	<i>ARPC2</i>	71,90
<i>LOC729799</i>	30,52	<i>DAPL1</i>	55,37	<i>SUMO3</i>	65,99	<i>RPL24</i>	71,91
<i>LOC100288263</i>	30,63	<i>PSCA</i>	55,47	<i>HINT2</i>	66,02	<i>CDK5RAP2</i>	71,94
<i>ZNF548</i>	30,70	<i>ZNF761</i>	55,57	<i>NDUFA1</i>	66,02	<i>PIH1D1</i>	71,98
<i>LOXL4</i>	30,72	<i>CLUAP1</i>	55,58	<i>RNASET2</i>	66,02	<i>RPS18</i>	72,01
<i>NDUFA6</i>	30,75	<i>RPS6KA6</i>	55,59	<i>C20orf3</i>	66,03	<i>MRPL15</i>	72,02
<i>NOL9</i>	30,76	<i>FAM43A</i>	55,62	<i>PDIA3</i>	66,04	<i>LOC647979</i>	72,04
<i>UBR3</i>	30,85	<i>ID1</i>	55,65	<i>TMEM18</i>	66,07	<i>ST14</i>	72,05
<i>ADCY6</i>	31,14	<i>SNTN</i>	55,66	<i>ATP5G1</i>	66,08	<i>MAGOHB</i>	72,06
<i>LOC100133775</i>	31,15	<i>SSRP1</i>	55,71	<i>MRPL38</i>	66,08	<i>CST3</i>	72,14
<i>CADPS</i>	31,20	<i>SPON2</i>	55,73	<i>DCTPP1</i>	66,11	<i>RPL35A</i>	72,14
<i>SLC22A3</i>	31,41	<i>RPL15</i>	55,98	<i>TRAPPC2P1</i>	66,12	<i>EPS8L2</i>	72,19
<i>RALGAPB</i>	31,45	<i>RPL36A</i>	56,35	<i>PKP1</i>	66,13	<i>TRIM27</i>	72,21
<i>LOC100287932</i>	31,52	<i>IFLTD1</i>	56,36	<i>FAM129A</i>	66,15	<i>MRPL21</i>	72,24
<i>POLR3A</i>	31,54	<i>RPL18</i>	56,42	<i>NSMCE1</i>	66,15	<i>ATP5L2</i>	72,25
<i>SLC39A9</i>	31,76	<i>DUSP1</i>	56,48	<i>ERGIC3</i>	66,18	<i>MYL6</i>	72,27
<i>LOC644397</i>	31,84	<i>CHP</i>	56,51	<i>TIMP3</i>	66,19	<i>BEX4</i>	72,31
<i>SCPEP1</i>	31,89	<i>LOC440910</i>	56,52	<i>PXMP2</i>	66,24	<i>TSPAN1</i>	72,35
<i>LMAN1</i>	31,92	<i>TGM4</i>	56,60	<i>WDR46</i>	66,24	<i>ETFB</i>	72,45
<i>DHX15</i>	32,02	<i>DST</i>	56,76	<i>SPDEF</i>	66,25	<i>CADM1</i>	72,47
<i>CCNG1</i>	32,07	<i>RPS19</i>	56,77	<i>EEF1A1</i>	66,26	<i>DEPDC6</i>	72,50
<i>COIL</i>	32,09	<i>ORM2</i>	56,84	<i>NR1H3</i>	66,27	<i>SPG21</i>	72,51
<i>TOX4</i>	32,16	<i>SFTPA2</i>	56,89	<i>DPT</i>	66,29	<i>LGALS1</i>	72,55
<i>EPB41L5</i>	32,18	<i>ETV4</i>	56,95	<i>RPL8</i>	66,35	<i>COX5B</i>	72,57
<i>CAB39</i>	32,26	<i>ETV4</i>	56,99	<i>ARL6IP5</i>	66,36	<i>HSPE1</i>	72,59
<i>TMCO1</i>	32,38	<i>RPL10</i>	57,15	<i>CNN3</i>	66,37	<i>RPS27L</i>	72,62
<i>ZCCHC4</i>	32,46	<i>IGLL5</i>	57,21	<i>CDK2AP1</i>	66,44	<i>CHMP4A</i>	72,63
<i>UBC</i>	32,66	<i>CTGF</i>	57,22	<i>NME1</i>	66,46	<i>CLDN7</i>	72,64
<i>NEAT1</i>	32,86	<i>ZNF577</i>	57,23	<i>NHP2L1</i>	66,49	<i>STUB1</i>	72,66

**Cont./ Anexo 3**

SLC25A3	32,88	TARP	57,46	TSC22D1	66,54	GADD45B	72,68
SLC44A4	32,97	TM4SF1	57,53	XRCC5	66,57	RPL10A	72,74
PHC3	33,10	UST	57,55	TIMM8B	66,61	TOP1MT	72,75
ANXA1	33,21	PSMC3	57,85	ANKRD37	66,63	COX6A1	72,77
NMT2	33,21	RPS4Y1	57,90	MRPS2	66,65	LOC400590	72,77
CADPS	33,36	NAT10	58,03	GATC	66,66	PDCD5	72,83
IFNAR1	33,49	SRSF9	58,16	SAP130	66,67	TMEM50B	72,86
SEC61G	33,55	EIF4A1	58,19	SLC33A1	66,67	C5orf32	72,94
GJC1	33,57	DAZAP1	58,27	CLDN7	66,72	ELF3	73,04
MTRNR2L8	33,61	PGAM1	58,27	REPS2	66,72	ATOX1	73,10
FAT3	33,76	ZG16B	58,31	STYX	66,72	CCDC56	73,16
INTS6	33,79	IGFBP7	58,33	ATP6V0C	66,74	NUP210	73,17
CTNNAL1	33,97	C2orf72	58,41	HMGCS2	66,74	KRTCAP2	73,19
FIG4	33,98	RPS23	58,59	GCAT	66,75	DYNLL1	73,25
METTL3	34,05	CCBE1	58,60	SIDT2	66,77	LRRC47	73,25
ZZEF1	34,09	TUFM	58,60	PUF60	66,78	AKR7A2	73,37
HSPA9	34,14	SAA1	58,63	TAF7	66,78	RPL32	73,48
YEATS2	34,16	TMED3	58,84	CPE	66,80	TBK1	73,51
LOC100127894	34,18	TEX2	59,02	AFTPH	66,85	COX6C	73,77
MTRNR2L6	34,21	USP7	59,08	AQP3	66,87	RPL26L1	73,80
SPATA20	34,36	EDF1	59,15	RPSA	66,88	KDELRL1	73,83
GALNT7	34,38	SOD2	59,15	ERGIC3	66,89	NR2C2AP	73,86
RFPL1S	34,39	TGM2	59,35	SRSF2	66,89	RBM47	73,91
RAB11FIP1	34,46	POTEM	59,45	ZBTB4	66,98	PYGO2	73,96
ZC3H7A	34,55	GRAMD4	59,50	KIF5B	67,01	ATPIF1	74,04
ZMAT3	34,56	C15orf24	59,63	UBL5	67,01	LOC100293228	74,05
TCP1	34,59	TMEM178	59,66	ACTB	67,04	OSTC	74,05
RPL19	34,63	MRPS18A	59,71	RPS15A	67,10	AHNAK	74,06
EIF3L	34,69	LOC647979	59,73	CHD3	67,11	RAB7A	74,10
VPS39	34,74	DDX41	59,76	MGP	67,14	EMG1	74,13
UBB	34,82	IFI6	59,83	ATP6AP1	67,18	RPL30	74,15
ZC3HAV1	34,96	CPB1	59,87	MRPL20	67,27	RPL36	74,18
TPM2	35,04	MMP7	59,90	RPL36	67,29	SRSF3	74,18
PLA2G7	35,08	PCOTH	59,94	FKBP2	67,30	RNF5	74,29
SYS1	35,12	SLC9A3R1	60,06	ZNF593	67,37	SNHG6	74,32
CDR2	35,19	LOC100287015	60,08	LOC100294411	67,46	SUMO2	74,34
HUWE1	35,22	PSMA6	60,20	SRSF7	67,49	MRPL27	74,38
STEAP2	35,25	SOD2	60,23	ADPRHL2	67,50	RHOT2	74,44
SLC44A4	35,27	LOC400643	60,27	C14orf166	67,52	ABHD12	74,47
SCGN	35,28	FAM103A1	60,31	TRAFD1	67,55	LOC400590	74,56
TXNIP	35,29	CD63	60,33	KRTCAP3	67,57	HAX1	74,66
SOBP	35,64	KCTD10	60,51	RPS8	67,59	NDUFB2	74,67
LIMK2	35,78	MESP1	60,58	PCNT	67,64	C17orf49	74,72
MGEA5	35,83	TTC3	60,61	MCM7	67,65	ENO1	74,78
LOC284551	35,87	PDLIM2	60,75	TRADD	67,65	C6orf108	74,85
LOC388692	35,95	MRPL43	60,79	SLC15A2	67,71	RPL37	74,94
FOLH1	36,01	ZDHHC3	60,86	DSTN	67,78	RPS5	74,94
CKAP5	36,23	EIF1	60,90	BACE2	67,81	LRPAP1	74,99
EEF1B2	36,36	PTPRF	60,99	ANKZF1	67,83	MCM3AP	74,99

**Cont./ Anexo 3**

LOC100128252	36,37	TMEFF2	61,00	C6orf129	67,94	SNRPG	75,16
CCDC115	36,51	EML4	61,06	MAP4	67,95	ESRP2	75,18
USP9X	36,52	IER2	61,10	PSMD2	68,05	ROMO1	75,23
ADPGK	36,53	LARP6	61,16	DDAH2	68,08	TPD52	75,23
PAPLN	36,53	PLA2G12A	61,17	GNS	68,10	CAMTA1	75,45
NVL	36,56	PRIM2	61,21	H2AFJ	68,12	LOC400590	75,45
SF3A3	36,56	SPARC	61,49	COX7A2	68,21	RPL13	75,50
SLC25A6	36,73	PHB2	61,50	ACPP	68,24	RPS12	75,58
ENTPD4	36,76	ZNF829	61,55	YWHAQ	68,25	TUBB2A	75,60
SEPT8	36,79	LTF	61,61	TRAM1	68,32	UQCRQ	75,62
ADSL	36,81	MBOAT2	61,64	MAP4	68,34	DYNLL1	75,94
FOLH1	36,93	RGS11	61,65	HMG5	68,36	ITGAE	76,13
SC5DL	36,95	HEBP2	61,67	PRAC	68,42	UFM1	76,16
TCP1	36,95	C11orf31	61,70	MLL2	68,50	FUNDC2	76,26
EIF2AK1	37,01	SLC15A2	61,70	TMC4	68,53	RPS17	76,26
PPP1R11	37,09	UXT	61,71	AMACR	68,56	PARK7	76,30
PRKAR1A	37,24	C10orf116	61,76	TMEM93	68,56	UBE2E2	76,34
CWH43	37,34	WWC1	61,76	C22orf13	68,57	TFF3	76,39
CEP350	37,38	COX6B1	61,78	CBS	68,58	RPLP1	76,53
TRPM7	37,49	AHCYL1	61,80	FBN1	68,60	RPL29	76,67
SLFN5	37,55	PTDSS1	61,91	PCOLCE	68,63	APRT	76,77
CLMN	37,78	FADD	61,99	FAM3B	68,64	JUN	76,88
KIAA1310	37,78	TUBA1A	62,10	ACADSB	68,66	NFKBIZ	76,90
CADPS	37,85	SCGB1D2	62,16	NDUFV1	68,68	DAD1	76,96
UBE4A	37,87	NARS	62,20	MRPL30	68,75	ZFP36	77,04
C1orf63	37,91	SYNPO	62,20	CTSF	68,80	CCDC28A	77,08
LOC100293563	38,16	GCN1L1	62,25	SF3B1	68,80	TMEM9	77,17
MAP2K1	38,38	LOC100290254	62,28	PSMA2	68,81	REEP5	77,25
FOLH1	38,70	ACTG1	62,34	GPR125	68,82	RPS29	77,38
LOC100288578	38,80	PTPRF	62,34	CLTC	68,84	RPL28	77,52
MDH1B	38,80	PLA2G5	62,45	EDEM1	68,90	ACTR3	77,71
EPM2AIP1	39,04	TXN	62,45	MDH2	68,91	WBSCR22	77,75
LMX1B	39,22	VPS4A	62,46	NCRNA00275	68,91	RPL34	77,77
MKRN1	39,42	NDUFB10	62,49	RPL22	68,91	S100A6	77,77
STEAP2	39,53	RAB3D	62,52	HIST1H4J	68,95	GSS	77,87
CSTF2T	39,73	MAVS	62,54	ATP11B	68,97	C2orf24	77,88
LOC100288109	39,90	SOD1	62,55	POLR2F	68,97	C6orf108	77,89
SAT1	39,90	CRK	62,66	FAM82B	68,98	RPL41	78,07
CAMK2N1	39,99	HLA-B	62,66	PMM1	68,98	KLF4	78,08
PIK3C2A	40,00	SNRNP200	62,68	SFXN3	68,99	RPL38	78,09
SGMS1	40,04	KRT18	62,77	TXNDC17	69,02	RPL13	78,16
ZRANB2	40,10	PTPN18	62,92	TRIB1	69,03	UQCR11	78,20
COPB2	40,28	PTPRM	62,96	FAM158A	69,05	RPL39	78,39
PLA2G7	40,30	MTFP1	63,03	TXNL1	69,05	JUND	78,53
PPIB	40,42	ADAMTS1	63,04	PPP1R2	69,09	RPL8	78,77
RPL5	40,49	FKBP2	63,07	FOS	69,10	RPL32	79,23
SURF4	40,50	FST	63,13	LASS5	69,10	APRT	79,59
HLA-DMB	40,52	CREB3L4	63,23	LOC100134359	69,10	ITM2C	79,63



**Cont./ Anexo 3**

JAG1	40,85	PXDN	63,28	RPL27A	69,12	GNG5	79,67
RPS4X	40,91	RPS14	63,34	POLR2L	69,13	RPL23	79,94
RPL4	40,97	PLA1A	63,38	EFS	69,16	PTMA	79,95
NVL	40,99	UBA52	63,40	FAM3B	69,23	EIF3K	80,05
NBPF15	41,01	MAPKAPK3	63,45	SWAP70	69,25	COX7A2	80,13
COL4A3BP	41,12	SEPT7	63,51	MSMB	69,26	AK2	80,23
RPL11	41,19	PSMB4	63,56	PPP1R7	69,28	C2orf28	80,45
MAOA	41,44	TRIB2	63,56	FKBP2	69,34	PRSS8	80,80
RN18S1	41,44	USP1	63,57	XRRA1	69,34	BRP44	80,88
PDLIM5	41,55	MGC70857	63,65	YIPF1	69,42	CHCHD2	81,24
FAAH	41,57	SAA1	63,65	ADI1	69,44	TMSB10	81,30
ANXA5	41,66	TSPAN8	63,67	PAQR7	69,53	GADD45G	81,41
HEATR6	41,67	YARS	63,67	DEAF1	69,58	BRP44	81,99
TRNT1	41,82	CAST	63,68	KRT19	69,58	RPS27	82,09
PCLO	41,97	RPS3	63,68	RHOC	69,59	RPS20	82,21
SAT1	41,98	PSMA5	63,71	MFSD7	69,60	RPS26	82,63
KLK3	42,20	AEBP1	63,80	RPL31	69,60	ETFB	82,66
GRM6	42,27	PTPRM	63,83	WDR17	69,60	RPL32	82,81
RPL7A	42,35	C11orf1	63,84	HIF1A	69,62	RPS10	82,89
LOC389834	42,36	AES	63,92	MRPS24	69,63	NDUFB4	83,01
LOC100131754	42,52	CNTNAP2	63,95	NAMPT	69,64	FTL	83,20
DDIT4	42,60	CXADRP2	64,09	LOC100291837	69,70	RPL34	83,37
F5	42,72	TAPBP	64,23	C19orf62	69,71	TRPM4	83,38
NBPF11	42,94	ZBTB16	64,24	EGR1	69,71	BRP44	83,41
ALS2	43,02	MRPL34	64,26	GPKOW	69,71	RPL41	83,96
KLK3	43,10	ZNF839	64,26	C11orf75	69,75	FABP5	84,20
UNC80	43,10	CYB5R1	64,27	YIF1A	69,76	TRPM4	85,33
SCARB2	43,12	C4orf3	64,28	NAA50	69,85	UBA52	85,47
NEFH	43,15	RPS16	64,38	XBP1	69,87	RPS15	86,01
UBL3	43,60	ATP6V1C1	64,40	NIT2	69,89	NFKBIZ	87,53
KLK3	43,65	HSBP1L1	64,43	GTF3A	69,90	C12orf57	89,17
EPCAM	43,67	LOC100128098	64,43	ZBTB4	69,90	RPS28	89,36
TM9SF2	43,85	B4GALT7	64,47	GNPAT	69,91	RPL35	90,41