

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA**

Programa de Pós Graduação em Ciências Biológicas (Bioquímica)

YURI JOSÉ DE CAMARGO BARROS MOREIRA

**Identificação de perfis de expressão de
RNAs codificadores e não codificadores
de proteína como preditores de
recorrência de câncer de próstata**

SÃO PAULO

Data do Depósito na SPG:

16/07/2010

YURI JOSÉ DE CAMARGO BARROS MOREIRA

Identificação de perfis de expressão de RNAs codificadores e não codificadores de proteína como preditores de recorrência de câncer de próstata

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

Área de Concentração: Bioquímica

Orientador: Sergio Verjovski-Almeida

SÃO PAULO

2010

Yuri José de Camargo Barros Moreira

Identificação de perfis de expressão de RNAs codificadores e não codificadores de proteína como preditores de recorrência de câncer de próstata

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

Aprovado em: _____

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

DEDICATÓRIA

A meu pai, a quem devo desde pequeno a curiosidade científica.

A minha mãe, exemplo de força e dignidade.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer ao Professor Sergio Verjovski-Almeida, pela orientação no doutorado.

Ao Professor Eduardo, pelas críticas e sugestões durante a realização do projeto.

Aos queridos amigos, Lauren, Ângela, Kátia, Camila, Santiago, Vinícius, Helder. Pelas trocas de ideias, problemas, soluções, pelos financiamentos dos cafés... Pela paciência (às vezes nem tanta...), e interesse por minhas aleatoriedades nem sempre (quase nunca) pertinentes.

A nossa querida secretária (e segunda mãe) Ana Paula, sempre efficientíssima. Nossa vida acadêmica fica muito mais fácil (e mais feliz) com ela.

Aos técnicos Denise, Renato e Jefferson, pelo apoio como técnicos e pelo apoio como amigos.

Aos outros membros do laboratório, Ana Ayupe, Ana Tahira, Felipe, Giulliana, Esther e Carlos, pela amizade.

Aos ex-colegas de laboratório e amigos de sempre: Thiago, Julio, Ana Cláudia.

A todos os meus familiares que sempre me apoiaram e me incentivaram.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro.

A todos os funcionários do Instituto de Química pelos serviços prestados.

RESUMO

Moreira, Y.J.C.B.M **Identificação de perfis de expressão de RNAs codificadores e não codificadores como preditores de recorrência de câncer de próstata**. 2010. 158p. Tese de Doutorado – Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

O câncer de próstata é o quinto tipo mais comum de câncer no mundo e o mais comum em homens. Fatores clínicos e anatomopatológicos atualmente usados na clínica não são capazes de distinguir entre a doença indolente e a agressiva. Existe uma grande necessidade de novos marcadores de prognóstico, a fim de melhorar o gerenciamento clínico de pacientes de câncer de próstata.

Além das anormalidades em genes codificadores de proteínas, alterações em RNAs não codificadores (ncRNAs) contribuem para a patogênese do câncer e, portanto, representam outra fonte potencial de biomarcadores de câncer de próstata. Entretanto, até o momento, poucos estudos de perfis de expressão de ncRNAs foram publicados.

Este projeto teve como principal objetivo identificar perfis de expressão de genes codificadores e não codificadores de proteína correlacionados com recorrência de tumor de próstata, a fim de gerar um perfil prognóstico com potencial uso como biomarcadores e elucidar o possível papel de ncRNAs no desenvolvimento do câncer. Para isso, foram analisados os perfis de expressão de genes codificadores e não codificadores de proteína de um conjunto de 42 amostras de tecido tumoral de câncer de próstata de pacientes de amostras de pacientes submetidos à prostatectomia radical, com longo acompanhamento clínico (cinco anos) e conhecida evolução da doença. Nós utilizamos microarranjos por nós desenhados e fabricados pela Agilent sob encomenda, interrogando aproximadamente 18.709 transcritos não codificadores longos (>500 nt), sem evidência de *splicing*, que mapeiam em regiões intrônicas dentro de 5.660 *loci* genômicos. Os dados de expressão foram extraídos de cada arranjo, normalizados entre todas as 42 amostras de pacientes. Usando uma estratégia de

múltipla amostragem, foi identificado um perfil de expressão de mau prognóstico, contendo 51 transcritos intrônicos não codificadores de proteína. O perfil prognóstico de ncRNAs foi aplicado a um conjunto teste independente de 22 pacientes, classificando corretamente 82% das amostras. Uma análise de Kaplan-Meier dos pacientes do conjunto teste indicou que as curvas de sobrevida dos grupos de alto e baixo risco foram significativamente distintas (*Log-rank test* $p = 0,0009$; *Hazard ratio* = 23,4, 95% CI = 3,62 a 151,2), confirmando assim que este classificador é útil para identificar pacientes com alto risco de recorrência. Além disso, estas descobertas indicam um potencial papel destes RNAs intrônicos não codificadores na progressão do tumor de próstata e apontam para os RNAs intrônicos como potenciais novos marcadores de câncer.

Palavras-chave: câncer de próstata, marcadores moleculares, RNAs não codificadores intrônicos, transcrição antissenso, microarranjos de DNA, recorrência bioquímica.

ABSTRACT

Moreira, Y.J.C.B.M **Identification of protein-coding and non-coding RNA expression profiles as prognostic marker of prostate cancer biochemical recurrence.** 2010. 158p. PhD Thesis – Graduate Program in Biochemistry. Instituto de Química , Universidade de São Paulo, São Paulo.

Prostate cancer is the fifth most common type of cancer in the world, and the most common in men. Clinical and anatomic-pathological factors currently used in clinic are not able to distinguish between the indolent and the aggressive disease. There is a major need of new prognostic markers in order to improve the clinical management of prostate cancer patients.

Apart from abnormalities in protein-coding genes, changes in non-coding RNAs (ncRNAs) contribute to the pathogenesis of cancer and thus represent another potential source of prostate cancer biomarkers. However, few studies of expression profiles of ncRNAs have been published.

This project aimed to identify expression profiles of protein-coding and non-coding genes correlated to prostate cancer biochemical recurrence. For this, we analyzed the expression profile of 42 prostate cancer samples from patients undergoing radical prostatectomy, with long follow-up (five years), and know disease outcome. We used a custom microarray designed by us and printed by Agilent, that probes 18,709 long (>500 nt) ncRNAs mapping to intronic regions within 5,660 genomic *loci*. The expression data were extracted from each array and normalized across all 42 samples. Using a multiple random sampling validation strategy, we identified an expression profile of poor prognosis, comprising 51 ncRNAs. The prognostic profile of ncRNAs was applied to an independent test set of 22 patients, correctly classifying 82% of the samples. A Kaplan-Meier analysis of the test set of patients indicated that the survival curves of high and low risk groups were significantly different (Log-rank test $p = 0.0009$, Hazard ratio = 23.4, 95% CI = 3.62 to

151.2) thus confirming that this classifier is useful for identifying patients at high risk of recurrence. Furthermore, these findings indicate a potential role of these intronic non-coding RNAs in the progression of prostate tumors and points to the intronic ncRNAs as potential new markers of cancer.

Keywords: prostate cancer, molecular markers, intronic non-coding RNAs, antisense transcription, DNA microarray, biochemical recurrence.

Lista de abreviaturas e siglas

AS-PIN - *Antisense Partially Intronic*,

AS-TIN - *Antisense Totally Intronic Transcript*

ASTRO - *American Society for Therapeutic Radiology and Oncology*

BPH - Hiperplasia Prostática Benigna

cDNA - DNA complementar

CI - Intervalo de Confiança

ComBat - *Combining Batches*

cRNA - RNA complementar amplificado

CTP - Citosina Tri Fosfato

Cy3- Cianina 3

Cy5 - Cianina 5

DRE - Exame Digital Retal

EBR - Radioterapia De Radiação Externa

ENCODE - Encyclopedia Of DNA Elements

EST - *Expressed Sequence Tags*

FDR - Taxa De Falsa Descoberta

FE - Feature Extraction

GEO - *Gene Expression Omnibus*

GS - *Gleason Score*

HG-PIN - Neoplasia Prostática Intraepitelia De Alto Grau

INCA - Instituto Nacional de Câncer

INCiTO - Instituto Nacional de Ciência e Tecnologia em Oncogenômica

IPA - Ingenuity Pathways Analysis

LOOCV - *Leave-One-Out-Cross-Validation*

LOWESS - *Local Weighted Scatterplot Smoothing*

ncRNAs - RNAs Não Codificadores

NSCLC - Carcinoma Pulmonar De Não Pequenas Células

nt - Nucleotídeos

ORESTES - Orf ESTs

PC - Câncer de Próstata

PCGUP - *Prostate Cancer Guidelines Update Panel*

PIN - transcritos não codificadores parcialmente intrônicos

piRNA Piwi-interacting RNA

PSA - antígeno específico da próstata

RIN - *RNA Integrity Number*

RMS - Rabdomiossarcoma

B-CLL - Leucemia Linfocítica Crônica De Células B

SAM - Análise De Significância Para Microarranjos

TIN - Transcritos Não Codificadores Totalmente Intrônicos

TMA - *Tissue Microarray Analysis* -

ULN - *Upper Limit Of Normal*

VACURG - *Veterans Administration Cooperative Urological Research Group*

ÍNDICE

1. Introdução	18
1.1 Aspectos gerais do câncer de próstata (PC)	18
1.1.1 Anatomia da próstata	18
1.1.2 Estatísticas geográficas do câncer de próstata	21
1.1.3 Fatores de risco	22
1.2 Sintomas	23
1.3 Diagnóstico	24
1.2.1 Toque retal	24
1.2.2 PSA	24
1.2.3 Gleason <i>Score</i>	26
1.4 Tratamento	30
1.5 Recorrência Bioquímica	32
1.5.1 Definição	32
1.6 Definições de carcinomas clinicamente indolentes	35
1.7 Marcadores Moleculares De Câncer de Próstata	36
1.7.1 Marcadores epigenéticos	36
1.7.2 Marcadores de Prognóstico Após a Cirurgia	37
1.8 Uso de Microarranjos para Prognóstico de Câncer de Próstata	39
1.8.1 Classificação de subgrupos clinicamente relevantes	42
1.8.2 Genes como biomarcadores de câncer de próstata	42
1.8.3 Assinaturas de expressão gênica associadas ao desfecho de câncer de próstata	43
1.9 RNAs Não Codificadores (ncRNAs)	44
1.9.1 Transcrição Intrônica	47
2 Objetivo Geral	51
2.1 Objetivos Específicos	51
3 Material e Métodos	53
3.1 Amostras de tecido	53
3.1.1 Critérios de inclusão	53
3.2 Extração do RNA das amostras tumorais	53
3.3 Purificação do RNA e tratamento com DNase	54
3.4 Avaliação do rendimento e da integridade do RNA obtido	55
3.5 Amplificação do RNA contendo cauda poli-A	55
3.6 Desenho do oligoarranjo	58
3.7 Hibridização	58
3.8 Processamento de Imagem	58

3.9	Processamento de Dados	59
3.10	Análise estatística:.....	60
3.11	Classificação de um conjunto independente usando o perfil de mau prognóstico.	62
3.12	Processamento das amostras do conjunto teste.....	62
3.12.1	Normalização por quantil.....	63
3.12.2	Ajuste de lotes (ComBat).....	63
3.12.3	Normalização intergênica (<i>Z-score</i>).....	63
3.13	Classificação das amostras do conjunto teste.....	64
3.14	Análise de sobrevida	64
3.15	Anotação funcional dos transcritos identificados	64
3.15.1	Análises de redes gênicas.....	65
4	Resultados.....	66
4.1	Amostras de tecido tumoral prostático e obtenção de RNA para experimentos de hibridização	66
4.1.1	Dados clínico-patológicos dos pacientes utilizados no estudo.....	68
4.2	Amplificação.....	73
4.3	Desenho do oligoarranjo Íntron-Éxon	77
4.4	Hibridização e obtenção das imagens.....	78
4.5	Análise dos dados de expressão.....	80
4.5.1	Geração do Grupo de Treinamento e do Grupo Teste.....	80
4.5.2	Identificação dos transcritos expressos.....	80
4.5.3	Análise de expressão	84
4.5.4	Normalização dos dados de expressão.....	85
4.5.5	Análise de Dados – Agrupamentos Hierárquicos	86
4.5.6	Genes codificadores de proteína diferencialmente expressos em recorrência de câncer de próstata.....	91
4.5.7	Análise de funções gênicas.....	97
4.6	Identificação de um perfil prognóstico de recorrência baseado na expressão de ncRNAs.....	102
4.7	Teste independente do perfil de expressão de ncRNA como um marcador de prognóstico de recorrência em PC.....	110
4.8	Análise de sobrevida das amostras do conjunto teste.....	111
4.9	Comparação entre os ncRNAs do perfil de genes codificadores de proteína diferencialmente expressos.....	113
5	Discussão.....	120
6	Conclusão.....	132
7	Referências.....	134

Lista de Anexos	152
Súmula curricular do autor	154

SUMÁRIO DE FIGURAS

Figura 1: Anatomia da próstata normal (modificado de Holland 1997).....	20
Figura 2: Regiões da próstata (Holland 1997).	20
Figura 3: Corte histológico da próstata normal (McNeal 1988).	20
Figura 4: Incidência e mortalidade de câncer no mundo em 2002 (Parkin, Bray et al. 2005).22	
Figura 5: Graus de Gleason: Desenho padrão (Gleason 1966).....	27
Figura 6: Método de amplificação de cRNA e marcação com fluoróforos (Agilent Technologies).....	56
Figura 7: Perfil analítico típico de uma amostra de RNA total tratado com DNase I obtido através de eletroforese no Agilent 2100 Bioanalyzer (Agilent Technologies).....	66
Figura 8: Desenho do microarranjo de oligo-nucleotídeos íntron-éxon..	77
Figura 9: Amostra de RNA de tumor de próstata marcada com Cy3 (verde) e Cy5 (vermelho) e hibridizada com a lâmina 37k intron-éxon da Agilent.....	79
Figura 10: Expressão intrônica e exônica em PC.	82
Figura 11: Diagrama de Venn com a sobreposição da transcrição intrônica complementar aos genes exônicos expressos, do total de 6363 genes.....	83
Figura 12: Diagrama de Venn mostrando a sobreposição da expressão, para os 6566 locus gênicos analisados, de regiões exônicas e de regiões totalmente intrônicas senso (S-TIN) e anti-senso (S-TIN).	84
Figura 13: Normalização por quantil.....	86
Figura 14: Agrupamento hierárquico dos ncRNAs intrônicos (A) e dos genes codificadores de proteína (B) para as amostras do conjunto de treinamento. Dados normalizados por quantil dos transcritos válidos, antes do processamento com o <i>script</i> ComBat.R.....	88
Figura 15: Agrupamento hierárquico dos ncRNAs intrônicos (A) e dos genes codificadores de proteína (B) para as amostras do conjunto de treinamento. Dados normalizados por quantil dos transcritos válidos, após do processamento com o <i>script</i> ComBat.R.....	90
Figura 16: Gráfico dos resultados da análise SAM – Significance Analysis of Microarray ...	92
Figura 17: Esquematização da estratégia de múltipla amostragem.	103
Figura 18: Taxa Classificação errada do desfecho dos pacientes nos conjunto de validação em função dos diferentes tamanhos dos conjuntos de treinamento correspondentes.	105
Figura 19 A-B: Perfil de expressão de ncRNAs de mau prognóstico.....	109
Figura 20: Análise de sobrevida para os pacientes no conjunto teste. Curvas de Kaplan-Meier para os grupos de alto e baixo risco de recorrência, definidos pela correlação de cada amostra com o perfil de mau prognóstico de ncRNAs do conjunto de treinamento.	111
Figura 21: A: Curva de Kaplan-Meier do conjunto teste de pacientes de acordo com os níveis pré-operatórios de PSA (alto ≥ 10 ng/mL; baixo < 10 ng/mL). B: Curva de Kaplan-Meier de acordo com o <i>Gleason Score</i> ($GS \geq 7$ ou $GS < 7$).	112
Figura 22: Perfil de expressão de genes codificadores de proteína diferencialmente expressos segundo a estratégia de SAM.....	114
Figura 23: Análise de sobrevida para os pacientes no conjunto teste. Curvas de Kaplan-Meier para os grupos de alto e baixo risco de recorrência, definidos pela correlação de cada amostra com o perfil de mau prognóstico baseado na expressão dos genes codificadores de proteína	

diferencialmente expressos entre pacientes com recorrência e sem recorrência do conjunto de treinamento identificados pela estratégia de SAM.	115
Figura 24: Perfil de expressão de genes codificadores de proteína pela estratégia reamostragem múltipla aleatória.....	116
Figura 25: Análise de sobrevida para os pacientes no conjunto teste. Curvas de Kaplan-Meier para os grupos de alto e baixo risco de recorrência, definidos pela correlação de cada amostra com o perfil de mau prognóstico baseado na expressão dos genes codificadores de proteína diferencialmente expressos entre pacientes com recorrência e sem recorrência do conjunto de treinamento identificados pela estratégia de reamostragem múltipla aleatória.	119
Figura 26: Rede gênica mais enriquecida com genes diferencialmente expressos entre amostras de pacientes com recorrência e sem recorrência.....	127
Figura 27: Segunda rede gênica mais enriquecida com genes diferencialmente expressos entre amostras de pacientes com recorrência e sem recorrência.	128
Figura 28: Via de sinalização do câncer de próstata. (<i>Ingenuity Pathway Analysis</i>).	129

SUMÁRIO DE TABELAS

Tabela 1: Definição de carcinoma clinicamente indolente	36
Tabela 2: Amostras recebidas do Banco de Tumores do Hospital Sírio-Libanês.....	67
Tabela 3: Dados clínicos e anatomo-patológicos para cada um dos 42 pacientes usados no conjunto de treinamento.....	69
Tabela 4: Dados clínicos e anatomo-patológicos para cada um dos 22 pacientes usados no conjunto teste.	71
Tabela 5: Comparação da proporção das características clinico-patológicas dos pacientes usados no estudo com as características observadas na população americana (%).	73
Tabela 6: Quantidade de cRNA amplificado e incorporação de fluoróforos para as primeiras 31 amostras processadas.	75
Tabela 7: Correlação entre a qualidade do RNA inicial (razões 28S/18s e RIN) e os resultados obtidos após a amplificação (cRNA amplificado e incorporação de fluoróforo), indicando uma correlação média ente a qualidade de RNA inicial e o total de cRNA obtido após a amplificação.	76
Tabela 8: Genes codificadores de proteína diferencialmente expressos entre amostras de tumor de próstata de pacientes com recorrência	93
Tabela 9: Categorias dos genes codificadores de proteína diferencialmente expressos entre pacientes com recorrência e sem recorrência.....	98
Tabela 10: ncRNAs que compõem o perfil de expressão de mau prognóstico de recorrência bioquímica de câncer de próstata dentro de 5 anos após a prostatectomia radical.	106
Tabela 11: Genes codificadores de proteína identificados pela estratégia de múltipla amostragem.	117
Tabela 12: Sobreposição de genes diferencialmente expressos entre pacientes com recorrência e sem recorrência nos 5 trabalhos da literatura e com o presente trabalho.	124
Tabela 13: Redes gênicas estatisticamente enriquecidas. ↑ = Aumentado em pacientes com recorrência. ↓ = Diminuído em pacientes com recorrência.	125

1. Introdução

1.1 Aspectos gerais do câncer de próstata (PC)

O câncer é definido por um crescimento anormal de células causado por múltiplas alterações na expressão de seus genes, levando a um balanço desregulado da proliferação e morte celular, finalmente evoluindo em uma população de células que podem invadir tecidos e metastatizar em sítios distantes (Ruddon, 2007). Este crescimento anormal é resultado da mutação em proto-oncogenes e genes supressores tumorais, que regulam a proliferação, diferenciação e sobrevivência das células (Lieberman *et al.*, 2009).

O câncer de próstata é uma doença complexa, envolvendo fatores genéticos, epigenéticos e hormonais (Hofmann *et al.*, 2003). Apesar dos recentes progressos feitos na biologia molecular, a patogênese do câncer de próstata permanece pouco entendida (Hofmann *et al.*, 2003; Cardiff *et al.*, 2010). Isto é reflexo de vários fatores, que incluem a composição complexa da glândula prostática e a natureza heterogênea e multifocal do câncer de próstata (Ware, 1994; Andreoii *et al.*, 2010).

1.1.1 Anatomia da próstata

A próstata está situada na pélvis, rodeada pelo reto posteriormente e pela bexiga superiormente (Figura 1). Existem quatro principais zonas dentro da próstata normal: a zona periférica, com 70% de tecido glandular; a zona central, com 20% de tecido glandular, zona de transição, com 75% de tecido glandular, e o estroma fibromuscular anterior (Figura 2).

A zona periférica estende postero-lateralmente ao redor da glândula a partir do ápice até a base e representa o local mais comum para o desenvolvimento de carcinomas na próstata. A zona central rodeia o aparato ductal ejaculatório e perfaz a maioria da base da próstata. A zona de transição se constitui de dois pequenos lóbulos que confinam a uretra prostática e representam a região onde a Hiperplasia Prostática Benigna (BPH)

primariamente origina-se. Já o carcinoma pode se apresentar em todas as regiões da próstata. No entanto, comparada aos cânceres da zona periférica, carcinomas que se originam na zona de transição tem sido sugeridos como de menor potencial de malignidade; contudo, outros estudos sugerem que não existe diferença no desfecho quando normalizado por graduação e estágio (Greene *et al.*, 1991; Reissigl *et al.*, 1997).

Não existe nenhuma cápsula anatômica verdadeira que rodeia a próstata, similar ao que é observado no rim. O tecido muscular do estroma prostático gradativamente se estende em tecido fibroso, que por fim termina em tecido conectivo frouxo e tecido adiposo (McNeal, 1988a).

A próstata é composta de glândulas ramificadas, com dutos que são alinhados com as células epiteliais secretoras e as células basais (Figura 3) (McNeal, 1988b). Células neuroendócrinas dispersas estão também presentes e especula-se que forneçam uma função parácrina da glândula (di Sant'Agnesse, 1998). As células epiteliais secretoras representam o tipo celular mais comum na glândula, são andrógeno-dependentes para crescer, e produzem o antígeno específico da próstata (*Prostate Specific Antigen* - PSA) e fosfatase ácida prostática (Isaacs, 1983). A camada basal não é dependente de andrógeno para o crescimento e especula-se que contenha populações de células-tronco para as células epiteliais prostáticas. Ao redor da glândula está o estroma, que inclui fibroblastos, tecido muscular liso, nervos e linfócitos. As interações estroma-epitélio permanecem mal compreendidas, mas alguns estudos sugerem que o estroma produza múltiplos fatores de crescimento importantes para o crescimento e desenvolvimento da próstata normal, bem como do câncer de próstata (Ware, 1993; Chung, 1995; Scher *et al.*, 1995).

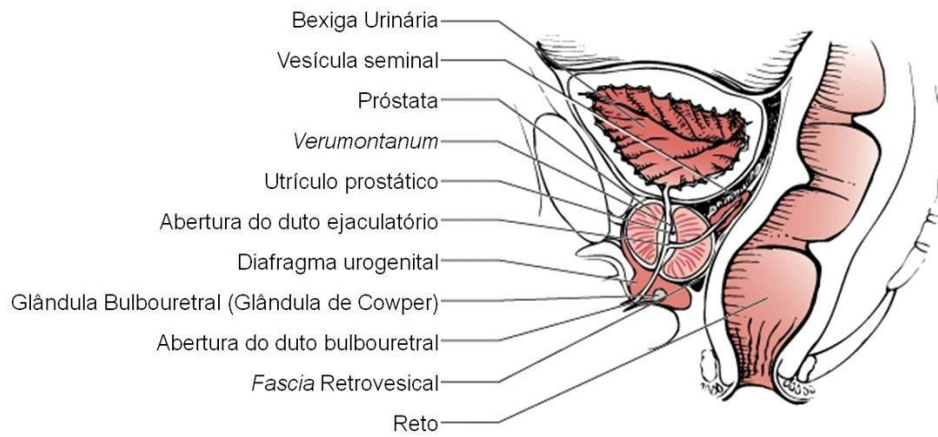


Figura 1: Anatomia da próstata normal (modificado de Holland 1997).

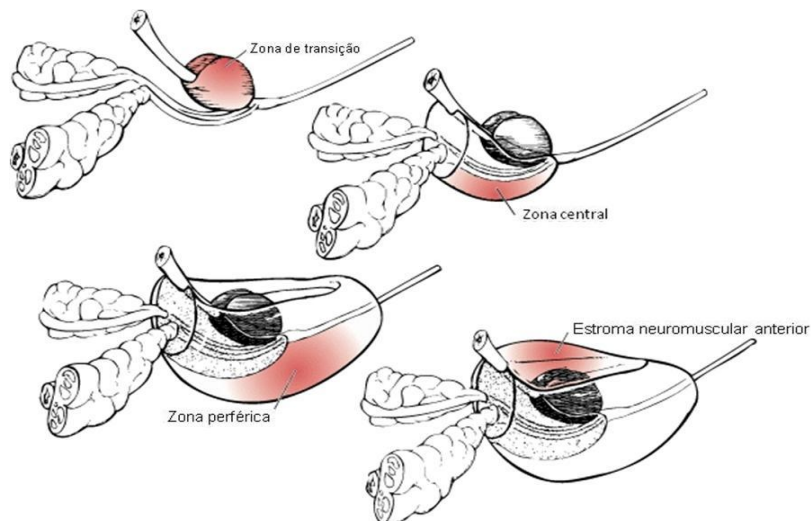


Figura 2: Regiões da próstata (modificado de Holland, 1997).

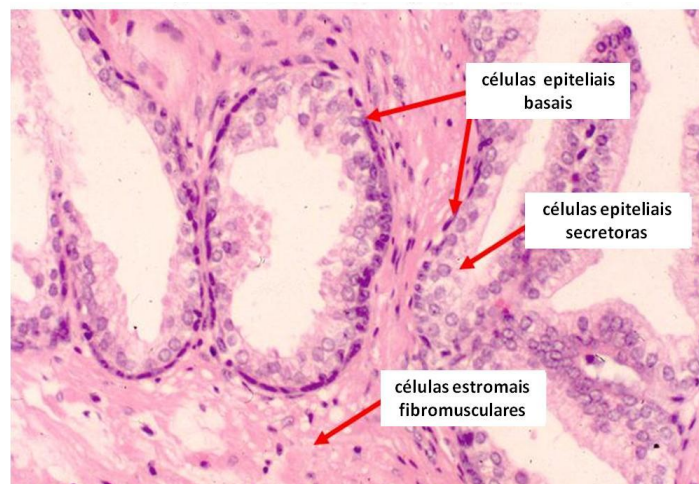


Figura 3: Corte histológico da próstata normal (modificado de McNeal, 1988a).

Por ser uma glândula, a próstata é afetada pelos hormônios sexuais masculinos. Esses hormônios estimulam a atividade da próstata e a substituição das células da próstata, uma vez que essas sejam destruídas. É também dentro da próstata que ocorre a transformação da testosterona em diidrotestosterona, que controla o crescimento da glândula. A função da próstata é produzir um líquido, o fluido seminal que é eliminado durante o ato sexual juntamente com os espermatozoides, os quais têm origem nos testículos. Este líquido produzido pela próstata é importante para a vitalidade dos espermatozoides na fecundação. O fluido das glândulas é produzido no epitélio (camada de células epiteliais).

A atividade secretora e proliferativa da próstata é totalmente dependente de andrógeno. Recentes estudos relatam que fatores de crescimento presentes na próstata são particularmente importantes, em relação tanto à proliferação estromal quanto glandular. Um desequilíbrio hormonal no sistema de produção do fator de crescimento, de dependência de andrógeno e estrogênio e mecanismos gerais de envelhecimento celular devem ser levados em consideração em alterações patológicas da próstata, como Hiperplasia Prostática Benigna (BPH) e Câncer de Próstata (PC) (Aumuller, 1989).

1.1.2 Estatísticas geográficas do câncer de próstata

O PC é o quinto tipo mais comum de câncer no mundo e o segundo mais comum em homens, sendo 11,7% dos novos casos: 19% em países desenvolvidos e 5,3% em países em desenvolvimento (Figura 4) (Parkin *et al.*, 2005). Os dados mundiais mais recentes sobre incidência e morte por câncer são de 2002, quando houve 679.000 novos casos de PC no mundo. A incidência e mortalidade variam em diferentes partes do globo, sendo maior nos países ocidentais (Hsing *et al.*, 2000). Enquanto o PC pode se mostrar fatal, 93% dos homens diagnosticados com a doença irão sobreviver por pelo menos cinco anos, e 72,1% irão sobreviver por pelo menos dez anos (Jemal *et al.*, 2009).

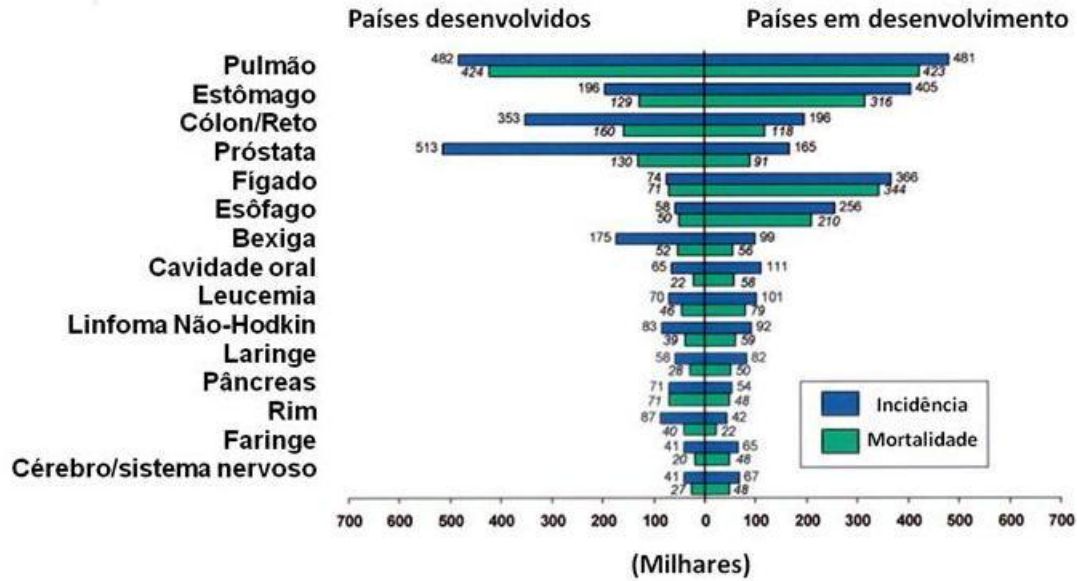


Figura 4: Incidência e mortalidade de câncer no mundo em 2002 (Parkin *et al.*, 2005).

As taxas de morte devido a PC têm diminuído em muitos países desenvolvidos, incluindo Estados Unidos, Reino Unido e Canadá. Em contraste, as taxas de morte estão subindo em alguns países asiáticos, como Japão e Cingapura. Enquanto a diminuição das taxas de morte devido a PC nos países ocidentais tem sido atribuída em parte à melhora no tratamento e na detecção precoce, o aumento nos países asiáticos pode ser reflexo da ocidentalização, incluído o aumento do consumo de gordura animal. No Brasil, segundo o Instituto Nacional de Câncer (INCA), o PC foi o tipo de câncer mais prevalente em homens, com um número de casos novos de PC estimado para o ano de 2010 de 52.350 (INCA, 2009).

1.1.3 Fatores de risco

Mais de 75% dos casos de PC diagnosticados, no mundo, ocorrem a partir dos 65 anos (Parkin *et al.*, 2005). Enquanto a probabilidade de desenvolver PC até os 39 anos de idade é de apenas 1 em 10.553, para homens com mais de 70 anos é de 1 para 7. A incidência é também maior em afro-americanos que em caucasianos (Perrone *et al.*, 2000), sendo que outros estudos indicam que predisposições familiares podem ser responsáveis por 5% a 10%

dos casos de PC (Hayes *et al.*, 1995). Um grande estudo com gêmeos monozigóticos realizado na Escandinávia demonstrou que fatores hereditários são responsáveis por aproximadamente 42% do risco de PC (Lichtenstein *et al.*, 2000). O peso ao nascer também foi associado ao risco de PC em gêmeos dizigóticos (Cnattingius *et al.*, 2009).

Um dos fatores mais relacionados tanto à prevenção do câncer de próstata quanto ao aumento da agressividade do câncer é a dieta. Alimentos de origem vegetal, ricos em compostos como micronutrientes (vitaminas C e E) e fito-químicos (ex. carotenoides, flavonoides e isotiocianatos) tem sido mostrados como possuindo certas propriedades anticarcinogênicas (Wattenberg *et al.*, 1985; Steinmetz *et al.*, 1991; Sies, 1997; Kirsh *et al.*, 2007). Dietas ricas em carne vermelha e gorduras, por outro lado, estão associadas ao aumento do risco de PC (Felton *et al.*, 1986; West *et al.*, 1991; Shirai *et al.*, 1997; Bairati *et al.*, 1998). Entretanto, os estudos clínicos relacionando ingestão de carne vermelha e PC são inconclusivos (Norrish *et al.*, 1999; Cross *et al.*, 2005; Tang *et al.*, 2007). A obesidade também tem sido correlacionada à incidência e agressividade do PC (Demark-Wahnefried *et al.*, 2007). Um alto índice de massa corporal também é associado com maior agressividade e um pior desfecho da doença (Giovannucci *et al.*, 2007).

1.2 Sintomas

Apesar das altas taxas de morte, o PC pode se manter sem sintomas por anos. Com o avanço da doença, pode-se apresentar fluxo urinário fraco ou interrompido, incapacidade de urinar ou dificuldade de iniciar ou parar o fluxo urinário, necessidade de urinar frequentemente, especialmente à noite, sangue na urina, dor ou queimação ao urinar. Dor contínua na região inferior das costas, pélvis ou coxa pode ser uma indicação da metástase da doença. Muitos desses sintomas, no entanto, são similares aos causados por condições benignas (Mansson *et al.*, 1999).

1.3 Diagnóstico

1.2.1 Toque retal

O toque retal é o meio mais simples para o diagnóstico do tumor de próstata e o principal método de diagnóstico em homens acima de 40 anos. Esse método é altamente específico, porém, pouco sensível e cerca de 2/3 dos tumores detectados deste modo se encontram em estágios avançados (Ohori *et al.*, 1994).

1.2.2 PSA

O antígeno específico da próstata (*Prostate Specific Antigen – PSA*) é uma glicoproteína produzida quase exclusivamente pelo componente epitelial da glândula prostática (Barry, 2001). Membro da família gênica das calicreínas, o PSA é uma serina protease de 33 kilodaltons secretada pelo epitélio da próstata e pela superfície das glândulas periuretrais. É funcionalmente significante no seu papel de liquefação do coágulo seminal, para permitir a liberação dos espermatozoides (Lilja, 1985).

Homens com doenças na próstata, incluindo adenocarcinoma, podem ter altos níveis de PSA sorológico por causa do aumento da produção de PSA, bem como de distorções na glândula, que permitem ao PSA um maior acesso à circulação (Barry, 2001). As medidas do PSA se tornaram o evento mais comum levando ao diagnóstico de carcinoma de próstata e podem ser o teste clínico de câncer mais utilizado (Hernandez *et al.*, 2004).

As primeiras investigações de antígeno específico na próstata humana foram conduzidas por Ablin *et al.* em 1970 (Ablin *et al.*, 1970). Outros investigadores independentemente descobriram antígenos prostáticos no plasma seminal (Hara *et al.*, 1971; Li *et al.*, 1973). Subsequentemente, Sensagaugh isolou e caracterizou PSA a partir do plasma seminal humano enquanto procurava por um potencial marcador que pudesse ser usado na

investigação de crimes de estupro (Sensabaugh, 1978). Em 1979, Wang et al. isolaram e purificaram o PSA, caracterizando-o como um antígeno específico de tecido de próstata (Wang *et al.*, 1979).

O amplo uso de PSA na clínica não começou até meados da década de 1980. As primeiras observações tiveram aplicações importantes na clínica, e incluíam: 1) uma diminuição nos níveis de PSA depois da terapia hormonal parecia ser correlacionada com a resposta ao tratamento (Stamey *et al.*, 1987; Ferro *et al.*, 1989; Hudson *et al.*, 1989); 2) um aumento no nível de PSA parecia preceder e anunciar a recorrência da doença (Stamey *et al.*, 1987; Oesterling *et al.*, 1988; Lange *et al.*, 1989); e 3) após a prostatectomia radical, PSA deve ser indetectável; caso contrário, a recorrência da doença é constatada (Stamey *et al.*, 1987; Oesterling *et al.*, 1988; Lange *et al.*, 1989).

Vários destes primeiros investigadores rejeitaram a possibilidade de usar o PSA para triagem por causa da sobreposição substancial nos valores de PSA entre pacientes com e sem carcinoma e da baixa especificidade resultante (Guinan *et al.*, 1987; Stamey *et al.*, 1987; Barak *et al.*, 1989). Reconhecendo a natureza crítica da especificidade em uma ferramenta de triagem para carcinoma de próstata, Guinan et al. estabeleceram um limite superior do normal (*upper limit of normal* - ULN) para o PSA de 24 ng/ml num estudo caso-controle (Guinan *et al.*, 1987). Este valor foi selecionado por que era a média mais dois desvios padrão de um grupo de homens controles que tinham: 1) um diagnóstico histológico de BPH ou 2) não apresentavam nenhum sintoma geniturinário. É importante destacar que nenhum dos pacientes havia passado por uma biópsia da próstata. É interessante notar que, entre os pacientes assintomáticos neste estudo, o ULN foi de 3,59 ng/mL. Em um estudo semelhante, o nível sorológico de 10 ng/mL foi estabelecido como ULN, apesar da descoberta de que 90% dos sujeitos aparentemente saudáveis apresentavam níveis de PSA < 2 ng/mL (Barak *et*

al., 1989). Atualmente, o nível de PSA < 2,5 ng/mL é considerado o máximo normal (Smith *et al.*, 2006).

1.2.3 Gleason Score

O sistema de Gleason é o método dominante de graduação do tumor de próstata ao redor do mundo na pesquisa e na prática diária. Esta classificação foi desenvolvida por Dr. Donald F Gleason e membros da *Veterans Administration Cooperative Urological Research Group* (VACURG) (Gleason, 1966). Uma das maiores vantagens do sistema de graduação de Gleason é que ele foi testado em um grande conjunto populacional, com grande acompanhamento clínico que incluía o uso da sobrevida como ponto final (Humphrey, 2004).

O sistema de graduação de Gleason é baseado inteiramente no padrão histológico, no arranjo das células do carcinoma e na extensão da diferenciação glandular em secções de tecido prostático coradas com hematoxilina-eosina (Bostwick, 1990). Nove padrões de crescimento foram consolidados em cinco graus (Figura 5).

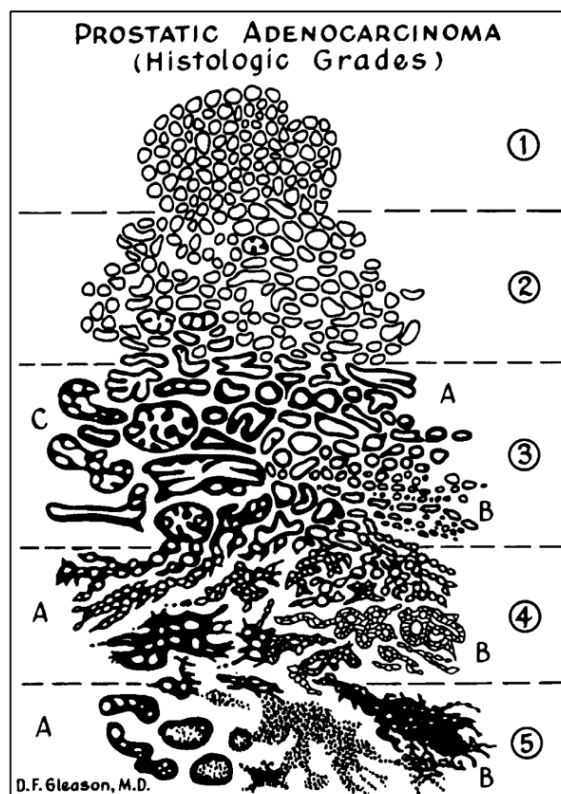


Figura 5: Graus de Gleason: Desenho padrão (Gleason, 1966).

Os cinco padrões básicos de graduação são usados para gerar um *score* histológico, que podem variar de 2 a 10, adicionando-se o grau do padrão primário com o grau do padrão secundário. O padrão primário é o que é predominante em área, pela simples inspeção visual. O padrão secundário é o segundo padrão mais comum. Se somente um padrão está presente na amostra de tecido, o grau é multiplicado por dois para resultar no *score*. De acordo com a abordagem de Gleason de 1977, se o segundo grau representa menos de 3% do total do tumor, ele é ignorado, e o primeiro grau é novamente dobrado para resultar no *Gleason Score* (GS) (Tannenbaum, 1977).

A adição dos dois padrões mais comuns para a geração do GS essencialmente toma a média do padrão primário e secundário. Este procedimento é único na graduação de cânceres humanos, onde, para a maioria das malignidades, é o pior grau que determina o desfecho do paciente. Particularmente para o carcinoma de próstata, quando existem dois

padrões de Gleason diferentes, as taxas de morte são intermediárias entre as taxas para pacientes com apenas a forma pura de cada um dos dois graus (Gleason, 1992).

O padrão de Gleason 1 (grau 1) é um padrão raro de células com crescimento muito bem diferenciado, com grande empacotamento, porém separadas, uniformes e ovais. No estudo original com 2911 casos de carcinoma prostático nas séries VACURG, este padrão foi encontrado em 3,5% dos casos (Bostwick, 1990). O carcinoma de padrão 1 é mais frequentemente descoberto de forma incidental na zona de transição.

O padrão de Gleason 2 forma massas celulares menos definidas que não são circunscritas como as do padrão 1. Comparadas ao padrão 1, as células glandulares do padrão 2 mostram certo grau de separação pelo estroma, com uma média de distância de separação menor do que o diâmetro das células glandulares. Outra característica distintiva do padrão 2 é o aumento da variabilidade do tamanho e da forma das glândulas. Existe, entretanto, uma capacidade limitada de infiltração no estroma. O padrão 2 é geralmente encontrado juntamente com o padrão 3, produzindo um Gleason *Score* de 5. A composição do padrão 2 com os padrões 4 e 5 é extremamente rara.

O padrão de Gleason 3 é moderadamente diferenciado e é o padrão mais comum de crescimento em carcinoma prostático (Humphrey *et al.*, 1996). O padrão 3 tem sido caracterizado como tendo três aparências distintas, designadas padrões 3A, 3B e 3C. Sugere-se que exista uma agressividade aumentada do carcinoma de padrão 3 progressivo do padrão 3A para 3B para 3C (Bostwick, 1990), mas nenhum dado publicado corrobora essa ideia. O padrão 3A é caracterizado por glândulas com poder de infiltração, de tamanho médio, únicas e com formas e espaçamento irregulares. O padrão 3B é essencialmente o mesmo do 3A, exceto que as glândulas são menores. O padrão 3C é composto de cilindros expandidos ou dutos com massas de tumor cribriforme ou papilar (Bostwick, 1990; Gleason, 1992). O padrão de Gleason 3 é, na grande maioria dos casos, encontrado de forma pura, de maneira

que o *score* mais comum é $3+3=6$. É também comum identificar padrão 3 em combinação com padrão 2, para um *score* 5, ou em conjunto com padrão 4 para um *score* 7. Nas séries de Gleason, uma composição com o padrão 8 para formar um *score* 8 era mais comum do que uma mistura com o padrão 4 para formar um *score* 7 (Tannenbaum, 1977). Atualmente, entretanto, o Gleason *score* 7 é mais comum que o 8 (Humphrey, 2004). Em geral, GS entre 5-7 contendo o padrão 3 são os graus histológicos mais comuns de carcinoma prostático (Humphrey, 2004).

O padrão 4 é de alto grau e com crescimento do carcinoma pouco diferenciado, com massas infiltradas de forma irregular. O padrão 4 é mais comumente combinado ao padrão 3, formando um *score* 7, o qual é atualmente um dos *scores* mais comumente encontrados. Doença puramente de alto grau, onde existe somente o grau 4 ou este misturado ao grau 5 para produzir *scores* de 8 e 9, são menos comuns. Gleason *score* 8 é o mais comum da série, e incluem cânceres de próstata de alto estágio, clinicamente confinados. O padrão 4 é raramente encontrado com os padrões 1 e 2.

O padrão 5 é o mais pobremente diferenciado, e apresenta duas formas: 5A e 5B. O padrão 5A é semelhante ao tipo *comedo* de câncer de mama, massas arredondadas lisas, cordas ou cilindros de carcinoma. A necrose é tipicamente central, cercada por massas de carcinoma papilares, cribriformes, ou sólidos. A aparência pode ser semelhante a 3C, exceto pela presença de necrose. O padrão 5B é composto por camadas irregulares de células de adenocarcinoma anaplásicas. Esta forma é tão pobremente diferenciada que apenas uma pequena quantidade de lúmen glandular pode estar presente para indicar um adenocarcinoma, ao invés de outra neoplasia pobremente diferenciada de tipo celular desconhecido. O padrão 5 é mais frequentemente encontrado com padrão 3, resultando em *score* 8. Menos comum é o padrão 5 puro ou misturado ao padrão 4, para produzir padrões 10 e 9, respectivamente. Gleason *Scores* 9 e 10 são vistos associados com altos estágios clínicos da doença, com

cânceres detectáveis clinicamente fora da próstata, ou localmente ou em forma de doença metastática (Tannenbaum, 1977).

1.4 Tratamento

Uma vez que diferentes tratamentos têm diferentes efeitos, é importante destacar os diferentes tipos de terapia atualmente disponíveis. Tipicamente, alguma forma de cirurgia ou radiação é usada quando o câncer é de baixa graduação e confinado a glândula (doença localizada). Opções de tratamento hormonal e quimioterápico são ordinariamente usados quando a malignidade se espalha da próstata para outras áreas do corpo (isto é, na doença avançada) (Eton *et al.*, 2002). Tratamentos para PC localizado podem resultar na cura, mas aqueles para doença avançada são principalmente usados para retardar a progressão ou para paliar os sintomas (Eton *et al.*, 2002). Prostatectomia radical, incluindo a abordagem de “*nerve sparing*”, envolve a remoção cirúrgica da próstata inteira e é usada principalmente quando se acredita que o câncer não se espalhou para fora da glândula. Ressecção transuretral da próstata envolve a remoção cirúrgica do tecido prostático que circunda a uretra e é mais usada quando o homem não pode sofrer a cirurgia radical devido a outras complicações de saúde ou por idade avançada. Criocirurgia, que envolve congelar áreas malignas da próstata com sondas metálicas resfriadas, é usada por uma pequena proporção de homens que estão interessados em uma forma de cirurgia menos agressiva.

Opções não cirúrgicas para doença localizada incluem radiação interna, externa e “*watchful waiting*” (espera vigilante). Terapia de radiação é comumente usada quando o PC está localizado na próstata e na região adjacente. Radioterapia de radiação externa (*external beam radiotherapy*, EBR), faz uso de raios X de alta energia ou de partículas radioativas geradas fora do corpo e focadas na malignidade. O EBR pode cobrir grandes áreas da pélvis ou com estereotaxia, conforme a região especificada, reduzindo assim a chance de dano no

tecido adjacente. Radiação interna, ou braquiterapia, utiliza pequenos *pellets* que são implantados diretamente na próstata.

Finalmente, para alguns homens com doença localizada, o curso apropriado de ação depois de um diagnóstico de PC é a observação, ou “espera vigilante” (Fleming *et al.*, 1993). Isto é, nenhum tratamento imediato é administrado; ao invés disso, o câncer é observado e monitorado a procura de sinais de progressão. Esta é uma escolha razoável para homens mais velhos com a doença em estágio inicial, de baixa graduação e com pouca ou nenhuma complicação, pois o PC pode ser uma malignidade de crescimento lento.

A prostatectomia radical permanece como o principal tratamento para PC localizado (Goldstraw, 2006). O relato mais significativo comparando a cirurgia com estratégias de espera vigilante foi produzido por Bill-Axelson *et al.* (Bill-Axelson *et al.*, 2005) em 2005. Eles relatam os resultados finais do grupo de estudo de PC da Escandinávia – um estudo randomizado prospectivo seguindo cerca de 700 homens com PC inicial (moderadamente e bem diferenciado) em um período de 10 anos (1989-1999). Durante um período mediano de 8,2 anos, a morte por PC ocorreu em 14,4% dos pacientes para os quais foi utilizada a estratégia de espera vigilante, contra 8,6% dos pacientes que se submeteram a cirurgia. Este é o único relato que mostra um benefício na sobrevida com cada modalidade de tratamento em separado. No momento, não há consenso sobre qual opção terapêutica é melhor para homens com PC localizado. Assim, a escolha de um tratamento é baseada em outros fatores além das vantagens de sobrevida, e podem incluir o efeito da terapia na qualidade de vida do paciente.

Para homens com PC avançado, as terapias hormonais e quimioterapia são principais opções (Rosenthal *et al.*, 2010). Terapias hormonais ajudam a diminuir os níveis de andrógenos circulantes. Andrógenos podem promover o crescimento de células malignas do PC. Existem terapias cirúrgicas e não cirúrgicas. Orquiectomia envolve a remoção cirúrgica

dos testículos. Isto fisicamente elimina a fonte primária de andrógeno no corpo. Hormônios farmacológicos podem alcançar essencialmente o mesmo fim sem a cirurgia. Análogos do hormônio liberador de gonadotrofina (LHRH) são usados para interromper a produção de testosterona. Outras drogas clássicas, antiandrógenos, são usadas para bloquear a capacidade do corpo de usar o andrógeno. Antiandrógenos podem ser combinados tanto com orquiectomia ou com análogo de LHRH juntos para bloquear completamente a atividade do andrógeno.

Numerosos estudos indicam que, embora a retirada de andrógeno forneça uma terapia significativa para a maioria dos pacientes, ela nunca é curativa (Denmeade *et al.*, 2002). Estes resultados são consistentes com o fato de que o câncer de próstata é composto de uma coleção heterogênea de células – dependentes e independentes de andrógeno. A terapia de retirada de andrógeno, independente de quão cedo é aplicada, não elimina as células do tipo independente de andrógeno. Esta percepção levou a estratégias de tratamento alternativas que procuram minimizar a duração da retirada de andrógeno, ou por adiar a terapia até que os pacientes tenham uma clara evidência de metástase, ou aplicando de maneira intermitente (Denmeade *et al.*, 2002). Finalmente, a quimioterapia é uma opção para alguns poucos pacientes que apresentam PC que se metastatizou e para quem a terapia hormonal falhou. Da mesma maneira que a terapia hormonal, o objetivo da quimioterapia é parar o espalhamento da doença e reduzir as complicações físicas, como dores nos ossos.

1.5 Recorrência Bioquímica

1.5.1 Definição

Entre pacientes com PC localizado, a medida do nível de PSA no soro para monitorar a recorrência depois do tratamento e para avaliar o efeito da terapia local definitiva tem sido amplamente aceita (Trapasso *et al.*, 1994; Zincke *et al.*, 1994). Um nível de PSA

detectável depois da prostatectomia, ou um aumento de PSA após a cirurgia tem sido chamado de “recorrência bioquímica” e é geralmente aceito como uma evidência da recorrência do câncer (Lange *et al.*, 1989; ASTRO, 1997).

Infelizmente, definir a falha bioquímica após o tratamento para PC localizado é problemático. Dependendo do tipo de terapia, os níveis de PSA que poderiam ser definidos como indicativos da doença variam. Após a prostatectomia radical, o nível de PSA sorológico deve ser indetectável dentro de seis semanas após a cirurgia, pois a fonte de produção foi removida. Assim, um nível detectável de PSA após a prostatectomia radical implica em tecido prostático residual e é mais provavelmente residual do que PC recorrente (Moul, 2000). No caso de terapia de radiação, onde o tratamento não elimina consistentemente todo o epitélio prostático, os níveis de PSA podem não diminuir a níveis indetectáveis e pode levar até 2 a 3 anos para alcançar o limite normal pós-tratamento (Cookson *et al.*, 2007). Assim, por causa do potencial resíduo epitelial funcional e sua capacidade de produzir PSA, não há um consenso significativo em relação ao tempo e nível de PSA que constituem falha no tratamento após a terapia de radiação (Vicini *et al.*, 2000; Critz, 2002). Uma definição para a falha de PSA após a terapia de radiação, desenvolvida pela *American Society for Therapeutic Radiology and Oncology* (ASTRO), descreve a falha bioquímica como três aumentos consecutivos de PSA (ASTRO, 1997). Apesar das tentativas de padronizar a definição de recorrência bioquímica entre os pacientes tratados com terapia de radiação, o critério da ASTRO tem sido alvo de debate (Gretzer *et al.*, 2002; Coen *et al.*, 2003; Horwitz *et al.*, 2005; Kuban *et al.*, 2005).

A importância de definir a recorrência bioquímica e identificar propriamente pacientes com falha antecipada do tratamento usando uma terminologia e definição consistentes não pode ser subestimada. Primeiro, um marcador preciso de falha no tratamento antecipado permite aos clínicos e pacientes recém-diagnosticados selecionarem terapias que

são mais prováveis de resultar no controle durável do câncer (Cookson *et al.*, 2007). Segundo, pacientes com falha bioquímica podem ser candidatos à terapia de resgate antecipada e, assim, é possível identificar corretamente pacientes com recaída bioquímica antes que a progressão clínica ocorra. Isto permitiria a aplicação de terapias efetivas para alterar a história natural do câncer recorrente (Cookson *et al.*, 2007). Finalmente, somente usando-se definições padronizadas de falha bioquímica é possível comparar de maneira confiável séries de pacientes recebendo uma dada terapia. É frequentemente desejável comparar diferentes tratamentos usando um substituto ao desfecho do PSA, mas, no presente, não há métodos de se conseguir tal confiabilidade.

Cookson *et al.* observaram a falta de uma padronização da definição de recorrência bioquímica baseados na avaliação da literatura relacionada a desfechos de PC T1-T2N0M0 (Cookson *et al.*, 2007). Os autores encontraram 53 diferentes variações da definição de progressão de PSA ou não progressão. Existem, entretanto, certas tendências que podem ser delineadas a partir dos dados existentes em relação à definição de recorrência de PSA após a prostatectomia radical. Primeiro, um período de “lavagem” após a prostatectomia radical de pelo menos seis semanas deve ser usado após obter o primeiro valor pós-operatório de PSA; Segundo, qualquer valor de PSA detectável após a prostatectomia radical deve ser repetido para confirmação e para eliminar erros laboratoriais; terceiro, um valor de corte entre 0,2 e 0,4 ng/mL parece ser o mais preciso para definir a recorrência bioquímica que irá por fim se traduzir na falha clínica após a prostatectomia radical. A seleção de valores menores aumenta a sensibilidade do marcador para a correlação com a recorrência da doença, enquanto maiores valores aumentam a especificidade (Cookson *et al.*, 2007).

Dado o alto grau de variabilidade que existe na literatura de PC em relação à recorrência bioquímica, o *Prostate Cancer Guidelines Update Panel* PCGUP recomenda uma definição padrão de recorrência bioquímica após a prostatectomia radical. É recomendado

que a recorrência bioquímica (PSA) após a prostatectomia radical seja definida como um nível de PSA sorológico acima de 0,2 ng/mL, com um segundo nível confirmatório > 0,2 ng/mL (Cookson *et al.*, 2007). O primeiro PSA deve ser obtido entre seis semanas e três meses após a terapia. A data da falha deve ser definida como a data do primeiro nível de PSA uma vez que este valor foi confirmado. O PCGUP reconhece que níveis maiores de PSA têm mais sensibilidade na detecção de recorrência da doença e progressão, mas considera que estes valores fornecem alta sensibilidade para recorrência bem como grande generabilidade, não confiando em ensaios ultrasensíveis de PSA. O painel também acredita que é apropriado reportar desfechos bioquímicos usando outros limites de corte de PSA, mas encoraja fortemente o uso de um padrão de 0,2 ng/mL para que comparações futuras possam ser feitas (Cookson *et al.*, 2007).

1.6 Definições de carcinomas clinicamente indolentes

Com o objetivo de prever o comportamento de um tumor individual em um paciente individual, um grande número de critérios tem sido desenvolvido. Em uma abordagem, Eastman *et al.*, e D'Amico *et al.* desenvolveram o conceito de grupos de “alto”, “intermediário”, e “baixo” risco (Wilt *et al.*, 1994; D'Amico *et al.*, 1998; Eastham *et al.*, 2002). A maioria dos sistemas combina PSA, Exame Digital Retal (*Digital Rectum Examination* DRE), informações do estadiamento e gradação tumoral. Infelizmente, no contexto baseado em população, estes sistemas não avançaram porque a maioria dos pacientes é categorizada como “baixo risco”. A Tabela 1 lista um número de definições para tumores clinicamente insignificantes.

Tabela 1: Definição de carcinoma clinicamente indolente

Série	Definição
(Smith <i>et al.</i>, 1994)	Impalpável, localizado, sem grau de Gleason 3, 4, ou 5
(Epstein <i>et al.</i>, 1994)	Volume tumoral < 0,2 cm ³ , confinado, sem grau de Gleason 4 ou 5
(Stamey <i>et al.</i>, 1993)	Volume tumoral < 0,5 cm ³
(Dugan <i>et al.</i>, 1996)	Volume tumoral < 20 cm ³ no momento da morte esperada, soma de Gleason < 1/10 da idade do paciente.

1.7 Marcadores Moleculares De Câncer de Próstata

1.7.1 Marcadores epigenéticos

Em PC, um grande número de estudos demonstrou que a detecção de hipermetilações aberrantes tem potencial para a detecção e prognóstico da doença (referências discutidas abaixo). A maioria destes estudos restringe sua análise ao tecido prostático tumoral, linfonodos pélvicos, tecido prostático não canceroso e linhagens celulares de próstata. A associação com a doença envolve a comparação de porcentagem de metilação de regiões promotoras alvo entre espécimes com e sem a doença. Tipicamente, estes estudos implicaram em um grande número de diferentes genes, e o objetivo tem sido identificar subconjuntos limitados de genes hipermetilados (ou, em alguns casos, hipometilados) que coletivamente fornecem diagnóstico adequado ou utilidade prognóstica (Reynolds *et al.*, 2007). Por exemplo, o índice de metilação, definido como a razão entre genes hipermetilados em relação ao total de genes analisados, foi mostrado ser correlacionado com mau

prognóstico, embora nenhum gene individual tenha fornecido correlação suficiente (Maruyama *et al.*, 2002). No presente momento, a hipermetilação do gene Glutathione S-Transferase PI (*GTSP1*) tem sido demonstrada ser a mais comum alteração epigenética em PC, estando presente em mais de 90% dos casos (Maruyama *et al.*, 2002; Jeronimo *et al.*, 2003; Singal *et al.*, 2004; Tokumaru *et al.*, 2004; Woodson *et al.*, 2004). Um estudo com 118 carcinomas de próstata demonstrou que *GSTPI* medido em combinação com um segundo gene, *Adenomatosis polyposis coli* (*APC*), forneceu uma sensibilidade de 98,3% para o carcinoma de próstata, com 100% de especificidade (Jeronimo *et al.*, 2004). Além de análises de tecido, análises de metilação de marcadores tem sido estendida para o soro, visando o desenvolvimento de testes de diagnóstico e prognóstico não invasivos. Neste caso, a suposição é de que o DNA circulante originário do carcinoma de próstata pode ser detectado e discriminado contra um DNA de fundo, originado de outros tecidos. Estas análises requerem alta sensibilidade e alta precisão analítica para um desempenho robusto (Reynolds *et al.*, 2007).

1.7.2 Marcadores de Prognóstico Após a Cirurgia

O prognóstico após o diagnóstico do PC é altamente variável, e a patogênese da doença não é bem compreendida (Jeronimo *et al.*, 2004). Para um paciente com a doença confinada ao órgão, de natureza relativamente indolente, pode não haver impactos significativos por 15 anos (Siddiqui *et al.*, 2004). Um paciente de mais idade diagnosticado com câncer de baixo grau irá provavelmente morrer de outras causas antes que precise de tratamento para PC, e pode não haver nenhum efeito deletério na qualidade de vida (Buhmeida *et al.*, 2006). Infelizmente, alguns pacientes apresentam formas mais agressivas da doença, onde o tempo para a metástase pode ser de apenas dois anos (Siddiqui *et al.*, 2004). O espalhamento da metástase é responsável pela letalidade do câncer (Gopalkrishnan

et al., 2001). Estes pacientes podem apresentar a doença invasiva de estágio tardio. Neste estágio não existem opções curativas disponíveis para os pacientes, pois o tratamento radical é bem sucedido apenas em doenças confinadas ao órgão (Bickers *et al.*, 2009). Existe uma grande necessidade para novos marcadores de prognóstico, a fim de melhorar o gerenciamento clínico de pacientes de PC, uma vez que, até o momento, não existe nenhum marcador que possa distinguir entre doenças indolentes e agressivas (Gopalkrishnan *et al.*, 2001). Este marcador poderia permitir aos clínicos oferecer ao paciente a estratégia de tratamento mais apropriada, e permitir que formas indolentes da doença possam ser tratadas com espera vigilante, assim diminuindo o número de biópsias negativas e intervenções cirúrgicas desnecessárias. Poderia também apontar para um tratamento mais efetivo para formas mais agressivas da doença, onde a estratégia da espera vigilante poderia levar a um crescimento não detectado e o espalhamento do câncer. Isto poderia assegurar que o tratamento fosse restrito àqueles pacientes nos quais fossem benéficos (Bickers *et al.*, 2009). É também extremamente desejável para um teste clínico que ele seja capaz de prever a probabilidade do desfecho antes da cirurgia, a fim de se evitar procedimentos desnecessários e permitir planejar tratamentos pós-operatórios otimizados para minimizar a recorrência (Miyake *et al.*, 2008).

Atualmente, vários fatores são usados para avaliar pacientes com PC em relação ao prognóstico da doença. Estes podem ser avaliados somente pelo exame da próstata inteira, se e quando a próstata é removida. No presente, fatores clínico-patológicos que são usados para prever a falha do tratamento incluem o PSA, a graduação de Gleason, estágio patológico, status da margem cirúrgica, invasão da vesícula seminal e o volume tumoral (Miyake *et al.*, 2008). A classificação de material de biópsia pelo sistema de Gleason (Figura 5) é um dos preditores mais relevantes do desfecho da doença do paciente, pois é fortemente correlacionado com a progressão da doença e sobrevida (Sogani *et al.*, 1985; Gleason, 1992;

Bostwick, 1994). Entretanto, como previamente mencionado, a classificação da biópsia é difícil e, quando comparada à classificação de amostra após a prostatectomia radical, mais da metade das biópsias são encontradas sub ou sobreclassificadas comparado com a classificação de amostras de prostatectomia radical (Djavan *et al.*, 1998).

Foram propostos diversos nomogramas pré e pós-operatórios que utilizam PSA entre outros múltiplos parâmetros clínicos para predizer o estágio da doença, a possibilidade de recorrência (medida pelo valor de PSA após a cirurgia) e a chance de sobrevida livre da doença (Ross *et al.*, 2001). Devido às limitações do PSA como biomarcador, existe uma necessidade urgente para novos biomarcadores que possam ser usados como indicadores de prognóstico em PC para diferenciar de maneira efetiva entre doenças indolentes e agressivas. Ao se avaliar a eficiência de um marcador, frequentemente a recorrência bioquímica é usada como indicador de falha do tratamento. Cerca de 30-40% dos pacientes irão sofrer recorrência bioquímica após a prostatectomia radical (Amling, 2006). Pode não ser possível, para um único biomarcador, fornecer a informação prognóstica necessária sobre o paciente para se basear as opções de tratamento. Ao invés disso, painéis de biomarcadores podem ser usados para predizer de maneira mais precisa o estágio da doença e como ela irá progredir (Shariat *et al.*, 2008).

1.8 Uso de Microarranjos para Prognóstico de Câncer de Próstata

A geração de perfis de expressão com a tecnologia de microarranjos para a descoberta de biomarcadores é baseada na suposição que os padrões de expressão gênica são determinantes importantes do comportamento da célula cancerosa (Sorensen *et al.*, 2010). Usando a tecnologia de microarranjo, é possível mapear, em escala genômica, as complexas aberrações moleculares associadas com o desenvolvimento do câncer (Sorensen *et al.*, 2010). Estas descobertas podem ser correlacionadas com os dados clínicos a fim de identificar

moléculas, ou assinaturas moleculares, associadas com certos estágios tumorais, graus ou desfechos clínicos (Sorensen *et al.*, 2010).

Um microarranjo é um arranjo ordenado de milhares de moléculas de DNA (sondas) imobilizadas em uma lâmina de vidro (Quackenbush, 2002). Várias plataformas de microarranjos usando sondas de oligonucleotídeos em lugar de sondas de cDNA estão disponíveis comercialmente, e em geral produzem dados de alta qualidade, com reprodutibilidade superior aos microarranjos de cDNA (Sorensen *et al.*, 2010). Em um experimento típico, o RNA total extraído de uma amostra biológica é convertido em cDNA por transcrição reversa, marcado e hibridizado em microarranjos para a determinação da abundância de transcritos (sistema "uma cor") ou expressão relativa comparada com uma amostra referência marcada com uma cor diferente (sistema "duas cores") (Quackenbush, 2002).

A viabilidade de medir a expressão de dezenas de milhares de genes em um único experimento modificou o gargalo da geração de dados para a análise de dados. Existem duas estratégias fundamentais para analisar dados de microarranjos: A primeira abordagem, conhecida como classificação não supervisionada (ou classificação hierárquica), tem como objetivo identificar subgrupos de amostras com perfis de expressão gênica similares e é usada para descobrir relações previamente desconhecidas entre as amostras (ou genes) (Eisen *et al.*, 1998; Golub *et al.*, 1999). Uma vantagem deste método é que dados clínicos adicionais não são necessários. A segunda abordagem, chamada classificação supervisionada, é utilizada para identificação de genes diferencialmente expressos entre grupos pré-definidos de pacientes, por exemplo, com recorrência *versus* sem recorrência. Enquanto um conjunto de amostras (o conjunto de treinamento) é usado para identificar a assinatura de expressão ótima que pode discriminar entre os grupos, outro conjunto independente (conjunto de validação ou conjunto teste) é usado para avaliar quão precisamente a assinatura pode classificar amostras

que não foram agrupadas (van 't Veer *et al.*, 2002). Devido à complexidade dos dados de microarranjos, existe um risco inerente de se identificar genes ou *clusters* irrelevantes. Assim, a validação independente é essencial para a avaliação da confiabilidade dos dados (Sorensen *et al.*, 2010).

Estudos de perfis de expressão gênica centrados em PC identificaram inúmeros genes codificadores de proteína com expressão diferencial em, por exemplo, tecido não-malígnos versus PC (Luo *et al.*, 2001; Magee *et al.*, 2001; Singh *et al.*, 2002; Halvorsen *et al.*, 2005; Kristiansen *et al.*, 2005; Merrick *et al.*, 2006), Neoplasia Prostática Intraepitelial de alto Grau (HG PIN) versus PC (Ashida *et al.*, 2004), PC metastático versus localizado (Eisen *et al.*, 1998; LaTulippe *et al.*, 2002; Varambally *et al.*, 2002), tumor sensível a hormônio versus câncer resistente à castração (True *et al.*, 2006) e PC de alto grau versus grau moderado e baixo (Welsh *et al.*, 2001; Best *et al.*, 2003). Estas listas de genes têm sido fontes valiosas para a identificação de marcadores de PC, bem como para a descoberta de genes com papéis funcionais importantes. A variação entre os resultados de diferentes estudos tem levantado a preocupação sobre a confiabilidade dos dados de microarranjos para PC. Entretanto, grande parte da variação entre os estudos pode ser explicada pelo uso de diferentes plataformas de arranjos, contendo sondas de diferentes sequências para um mesmo gene, e tipos diferentes de sondas (cDNA *versus* oligonucleotídeos), variação na preparação das amostras (microdissecção a laser ou aparagem crua), e diferentes protocolos de marcação (ex. com ou sem amplificação adicional de RNA), bem como pelo número limitado de amostras de pacientes de PC analisados na maioria dos estudos. Na primeira meta-análise de quatro estudos independentes de dados de expressão de microarranjos, com geração de perfil de tecido dissecado usando arranjos de oligonucleotídeos ou de cDNA, Rhodes *et al.* relataram uma lista de 103 genes que eram consistentemente regulados negativamente e 50

genes que eram consistentemente regulados positivamente em PC comparado com tecido não maligno com taxa de falsa descoberta (FDR) de 0,1% (Rhodes *et al.*, 2002).

1.8.1 Classificação de subgrupos clinicamente relevantes

Vários estudos de geração de perfil de expressão com microarranjos publicados focaram a classificação molecular baseada nos padrões de expressão de mRNA. A hipótese por trás disto é que o perfil de expressão global pode refinar a classificação além daquela obtida por dados morfológicos. Usando classificação hierárquica não supervisionada, vários estudos demonstraram a viabilidade de se discriminar entre condições não malignas da próstata (por exemplo: normal, não maligno adjacente e BPH), PC localizado e PC metastático, somente baseando-se em perfis de expressão gênica de tecido dissecado grosseiramente (Dhanasekaran *et al.*, 2001; Welsh *et al.*, 2001; LaTulippe *et al.*, 2002; Lapointe *et al.*, 2004). Além destes trabalhos, em um estudo por Ashida *et al.*, a classificação não supervisionada de dados de expressão de microarranjos obtidas de tecidos microdissecados levou a uma boa separação entre Neoplasia Prostática Intraepitelial de alto grau (*High-grade prostatic intraepithelial neoplasia* - HG-PIN) e PC, apesar de seus perfis de expressão comparados com tecido normal serem altamente sobrepostos (Ashida *et al.*, 2004). Isto é consistente com HG-PIN sendo um precursor de PC, como corroborado por um estudo subsequente de 101 populações de células microdissecadas representando todos os principais estágios da progressão de PC (Tomlins *et al.*, 2007). Outro estudo de geração de perfil de expressão de tecido microdissecado de amostras de tumor verificou que PC sensível ou resistente a hormônio poderiam ser distinguidos por classificação não supervisionada (Tamura *et al.*, 2007).

1.8.2 Genes como biomarcadores de câncer de próstata

Um relato de Dhanasekaran et al. foi o primeiro a integrar os dados de perfis de expressão de PC usando microarranjos com análise de arranjos de tecido (*Tissue Microarray Analysis* - TMA) e ligação com dados clínicos e patológicos (Dhanasekaran et al., 2001). Por imunohistoquímica, o grupo identificou a hepsina (*HPN*; uma serina protease de membrana) e *PIMI* (uma serina/treonina quinase) como um forte marcador de recorrência bioquímica após a prostatectomia. De forma semelhante, em um estudo separado de Varambally et al., o grupo *polcomb* (*PcG*) *protein enhancer of zeste homolog 2* (*EZH2*) foi encontrado sendo regulado positivamente em PC metastático comparado com PC localizado, usando perfis de expressão com microarranjos, e uma alta expressão da proteína em PC clinicamente localizado foi mostrada como sendo associada com mau desfecho, por análise TMA (Varambally et al., 2002). Outros potenciais marcadores com valor prognóstico independentes foram identificados com abordagens semelhantes, incluindo *AZGP1*, *MUC1* (Lapointe et al., 2004), *hCAP-D3* (Lapointe et al., 2008), *CD166*, *MEMD*, *CD24* (Kristiansen et al., 2005), *SLC18A2* (Sorensen et al., 2009), *TEAD1* e *c-CBL* (Knight et al., 2008).

1.8.3 Assinaturas de expressão gênica associadas ao desfecho de câncer de próstata

O primeiro relato de uma assinatura associada a desfecho da doença em PC foi publicado em 2002 por Singh et al. (Singh et al., 2002), que usou um *k-nearest neighbour supervised machine-learning algorithm* para construir um modelo de cinco genes que predizem a recorrência de PSA. A assinatura foi eficiente para 90% dos testes usando *Leave-One-Out-Cross-Validation* (LOOCV). Em um relato separado, usando análise de significância para microarranjos (*Significance Analysis of Microarrays*, SAM), Lapointe et al. identificaram uma assinatura de 23 genes associada com a volta do tumor que também podia predizer a recorrência com exatidão razoável no conjunto de dados de Singh et al. (Lapointe et al., 2004). Uma terceira assinatura foi relatada por Bismar et al., que usou dados

de expressão proteica e análise discriminante para o desenvolvimento de um modelo (Bismar *et al.*, 2006). O poder preditivo deste modelo de 12 genes foi testado no nível de transcrito usando os dados de expressão de Glinsky *et al.*, resultando numa exatidão modesta de 64% (Glinsky *et al.*, 2004).

Em outro estudo, Varambally *et al.* usaram somente genes com expressão concordante de RNA e proteína em PC localizado versus PC metastático, e desenvolveram uma assinatura de 50 genes que podia não só prever a recorrência de PSA, mas também prever o desfecho para pacientes com glioma, câncer de pulmão e mama (Varambally *et al.*, 2005). Anterior a este trabalho, Yu *et al.* relataram uma assinatura de 70 genes para a predição de PC agressivo (definido pela presença de invasão local, metástase distante e/ou recorrência de PSA), que mostrou 78% de exatidão em um conjunto independente de pacientes (Yu *et al.*, 2004).

1.9 RNAs Não Codificadores (ncRNAs)

Além das alterações de expressão de genes codificadores de proteínas, alterações em RNAs não codificadores (ncRNAs) curtos, com < 30 bases, os microRNAs, contribuem para a patogênese do câncer (Esquela-Kerscher *et al.*, 2006) e, portanto, representam outra fonte potencial de biomarcadores de PC. Entretanto, até o momento, poucos estudos de perfis de expressão de ncRNAs foram publicados.

Recentes melhorias nas análises de larga-escala de expressão gênica levaram à descoberta de que ncRNAs curtos, e também longos (> 200 bases) representam a grande maioria das unidades transcricionais humanas (Rinn, Euskirchen *et al.* 2003; Bertone, Stolc *et al.* 2004; Johnson, Edwards *et al.* 2005; Birney, Stamatoyannopoulos *et al.* 2007). A caracterização de muitos destes transcritos tem levado a descoberta de novos RNAs regulatórios.

O aspecto mais interessante nos ncRNAs é sua implicação em um grande número de doenças humanas. A alteração de níveis de expressão ou alterações genéticas e epigenéticas afetando ncRNAs, acompanhando processos malignos, apoiam fortemente o papel funcional dos ncRNAs, especialmente quando genes de ncRNAs são interrompidos por aberrações cromossômicas (Szymanski *et al.*, 2005).

A alteração de expressão de certos ncRNAs tem sido associada com algumas formas de câncer. O primeiro e mais bem estudado ncRNA relacionado a câncer é o RNA *H19*. É também o primeiro ncRNA identificado como produto de um gene “*imprinted*”. O gene *H19* humano é localizado no cromossomo 11p15.5 adjacente ao *IGF2* (*insulin-like growth factor 2*). Estes dois genes são *imprinted* reciprocamente, com *IGF2* e *H19* mostrando expressão paterna e materna, respectivamente. Durante o desenvolvimento, *H19* é transcrito em níveis elevados na placenta, no embrião e na maioria dos tecidos fetais, mas após o nascimento a sua expressão é significativamente reduzida (Looijenga *et al.*, 1997). Tem sido proposto que padrões de expressão recíproca de *IGF2* e *H19* no curso do desenvolvimento refletem diferentes papéis desempenhados pelos produtos destes dois genes. *IGF2* é um fator de crescimento que estimula a proliferação celular enquanto que o RNA *H19* pode ser responsável pela diferenciação (Looijenga *et al.*, 1997). Alterações no status de metilação de regiões diferencialmente metiladas (DMR) acima do gene *H19*, resultando na perda do *imprinting* de *H19* e *IGF2* e da expressão bialélica de um dos genes pode causar crescimento celular maligno (Ulaner *et al.*, 2003).

Outro gene *imprinted* implicado em câncer é um RNA antisense expresso paternalmente *PEG8/IGF2AS*. Este RNA é transcrito do locus *IGF2*. *IGF2AS* mostra expressão significativamente aumentada em tumores de Wilms e em vários outros tumores fetais (Okutsu *et al.*, 2000). A superexpressão é restrita a células cancerosas e não é observada em tecido de rim normal adjacente. *IGF2AS* humano contém uma ORF

codificando uma proteína longa putativa de 273 aminoácidos. Esta *ORF* não é conservada em camundongo e nenhum produto proteico correspondente foi identificado, o que sugere que não é traduzido (Okutsu *et al.*, 2000). A superexpressão específica de ncRNAs foi também descrita como um bom marcador de vários outros tumores em adultos. Células de carcinoma de cólon mostraram níveis significativamente elevados do transcrito gênico *OCC-1* (*overexpressed in colon carcinoma 1*). *OCC-1* é localizado no cromossomo 12q24.1 e produz dois transcritos, de 1,2 e 1,3 kb de tamanho, que diferem na sua porção terminal 5' e 3'. *OCC-1* mostra padrões de expressão tecido-específicos e é ausente ou expresso em níveis muito baixos em mucosa normal (Pibouin *et al.*, 2002).

Um dos mais longos genes de ncRNAs humanos é o gene *BCMS* (*B-cell neoplasia-associated with multiple splicing*) localizado na região cromossômica 13q14.3. Deleções afetando *BCMS* foram encontradas em mais da metade dos casos de leucemia linfocítica crônica de células B (B-CLL) e em mais de 60% dos linfomas de células do manto. *BCM* abrange aproximadamente 560 kb e é composto por pelo menos 50 éxons. O *splicing* alternativo do transcrito primário gera um grande número de produtos maduros, que mostram uma distribuição tecido-específica. De maneira interessante, nenhuma das variantes de *splicing* descritas atualmente tem potencial codificador e estas são, provavelmente, o produto funcional final (Wolf *et al.*, 2001). O papel destes transcritos e o seu potencial envolvimento na supressão tumoral não são conhecidos.

A expressão de ncRNAs pode variar entre diferentes tipos de células cancerosas. Em carcinoma pulmonar de não pequenas células (NSCLC), a metástase foi associada à expressão do gene *MALAT-1* (*metassociated in lung adenocarcinoma transcript 1*), gerando um ncRNA de 8 kb. A associação específica do *MALAT-1* com formas metastáticas de câncer pulmonar o torna um bom candidato para o diagnóstico de pacientes de câncer de pulmão (Ji *et al.*, 2003). Em rabiomiossarcoma (RMS), os dois subtipos histológicos também podem ser

distintos baseados na expressão do gene *NCRMS* (*noncoding RNA in RMS*). O gene *NCRMS* consiste de pelo menos 11 éxons com *splicing* alternativo mapeados no cromossomo 12q21. *NCRMS* mostra expressão elevada em RMS alveolar, mas não no subtipo embrionário do RMS. A expressão de *NCRMS* pode indicar uma desregulação da expressão gênica dentro de uma grande região cromossômica, incluindo diversos genes associados com desenvolvimento muscular, reguladores miogênicos *Myf5* e *Myf6*, e o fator de crescimento *Igf2*. Isto pode ser um agente causador do desenvolvimento tumoral. Padrões específicos de expressão de *NCRMS* foram observados em neuroblastoma e sarcoma sinovial, o que sugere que estes tumores podem ter uma etiologia comum (Chan *et al.*, 2002).

Em tumores de próstata, existem dois ncRNAs que são superexpressos em câncer comparado com tecido saudável. O RNA DD3 é produto do gene *DD3-PCA3* (*prostate cancer antigen 3*) mapeado no cromossomo 9q21-22, e sua transcrição parece ser restrita à próstata. Níveis elevados deste transcrito foram observados em 90% dos casos de tumor de próstata (Bussemakers *et al.*, 1999), e foi mostrado ser um marcador específico, que pode ser usado para diagnóstico de câncer de próstata (de Kok *et al.*, 2002). Outro ncRNA específico de próstata que é regulado positivamente em tecido tumoral e responsivo a andrógeno é o *PCGEM1* (Srikantan *et al.*, 2000). A superexpressão de *PCGEM1* está correlacionada com proliferação e formação de colônia, o que sugere o seu envolvimento na regulação do crescimento celular (Petrovics *et al.*, 2004).

1.9.1 Transcrição Intrônica

A crescente caracterização da existência de transcrição intrônica de ncRNAs generalizada no genoma humano (Rinn *et al.*, 2003; Kampa *et al.*, 2004) e a correlação da transcrição intrônica com o grau de malignidade em PC (Reis, E. M. *et al.*, 2004) sugerem a necessidade de um amplo estudo de possíveis alterações de expressão nesse grupo de

transcritos. Estima-se que cerca de 98% dos transcritos de eucariotos são não codificadores (Mattick, 2001). Isso inclui íntrons e transcritos de genes não codificadores de proteínas, sendo estes cerca de 50%-75% de todos os transcritos em eucariotos superiores (Mattick, 2001; Shabalina *et al.*, 2001).

Estudos usando microarranjos com sondas de DNA genômico ao longo da sequência completa dos cromossomos 21 e 22 (Rinn *et al.*, 2003; Kampa *et al.*, 2004) e, posteriormente, os *tilling arrays* com todo o genoma humano (Bertone *et al.*, 2004) mostraram que existem 10 vezes mais *loci* com atividade transcricional fora dos éxons, em regiões intrônicas ou intergênicas de genes anotados, sendo que quase metade dos transcritos intrônicos eram expressos antissenso ao seu respectivo gene. Além disso, estima-se que cerca de 20% dos transcritos em humanos podem formar pares senso antissenso (Chen *et al.*, 2004).

Os estudos iniciais do projeto ENCODE, que analisaram apenas 1% do genoma humano, mostram que de 74% a 93% das bases são transcritas (Birney *et al.*, 2007), sendo a maioria fora dos 3% de bases que codificam proteínas. As funções dos RNAs não codificadores curtos, como microRNA e piRNA, estão sendo estudadas por vários grupos de pesquisa (O'Donnell *et al.*, 2007; Zhang *et al.*, 2007). RNAs não codificadores longos tem recebido menos atenção da comunidade científica, baseada na percepção provavelmente equivocada de que todos eles resultam de RNAs imaturos (Louro *et al.*, 2009). Em 2005, nosso grupo fez um levantamento *in silico* das principais descobertas dos projetos de análise de expressão em larga escala que levaram à identificação de ncRNAs de transcrição independente (Reis *et al.*, 2005). Na maioria dos casos, a informação sobre esta classe de ncRNAs estava escondida entre os dados de expressão em larga-escala dos diferentes tipos de transcritos desconhecidos. Usando abordagem *in silico* baseada no mapeamento genômico e *clustering* de EST (*Expressed Sequence Tags*), juntamente com experimentos de microarranjos combinando íntrons e éxons, nós fomos capazes de apontar para regiões

intrônicas como fontes-chave de potenciais ncRNAs reguladores (Nakaya *et al.*, 2007; Louro *et al.*, 2008). O catálogo atualizado de transcrição intrônica humano e de camundongo descreve 78.147 e 39.660 *contigs* longos de ESTs, respectivamente, dos quais a maioria deve representar novos ncRNAs (Louro *et al.*, 2008). Além disso, nosso trabalho revelou que aproximadamente 81% e 70% de todos os genes codificadores de proteína que sofrem *splicing* contém íntrons transcricionalmente ativos (Louro *et al.*, 2008).

Oncogenes e genes supressores tumorais como N-myc (*MYCN*) e p53 (*TP53*) foram identificados há muito tempo atrás como apresentando ncRNAs transcritos em sua região intrônica (Khochbin and Lawrence 1989; Krystal, Armstrong *et al.* 1990), apesar de que uma correlação entre esses ncRNAs e câncer não tenha sido estudada. Mais recentemente, expressão aberrante de ncRNAs intrônicos tem sido correlacionada com o grau de malignidade do câncer de próstata (Reis, E. M. *et al.*, 2004).

Durante o desenvolvimento do projeto temático “Identificação de marcadores moleculares para diagnóstico e prognóstico em câncer utilizando microarranjos de DNA”, o nosso grupo identificou vários transcritos longos intrônicos não codificadores, cujos níveis de expressão estavam correlacionados ao grau de diferenciação de tumores de próstata (Reis, F. *et al.*, 2004). Naquele trabalho foram estudadas 27 amostras de tumores de próstata, sendo construído em nosso laboratório um microarranjo com quatro mil sondas de cDNA (Reis, E. M. *et al.*, 2004), selecionadas a partir de aproximadamente um milhão de clones de cDNA que foram gerados e sequenciados pelo Projeto Genoma do Câncer com a técnica ORESTES (Dias Neto *et al.*, 2000). Dos transcritos que selecionamos, cerca de mil eram transcritos “*no-match*”, representando fragmentos de RNAs não-codificadores mapeados no genoma em regiões de íntrons de genes conhecidos. Naquele estudo verificou-se que uma fração consideravelmente maior (6,6%) dos transcritos intrônicos estava correlacionada com o grau de diferenciação do tumor de próstata (através da graduação de Gleason), quando comparadas

com transcritos de regiões genômicas não anotadas (2%). Dentre os 12 transcritos mais correlacionados com a diferenciação do tumor, seis (50%) são mensagens intrônicas antissenso (Reis, E. M. *et al.*, 2004).

Em seguida, demonstramos que a expressão de uma série de transcritos intrônicos pode ser modulada por andrógeno (Louro *et al.*, 2007) em uma linhagem de próstata. Na etapa seguinte, estudamos perfis de expressão intrônica em tecido humano normal de fígado, rim e duas amostras de tumor de próstata (Nakaya *et al.*, 2007). Para isso usamos o mesmo desenho do microarranjo descrito na presente tese, contendo cerca de 30 mil sondas de oligonucleotídeos 60-meros para RNAs intrônicos. Surpreendentemente, observamos que os mais abundantes desses transcritos de ncRNAs são gerados em *loci* gênicos onde são transcritos genes codificadores de proteínas envolvidos com a regulação da transcrição, em todos os três tecidos estudados (Nakaya *et al.*, 2007). Para as sequências exônicas dos genes codificadores de proteínas não observamos enriquecimento de genes relacionados com a regulação da transcrição entre os mais abundantes, o que sugere um papel especial dos transcritos intrônicos produzidos nos *loci* dos genes reguladores de transcrição na regulação da expressão gênica. Também observamos padrões de expressão intrônica tecido-específicos, comparando os três tecidos estudados (Nakaya *et al.*, 2007).

2 Objetivo Geral

Identificar perfis de expressão de genes codificadores e não codificadores de proteína correlacionados com recorrência de tumor de próstata, a fim de gerar um perfil prognóstico com potencial uso como biomarcador, e elucidar o possível papel de ncRNAs no desenvolvimento do câncer de próstata.

2.1 Objetivos Específicos

- Extração e purificação de RNA total de tecido tumoral de câncer de próstata de pacientes de amostras de pacientes submetidos à prostatectomia radical, com longo acompanhamento clínico (cinco anos) e conhecido desfecho da doença.
- Amplificação do RNA contendo cauda poli-A e marcação com fluoróforos, a partir do RNA total purificado de cada amostra.
- Hibridização do cRNA amplificado com o microarranjo por nós desenhado, contendo aproximadamente 37 mil sondas de oligonucleotídeos simples-fita, das quais cerca de 32 mil representam a fita senso e a fita antissenso de aproximadamente 16.000 RNAs intrônicos longos não codificadores, sem evidência de *splicing*.
- Extração e processamento das imagens dos microarranjos para geração de dados de expressão.
- Filtragem e normalização dos dados de expressão dos genes codificadores e intrônicos não codificadores de proteína.
- Busca por genes estatisticamente diferencialmente expressos entre amostras de pacientes com e sem recorrência do tumor.

- Análise funcional *in silico* dos genes diferencialmente expressos, verificando enriquecimento de categorias de ontologias gênicas e de redes gênicas relevantes na progressão tumoral.
- Busca por perfis de expressão de ncRNAs intrônicos estatisticamente correlacionados com o desfecho da doença, que possam ser preditores do desfecho da doença.
- Comparação dos resultados obtidos com estudos de perfil de expressão disponíveis na literatura e relevantes em câncer de próstata.
- Validação dos perfis preditores identificados, usando um grupo externo independente de amostras.

3 Material e Métodos

3.1 Amostras de tecido

As amostras de tecido tumoral foram obtidas de um banco de tecidos mantido pelo Laboratório de Patologia Cirúrgica e Molecular do Hospital Sírio-Libanês, em São Paulo. As amostras de tecido foram reexaminadas por um patologista a fim de ter certeza da presença de pelo menos 75% de tumor. Todas as amostras foram coletadas entre 1998 e 2001, com consentimento dos pacientes submetidos à cirurgia. Todas as amostras foram revisadas por um único uropatologista. Microsecções foram obtidas a partir dos blocos congelados e aproximadamente 50 mg de tecido foi utilizada para a extração de RNA.

3.1.1 Critérios de inclusão

Todos os pacientes incluídos neste estudo não receberam nenhum tratamento adjuvante após a prostatectomia radical enquanto eles permaneceram sem recorrência. Um conjunto de treinamento de 43 pacientes foi obtido do banco: o conjunto é composto de 16 amostras de pacientes com recorrência (38% do conjunto de treinamento), definidas como tendo um aumento dos níveis de PSA para $\geq 0,2$ ng/mL em qualquer período durante os 5 anos de acompanhamento depois da prostatectomia, seguidas de outra medida de PSA $\geq 0,2$ ng/mL confirmatória; e 26 amostras de pacientes sem recorrência (62% do conjunto de treinamento), determinado pelo período de acompanhamento de 5 anos. Um conjunto adicional independente de 22 amostras foi obtido e usado em um teste de classificação cega a fim de testar o perfil prognóstico.

3.2 Extração do RNA das amostras tumorais

O RNA total foi extraído a partir do tecido tumoral congelado através de homogeneização em tampão baseado em Isotiocianato de guanidina (Trizol; Invitrogen,

Carlsbad, CA). O reagente Trizol foi homogeneizado com tecido macerado sob nitrogênio líquido na proporção de 1 mL do reagente para cada 50 mg de tecido. Após incubação por 5 minutos a temperatura ambiente, são adicionados 200 µL de clorofórmio. Após uma nova incubação a temperatura ambiente por 5 minutos, é realizada uma centrifugação dos tubos a 11.000 x g por 15 minutos a 2°C, a fim de separar a fase aquosa, que contém o RNA, dos demais componentes. Após a centrifugação, a fase aquosa é separada e colocada em um novo tubo, onde são adicionados 500 µL de isopropanol e seguido incubação por 10 minutos a temperatura ambiente.

Após nova centrifugação, o sobrenadante é removido e descartado. Em seguida, adiciona-se ao tubo 1 mL de etanol 75% e segue-se nova centrifugação a 7.000 x g por 5 minutos a 2°C. Após remoção do sobrenadante, o precipitado é dissolvido em 30 µL de água previamente tratada com DEPC.

3.3 Purificação do RNA e tratamento com DNase

Em seguida, o RNA é submetido a uma etapa de purificação e tratamento com RNeasy Mini Kit (Qiagen), de acordo com o protocolo do fabricante. Inicialmente o volume total da amostra de RNA é ajustado para 100 µL (adição de 70 µL de H₂O DEPC). Em seguida são adicionados 350 µL de tampão RTL, homogeneizado e adicionado 250 µL de etanol 100% e homogeneizado novamente por pipetagem. Em seguida os 700 µL resultantes são transferidos para a coluna de captura do RNA, montada dentro de um tubo coletor de 2 ml. A coluna com o RNA é centrifugada por 30 segundos a velocidade máxima na microcentrífuga Eppendorf modelo 5417R. O filtrado é dispensado, mantendo-se o tubo coletor. Em seguida são adicionados 350 µL de tampão RW1 à coluna, seguida de nova centrifugação por 30 segundos a velocidade máxima. O filtrado é descartado. Em seguida é adicionado à coluna uma solução contendo 10 µL de solução de DNase I e 70 µL de tampão

RDD. A coluna contendo DNase é incubada a temperatura ambiente por 15 minutos. Em seguida são pipetados 350 µL de tampão RW1 à coluna e se centrifuga a velocidade máxima por 30 segundos na microcentrífuga Eppendorf modelo 5417R. O filtrado é descartado. As etapas de adição de RW1 e centrifugação são repetidas. Em seguida, a coluna é transferida para um novo tubo coletor e são adicionado 30 µL de H₂O DEPC à coluna. Após a incubação por 1 minuto, segue-se uma centrifugação por 1 minuto à velocidade máxima na microcentrífuga Eppendorf modelo 5417R. O filtrado, contendo o RNA purificado, é armazenado.

3.4 Avaliação do rendimento e da integridade do RNA obtido

O RNA obtido é analisado quanto à sua qualidade utilizando-se o equipamento Bioanalyzer 2100 (Agilent Technologies), que realiza uma eletroforese capilar de alta tensão utilizando as amostras de RNA. A integridade das amostras foi avaliada com o software 2100 Expert (Agilent). Este programa atribui um valor de integridade de RNA (*RNA Integrity Number*, RIN), que permite uma estimativa da integridade através de todo traçado eletroforético da amostra, e não apenas da razão entre os RNA ribossomais. Apenas amostras que tinham uma boa integridade (RIN > 7) foram utilizadas nas etapas de amplificação, marcação e hibridização.

A quantificação das razões de medida óptica realizadas a 260 nm (pico de absorbância do RNA) e 280 nm (pico de absorbância de proteínas), foram realizadas utilizando-se o espectrofotômetro NanoDrop (Thermo Scientific).

3.5 Amplificação do RNA contendo cauda poli-A

A amplificação e marcação com fluoróforos foi realizada com iniciadores de oligo d-T ligados ao promotor da enzima T7 RNA polimerase, utilizando o Kit de

amplificação *Low RNA Input Fluorescent Linear Amplification Kit* (Agilent Technologies) e 300 ng de RNA total, de acordo com as instruções do fabricante (Figura 6). Com essa abordagem, nós medimos as mensagens codificadoras de proteínas e ncRNA que apresentam uma cauda poli-A 3'.

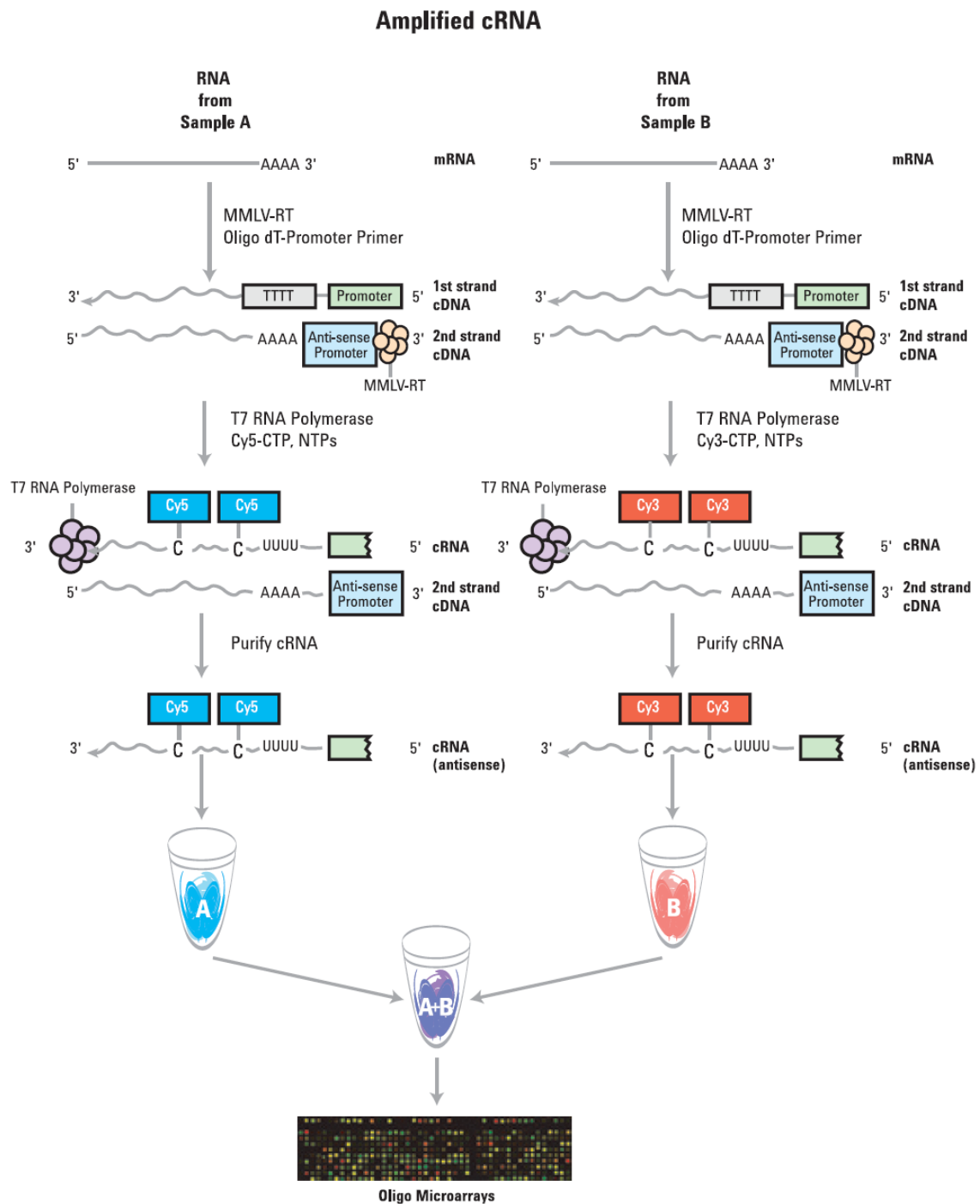


Figura 6: Método de amplificação de cRNA e marcação com fluoróforos (Agilent Technologies).

Cada amostra de RNA de PC foi amplificada separadamente em duplicata com Cy3 ou Cy5. O método usa T7 polimerase, que simultaneamente amplifica o material alvo gerando o chamado cRNA (RNA complementar) e incorpora CTPs marcados com cianina 3 (Cy3) ou cianina 5 (Cy5). Há normalmente uma amplificação na ordem de 100 vezes do RNA com cauda poli-A com o uso deste kit. 300 ng de RNA total são usados em cada reação, em um volume de 8,3 µL. São adicionados 1,2 µL de iniciador oligo-dT ligado ao promotor de T7, e 2 µL da solução de RNAs “spikes”, sendo *Spike A* para RNA que será marcado com Cy3, e *Spike B* para o RNA que será marcado com Cy5. O RNA *spike* é de planta, em uma mistura de 10 RNAs diferentes com proporções conhecidas, e o oligoarranjo construído pela Agilent possui sondas para estes RNAs, que funcionam como controle da amplificação e marcação. Os tubos são levados ao termociclador para a desnaturação do RNA, a 65°C por 10 minutos. Em seguida, é adicionado 8,5 µL do tampão de transcrição reversa cDNA Master Mix (4 µL 5 X *First Strand Buffer* previamente aquecido por 4 minutos a 80 °C; 2 µL 0.1 M DTT; 1 µL 10 mM dNTP mix; 1 µL MMLV-RT; 0,5 µL RNase OUT) e a reação de transcrição reversa se dá no termociclador por 2 horas a 40 °C. Em seguida, é adicionado 60 µL do tampão de transcrição *Transcription Master Mix* (15,3 µL H₂O DEPC; 20 µL 4x *Transcription Buffer*; 6 µL 0.1 DTT; 8 µL NTP mix; 6,4 µL 50% PEG previamente aquecido por 1 minuto a 40°C; 0,5 µL RNase OUT; 0,6 µL pirofosfatase inorgânica; 0,8 µL T7 RNA polimerase; 2,4 µL Cianina 3 ou 5). A reação de transcrição se dá no termociclador a 40°C por 2 horas.

Em seguida, o RNA amplificado é purificado utilizando-se o *RNeasy amplification Kit*, como descrito para a purificação do RNA total, mas sem a etapa de digestão com DNase. A incorporação do corante é confirmada medindo-se a absorbância do corante utilizando o espectro-fotômetro NanoDrop (NanoDrop Technologies). Como

recomendado pelo fabricante, cRNA amplificado com uma incorporação mínima de 7 pmol fluoróforo / micrograma cRNA foi usada para hibridização.

3.6 Desenho do oligoarranjo

Sondas contendo fragmentos de DNA fita-simples de 60 nucleotídeos, cada um com sequência única representando um gene de interesse, bem como sondas controles positivos e negativos, foram desenhadas estritamente de acordo com as especificações da Agilent, ou seja, com conteúdo C-G menor que 60%, ausência de bases repetidas consecutivamente mais do que 7 vezes, e com bias para a região 3' do transcrito (Hughes *et al.*, 2001). Os oligoarranjos foram impressos pela Agilent. A anotação das sondas e suas sequências estão depositadas no banco público do NCBI (número de acesso GEO: GPL4051).

3.7 Hibridização

Nós usamos hibridização *self-self*, na qual os microarranjos são co-hibridizados com cada uma das réplicas de cRNA marcadas com Cy3 e Cy5. De acordo com o protocolo *Agilent's ISH Kit Plus*, nós usamos 750 ng de cRNA marcado com Cy5 e Cy3 por amostra e hibridizamos por 17 horas a 65°C, seguido de lavagem com o *Agilent SSPE wash protocol v. 2.1*.

3.8 Processamento de Imagem

Os arranjos foram digitalizados com uma resolução de 5- μ m usando um *scanner* GenePix 4100B (Molecular Devices Corporation, Sunnyvale, CA). Configurações do tubo fotomultiplicador foram ajustadas automaticamente para se obter valores entre 0,95 e 1,05 para as razões de intensidades globais entre os dois canais (Cy3 e Cy5). Intensidades de fluorescência foram extraídas usando o *Feature Extraction software* (FE). O programa

identifica o local de cada sonda (*spot*), obtem a intensidade daquela região registrada pelo scanner, e aplica a subtração da intensidade do fundo local (em torno do *spot*). Em seguida, corrige a intensidade do conjunto de spots da lâmina, utilizando o método padrão LOWESS (*Local Weighted Scatterplot Smoothing*), que compensa para a incorporação desigual de corante e a sensibilidade diferente da medida para os diferentes comprimentos de onda de cada corante.

3.9 Processamento de Dados

Programas adicionais caseiros foram desenvolvidos para processar os dados utilizando o ambiente do projeto R (R_Development_Core_Team, 2006). Sondas que não foram consideradas expressas pelo programa FE, isto é, com valores de intensidade não estatisticamente acima do *background* foram excluídas. A fim de preservar potenciais marcadores que poderiam ser expressos em apenas um grupo de amostras, foram consideradas para análise todos os ncRNAs e genes codificadores de proteínas expressos em pelo menos 90% das amostras de um dos dois grupos: no conjunto de pacientes com recorrência (14/16), ou no conjunto de pacientes sem recorrência (23/26). Esta abordagem assegura que os genes que tem expressão completamente ausente em um dos dois grupos fossem mantidos na análise.

Para cada sonda, foi utilizada a média geométrica das duas replicas (Cy3 e Cy5) da mesma amostra, em escala log, isto é, $\frac{\log_2(\text{canal verde}) + \log_2(\text{canal vermelho})}{2}$. Para tornar as lâminas comparáveis e reduzir artefatos experimentais, utilizamos o método de normalização intra-lâmina denominado quantil (Bolstad *et al.*, 2003), disponível no pacote *affy* (Gautier *et al.*, 2004) do projeto R (R_Development_Core_Team, 2006). Neste método, as medidas de cada lâmina são ordenadas (*ranked*) segundo os valores de intensidade. Em seguida, para cada lâmina e para cada posição no ranking este valor é substituído pela média das

intensidades, obtida para cada posição no ranking, entre todas as lâminas de treinamento; obtêm-se assim um conjunto único de dados (média das intensidades entre todas as lâminas), sendo um valor médio para cada posição do ranking. Este ranking de intensidades médias gerado com as amostras do grupo de treinamento foi utilizado para a normalização das amostras do teste independente, como discutido mais adiante, no processamento dessas amostras. Por esse método, é mais relevante a posição do *spot* no *ranking* do que a sua intensidade bruta. Com esse critério, podem-se comparar diferentes experimentos com faixas de intensidades diferentes.

Para ajustar para efeitos de lote entre os experimentos, foi usado um método que é efetivo para números pequenos (<10) de amostras por lote experimental (Johnson *et al.*, 2007) para ajustar todo o conjunto de treinamento. Os dados de expressão foram depositados no banco de dados GEO do NCBI com o número de acesso GSE18918.

3.10 Análise estatística

No conjunto de treinamento com 42 amostras, foi aplicado para os genes codificadores de proteína um teste de significância específico para microarranjos chamado SAM – *Significance Analysis of Microarray* (Tusher *et al.*, 2001). Este teste leva em consideração, além dos valores de significância tradicionais, como o p-valor do teste-t, o número de testes realizados – um para cada medida, e a probabilidade de se obter genes com valores altos de significância (ex. p valor menor que 0,005), mas resultantes do grande número de perguntas (10000 spots x 0,05 = 50 genes), a chamada taxa de falsa descoberta ou *False Discovery Rate* (FDR). Aqui foram utilizadas 1000 permutações.

Para a identificação de transcritos intrônicos não codificadores correlacionados com a recorrência em PC nós usamos a abordagem de Amostragem Múltipla Aleatória de Treinamento-validação (Michiels *et al.*, 2005), descrita a seguir: Transcritos que mostraram

pouca ou nenhuma variação entre as amostras foram eliminados, restando-se 1803 ncRNAs com coeficiente de variação $> 0,03$. O conjunto de treinamento (de tamanho $N = 42$ pacientes) foi dividido usando uma abordagem de amostragem em 500 conjuntos aleatórios de treinamento (cada um com tamanho n), tendo $n/2$ pacientes de cada desfecho, e 500 conjuntos de validação associados (de tamanho $N-n$), conforme descrito (Michiels *et al.*, 2005). Em nossa análise, $n = 10, 14, 18, 22, 26$ ou 30 . O número n máximo de pacientes no conjunto de treinamento ($n=30$) foi determinado de maneira que existisse no mínimo 1 paciente de cada desfecho no conjunto (de tamanho $N-n=12$) de validação; ou seja, como temos 16 amostras de pacientes com recidiva, teremos o número máximo $n/2=15$ (ou seja, máximo $n=30$), situação em que 15 pacientes com recidiva estão no conjunto de treinamento de $n=30$, e sobra 1 paciente com recidiva para o conjunto de validação, junto com mais 11 pacientes sem recidiva (para um conjunto de validação de 12). A assinatura preditora de cada um dos conjuntos de treinamento foi definida como a lista dos 100 ncRNAs que foram mais altamente correlacionados com o desfecho dos pacientes, mostrado pelo coeficiente de correlação de Pearson. Os perfis preditivos de alto ou baixo risco foram obtidos tomando-se o valor de expressão médio de cada um dos 100 ncRNAs entre os $n/2$ pacientes de cada desfecho. Em seguida, a classificação dos pacientes nos conjuntos de validação correspondentes (cada um com $N-n$ pacientes) foi realizada pela regra de predição do centroide mais próximo (Michiels *et al.*, 2005), isto é, cada paciente foi colocado no grupo de alto ou baixo risco de acordo com o perfil médio preditivo com o qual a amostra tinha maior correlação. A proporção de classificação errada de cada conjunto de validação associado foi determinada e a média da taxa de erro de classificação foi calculada, bem como o intervalo de confiança de 95%. Os genes discriminantes foram definidos como aqueles inclusos em mais de 50% das 500 assinaturas moleculares (Michiels *et al.*, 2005); para $n=30$, um conjunto de 51 ncRNAs discriminantes foi obtido. Os valores de expressão médio destes últimos 51

ncRNAs no conjunto de pacientes com recorrência dentro do conjunto de treinamento de 42 amostras foi usada para calcular um perfil de mau prognóstico. Este perfil final médio foi usado para classificar o conjunto independente de 22 amostras, descrito a seguir.

3.11 Classificação de um conjunto independente usando o perfil de mau prognóstico.

Um conjunto independente composto por 22 amostras foi usado para avaliar a exatidão do classificador de prognóstico. Medidas brutas de expressão do conjunto teste foram processadas pela normalização de quantil, seguida do ajuste de lotes.

3.12 Processamento das amostras do conjunto teste

Uma vez criados os perfis de classificação com as amostras de treinamento, estes perfis foram aplicados às amostras independentes do conjunto externo de teste. Estas amostras não foram processadas juntamente com as amostras de treinamento. Assim, a fim de que as amostras de validação fossem comparáveis com as de treinamento, foi necessária a criação de um protocolo de processamento destas amostras que fosse composto dos mesmos passos de processamento do conjunto de treinamento (normalização e ajuste de lotes), mas sem alterar os valores já calculados dos parâmetros obtidos no conjunto de treinamento.

Como descrito anteriormente, existem passos do processamento dos valores de intensidade que utilizam como referência os valores médios de expressão ao longo do conjunto de amostras analisado. Se o conjunto de amostras é alterado, todos os valores são alterados também. Assim, foi necessário que, para cada passo, as amostras de teste se submetessem ao mesmo processamento das amostras de treinamento, mas que não fossem utilizadas como referência para os fatores de transformação. A seguir estão descritos os passos do processamento que envolvem essas transformações e como cada um deles foi realizado.

3.12.1 Normalização por quantil.

Inicialmente os valores de intensidade de cada *spot* em cada lâmina são ordenados (*ranked*) segundo o valor de intensidade dentro da lâmina.

Estes valores de cada lâmina são substituídos pela média de intensidade obtida para determinada posição no ranking, valor este que havia sido obtido no conjunto de treinamento, conforme discutido anteriormente. Para o processamento das amostras de teste, houve apenas a substituição dos valores de cada gene em sua posição no *ranking* naquela lâmina por aquele valor médio daquela posição no *ranking*, obtido com o conjunto de treinamento, sem um novo cálculo de médias. Desta maneira, o conjunto de valores das amostras de teste se torna o mesmo do das amostras de treinamento.

3.12.2. Ajuste de lotes (ComBat).

Para retirar o efeito de lote das amostras de teste, cada um dos 5 lotes deste conjunto foi processado separadamente, sendo cada um processado junto com os 9 lotes do conjunto de treinamento. Desta maneira, o efeito de lote é corrigido, sem alterar os valores das amostras de treinamento que já foram processadas. Além disso, o processamento de um lote de teste por vez, frente aos lotes de treinamento, permite que novos lotes de amostras sejam processados no futuro, sem alterar o valor das amostras de treinamento e de teste já processadas.

3.12.3 Normalização intergênica (*Z-score*).

A normalização intergênica é usada para comparar os genes entre as amostras. Nesta normalização, os valores de intensidade são considerados como desvios acima ou abaixo da média daquele gene entre todas as amostras.

Neste caso, as amostras de teste foram subtraídas pelas médias e divididas pelos desvios previamente gerados com o conjunto de treinamento, mais uma vez, portanto, sofrendo a transformação necessária sem serem utilizadas para gerar os valores de treinamento referência (médias e desvios).

3.13 Classificação das amostras do conjunto teste

Finalmente, as amostras do conjunto teste foram classificadas calculando-se o coeficiente de correlação entre cada amostra e o perfil médio de mau prognóstico obtido com o conjunto de treinamento. Uma amostra teste foi designada como sendo do grupo de alto ou baixo risco de recorrência quando o coeficiente de correlação calculado era ou positivo ou negativo, respectivamente.

3.14 Análise de sobrevida

A análise de Kaplan-Meier para o conjunto independente de teste foi feita utilizando o programa GraphPad Prism (versão 5.0) e os dados de seguimento clínico de cada paciente do conjunto.

3.15 Anotação funcional dos transcritos identificados

Anotações gênicas da função dos genes de interesse e de cruzamento das listas de genes identificados com outros trabalhos de expressão gênica em câncer de próstata foram feitas utilizando-se a ferramenta *Ingenuity Pathways Analysis* (Ingenuity Systems) (Calvano *et al.*, 2005).

3.15.1 Análises de redes gênicas

Os números de acesso dos genes codificadores expressos, bem como os dados de *fold change* e a taxa de falsa descoberta (FDR), gerados a partir da análise de SAM entre pacientes com recorrência e sem recorrência, foram importados no *software Ingenuity Pathway Analysis 7.1*. A análise foi feita com FDR <5% como critério de corte para genes diferencialmente expressos. Cada identificador (número de acesso do gene) foi associado aos dados daquele gene na base de conhecimento do programa. Estes genes foram usados como ponto inicial para a geração de redes biológicas. A análise de redes biológicas foca nas relações funcionais entre genes já conhecidas e documentadas na base de dados, olhando para os genes que foram considerados significativamente alterados. O programa cria redes destes genes com funções similares e interações já reportadas. Um *score* é computado para cada rede de acordo com o número de genes considerados diferencialmente expressos que estão presentes na rede. Este *score* reflete o logaritmo negativo, em base 10 do p-valor que indica a chance dos genes diferencialmente expressos encontrados juntos na rede gênica sejam devido ao acaso. *Scores* de 3 ou mais indicam 99,9% de nível de confiança que a rede não foi gerada por acaso. Este *score* foi utilizado como critério para identificar redes gênicas enriquecidas com genes de expressão alterada na condição estudada (recidiva).

4 Resultados

4.1 Amostras de tecido tumoral prostático e obtenção de RNA para experimentos de hibridização

Todas as amostras coletadas foram enviadas ao nosso laboratório, congeladas em gelo seco. As amostras foram então processadas para extração de seu RNA total utilizando-se o método do tiocianato de guanidina/fenol/clorofórmio (ver seção Material e Métodos).

Após este procedimento seguiu-se avaliação da quantidade, pureza e integridade do RNA total obtido através de espectrofotometria (NanoDrop) e eletroforese capilar (Bioanalyzer 2100), respectivamente. A Figura 7 apresenta o resultado típico da análise por eletroforese capilar de RNA total isolado de amostras de tecido tumoral de próstata processadas em nosso laboratório.

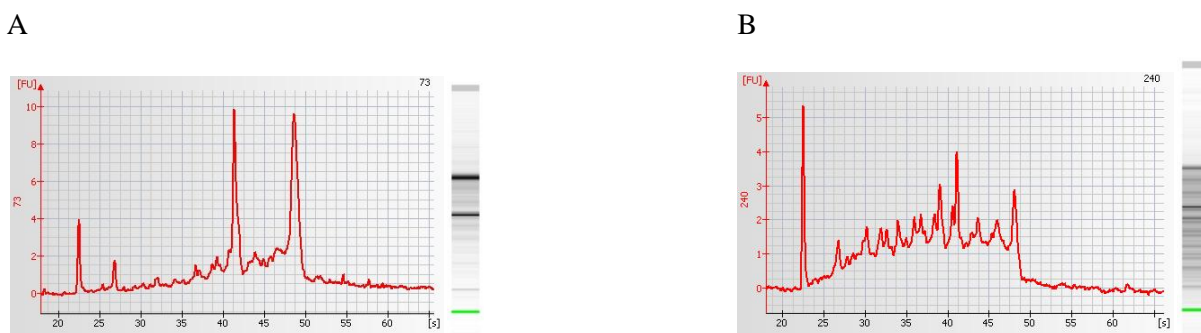


Figura 7: Perfil analítico típico de uma amostra de RNA total tratado com DNase I obtido através de eletroforese no Agilent 2100 Bioanalyzer (Agilent Technologies). Os painéis apresentam o eletroferograma obtido a partir de 100 ng de RNA total isolado de tecido A: Perfil de uma amostra de boa qualidade (RIN 7,9 e razão 28S/18S 1,2). Note as duas bandas mais fortes (RNA ribossomais 18S e 28S) representadas pelos picos em 41 e 49 s. B: Perfil de uma amostra de má qualidade (RIN 4,8 e razão 29S/18S 0,7) com as bandas de RNA ribossomais já degradadas.

O conjunto de amostras de tecido tumoral prostático de pacientes utilizados neste trabalho é fornecido pelo banco de tumores do Laboratório de Patologia Celular e Molecular – Hospital Sírio-Libanês. A previsão inicial para a primeira etapa era a realização de experimentos de hibridização de um conjunto com 61 amostras, recebidas em 2005 por nosso laboratório. Estas amostras tiveram o RNA total extraído e avaliado em relação à quantidade

e qualidade pela aluna de doutorado Camila Egidio. Entretanto, após a extração de RNA dessas amostras, verificou-se que das 61 amostras recebidas, 29 amostras ($29/61 = 50\%$) não forneceram RNA de qualidade suficiente para os experimentos de hibridização, isto é, com o valor de *RNA Integrity Number* (RIN) acima de 7 e a razão das bandas ribossomais 28S/18S maior ou igual a 2 (Tabela 2). Assim, foi necessária a retirada de novas amostras do Banco de Tumores do Hospital Sírio-Libanês, com as quais fizemos novas extrações de RNA.

Tabela 2: Amostras recebidas do Banco de Tumores do Hospital Sírio-Libanês.

	Total	Utilizadas para experimentos de microarranjos	Com RNA degradado (RIN < 7)
Amostras recebidas em 2005	61	32	29
Amostras recebidas em 2006 e 2007	51	32	18
Total	112	64	47

Das 51 novas amostras recebidas, 18 (35%) resultaram em RNA de baixa qualidade e não foram utilizados para amplificação e hibridização (Tabela 2). Esta porcentagem de perda já era prevista devido ao longo tempo de armazenamento dessas amostras, e pelas eventuais perdas durante o processo de extração.

Fizemos uma importante modificação na rotina de processamento das amostras, para as amostras recebidas em 2007. As amostras de RNA iniciais do conjunto de 62 amostras haviam sido extraídas e armazenadas durante um longo intervalo de tempo até a utilização para os experimentos de microarranjos. Verificamos que esta é uma rotina inadequada, pois o tempo entre a extração do RNA e seu processamento final é muito grande, o que poderia acarretar em uma eventual degradação do RNA. Optou-se então, pelo processamento completo (do recebimento das amostras à hibridização e análise dos dados) de

conjuntos de seis a dez amostras por ensaio, o máximo a ser processado em conjunto, uma vez que contamos com seis câmaras de hibridização. Além disso, em cada lote de marcação e hibridização incluímos amostras com e sem recidiva, de forma balanceada. Assim, constatando-se que alguma amostra de um grupo específico estava degradada, o desbalanceamento deste poderia ser compensado numa hibridização subsequente.

4.1.1 Dados clínico-patológicos dos pacientes utilizados no estudo.

Estes pacientes formam uma representação ampla de Grau de Gleason, idade na cirurgia e níveis de PSA pré-operatorio (Tabela 3 e Tabela 4). Nenhum paciente recebeu tratamento adjuvante após a prostatectomia radical. A recorrência bioquímica foi definida como um aumento nos níveis de PSA sanguíneo para $\geq 0,2$ ng/mL em qualquer momento durante o acompanhamento de 5 anos. O tratamento adjuvante foi administrado somente após a constatação da recorrência.

Tabela 3: Dados clínicos e anatomo-patológicos para cada um dos 42 pacientes usados no conjunto de treinamento.

No. no Banco	Recorrência em 5 anos de acompanhamento	Idade (anos)	Gleason Score	PSA pré-operatório	Estágio clínico	Volume tumoral (mm³)	Envolvimento da cápsula	Envolvimento extra-capsular	Invasão da vesícula seminal	Metástase linfonodal	Estágio patológico	Data da cirurgia	Acompanhamento (meses)	PSA na recorrência	Observações
50	-	54	7	5,5	T1c	7	-	-	-	-	pT2	19/03/1998	60		
54	-	65	5	14,6	T2b	7	+	-	-	-	pT2	30/03/1998	60		
62	-	70	7	8	T2a	34	+	+	-	-	pT3	16/04/1998	60		
102	+	60	5	37	T1c	34	-	-	-	-	pT2	28/07/1998	17,0	0,4	
108	+	76	5	9,3	T2b	14	+	+	-	-	pT3	18/08/1998	47,4	0,6	
125	-	70	5	3	T1c	4,8	-	-	-	-	pT2	03/10/1998	60		
131	+	69	9	7,69	T2a	13	+	-	-	-	pT2	22/10/1998	43,2	0,3	
183	-	67	9	7	T1c	10	+	-	-	-	pT2	02/02/1999	60		
195	+	60	5	6,9	T1c	23	+	-	-	-	pT2	27/02/1999	60	0,5	
201	+	60	8	11,3	T2b	15	+	+	-	-	pT3	16/03/1999	43,4	1,8	
202	+	53	6	6,2	T2a	5	+	+	-	-	pT3	18/03/1999	49,4	0,5	
221	-	50	5	3,7	T2a		-	-	-	-	pT2	11/05/1999	60		
223	-	58	8	11	T2b	14	+	+	-	-	pT3	11/05/1999	60		
240	+	67	7	22	T2a	4	-	-	-	-	pT2	19/06/1999	18,3	0,5	
250	-	65	7	6,8	T2a	10	+	+	-	-	pT3	06/07/1999	60		
254	+	68	8	5,6	T1c	4	+	+	-	-	pT3	08/07/1999	41,7	1,1	
261	-	75	8	13,6	T2a	18	+	+	-	-	pT3	03/08/1999	30		Morreu de outras causas
273	+	58	7	8,3	T1c	12	+	-	-	-	pT2	17/08/1999	42,3	0,5	
277	-	57	8	18,1	T1c	8	-	-	-	-	pT2	24/08/1999	60		
284	+	84	7	4,8	T2a	17	+	+	-	-	pT3	09/11/1999	18,7	0,4	
293	-	72	8	23	T2a	10	+	+	-	-	pT3	09/09/1999	60		
297	-	49	6	10,2	T2c	11,8	+	-	-	-	pT2	25/09/1999	60		
303	+	71	8	12	T1c	3	+	+	-	-	pT3	07/10/1999	50,7	1,5	
320	+	51	5	7,3	T1c	20	+	+	-	-	pT3	18/11/1999	24,8	0,3	
321	-	60	7	4,2	T2a	12	-	-	-	-	pT2	23/04/1999	60		

No. no Banco	Recorrência em 5 anos de acompanhamento	Idade (anos)	Gleason Score	PSA pré-operatório	Estágio clínico	Volume tumoral (mm³)	Envolvimento da cápsula	Envolvimento extra-capsular	Invasão da vesícula seminal	Metástase linfonodal	Estágio patológico	Data da cirurgia	Acompanhamento (meses)	PSA na recorrência	Observações
334	-	73	5	6,2	T2a	14	-	-	-	-	pT2	07/12/1999	60		
339	-	55	7	6,5	T1c	14	-	-	-	-	pT2	04/01/2000	60		
340	+	76	7	12	T2b	15	+	+	-	-	pT3	04/01/2000	20,5	4,8	
385	-	57	7	10	T1c	6	+	-	-	-	pT2	08/04/2000	47,3		
421	+	67	8	31	T2b	42	+	+	+	-	pT3	04/07/2000	2,8	0,3	
422	-	60	7	21	T2b	2	-	-	-	-	pT2	02/07/2000	60		
425	-	74	5	2,1	T1c	4	-	-	-	-	pT2	15/07/2000	41,5		
426	+	55	6	5,2	T1c	5	+	-	-	-	pT2	15/07/2000	27	0,2	
437	-	66	6	12	T2a	1	-	-	-	-	pT2	15/08/2000	60		
453	-	65	8			10	+	+	-	-	pT3	23/05/2001	60		
454	-	55	6	5,2	T1c	12	-	-	-	-	pT2	23/05/2001	47,6		
455	-	63	7	2		2	-	-	-	-	pT2	12/08/2001	60		
459	-	50	8			29	+	+	-	-	pT3	16/06/2001	60		
461	-	74	7	6,8	T1c	16	-	-	-	-	pT2	30/06/2001	60		
463	-	79	8			64	+	+	+	-	pT3	04/07/2001	60		
469	+	71	8	5,4	T1c	5	+	+	-	-	pT3	17/07/2001	20,7	0,26	
471	-	64	6	18,2	T1c	40	+	-	+	-	pT3	25/07/2001	60		

Tabela 4: Dados clínicos e anatomo-patológicos para cada um dos 22 pacientes usados no conjunto teste.

No. no Banco	Recorrência em 5 anos de acompanhamento	Idade (anos)	Gleason Score	PSA pré-operatório	Estágio clínico	Volume tumoral (mm ³)	Envolvimento da cápsula	Envolvimento extra-capsular	Invasão da vesícula seminal	Metástase linfonodal	Estágio patológico	Data da cirurgia	Acompanhamento (meses)	PSA na recorrência	Observações
28	-	47	4	3,1	T1c	47	-	-	-	-	pT2	17/01/1998	60		
35	-	64	4	12	T2c	2	-	-	-	-	pT2	03/02/1998	37,4		Morreu de outras causas
36	-	63	4	6,2	T2a	39	+	-	-	-	pT2	02/10/1998	51,8		
55	-	72	8	8,1	T2a	26	+	+	-	-	pT3	31/03/1998	60		
68	+	58	8	6,2	T2a	27	+	+	+	-	pT3	24/04/1998	33,133	0,5	
71	-	65	7	7	T2c	23	+	+	-	-	pT3	05/12/1998	60		
73	-	61	5	5,6	T1c	8	-	-	-	-	pT2	14/05/1998	60		
75	-	74	5	19,1	T2a	24	+	+	-	-	pT3	05/01/1998	60		
84	-	54	8	7,4	T2c	36	+	-	-	-	pT2	16/06/1998	60		
100	-	66	5	17,9	T2b	17	+	-	-	-	pT2	23/07/1998	60		
121	+	52	6	7,8	T1c	4	-	-	-	-	pT2	22/09/1998	56	0,2	
122	+	74	6	18,9	T2c	43	+	+	-	-	pT3	29/09/1998	56,8	0,7	
148	+	65	9	37	T1c	38	+	-	-	-	pT2	17/11/1998	17,7	0,4	
155	+	58	5	10	T2a	16	-	-	-	-	pT2	05/12/1998	54,6	0,2	
171	-	52	8	12,2	T1c	1,7	-	-	-	-	pT2	09/01/1999	59,5		
176	-	44	8	8,6	T2a	5	+	+	-	-	pT3	16/01/1999	54,2		
182	-	64	9	21	T2a	11	+	-	-	-	pT2	28/01/1999	53,8		
187	-	71	8	15,5	T2b	13	-	-	-	-	pT2	05/02/1999	58,6		
189	-	60	7	13,8	T1c	1,4	-	-	-	-	pT2	09/02/1999	54,4		
212	-	62	8	6,7	T2b	27	-	-	-	-	pT2	04/10/1999	60		
217	-	74	8	24	T2b	5	+	+	-	-	pT3	22/04/1999	28,267		Morreu de outras causas
295	+	66	7	12,6	T2b	17,4	+	+	-	-	pT3	23/09/1999	52	6	

A idade média apresentada pelo grupo de pacientes no momento do diagnóstico é de 63 anos, ligeiramente menor que a média apresentada pelos pacientes com câncer de próstata nos EUA (Scosyrev *et al.*, 2010) (Tabela 5). Enquanto que mais da metade dos pacientes nos EUA (60%) são diagnosticados com mais de 65 anos, sendo quase um quarto deles com mais de 75 anos, a maior parte dos pacientes utilizados neste estudo foi diagnosticada com câncer com menos de 65 anos (Tabela 5).

O nível de PSA no sangue no momento do diagnóstico para os pacientes do estudo é maior que o observado na população americana (11 ng/mL contra 6 ng/mL). O Score de Gleason também é maior para os pacientes do presente estudo em relação à população americana: enquanto apenas 14% dos pacientes americanos foram diagnosticados com GS maior que 7, esta proporção para os pacientes do presente estudo é de 37,5% (Tabela 5). Estes últimos dois aspectos podem indicar uma menor prevenção da doença de câncer de próstata no Brasil, quando comparado aos EUA já que os tumores são detectados com estágio mais avançado.

Tabela 5: Comparação da proporção das características clínico-patológicas dos pacientes usados no estudo com as características observadas na população americana (%).

Característica	População Americana 2004-2006 (ACS, 2010; Scosyrev <i>et al.</i> , 2010)	Conjunto estudado (%) n=64
Idade		
Média	67	63
18-64	40	51,6
65-74	35,3	32,8
75+	24,8	7,8
PSA, ng/mL		
%Média	6,5	11
≤2,5	7,0	3,1
2,6-4	6,7	4,7
4,1-6,9	32,5	25,0
7-10	15,7	17,2
10,1-20	12,9	28,1
>20	10,9	12,5
Desconhecido	14,3	4,7
Gleason Score		
2-6	46,3	37,5
7	33,0	25,0
8-10	14,2	37,5
Desconhecido	6,5	0,0
Estágio Tumoral		
T1	52,2	37,5
T2	39,5	56,3
T3	2,3	0,0
T4	1	0,0

4.2. Amplificação

As 64 amostras de RNA total foram amplificadas e marcadas com o kit “*Low RNA Input Fluorescent Linear Amplification kit*” (Agilent), seguindo o protocolo sugerido pelo fabricante (Figura 6). Foi utilizado um método indireto de estudo, com uma amostra por microarranjo, marcada com ambos fluoróforos. Já foi demonstrado que esta estratégia fornece uma alternativa confiável de experimentos baseados em hibridização, e que se constitui em uma opção adequada para análise de grandes quantidades de amostras (Peixoto *et al.*, 2006).

Cada amostra de RNA foi amplificada e marcada tanto com Cy3-CTP e Cy5-CTP independentemente. Controles positivos de RNA ("*spikes*") foram adicionados à reação de amplificação.

No início das amplificações e marcações observamos quantidades inferiores de cRNA amplificado em comparação ao esperado para o kit, por volta de 5 µg de cRNA (Tabela 6).

Tabela 6: Quantidade de cRNA amplificado e incorporação de fluoróforos para as primeiras 31 amostras processadas.

Nº BANCO	Total Cy5 RNA Amplificado (µg)	Total RNA amplificado (µg)	Incorporacao Cy5 (pmol Dye/µmol RNA)	Incorporação Cy3 (pmol Dye/µmol RNA)
28	2,22	2,951	11,64	4,62
35	2,254	2,001	8,19	6,17
36	1,043	1,216	4,92	5,21
50	4,534	4,009	12,7	19,83
54	2,143	1,708	10,98	8,61
55	1,683	2,006	6,08	20,28
68	5,07	4,683	17,16	9,11
71	2,788	2,539	13,01	25,82
73	4,4	4,056	12,1	25,29
75	2,385	2,999	12,49	12,35
84	0,84	1,098	5,16	5,87
100	1,338	1,324	20,92	19,38
102	1,239	1,283	13,4	9,46
108	0,915	1,446	13,62	6,02
121	0,922	0,798	5,95	14,28
122	1,372	0,966	6,31	10,6
125	1,217	1,282	12,01	5,6
148	1,95	1,313	13,14	12,72
155	0,964	1,282	13,19	9,72
171	3,684	3,963	13,15	7,24
176	0,658	0,675	7,51	15,97
182	1,334	1,534	11,19	4,9
183	1,341	1,077	11,79	11,74
187	1,359	1,864	11,39	3,96
189	2,084	2,387	13,56	15,38
212	2,052	2,004	14,89	11,96
217	1,603	1,062	13,41	8,22
223	1,822	0,895	9,26	15,09
295	2,141	1,927	13,92	11,83
320	9,229	10,552	14,8	13,11
421	3,897	4,24	4,19	19,16

A fim de verificar quais fatores estavam interferindo na variação da quantidade de cRNA amplificado e marcado, foram calculadas correlações entre os valores da qualidade de RNA, ou seja, a razão das bandas ribossomais 28S e 18S e o RIN e os resultados obtidos para as ampliações do RNA, ou seja, a quantidade de cRNA total obtido e a incorporação do fluoróforo (em pmol de fluoróforo incorporado por microgramas de cRNA), como mostra a Tabela 7. Conforme esperado, existe uma correlação média (0,5) entre a qualidade do RNA inicial e a quantidade de cRNA obtido. Já a quantidade de fluoróforo incorporado parece não estar correlacionada com a qualidade do RNA inicial.

Tabela 7: Correlação entre a qualidade do RNA inicial (razões 28S/18s e RIN) e os resultados obtidos após a amplificação (cRNA amplificado e incorporação de fluoróforo), indicando uma correlação média ente a qualidade de RNA inicial e o total de cRNA obtido após a amplificação.

Critério de Qualidade	Correlação com Quantidade cRNA total obtido	Correlação com a Incorporação de Fluoróforo
28/18S	0,503747	0,275898
RIN	0,478107	0,062324

Esses resultados indicam a necessidade de utilização de RNA inicial de ótima qualidade (RIN >7 e razão 28S/18S > 1) para a amplificação.

4.3 Desenho do oligoarranjo Íntron-Éxon

Em 2007 nós identificamos sítios transcricionalmente ativos de mensagens não codificadoras totalmente intrônicas (TIN), ou parcialmente intrônicas (PIN) dentro de genes codificadores de proteínas (Nakaya *et al.*, 2007). Baseados nessa informação, nós desenhamos um microarranjo de oligonucleotídeos com 44 mil sequencias combinando aleatoriamente genes codificadores de proteínas com os transcritos intrônicos correspondentes (Figura 8).

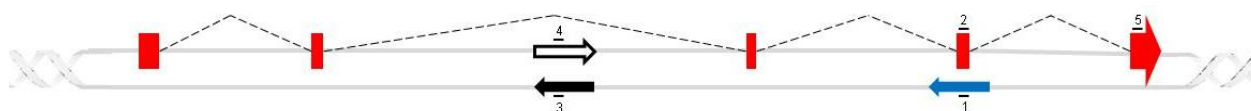


Figura 8: Desenho do microarranjo de oligo-nucleotídeos íntron-éxon. Sonda 5: sondas 60-mero disponíveis comercialmente pela empresa Agilent, representando um éxon de gene conhecido codificador de proteínas; Sondas 3 e 4: pares de sondas reverso-complementares para cada uma das duas possíveis fitas dos transcritos totalmente intrônicos; Sonda 1: sondas para transcritos parcialmente intrônicos antissenso. Sonda 2: sonda representando cada éxon que se sobrepõe, na fita oposta, a cada transcrito parcialmente intrônico.

Isto permitiu uma detecção em larga escala da expressão intrônica, de maneira gene-orientada. Um total de 8.780 sondas do conjunto de sondas 60-meros disponíveis comercialmente pela empresa Agilent foram usados (Figura 8, sonda 5), representando diferentes éxons, em 6.954 genes codificadores de proteínas únicos (Figura 8, sonda 2). Um par de sondas reverso-complementares para cada um dos 7.135 transcritos totalmente intrônicos também foi desenhada (Figura 8, sondas 3 e 4), de maneira a detectar independentemente a transcrição em um dado *locus*. Sondas para 4.439 transcritos parcialmente intrônicos antissenso (Figura 8, sonda 1) também foram desenhadas. Uma sonda representando cada éxon que se sobrepõe a cada transcrito parcialmente intrônico, na fita oposta, também foi desenhada (Figura 8, sonda 2).

4.4 Hibridização e obtenção das imagens.

Cada lâmina foi hibridizada com uma amostra, marcada tanto com Cy5 quanto com Cy3. Esta abordagem de auto-hibridização usada aqui tem sido descrita na literatura para classificar genes como diferencialmente expressos em experimentos que utilizam os valores de intensidades absolutos de cada amostra (Yang *et al.*, 2002; Papini-Terzi *et al.*, 2005). Existe na literatura uma discussão extensa, que compara as análises que usam diretamente as intensidades (como feito neste trabalho) com as que usam a relação de intensidades entre a expressão da amostra em estudo e uma referência externa; esta comparação tem mostrado um desempenho superior da estratégia que usa diretamente as intensidades (Koide *et al.*, 2004; Papini-Terzi *et al.*, 2005). Os processos de normalização aplicados aqui usam como referência os valores de intensidade do conjunto inteiro de amostras usadas na análise, ao invés de usar um pool de RNA referência de todas as amostras (Bolstad *et al.*, 2003; Johnson *et al.*, 2007). Esta abordagem é particularmente apropriada quando novas amostras serão adicionadas à análise no futuro; a abordagem exclui o pré estabelecimento de um pool pré determinado de RNA no início dos experimentos, que podem variar ao longo do tempo de trabalho.

Depois da hibridização, as lâminas foram lavadas com soluções específicas recomendadas pelo kit de hibridização, e os valores de intensidade de cada *spot* do microarranjo foram obtidos através da leitura da lâmina (GenePix – Molecular Devices). Uma imagem típica do microarranjo pode ser vista na Figura 9.

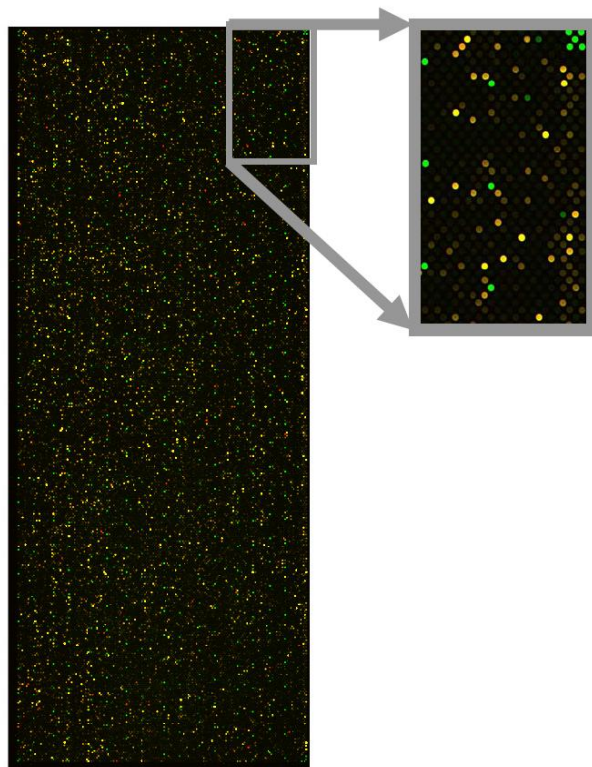


Figura 9: Amostra de RNA de tumor de próstata marcada com Cy3 (verde) e Cy5 (vermelho) e hibridizada com a lâmina 37k intron-éxon da Agilent. Note que, como esperado, a maior parte dos *spots* apresenta uma coloração amarela. Alguns controles positivos da hibridização estão em verde. *Spots* para os controles positivos adicionados à amplificação (*Spikes*) estão em vermelho. A ampliação mostra uma região do *chip* em maior detalhe.

No início do projeto as imagens eram processadas com o *software* ArrayVision® (GE Healthcare). Este *software* já estava disponível em nosso laboratório e é amplamente utilizado no processamento de imagens do nosso microarranjo doméstico de quatro mil sondas de cDNA. Em 2007, foi possível a aquisição do *software* Feature Extraction (Agilent Technologies), desenvolvido especificamente para o processamento de imagens dos oligoarranjos Agilent, e passamos a utilizá-lo para a extração dos dados. Este *software*, além da extração dos valores de intensidade de cada *spot*, possui fatores de correção específicos para as lâminas Agilent. Nós observamos que, para *spots* controles, havia um valor de intensidade maior de um dado controle no centro da lâmina quando comparado com o mesmo controle presente na periferia da lâmina. Este efeito é denominado “efeito de domo”, e é produzido pela bolha de hibridização, característica do método de hibridização Agilent. É

deixada uma pequena bolha de ar na câmara, que percorre a borda da lâmina durante a hibridização, enquanto a lâmina é girada no hibridizador rotatório. Assim, a periferia da lâmina passa por momentos sem a solução (presença da bolha), ao passo que o centro da lâmina está sempre em contato com a solução. Além deste efeito, o *software* FE também corrige diferenças de sensibilidade de detecção com dois fluoróforos diferentes (Cy3 é detectado pelo scanner com sensibilidade maior que Cy5, para baixas intensidades). Antes do uso do programa FE, o ajuste para o “efeito domo” não era executado, e este último ajuste, relacionado a sensibilidade diferencial de detecção entre os dois corantes era feito por programas específicos após a extração de dados.

4.5 Análise dos dados de expressão.

4.5.1 Geração do Grupo de Treinamento e do Grupo Teste

Inicialmente, antes de qualquer processamento dos valores brutos, as 64 amostras foram divididas nos dois conjuntos de análise: I, um conjunto de treinamento, contendo 42 amostras (Tabela 3), sendo 16 com recidiva (38% do conjunto de treinamento) e 26 sem recidiva (62% do conjunto de treinamento), determinado pelo período de acompanhamento de 5 anos. Estas amostras foram utilizadas para todas as etapas de processamento e de análise de significância; e II, Um conjunto independente de teste, com 22 amostras (Tabela 4), sendo 16 sem recidiva e 6 com recidiva. Estas amostras de teste foram totalmente desconsideradas em todo o processo envolvendo a identificação dos genes diferencialmente expressos, e só foram utilizadas em uma classificação cega para testar o perfil prognóstico, como descrito mais adiante. Os dados descritos a seguir se referem às amostras do conjunto de treinamento.

4.5.2 Identificação dos transcritos expressos

O passo seguinte à extração dos valores foi a filtragem dos *spots* considerados “apagados”, ou seja, cujos níveis de intensidade são devidos à hibridização inespecífica, não

indicando a presença daquele transcrito na amostra. O critério utilizado para a definição de *spot* “apagado” foi o *flag* produzido pelo próprio *software* de extração FE. Este *flag* é dado segundo um teste t para identificar se o valor de intensidade obtido no *spot* está significativamente ($p < 0.05$) acima dos valores de *background* local (regiões ao redor dos *spots*). Foram considerados apenas *spots* positivos em pelo menos 90% das amostras do grupo com ou do grupo sem recidiva, não excluindo, portanto, possíveis genes presentes (expressos) em uma condição e ausentes (totalmente reprimidos) em outra. Além disso, foram retirados da análise todos os *spots* controle e os *spots* que foram excluídos após sucessivas revisões dos desenhos das sondas encontrados na lâmina (estas revisões são feitas para reanotar os transcritos intrônicos, contemplando o mapeamento dos transcritos nos íntrons de novos genes, recentemente identificados). Com estes critérios, foram utilizados 19195 *spots* do total de 31928 presentes na lâmina.

Primeiramente analisamos o número de sondas de genes codificadores de proteína que indicaram expressão. Identificamos um total de 13219 sondas que indicaram expressão de genes codificadores, sendo que destas, 8780 são sondas comerciais, desenhadas pela Agilent (chamadas por nós de “Exonic”) e 3354 sondas exônicas por nós desenhadas com complementaridade aos transcritos intrônicos anti-senso (chamadas por nós de “Exonic PIN-target”). O número de sondas que indicam a expressão é maior que o número de genes interrogados, pois o oligoarranjo foi desenhado com mais de uma sonda interrogando o mesmo gene codificador de proteína. Retirando-se a redundância, estas sondas indicam a expressão de 6611 genes codificadores de proteína únicos, do total de 6945 genes únicos interrogados no oligoarranjo (Figura 10).

Em seguida analisamos presença de expressão com sondas representando transcritos intrônicos. Nós identificamos a expressão de 2082 transcritos de regiões parcialmente complementares a éxon (*Antisense Partially Intronic*, AS-PIN), ou seja, 63,12%

do total presente no oligoarranjo (Figura 10). Em relação às regiões totalmente intrônicas (TIN), nós identificamos 3159 transcritos na fita senso de regiões totalmente intrônicas (*Sense Totally Intronic Transcript – S-TIN*), ou seja, 44,27% do total de transcritos interrogados no oligoarranjo, e 2718 em regiões intrônicas transcritas na fita antissenso (*Antisense Totally Intronic Transcript – AS-TIN*), ou seja, 38,09% do total de transcritos interrogados no oligoarranjo (Figura 10).

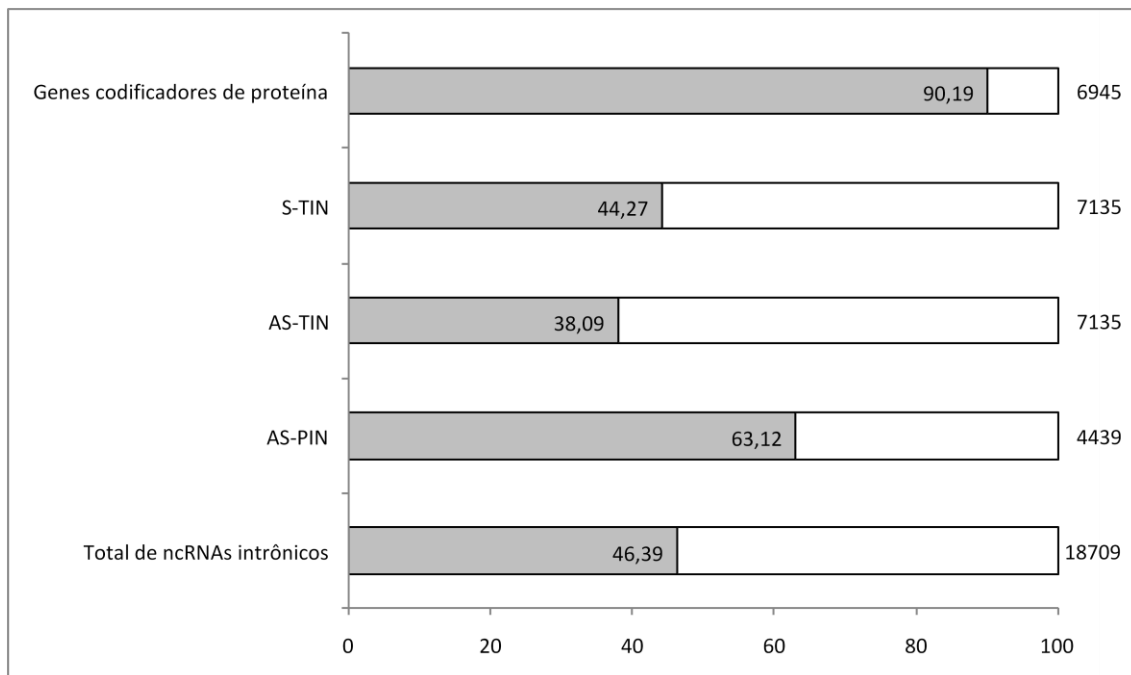
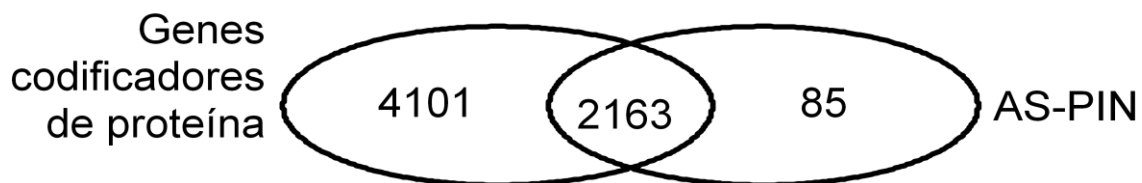


Figura 10: Expressão intrônica e exônica em PC. Número de transcritos expressos em próstata para as 42 amostras do conjunto de treinamento. Na extrema direita do gráfico encontram-se os totais de sondas presentes no oligoarranjo para cada categoria de transcrito: AS-PIN = Antissenso Parcialmente Intrônico; AS-TIN = Antissenso Totalmente Intrônico; S-TIN = Senso Totalmente intrônico. Cinza: Porcentagem de sondas que apresentaram expressão em pelo menos 23/26 (90%) das amostras sem recidiva ou 14/16 (90%) das amostras com recidiva.

Em seguida, nós verificamos se a expressão intrônica está correlacionada com a transcrição exônica da região correspondente, isto é, se os transcritos intrônicos estão presentes sempre de maneira concomitante à expressão do gene codificador de proteína do mesmo locus gênico. A Figura 11 mostra um diagrama de Venn com sobreposição da expressão dos transcritos intrônicos e dos transcritos exônicos correspondentes. De maneira

interessante, 35% do total de genes exônicos expressos apresentaram também a expressão antissenso complementar (AS-PIN). Do total de 6349 regiões gênicas analisadas, 4101 apresentaram expressão somente do gene codificador de proteína e 2163 apresentaram expressão tanto exônica quanto de AS-PIN. Foram identificados 85 genes que apresentaram a expressão do transcrito antissenso AS-PIN sem a expressão do gene codificador de proteína do locus correspondente, o que sugere que a expressão de AS-PIN independente do gene codificador de mesmo *locus*.



Total de genes na união: 6349

Figura 11: Diagrama de Venn com a sobreposição da transcrição intrônica complementar aos genes exônicos expressos, do total de 6363 genes.

Um padrão semelhante foi observado para os transcritos de regiões totalmente intrônicas (TIN), tanto na fita senso (S-TIN) quanto na fita antissenso (AS-TIN) como mostra o diagrama de Venn da Figura 12. Dos 6264 genes codificadores de proteína expressos, 682 apresentaram também a expressão de regiões totalmente intrônicas senso, 503 com a expressão da fita intrônica antissenso, e 1253 (20%) tiveram a expressão de regiões totalmente intrônica tanto senso quanto antissenso. De maneira interessante, 75 regiões gênicas apresentaram somente expressão na região intrônica senso (S-TINs) e 79 genes tiveram a expressão da região intrônica antissenso (AS-TINs), sem a expressão da fita intrônica complementar e do gene exônico correspondente (Figura 12). Estes resultados demonstram que, para os transcritos intrônicos interrogados do oligoarranjo, a expressão de apenas uma das duas fitas é rara.

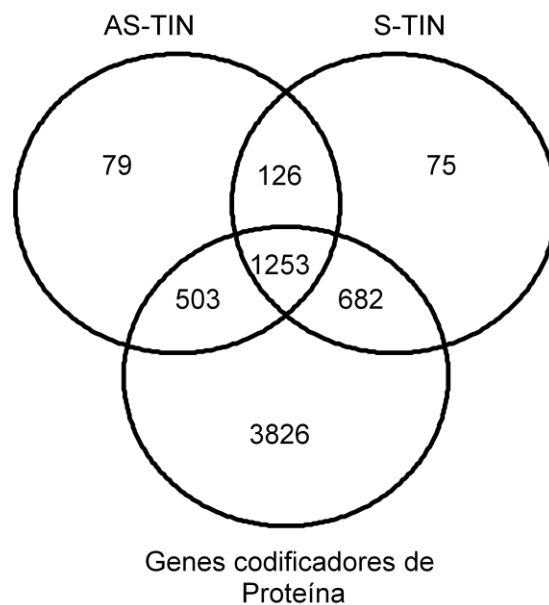


Figura 12: Diagrama de Venn mostrando a sobreposição da expressão, para os 6566 locus gênicos analisados, de regiões exônicas e de regiões totalmente intrônicas senso (S-TIN) e anti-senso (AS-TIN).

4.5.3 Análise de expressão

A fim de avaliar a qualidade dos dados analisados, nós analisamos a correlação entre os dados brutos obtidos para cada réplica técnica de cada amostra. Nós observamos uma alta correlação das medidas de expressão das réplicas técnicas Cy5 e Cy3 obtidas de cada amostra de paciente (media das réplicas técnicas $0,96 \pm 0,02$), mostrando alta reprodutibilidade da amplificação e marcação dos genes codificadores de proteína e dos ncRNAs. Nós também observamos uma alta correlação entre os dados de expressão ao longo das amostras (correlação media de expressão entre as amostras $0,90 \pm 0,04$). Uma vez que este arranjo customizado foi desenhado para interrogar genes não necessariamente envolvidos em PC, mas sim desenhado para sondar um conjunto de ncRNAs selecionados aleatoriamente do GeneBank, como previamente descrito na seção "Desenho do Oligoarranjo", era esperado uma pouca variação da expressão global entre os pacientes; nestas condições, a correlação

dos dados de expressão entre as amostras reflete a reprodutibilidade dos experimentos (Quackenbush, 2002).

Para cada sonda presente no oligoarranjo foi considerado, como valor de intensidade analisado do transcrito interrogado, o valor médio entre os dois canais (as réplicas marcadas com Cy5 e Cy3) de cada lâmina, em escala logarítmica de base 2, ou seja, $\frac{\log_2 \text{Cy3} + \log_2 \text{Cy5}}{2}$. Este valor é comumente chamado, em análises de microarranjos, de “A” e aparece frequentemente em gráficos de correção por intensidade, chamados "MxA plots". Para fins comparativos, os dados de expressão dos genes codificadores de proteína e de ncRNAs intrônicos foram analisados em separado.

4.5.4 Normalização dos dados de expressão.

A fim de tornar as lâminas comparáveis e de reduzir artefatos experimentais é necessário normalizar os valores de intensidade entre as lâminas para uma mesma escala. Existem diversos tipos de normalizações propostas para microarranjos. O principal desafio na escolha do método de normalização está em balancear os níveis de intensidade ao longo dos experimentos, mantendo os efeitos biológicos investigados (Do *et al.*, 2006). Optamos por utilizar o método do quantil (Bolstad *et al.*, 2003), no qual os valores de expressão, em cada lâmina, são ordenados (*ranked*) segundo a intensidade, ou seja, do mais intenso ao menos intenso, para cada amostra (Figura 13). Os valores então são substituídos pela média de intensidade obtida para determinada posição no ranking, obtendo-se assim um conjunto único de dados para todas as lâminas (Figura 13). Por esse método, é mais relevante a posição do gene no ranking do que a sua intensidade bruta. Com esse critério, podem-se comparar diferentes experimentos com escalas de intensidades diferentes. A normalização por quantil foi realizada com o pacote *affy* (Gautier *et al.*, 2004), disponível através do projeto R (R_Development_Core_Team, 2006).

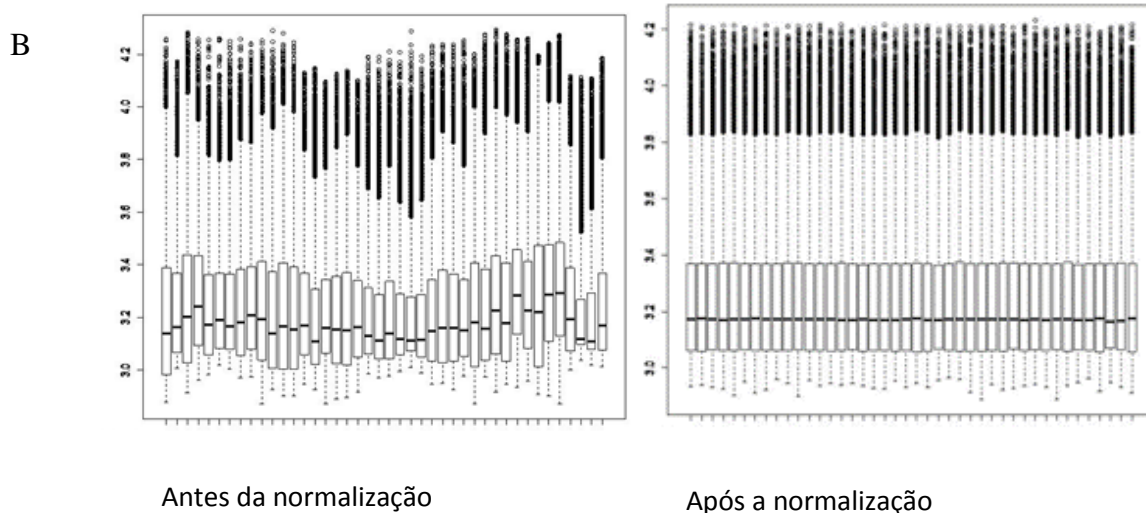
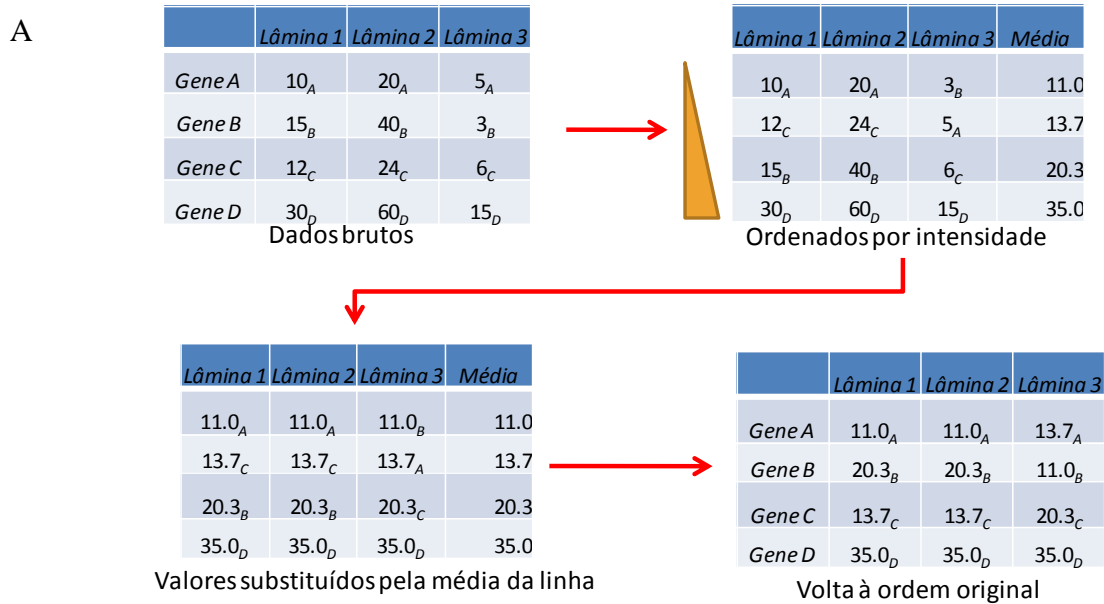


Figura 13: Normalização por quantil. A: Diagrama mostrando como é realizada a normalização: Os valores brutos de intensidade de cada amostras são ordenados independentemente. Em seguida é calculada o valor médio para cada posição do *ranking*. Este valor substitui o valor para uma determinada posição no *ranking*, para todas as amostras. B: Distribuição logarítmica das médias geométricas dos valores de cada *spot* (*Feature Extraction Processed Signal*) antes e após a normalização por quantil. O gráfico mostra um *boxplot* dos valores de intensidade antes e após a normalização por quantil. A caixa em cada coluna delimita o primeiro e o terceiro quartil do conjunto de valores de cada amostra. A linha tracejada indica a distribuição do 98º percentil do conjunto de valores. *Outliers*, ou seja, valores acima ou abaixo de três vezes a distância interquartílica, são marcados com pontos.

4.5.5 Análise de Dados – Agrupamentos Hierárquicos

Uma primeira análise foi realizada através de um cluster hierárquico dos *spots* válidos. A fim de tornar os transcritos comparáveis entre si, independentemente da intensidade, foi utilizado para cada transcrito, o número de desvios padrões acima ou abaixo

da média de cada gene ao longo dos experimentos, uma transformação chamada *Z-score* (Cheadle *et al.*, 2003). Este primeiro agrupamento das amostras (Figura 14) indicou claramente uma similaridade das amostras de acordo com o lote de hibridização em que cada amostra foi processada, e não com variações biológicas, como a recorrência do tumor. Este padrão foi observado tanto para os ncRNAs intrônicos (Figura 14 A) quanto para os genes codificadores de proteína (Figura 14 B).

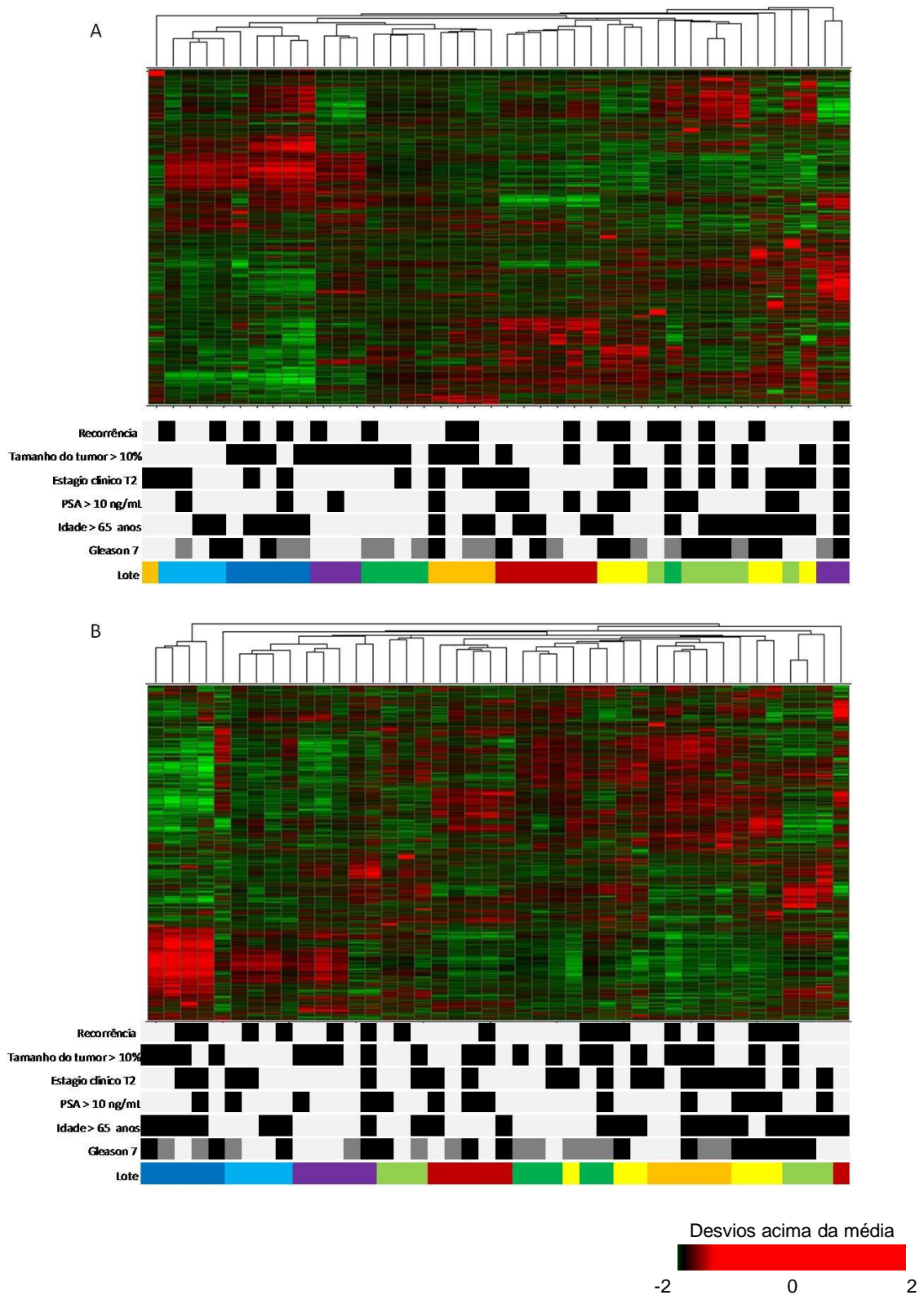


Figura 14: Agrupamento hierárquico dos ncRNAs intrônicos (A) e dos genes codificadores de proteína (B) para as amostras do conjunto de treinamento. Dados normalizados por quantil dos transcritos válidos, antes do processamento com o *script* ComBat.R. Cada linha vertical representa uma amostra, e cada linha horizontal um gene medido. A cor do gene é proporcional ao número de desvios padrão acima (vermelho) ou abaixo (verde) da média de cada gene ao longo de todas as amostras. Na barra marcada como “Lote”, cada amostra recebe a cor que representa o lote no qual a referida amostra foi processada. A existência de blocos contínuos de mesma cor indica que várias amostras processadas no mesmo lote se agruparam por terem perfis de expressão semelhantes.

Reduzir este efeito de agrupamento “por lote de hibridização” foi o maior desafio enfrentado neste estudo. Efeitos de lote foram observados desde os primeiros experimentos de microarranjos (Lander, 1999) e podem ser causados por muitos fatores, incluindo os lotes de kits de amplificação utilizados, a hora do dia em que o experimento é realizado, ou mesmo os níveis de ozônio atmosférico (Fare *et al.*, 2003; Branham *et al.*, 2007). Estes efeitos também são inevitáveis quando novas amostras são adicionadas progressivamente a um conjunto de dados pré-existente, como neste estudo. Alguns pesquisadores apresentaram métodos para o ajuste de efeitos de lote (Benito *et al.*, 2004; Alter *et al.*, 2006), mas estes métodos requerem muitas amostras (>25) em cada lote para melhor performance, e, além disso, podem remover variações biológicas reais dos dados.

Para corrigir este efeito, utilizamos um método recentemente proposto (Johnson *et al.*, 2007), que utiliza *Bayes Empírico* para o ajuste das amostras. Este método foi proposto especificamente para conjuntos de dados compostos por diversos lotes, com pequeno número de amostras em cada lote, como é o caso aqui descrito. Após o processamento dos dados pelo script `ComBat.R` (Johnson *et al.*, 2007) disponível em <http://statistics.byu.edu/johnson/ComBat/>, pode-se observar que o efeito de lote foi brutalmente reduzido (Figura 15).

Uma vez que este "efeito de lote" foi removido dos dados, foi possível observar que as amostras não se agrupam, segundo o valor de expressão global, por nenhuma das variáveis anatomico-patológicas: volume do tumor, estágio clínico, PSA pré-operatório, idade e Gleason Score (Figura 15). Da mesma forma, não foi identificado nenhum agrupamento pelo status de recorrência (Figura 15), tanto para os ncRNAs intrônicos (Figura 15A), quanto para os genes codificadores de proteína (Figura 15B).

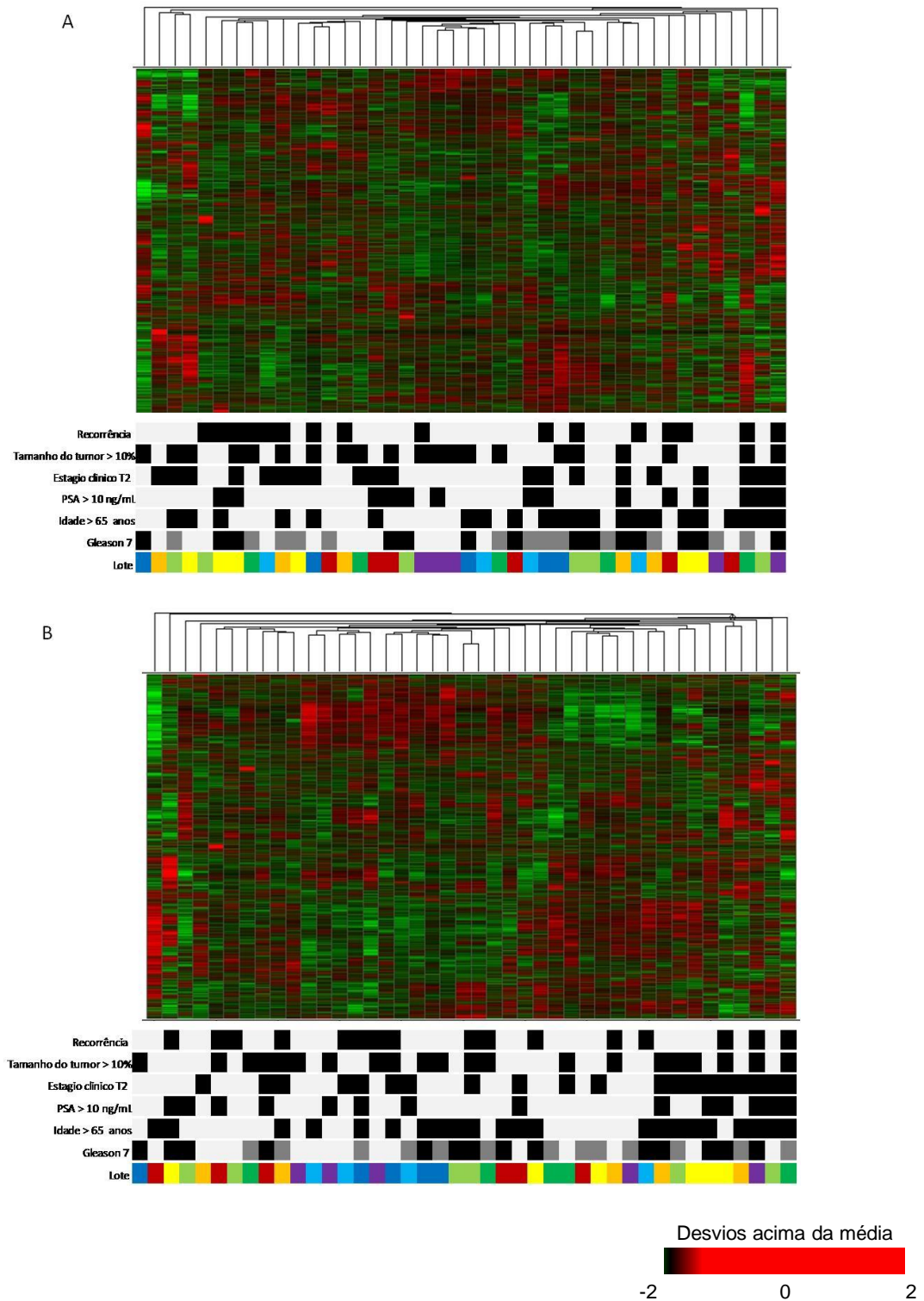


Figura 15: Agrupamento hierárquico dos ncRNAs intrônicos (A) e dos genes codificadores de proteína (B) para as amostras do conjunto de treinamento. Dados normalizados por quantil dos transcritos válidos, após o processamento com o *script* ComBat.R. Cada linha vertical representa uma amostra, e cada linha horizontal um gene medido. A cor do gene é proporcional ao número de desvios padrão acima (vermelho) ou abaixo (verde) da média de cada gene ao longo de todas as amostras. Na barra marcada como “Lote”, cada amostra recebe a cor que representa o lote no qual a referida amostra foi processada. A não existência de blocos contínuos de mesma cor indica que as amostras processadas no mesmo lote não mais se agruparam segundo seu perfil de expressão.

4.5.6 Genes codificadores de proteína diferencialmente expressos em recorrência de câncer de próstata

A fim de buscar por genes codificadores de proteína relacionados com a recidiva do câncer de próstata, foi aplicado um teste de significância específico para microarranjos chamado SAM – *Significance Analysis of Microarrays* (Tusher *et al.*, 2001) no conjunto de dados de genes codificadores de proteína entre 26 amostras de tumor de pacientes sem recidiva e as 16 amostras de pacientes com recidiva.

Este teste leva em consideração, além dos valores de significância tradicionais, como o p-valor do teste-t, o número de testes realizados – um para cada *spot*, e a probabilidade de se obter genes com valores altos de significância (ex. p valor menor que 0,005), mas resultantes do grande número de perguntas (40000 *spots* x 0,005 = 800 genes), a chamada taxa de falsa descoberta ou *false discovery rate*. A Figura 16 mostra um *plot* dos *scores* de significância observados entre os grupos com e sem recidiva *versus* o que seria observado aleatoriamente, computado através de permutações entre os grupos. Aqui foram utilizadas 1000 permutações.

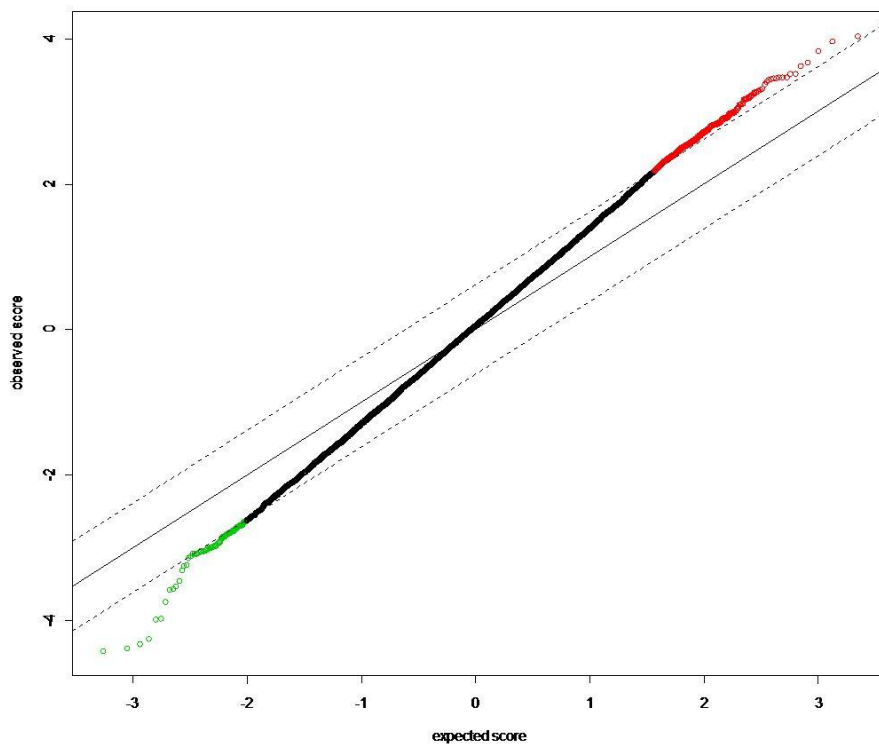


Figura 16: Grafico dos resultados da análise SAM – Significance Analysis of Microarray – mostrando o score de significância observado entre os grupos com e sem recidiva (*observed score*, eixo y) versus os valores esperados aleatoriamente (*expected score*, eixo x), com 500 permutações. Em vermelho e em verde estão os genes seleccionados significativamente mais (vermelho) e menos (verde) expressos no conjunto com recidiva em relação ao sem recidiva, com uma taxa de falso positivo (*false discovery rate*) de 5%.

Foram considerados significativamente diferencialmente expressos 98 genes codificadores de proteína com FDR de 5% (Tabela 8).

Tabela 8: Genes codificadores de proteína diferencialmente expressos entre amostras de tumor de próstata de pacientes com recorrência

Genes diminuídos em pacientes com recorrência					
Nome do Probe	Símbolo	Descrição	Fold Change	p value (Teste T de Student)	Taxa de falsa descoberta (FDR)
Q1_P02080	ACAA2	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	-1,59	<0,001	0,00
Q1_P08438	ACO2	aconitase 2, mitochondrial	-1,31	<0,001	4,76
Q1_P01584	CAPZA2	capping protein (actin filament) muscle Z-line, alpha 2	-1,56	<0,001	0,00
Q1_P22903	CCNDBP1	cyclin D-type binding-protein 1	-1,36	<0,001	0,00
Q1_P04615	DPYSL3	dihydropyrimidinase-like 3	-1,33	<0,001	4,76
Q1_P05498	LBH	limb bud and heart development homolog (mouse)	-1,77	<0,001	4,76
Q1_P07936	MARCH1	membrane-associated ring finger (C3HC4) 1	-1,23	<0,001	2,82
Q1_P22913	MFAP1	microfibrillar-associated protein 1	-1,32	<0,001	0,00
Q1_P04951	MS4A6A	membrane-spanning 4-domains, subfamily A, member 6A	-1,21	<0,001	4,76
Q1_P00300	OXCT1	3-oxoacid CoA transferase 1	-1,31	<0,001	0,00
Q1_P06655	PYGL	phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI)	-1,66	<0,001	0,00
Q1_P18071	RBPSUH	recombining binding protein suppressor of hairless (Drosophila)	-1,34	<0,001	0,00
Q1_P08889	RCHY1	ring finger and CHY zinc finger domain containing 1	-1,46	<0,001	0,00
Q1_P08294	SCRN1	secernin 1	-1,46	<0,001	0,00
Q1_P00855	UGP2	UDP-glucose pyrophosphorylase 2	-1,40	<0,001	0,00

Genes aumentados em pacientes com recorrência

Nome do Probe	Símbolo	Descrição	Fold Change	p value (Teste T de Student)	Taxa de falsa descoberta (% FDR)
Q1_P01028	AKAP1	A kinase (PRKA) anchor protein 1	1,33	<0,001	0,00
Q1_P07462	ANK3	ankyrin 3, node of Ranvier (ankyrin G)	1,39	0,006	3,18
Q1_P06830	ANKRD47	ankyrin repeat domain 47	1,11	<0,001	4,19
Q1_P06206	ANKS3	ankyrin repeat and sterile alpha motif domain containing 3	1,26	<0,001	0,00
Q1_P03774	ATP8A2	ATPase, aminophospholipid transporter-like, Class I, type 8A, member 2	2,21	<0,001	0,00
Q1_P15125	BAZ1A	bromodomain adjacent to zinc finger domain, 1A	1,32	<0,001	0,00
Q1_P08857	BOLA2	bolA homolog 2 (E. coli)	1,32	<0,001	1,81
Q1_P17118	C2orf33	chromosome 2 open reading frame 33	1,23	0,001	3,94
Q1_P18597	CANX	calnexin	1,27	<0,001	4,81
Q1_P02929	CBX5	chromobox homolog 5 (HP1 alpha homolog, Drosophila)	1,33	0,004	4,81
Q1_P08602	CCDC88	coiled-coil domain containing 88	1,41	0,002	4,19
Q1_P22428	CHEK1	CHK1 checkpoint homolog (S. pombe)	1,21	0,002	3,94
Q1_P16632	CNOT3	CCR4-NOT transcription complex, subunit 3	1,34	0,003	4,19
Q1_P23811	COL4A4	collagen, type IV, alpha 4	1,24	<0,001	3,94
Q1_P09147	COMTD1	catechol-O-methyltransferase domain containing 1	1,32	0,003	4,19
Q1_P07766	CPNE7	copine VII	1,25	0,001	3,94
Q1_P03468	CTBP2	C-terminal binding protein 2	1,36	<0,001	3,18
Q1_P17313	DDX27	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	1,25	0,004	4,81
Q1_P24537	DDX46	DEAD (Asp-Glu-Ala-Asp) box polypeptide 46	1,34	<0,001	1,81
Q1_P19898	DKC1	dyskeratosis congenita 1, dyskerin	1,23	<0,001	1,81
Q1_P05021	EIF4EBP1	eukaryotic translation initiation factor 4E binding protein 1	1,40	0,002	1,81
Q1_P16949	EPC2	enhancer of polycomb homolog 2 (Drosophila)	1,16	<0,001	4,19
Q1_P00163	FKBP5	FK506 binding protein 5	1,59	0,006	3,94
Q1_P14284	FNBP4	formin binding protein 4	1,30	0,002	1,81
Q1_P02601	FRYL	furry homolog-like	1,24	<0,001	1,81
Q1_P03886	GNL3L	guanine nucleotide binding protein-like 3 (nucleolar)-like	1,27	<0,001	1,81
Q1_P05701	GPR123	G protein-coupled receptor 123	1,15	0,001	3,94
Q1_P07211	GSK3B	glycogen synthase kinase 3 beta	1,20	<0,001	3,94

Nome do Probe	Símbolo	Descrição	Fold Change	p value (Teste T de Student)	Taxa de falsa descoberta (% FDR)
Q1_P23357	HELZ	helicase with zinc finger	1,35	0,005	3,94
Q1_P03374	HIPK2	homeodomain interacting protein kinase 2	1,48	<0,001	0,00
Q1_P19029	HNRPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1	1,33	0,001	1,81
Q1_P16706	ITSN2	intersectin 2	1,16	0,002	4,19
Q1_P15370	KIAA0256	KIAA0256 gene product	1,24	0,004	4,19
Q1_P21822	KIAA0319L	KIAA0319-like	1,33	0,003	3,94
Q1_P18069	SEL1L3	sel-1 suppressor of lin-12-like 3	1,22	0,002	4,19
Q1_P03952	KIAA0892	KIAA0892	1,17	0,001	4,19
Q1_P13865	LBR	lamin B receptor	1,18	0,001	4,81
Q1_P15762	LOC55565	hypothetical protein LOC55565	1,16	0,003	4,19
Q1_P19237	LUC7L2	LUC7-like 2 (<i>S. cerevisiae</i>)	1,50	0,002	1,81
Q1_P13393	MACF1	microtubule-actin crosslinking factor 1	1,27	0,003	3,18
Q1_P24767	MAP7	microtubule-associated protein 7	1,42	<0,001	0,00
Q1_P00472	MATN4	matrilin 4	1,25	<0,001	0,00
Q1_P22564	MON2	MON2 homolog (<i>S. cerevisiae</i>)	1,25	0,001	0,00
Q1_P01526	MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase	1,35	0,002	3,94
Q1_P05369	MYOHD1	myosin head domain containing 1	1,29	<0,001	0,00
Q1_P13439	NASP	nuclear autoantigenic sperm protein (histone-binding)	1,26	0,001	1,81
Q1_P02195	NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive	1,44	<0,001	1,81
Q1_P15680	NFATC2IP	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	1,31	<0,001	3,94
Q1_P00350	NNT	nicotinamide nucleotide transhydrogenase	1,28	<0,001	1,81
Q1_P02121	NOL5A	nucleolar protein 5A (56kDa with KKE/D repeat)	1,22	0,003	3,94
Q1_P03716	NOP5/ NOP58	nucleolar protein NOP5/NOP58	1,31	0,004	4,19
Q1_P06109	OSBPL10	oxysterol binding protein-like 10	1,19	<0,001	3,94
Q1_P22758	PARP2	poly (ADP-ribose) polymerase family, member 2	1,22	<0,001	3,18
Q1_P16915	PAX8	paired box gene 8	1,35	0,003	3,18
Q1_P05318	PCNXL2	pecanex-like 2 (<i>Drosophila</i>)	1,36	<0,001	0,00
Q1_P15624	PMM2	phosphomannomutase 2	1,36	<0,001	0,00
Q1_P17299	PRKCBP1	protein kinase C binding protein 1	1,51	<0,001	0,00
Q1_P00339	PTMA	prothymosin, alpha (gene sequence 28)	1,22	<0,001	3,94

Nome do Probe	Símbolo	Descrição	Fold Change	p value (Teste T de Student)	Taxa de falsa descoberta (% FDR)
Q1_P16260	<i>PTPN2</i>	protein tyrosine phosphatase, non-receptor type 2	1,26	0,001	1,81
Q1_P01189	<i>PTTG1IP</i>	pituitary tumor-transforming 1 interacting protein	1,25	<0,001	3,94
Q1_P19691	<i>RALGPS1</i>	Ral GEF with PH domain and SH3 binding motif 1	1,16	<0,001	1,81
Q1_P07576	<i>RPL8</i>	ribosomal protein L8	1,29	0,001	3,18
Q1_P01388	<i>SFRS2IP</i>	splicing factor, arginine/serine-rich 2, interacting protein	1,28	0,004	4,19
Q1_P24775	<i>SHPRH</i>	SNF2 histone linker PHD RING helicase	1,24	<0,001	3,18
Q1_P08371	<i>SMARCC1</i>	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1	1,20	<0,001	4,19
Q1_P00378	<i>SOX4</i>	SRY (sex determining region Y)-box 4	1,36	0,006	4,19
Q1_P04716	<i>SRGAP1</i>	SLIT-ROBO Rho GTPase activating protein 1	1,35	0,002	1,81
Q1_P15590	<i>SRRM2</i>	serine/arginine repetitive matrix 2	1,21	0,004	4,81
Q1_P08481	<i>TCEA1</i>	transcription elongation factor A (SII), 1	1,25	<0,001	1,81
Q1_P08503	<i>TCOF1</i>	Treacher Collins-Franceschetti syndrome 1	1,25	0,001	3,18
Q1_P22871	<i>TJP1</i>	tight junction protein 1 (zona occludens 1)	1,20	0,001	3,94
Q1_P02656	<i>TNRC6B</i>	trinucleotide repeat containing 6B	1,31	<0,001	1,81
Q1_P17790	<i>TOMM70A</i>	translocase of outer mitochondrial membrane 70 homolog A (<i>S. cerevisiae</i>)	1,28	0,004	4,19
Q1_P00427	<i>TPM1</i>	tropomyosin 1 (alpha)	1,36	0,002	4,19
Q1_P15052	<i>TPP2</i>	tripeptidyl peptidase II	1,21	0,001	3,94
Q1_P00953	<i>TRIM25</i>	tripartite motif-containing 25	1,27	<0,001	0,00
Q1_P22259	<i>TSPAN32</i>	tetraspanin 32	1,29	<0,001	4,19
Q1_P16199	<i>USP36</i>	ubiquitin specific peptidase 36	1,40	0,003	3,94
Q1_P05640	<i>USP48</i>	ubiquitin specific peptidase 48	1,12	<0,001	4,81
Q1_P02749	<i>VPS13A</i>	vacuolar protein sorting 13 homolog A (<i>S. cerevisiae</i>)	1,38	<0,001	0,00
Q1_P21765	<i>VPS13D</i>	vacuolar protein sorting 13 homolog D (<i>S. cerevisiae</i>)	1,12	<0,001	3,94
Q1_P02817	<i>WWC1</i>	WW and C2 domain containing 1	1,33	0,003	4,19
Q1_P21779	<i>ZUBR1</i>	zinc finger, UBR1 type 1	1,19	<0,001	4,19

4.5.7 Análise de funções gênicas

A fim de analisar como os genes codificadores de proteína identificados como diferencialmente expressos estão relacionados entre si, todos os 6611 genes expressos foram adicionados no contexto do conhecimento presente sobre vias e interações moleculares, usando o banco de dados do *software Ingenuity Pathway Analysis* (IPA). As funções biológicas foram atribuídas usando-se a base de dados referencia do IPA, representando mais de 9.600 classes de conceitos químicos e biológicos, consistindo de milhares de relações individuais entre proteínas, genes, complexos, células, tecidos, compostos químicos e doenças. Estas relações são baseadas em uma extração de dados contínua da literatura de domínio público, e cobrem > 15.700 genes humanos. Nossos dados indicam que as categorias que incluem o maior número de genes diferencialmente expressos foram *Cellular Growth and Proliferation*, *Cellular Assembly and Organization*, *Cell Morphology* e *Tissue Morphology* (Tabela 9).

Tabela 9: Categorias dos genes codificadores de proteína diferencialmente expressos entre pacientes com recorrência e sem recorrência.

Categoria	Número de genes	Genes	P valor	P valor corrigido*
Cellular Growth and Proliferation	22	<i>SOX4, PTPN2, TJP1, HNRNPA2B1, DKC1, MAP7, RCHY1, TPP2, ATP8A2, PTMA, EIF4EBP1, CHEK1, NASP, NFAT5, TCOF1, CTBP2, RBPJ, GSK3B, HIPK2, FKBP5, TRIM25, TSPAN32</i>	<0,04	<0,19
Cellular Assembly and Organization	14	<i>ANK3, TPM1, TJP1, PARP2, MAP7, BAZ1A, CANX, TPP2, PTMA, EIF4EBP1, MACF1, GSK3B, SMARCC1, HIPK2</i>	<0,05	<0,21
Cell Morphology	9	<i>ANK3, TPM1, MACF1, DPYSL3, MAP7, TCEA1, GSK3B, TPP2, CHEK1</i>	<0,03	<0,17
Tissue Morphology	9	<i>ANK3, PTPN2, TJP1, TCOF1, RBPJ, HIPK2, TPP2, EIF4EBP1, CHEK1</i>	<0,04	<0,19
Cellular Compromise	8	<i>ANK3, TPM1, PARP2, MAP7, SMARCC1, GSK3B, HIPK2, EIF4EBP1</i>	<0,04	<0,19
Embryonic Development	8	<i>MACF1, TCOF1, RBPJ, SMARCC1, GSK3B, HIPK2, TPP2, CHEK1</i>	<0,04	<0,19
Genetic Disorder	8	<i>MTRR, VPS13A, PAX8, TCOF1, DKC1, OXCT1, LBR, PMM2</i>	<0,03	<0,17
Small Molecule Biochemistry	8	<i>MTRR, UGP2, PARP2, ACO2, OXCT1, FKBP5, PMM2, EIF4EBP1</i>	<0,04	<0,19
Cellular Development	6	<i>MACF1, MAP7, RBPJ, GSK3B, HIPK2, CHEK1</i>	<0,04	<0,19
Nervous System Development and Function	6	<i>ANK3, DPYSL3, CANX, RBPJ, GSK3B, EIF4EBP1</i>	<0,04	<0,19
Carbohydrate Metabolism	6	<i>MTRR, UGP2, ACO2, PYGL, GSK3B, PMM2</i>	<0,04	<0,19
Cell Death	6	<i>TPM1, GSK3B, FKBP5, HIPK2, PTMA, CHEK1</i>	<0,04	<0,19
Gene Expression	6	<i>NFAT5, TJP1, CTBP2, RBPJ, CBX5, EIF4EBP1</i>	<0,04	<0,19
RNA Post-Transcriptional Modification	5	<i>NOP56, NOP58, DKC1, TCEA1, EIF4EBP1</i>	<0,04	<0,19

* P valor corrigido para múltiplo teste (Benjamini-Hochberg).

Categoría	Número de genes	Genes	P valor	P valor corregido*
Cell-To-Cell Signaling and Interaction	5	<i>ANK3, TJP1, COL4A4 (includes EG:1286), PTMA, CHEK1</i>	<0,05	<0,21
Cellular Function and Maintenance	5	<i>DPYSL3, RBPJ, TCEA1, PTMA, CHEK1</i>	<0,04	<0,19
Connective Tissue Development and Function	5	<i>PARP2, GSK3B, PTMA, CHEK1, EIF4EBP1</i>	<0,04	<0,19
Developmental Disorder	5	<i>TCOF1, DKC1, RBPJ, GSK3B, LBR</i>	<0,03	<0,17
Organismal Development	5	<i>PAX8, CTBP2, CANX, RBPJ, EIF4EBP1</i>	<0,04	<0,19
Cell Cycle	4	<i>TPM1, NASP, HIPK2, CHEK1</i>	<0,04	<0,19
DNA Replication, Recombination, and Repair	4	<i>PARP2, BAZ1A, SMARCC1, CHEK1</i>	<0,04	<0,19
Metabolic Disease	4	<i>MTRR, PAX8, OXCT1, PMM2</i>	<0,03	<0,17
Skeletal and Muscular Disorders	4	<i>TCOF1, GSK3B, LBR, PTMA</i>	<0,03	<0,17
Tissue Development	4	<i>MACF1, PARP2, HIPK2, CHEK1</i>	<0,03	<0,17
Post-Translational Modification	4	<i>PARP2, CAPZA2, GSK3B, EIF4EBP1</i>	<0,05	<0,21
Protein Synthesis	4	<i>TNRC6B, CAPZA2, GSK3B, EIF4EBP1</i>	<0,05	<0,21
Infection Mechanism	3	<i>RBPJ, TRIM25, EIF4EBP1</i>	<0,02	<0,13
Cancer	3	<i>TJP1, GSK3B, EIF4EBP1</i>	<0,04	<0,19
Hematological System Development and Function	3	<i>PTPN2, HNRNPA2B1, CHEK1</i>	<0,04	<0,19
Lipid Metabolism	3	<i>OXCT1, FKBP5, EIF4EBP1</i>	<0,04	<0,19
Neurological Disease	3	<i>ANK3, SOX4, VPS13A</i>	<0,04	<0,19

* P valor corregido para múltiplo teste (Benjamini-Hochberg)

Categoría	Número de Genes	Genes	P valor	P valor corregido*
Organ Development	3	<i>PAX8, RBPJ, GSK3B</i>	<0,03	<0,17
Organ Morphology	3	<i>PAX8, RBPJ, GSK3B</i>	<0,04	<0,19
Cell Signaling	3	<i>CAPZA2, GSK3B, EIF4EBP1</i>	<0,05	<0,21
Cellular Movement	3	<i>TPM1, NFAT5, HIPK2</i>	<0,04	<0,19
Connective Tissue Disorders	2	<i>LBR, PTMA</i>	<0,03	<0,17
Drug Metabolism	2	<i>MTRR, FKBP5</i>	<0,03	<0,17
Endocrine System Development and Function	2	<i>HIPK2, FKBP5</i>	<0,03	<0,17
Hematopoiesis	2	<i>PTPN2, CHEK1</i>	<0,04	<0,19
Nucleic Acid Metabolism	2	<i>UGP2, PMM2</i>	<0,02	<0,13
Skeletal and Muscular System Development and Function	2	<i>RBPJ, GSK3B</i>	<0,03	<0,17
Vitamin and Mineral Metabolism	2	<i>MTRR, GSK3B</i>	<0,03	<0,17
Hair and Skin Development and Function	2	<i>RBPJ, HIPK2</i>	<0,04	<0,19
Renal and Urological System Development and Function	2	<i>TJP1, HIPK2</i>	<0,04	<0,19
Cardiovascular System Development and Function	1	<i>GSK3B</i>	<0,04	<0,19
Cell-mediated Immune Response	1	<i>CHEK1</i>	<0,03	<0,17

* P valor corregido para múltiplo teste (Benjamini-Hochberg)

Categoría	Número de genes	Genes	P valor	P valor corregido*
Dermatological Diseases and Conditions	1	DKC1	<0,02	<0,13
Digestive System Development and Function	1	GSK3B	<0,02	<0,13
Endocrine System Disorders	1	PAX8	<0,02	<0,13
Gastrointestinal Disease	1	PARP2	<0,02	<0,13
Hematological Disease	1	LBR	<0,02	<0,13
Hepatic System Development and Function	1	GSK3B	<0,02	<0,13
Molecular Transport	1	UGP2	<0,02	<0,13
Organismal Injury and Abnormalities	1	GSK3B	<0,02	<0,13
Reproductive System Disease	1	TJP1	<0,02	<0,13
Tumor Morphology	1	EIF4EBP1	<0,02	<0,13
Inflammatory Disease	1	PTMA	<0,03	<0,17
Reproductive System Development and Function	1	RBPJ	<0,03	<0,17
Amino Acid Metabolism	1	PARP2	<0,04	<0,19
Antimicrobial Response	1	GSK3B	<0,04	<0,19
Lymphoid Tissue Structure and Development	1	PTPN2	<0,04	<0,19
Protein Degradation	1	TPP2	<0,04	<0,19
RNA Trafficking	1	EIF4EBP1	<0,04	<0,19
Renal and Urological Disease	1	NFAT5	<0,04	<0,19
Visual System Development and Function	1	RBPJ	<0,04	<0,19

* P valor corregido para múltiple teste (Benjamini-Hochberg)

4.6 Identificação de um perfil prognóstico de recorrência baseado na expressão de ncRNAs.

Uma vez que não foram identificados RNAs não codificadores de proteínas individualmente estatisticamente alterados entre os pacientes com e sem recidiva, através da abordagem de SAM, nós procuramos identificar um perfil de expressão de um conjunto de ncRNAs que pudesse estar correlacionado à recorrência e que fosse capaz de prever o desfecho em um conjunto independente de amostras. Para este fim, foi utilizada uma nova abordagem, mais robusta que as análises de significância anteriormente utilizadas. Essa abordagem, chamada múltipla reamostragem aleatória (*multiple random resampling*) (Figura 17) foi descrita em um artigo por Michaelis et al. (Michiels *et al.*, 2005), no qual os autores analisam, utilizando a técnica de múltipla amostragem, sete trabalhos publicados na literatura que identificaram perfis de expressão com uso prognóstico. Com esta abordagem, os autores identificaram que, para todos os trabalhos analisados exceto um, o limite superior do intervalo de confiança foi acima de 0.5, ou seja, que para 6 dos 7 trabalhos, os perfis gerados não classificaram melhor do que seria observado aleatoriamente (Michiels *et al.*, 2005).

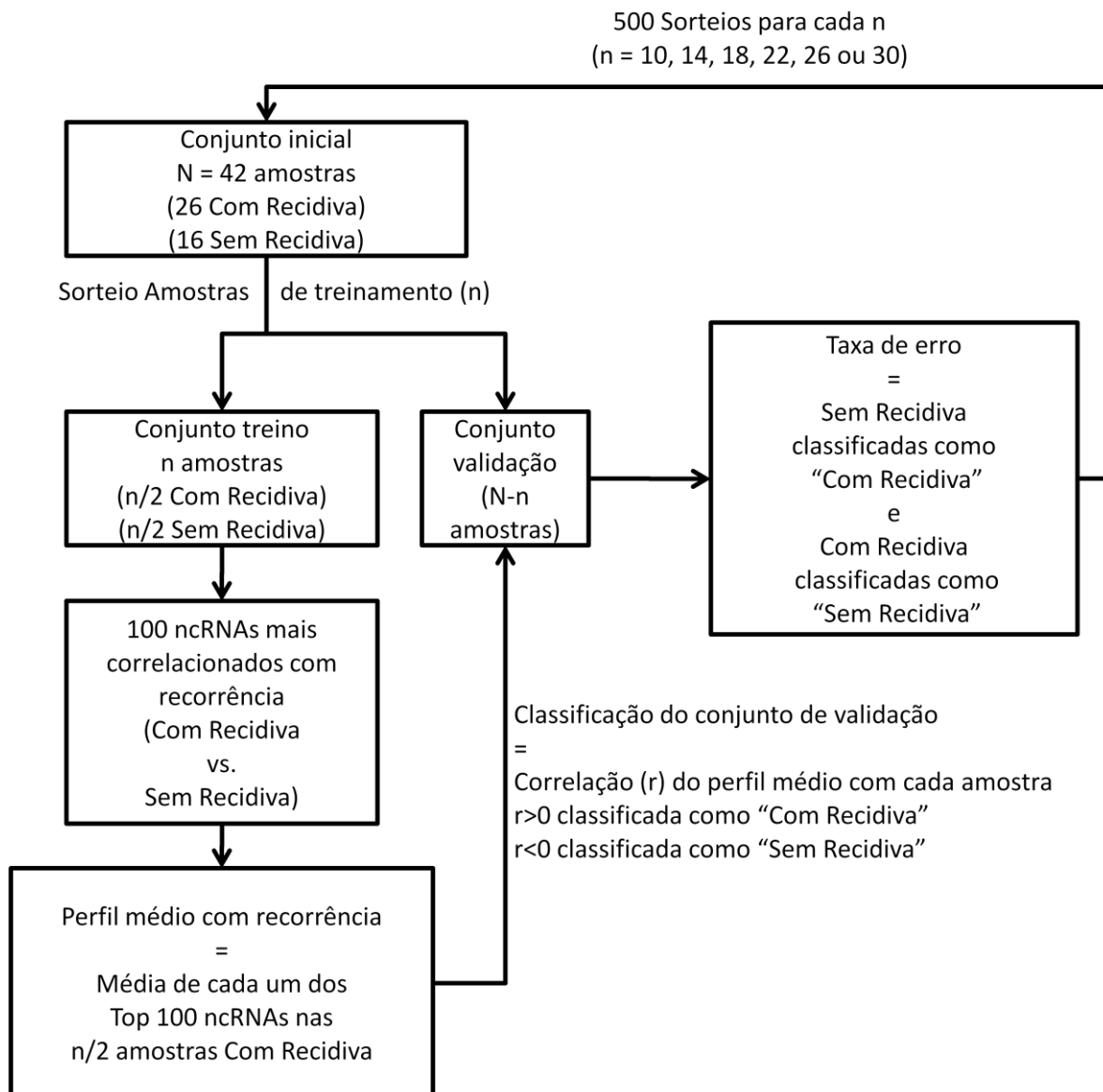


Figura 17: Esquematização da estratégia de múltipla amostragem. O conjunto de treinamento (de tamanho N = 42 pacientes) foi dividido usando uma abordagem de reamostragem 500 conjuntos aleatórios de treinamento (cada um com tamanho n), tendo n/2 pacientes tendo cada desfecho, e 500 conjuntos de validação associados (de tamanho N-n), conforme descrito (Michiels *et al.*, 2005). Em nossa análise, repetimos o processamento dos 500 conjuntos 6 vezes, cada vez com um valor diferente de n, tal que n = 10, 14, 18, 22, 26 ou 30. A assinatura preditora de cada um dos conjuntos de treinamento foi definida como a lista dos 100 ncRNAs que foram mais altamente correlacionados com o desfecho da doença dos pacientes, mostrado pelo coeficiente de correlação de Pearson. Os perfis preditivos de alto ou baixo risco foram obtidos tomando-se o valor de expressão médio de cada um dos 100 ncRNAs entre os n/2 pacientes de cada desfecho. Em seguida, a classificação dos pacientes nos conjuntos de validação correspondentes (cada um com N-n pacientes) foi realizada pela regra de predição do centroide mais próximo (Michiels *et al.*, 2005), isto é, cada paciente foi colocado no grupo de alto ou baixo risco de acordo com o perfil preditivo com o qual a amostra de validação tinha maior correlação. A proporção de classificação errada de cada conjunto de validação associado foi determinada e a média de classificação errada, bem como o intervalo de confiança de 95% foi calculado.

Para a múltipla reamostragem aleatória (Michiels *et al.*, 2005) foram eliminados transcritos que mostraram pouca ou nenhuma variação entre as amostras e retidos 1803 ncRNAs com coeficiente de variação $> 0,03$. Em seguida foram sorteadas n amostras para um conjunto de treinamento, sendo $n/2$ de cada desfecho (*outcome*), a partir do conjunto inicial de amostras N , e de um conjunto correspondente de $N-n$ amostras para a validação (Figura 17). O conjunto de treinamento é utilizado para a geração de um perfil classificador, compreendido dos 100 genes mais correlacionados (através da correlação de Pearson) com o desfecho da doença. Em seguida, são gerados dois perfis médios de expressão (um para cada desfecho da doença) destes 100 genes, que são em seguida utilizados para a classificação das amostras de do conjunto de validação. Cada amostra do conjunto de validação é classificada como pertencente ao grupo com qual tem maior correlação com o respectivo perfil médio. Por exemplo: uma amostra da validação é classificada como alto risco se o perfil de expressão da amostra tem maior correlação com o perfil médio das amostras que tiveram recorrência do conjunto de treinamento. Em seguida, é calculada uma taxa de erros do perfil classificador, contabilizando-se todos os erros para todas as amostras do conjunto de validação. Esse procedimento é repetido 500 vezes, gerando 500 diferentes conjuntos de treinamento e 500 conjuntos de validação associados, e a taxa média de erro, com intervalo de confiança de 95% para todos os conjuntos são computados. Além disso, foram utilizados diferentes tamanhos amostrais (n) de treinamento, sendo o n mínimo de 10 e o n máximo aquele que mantém pelo menos uma amostra de cada desfecho para ser deixada no conjunto de validação, neste caso, $n=30$, sendo 15 de pacientes com recorrência e 15 de pacientes sem recorrência.

Em nossa análise, foram utilizados $n = 10, 14, 18, 22, 26, 30$ para os conjuntos de treinamento, considerando os 100 transcritos intrônicos não codificadores mais correlacionados com a recorrência de PC para cada amostragem. Observamos que, para os

diferentes n dos conjuntos de treinamento, o limite superior de 95% se manteve sempre abaixo de 0,5, indicando que, para nosso conjunto amostral, a seleção de genes não foi aleatória (Figura 18).

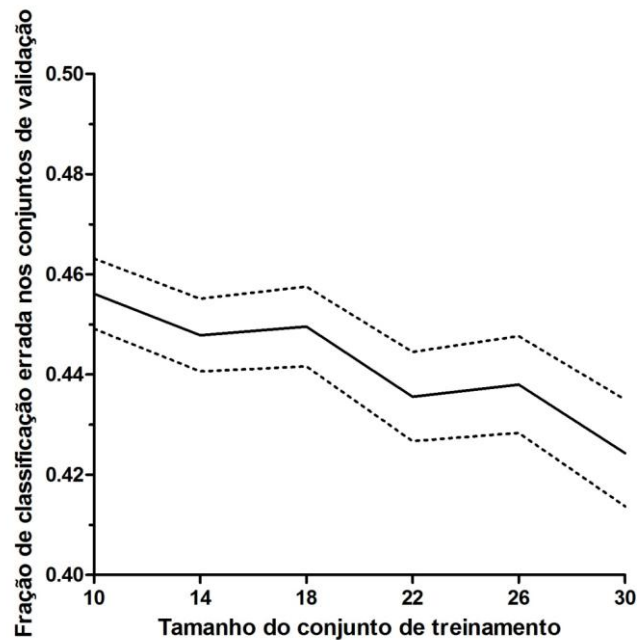


Figura 18: Taxa Classificação errada do desfecho dos pacientes nos conjunto de validação em função dos diferentes tamanhos dos conjuntos de treinamento correspondentes. A linha sólida indica a proporção média da taxa de classificação errada nos conjuntos de validação obtidos com as 500 reamostragens aleatórias do conjunto inicial de 42 pacientes. Linhas tracejadas mostram o intervalo de confiança de 95%.

Nós utilizamos o tamanho de 30 pacientes no conjunto de treinamento, que apresentou a menor proporção de classificação errada (Figura 18), para definir uma assinatura de mau prognóstico. Esta análise encontrou 51 ncRNAs que estavam entre os 100 mais correlacionados com o desfecho do PC em mais de 50% das reamostragens (Tabela 10).

Tabela 10: ncRNAs que compõem o perfil de expressão de mau prognóstico de recorrência bioquímica de câncer de próstata dentro de 5 anos após a prostatectomia radical.

	Símbolo do gene codificador hospedeiro	Coordenada genômica do ncRNA	Fita	Orientação em relação ao gene codificador hospedeiro	Fold Change	Veze em que o ncRNA aparece como top 100 nas 500 permutações
Q1_P27939	<i>ISE1.2</i>	chr12:119354926-119355389	-	antissenso	1,18	262
Q1_P27882	<i>ACTR6</i>	chr12:99114880-99115413	+	senso	1,17	303
Q1_P27205	<i>ANK3</i>	chr10:62135602-62136041	+	antissenso	1,38	393
Q1_P32933	<i>ANK3</i>	chr10:62135602-62136041	-	senso	1,31	274
Q1_P35761	<i>ATG7</i>	chr3:11475077-11475766	+	senso	1,27	339
Q1_P41598	<i>ATXN1</i>	chr6:16762895-16763606	-	senso	-1,24	355
Q1_P33475	<i>BICD1</i>	chr12:32394824-32395776	+	senso	1,31	379
Q1_P29734	<i>BLCAP</i>	chr20:35582213-35582933	+	antissenso	1,27	336
Q1_P37335	<i>BRAF</i>	chr7:139982218-139983360	-	senso	1,32	390
Q1_P35851	<i>CCBP2</i>	chr3:42884423-42885832	+	senso	-1,24	385
Q1_P32700	<i>CDC73</i>	chr1:189901199-189901888	-	antissenso	1,23	286
Q1_P10973	<i>CHD2</i>	chr15:91344662-91345102	-	antissenso	1,26	334
Q1_P28605	<i>CHD9</i>	chr16:51797153-51797773	-	antissenso	-1,26	353
Q1_P12059	<i>CTNNB1</i>	chr3:41252347-41255628	-	antissenso	-1,2	424
Q1_P40541	<i>CUTL2</i>	chr12:110020370-110021594	+	senso	1,16	342
Q1_P41193	<i>CYYR1</i>	chr21:26785519-26786681	-	senso	-1,24	302
Q1_P12706	<i>EPB41L2</i>	chr6:131228076-131229403	+	antissenso	1,3	387
Q1_P35353	<i>FLJ20701</i>	chr2:229791460-229792090	-	senso	-1,27	399
Q1_P30099	<i>GLB1</i>	chr3:33024047-33024521	+	antissenso	-1,21	274
Q1_P29449	<i>GLI2</i>	chr2:121274628-121274972	-	antissenso	1,21	329
Q1_P32160	<i>GRIN1</i>	chr9:137325309-137326363	-	antissenso	1,54	338
Q1_P28589	<i>GTF3C1</i>	chr16:27457846-27458360	+	antissenso	1,18	409
Q1_P12432	<i>HEXB</i>	chr5:74049421-74049983	-	antissenso	-1,18	341
Q1_P34245	<i>IGF1R</i>	chr15:97237840-97238729	+	senso	-1,27	498
Q1_P41967	<i>INVS</i>	chr9:100136377-100137579	+	senso	1,24	385
Q1_P32651	<i>IQWD1</i>	chr1:164740960-164742684	+	senso	1,31	349
Q1_P32045	<i>KIAA1529</i>	chr9:97138078-97139547	-	antissenso	-1,38	349
Q1_P10496	<i>LDHB</i>	chr12:21681062-21681575	+	antissenso	-1,21	494
Q1_P33193	<i>LDLRAD3</i>	chr11:35942918-35943629	+	senso	1,28	353
Q1_P40543	<i>LOC196463</i>	chr12:112265194-112265739	+	senso	-1,23	435
Q1_P28034	<i>LRCH1</i>	chr13:46053363-46053848	-	antissenso	1,25	494
Q1_P21624	<i>MAPK15</i>	chr8:144871983-144872761	-	antissenso	-1,17	344
Q1_P12244	<i>MFI2</i>	chr3:198218633-198219925	+	antissenso	-1,16	351
Q1_P33102	<i>MGMT</i>	chr10:131334009-131334572	+	senso	1,21	366

ID do probe	Símbolo do gene codificador hospedeiro	Coordenada genômica do ncRNA	Fita	Orientação em relação ao gene codificador hospedeiro	Fold Change	Veze em que o ncRNA aparece como top 100 nas 500 permutações
Q1_P36755	MYLIP	chr6:16247012-16247557	+	senso	1,2	286
Q1_P34308	NDUFAB1	chr16:23503305-23504122	-	senso	1,17	382
Q1_P34492	NLK	chr17:23484660-23485558	+	senso	1,3	295
Q1_P29327	NRXN1	chr2:50741152-50741731	+	antissenso	1,38	406
Q1_P40336	OSBPL5	chr11:3082563-3083241	-	senso	1,25	332
Q1_P41221	PCQAP	chr22:19255655-19256842	+	senso	1,29	485
Q1_P34252	PCSK6	chr15:99751290-99752769	-	senso	1,32	497
Q1_P28310	PPP1R13B	chr14:103299089-103299903	+	antissenso	-1,23	270
Q1_P21029	PREX1	chr20:46688746-46690107	+	antissenso	-1,22	480
Q1_P34147	RAB8B	chr15:61300748-61301146	+	senso	1,21	414
Q1_P37167	SEMA3C	chr7:80156286-80156761	-	senso	1,44	396
Q1_P34947	SH3YL1	chr2:226765-227409	-	senso	1,18	307
Q1_P20194	SLC39A13	chr11:47385240-47386014	-	antissenso	1,31	458
Q1_P40793	SPG7	chr16:88127278-88127700	+	senso	1,17	291
Q1_P11838	TASPI	chr20:13487293-13487861	+	antissenso	-1,24	350
Q1_P33396	TULP3	chr12:2894263-2895263	+	senso	1,18	276
Q1_P29207	ZNF577	chr19:57071395-57071911	+	antissenso	1,16	252

Os valores de expressão de cada gene foram normalizados, a fim de torná-los comparáveis, independentemente do nível de intensidade que possuam. Para isso foi realizada a normalização chamada *Z-score* (Cheadle *et al.*, 2003), na qual é calculada a média e o desvio padrão da expressão de cada gene ao longo das amostras. Em seguida, cada o valor de expressão de cada gene é subtraído pela média e este resultado é então dividido pelo desvio padrão. Desta maneira, o valor de expressão considerado é o número de desvios padrões acima ou abaixo da média de cada gene ao longo das amostras. Este processo gerou os valores de média e de desvios dos transcritos do perfil classificador para as amostras do conjunto de treinamento, que são posteriormente aplicados ao conjunto de validação, como está discutido posteriormente.

Foi calculado, em seguida, como perfil classificador, um “perfil médio” das amostras que não apresentaram recidiva, ou seja, o valor médio de cada *spot* para as amostras sem

recidiva. Em seguida, para a classificação das amostras, foi calculado o valor de correlação entre a expressão destes transcritos de cada amostra do conjunto de treinamento e o perfil médio com recidiva. Para valores positivos, a amostra foi classificada como pertencente ao grupo de mau prognóstico. Para valores negativos, a amostra foi classificada como bom prognóstico. Desta maneira, o perfil com transcritos intrônicos e classificou corretamente 17/19 (90%) das amostras com recidiva e 22/31 (70%) das amostras sem recidiva do conjunto de treinamento (Figura 19 A).

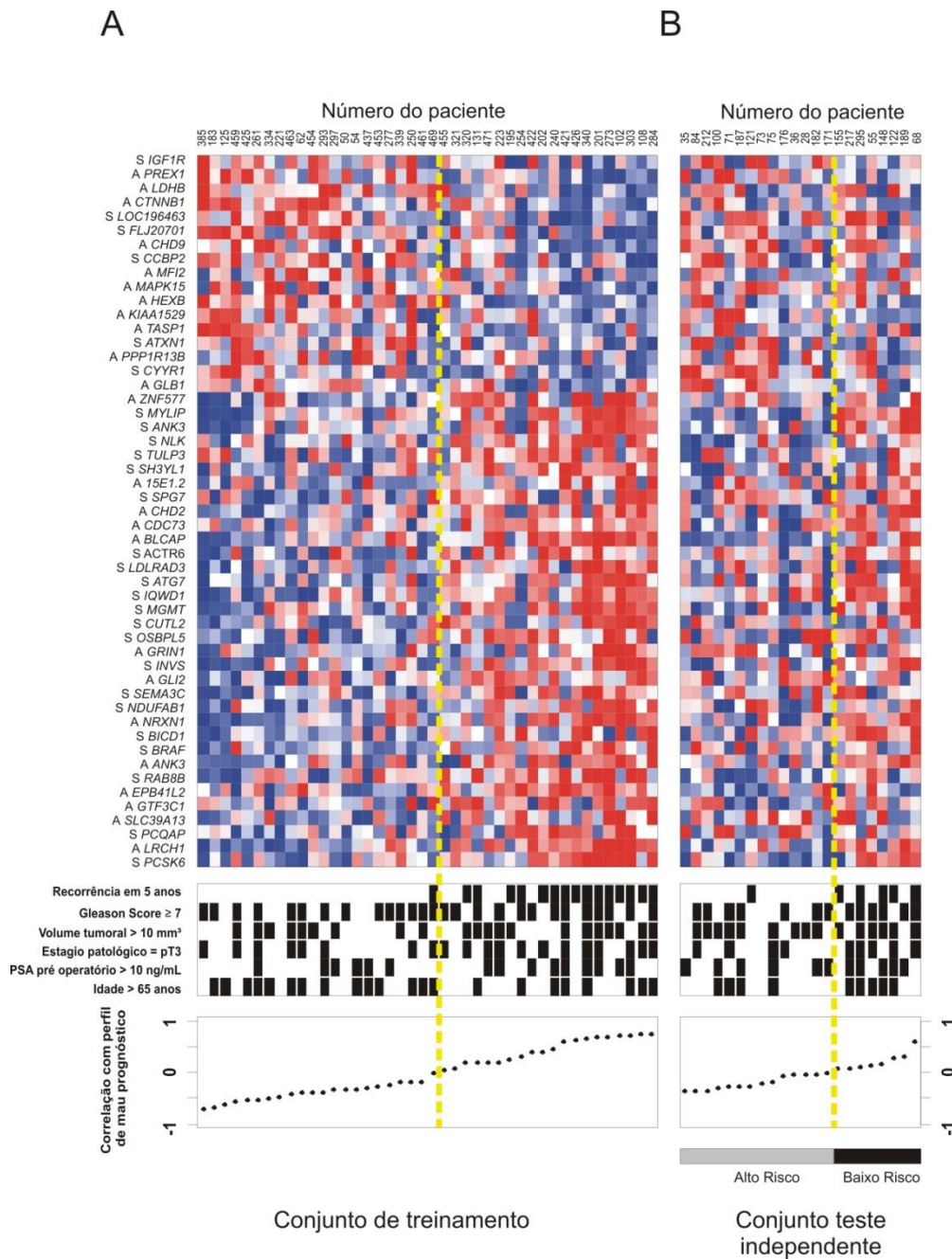


Figura 19 A-B: Perfil de expressão de ncRNAs de mau prognóstico A: ncRNAs correlacionados com recorrência bioquímica identificados no conjunto de treinamento de 42 pacientes pela estratégia de reamostragem múltipla aleatória. O *Heat-map* (painel superior) mostra o conjunto de 51 ncRNAs (linhas) correlacionados com recorrência bioquímica. Cada ncRNA (linha) em cada amostra de paciente (colunas) é representado pelo número de desvios padrão acima (vermelho) ou abaixo (verde) da média de expressão daquele ncRNA no conjunto de treinamento. O painel do meio mostra o status de recorrência de cada paciente dentro de 5 anos de acompanhamento (branco = livre de recorrência, preto: recorrente), juntamente com outros parâmetros clínicos associados com mau prognóstico (em preto): Gleason Score ≥ 7 , volume tumoral $> 10 \text{ mm}^3$, Estágio patológico = pT3, nível de PSA pré-operatório $\geq 10 \text{ ng/mL}$ e idade ≥ 65 anos. Painel inferior indica a correlação da expressão de cada paciente com a assinatura de mau prognóstico, definida como a média de expressão dos 51 ncRNAs nas 16 amostras com recorrência do conjunto de treinamento. B: Conjunto teste independente composto por 22 amostras de pacientes, que foram classificadas aplicando-se o perfil de expressão de mau prognóstico de ncRNAs. Os pacientes neste conjunto teste foram colocados no grupo de alto risco ou no grupo de baixo risco, de acordo com a correlação positiva ou negativa, respectivamente, do nível de expressão de cada paciente com o perfil médio de expressão de mau prognóstico.

4.7 Teste independente do perfil de expressão de ncRNA como um marcador de prognóstico de recorrência em PC.

A fim de testar o desempenho do perfil de expressão de mau prognóstico, nós fizemos uma "classificação cega" de um conjunto independente de 22 amostras de PC (Figura 19 B, painel superior).

Uma vez criados os perfis de classificação com as amostras de treinamento, eles foram aplicados às amostras independentes, do conjunto externo de teste. Estas amostras não foram processadas juntamente com as amostras de treinamento. Assim, a fim de que as amostras de teste sejam comparáveis com as de treinamento, foi necessária a criação de um protocolo de processamento destas amostras. Este protocolo é composto dos mesmos passos de processamento do conjunto de treinamento (normalização, ajuste de lotes e *Z-score*), mas sem alterar os valores já calculados dos parâmetros obtidos no treinamento, como descrito em "Material e Métodos". Os pacientes do conjunto de teste foram designados de alto risco de recorrência quando o perfil de expressão da amostra mostrou correlação positiva com o perfil de mau prognóstico (Figura 19 B, painel inferior).

Verificamos que o perfil gerado com transcritos intrônicos classificou corretamente 6/7 (85%) das amostras com recidiva e 8/11 (72%) das amostras sem recidiva do conjunto de teste (Figura 19 B).

É importante ressaltar que o perfil de expressão gênica de transcritos intrônicos é mais eficiente para prever a recorrência do que outras características clínicas das amostras do grupo de validação, atualmente usadas como fatores prognósticos de recorrência, tais como o nível de PSA acima de 10ng/ml de sangue, *Gleason Score* acima de 7, estadiamento tumoral e a idade do paciente (Figura 19 B, extrema direita).

Nós verificamos que apenas 1 paciente dos 14 que foram designados como do grupo de baixo risco teve recorrência bioquímica em 5 anos, enquanto que 5 dos 8 pacientes no grupo de alto risco tiveram recorrência no mesmo período (Figura 19 B, painel do meio). Assim, o perfil de expressão preditor teve uma acurácia final de 82%.

4.8 Análise de sobrevida das amostras do conjunto teste

Uma análise de Kaplan-Meier (Figura 20), com dados dos 22 pacientes do conjunto teste mostrou que as curvas de sobrevida livre de recorrência baseadas no perfil de mau prognóstico de ncRNAs eram significativamente distintas para os grupos de alto e baixo riscos (*Log-rank test* $p = 0,0009$; *Hazard ratio* = 23,4, Intervalo de confiança 95% = 3,62 a 151,2). Para pacientes classificados no grupo de baixo risco a expectativa de vida média foi de 42 meses, com probabilidade de 30% (25% a 39%) de sobrevida livre de recorrência em 60 meses após a cirurgia.

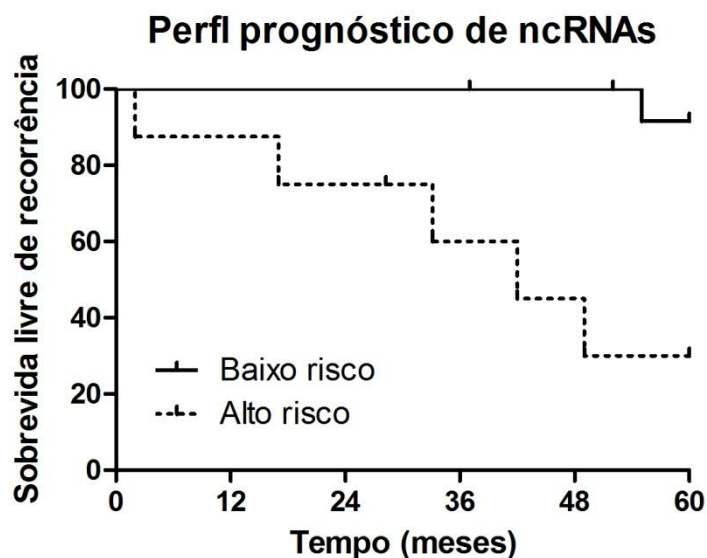
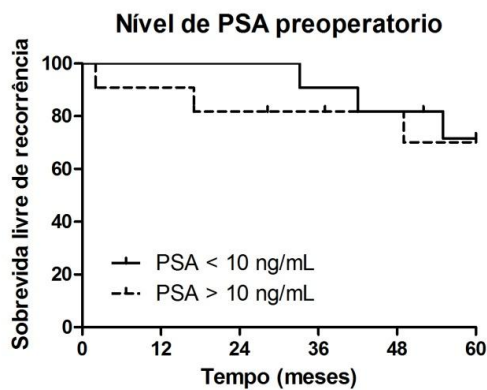


Figura 20: Análise de sobrevida para os pacientes no conjunto teste. Curvas de Kaplan-Meier para os grupos de alto e baixo risco de recorrência, definidos pela correlação de cada amostra com o perfil de mau prognóstico de ncRNAs do conjunto de treinamento.

Não houve uma relação evidente entre o perfil de expressão e os outros parâmetros clínicos ou patológicos mostrados na Figura 19 B além da recorrência, sugerindo que o perfil de expressão é um preditor independente de recorrência. Uma análise multivariada de Cox para avaliar o grau de independência não pode ser feita devido ao baixo número de amostras na coorte.

Para propósito de comparação, nós fizemos uma análise de sobrevida com o conjunto teste de 22 pacientes usando, como preditor de recorrência, o PSA pré-operativo acima de 10ng/mL (Figura 21A). A análise de Kaplan-Meier mostrou curvas sem diferença significativa (*Log-rank test p-value* = 0,79; *Hazard ratio* = 0,8025, CI 95% = 0,1599 a 4,027). A mesma falta de significância foi observada usando como preditor de recorrência o parâmetro *Gleason Score* ≥ 7 (Figura 21B) (*Log-rank teste p valor* = 1,00; *Hazard ratio* = 1,336, CI 95%, = 0,2640 a 6,764).

A



B

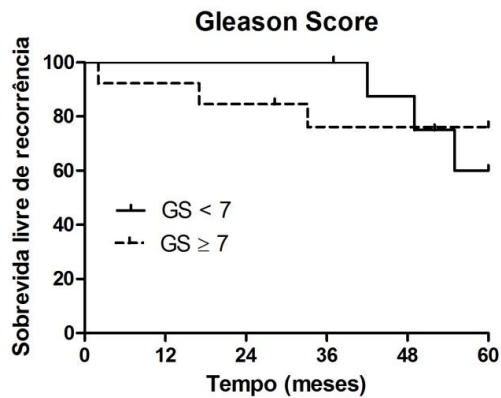


Figura 21: A: Curva de Kaplan-Meier do conjunto teste de pacientes de acordo com os níveis pré-operatórios de PSA (alto ≥ 10 ng/mL; baixo < 10 ng/mL). B: Curva de Kaplan-Meier de acordo com o *Gleason Score* ($GS \geq 7$ ou $GS < 7$).

4.9 Comparação entre os ncRNAs do perfil de genes codificadores de proteína diferencialmente expressos.

Os respectivos genes hospedeiros dos 51 ncRNAs do perfil prognóstico (Tabela 10) foram cruzados com a lista de genes codificadores de proteína encontrados como diferencialmente expressos pela análise de SAM, mostrada na Tabela 8. Não houve sobreposição entre as duas listas, exceto para o *locus* do gene *ANK3*. Ambos transcritos intrônicos (senso e antisenso) e exônico do *locus* do gene *ANK3* estão presentes no perfil prognóstico de ncRNAs (Tabela 8 e Tabela 10).

Foi realizada também a geração de um perfil de mau prognóstico baseado na média de expressão dos 98 genes codificadores identificados, no conjunto de 42 amostras de treinamento, como diferencialmente expressos pela análise de SAM, mostrada na Tabela 8. Para as 22 amostras do conjunto de teste (Figura 22) este perfil baseado nos genes codificadores de proteína conseguiu agrupar corretamente 4 das 6 amostras de pacientes com recidiva e 12 das 16 amostras sem recidiva, com uma acurácia final de 72% (16/22). A análise de sobrevida dos pacientes classificados como alto e baixo risco de acordo com este perfil prognóstico de genes codificadores de proteína (Figura 23) demonstrou que as curvas dos grupos de alto e baixo risco não chegam a ser estatisticamente distintas (*Log-rank test p-value* = 0,0588; *Hazard ratio* = 4,622, CI 95% = 0,84 a 25,43).

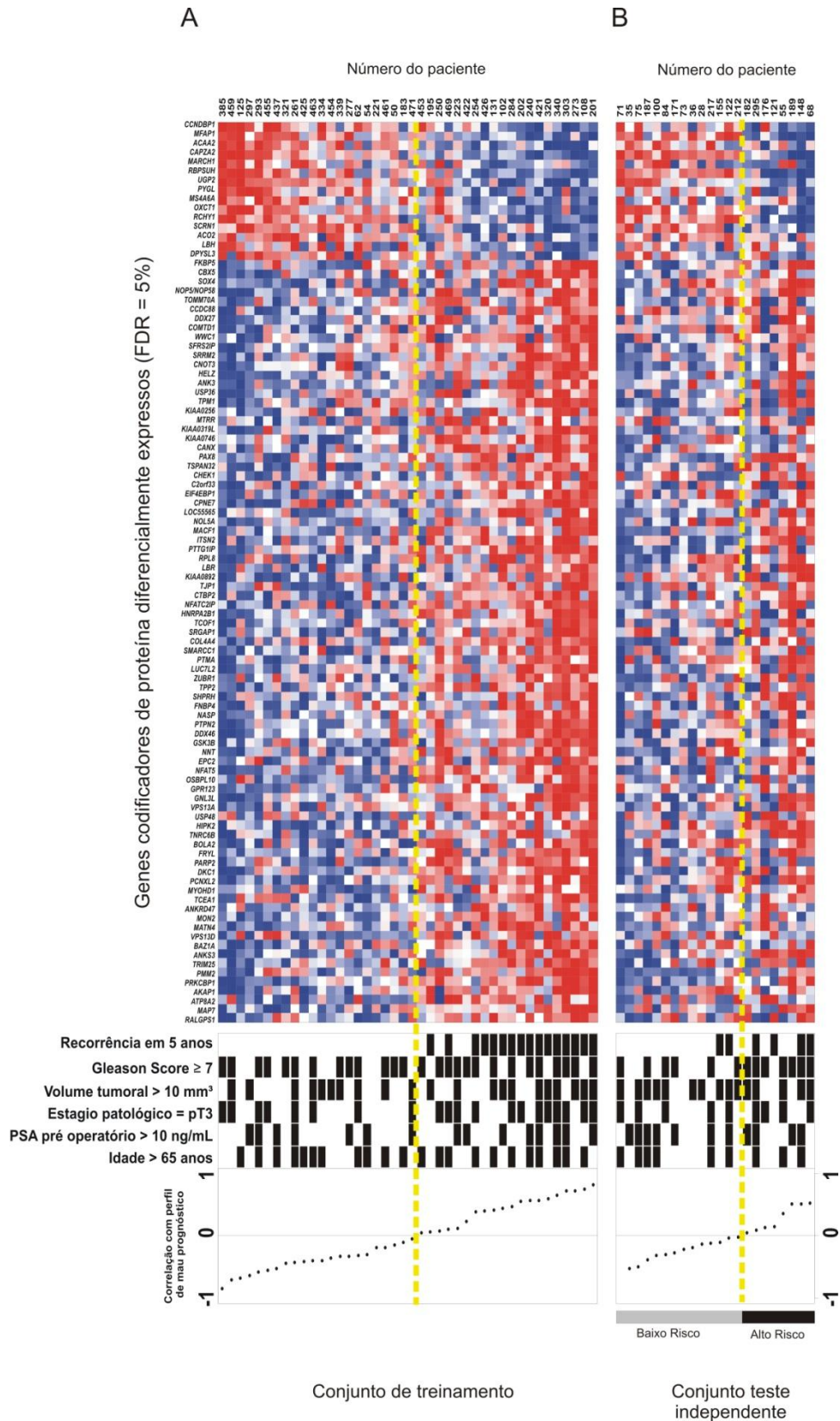


Figura 22: Perfil de expressão de genes codificadores de proteína diferencialmente expressos segundo a estratégia de SAM. A: Genes diferencialmente expressos entre pacientes com recorrência e sem recorrência do conjunto de treinamento de 42 pacientes. B: Conjunto teste independente composto por 22 amostras de pacientes, que foram classificadas aplicando-se o perfil de expressão de mau prognóstico de genes codificadores de proteína.

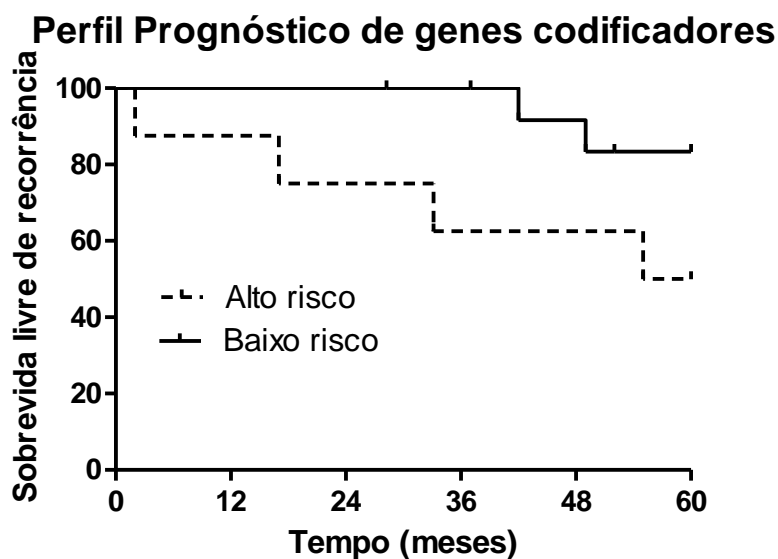


Figura 23: Análise de sobrevida para os pacientes no conjunto teste. Curvas de Kaplan-Meier para os grupos de alto e baixo risco de recorrência, definidos pela correlação de cada amostra com o perfil de mau prognóstico baseado na expressão dos genes codificadores de proteína diferencialmente expressos entre pacientes com recorrência e sem recorrência do conjunto de treinamento identificados pela estratégia de SAM.

A fim de comparar duas estratégias de seleção de genes para a geração do perfil classificador, (SAM contra a estratégia de reamostragem aleatória), nós realizamos a análise dos genes codificadores de proteína utilizando a estratégia de reamostragem aleatória, previamente utilizada para os ncRNAs intrônicos, utilizando os mesmos critérios já descritos. Com essa abordagem, nós identificamos 64 genes codificadores de proteína (Figura 24). Destes, 27 já haviam sido identificados com a estratégia de SAM e 37 foram novos genes identificados (Tabela 11). A análise de sobrevida dos pacientes classificados como alto e baixo risco utilizando este perfil de 63 genes demonstrou curvas estatisticamente distintas entre os dois grupos (*Log-rank test* $p = 0,0301$; *Hazard ratio* = 7,321, Intervalo de confiança 95% = 1,211 a 44,27) (Figura 25).

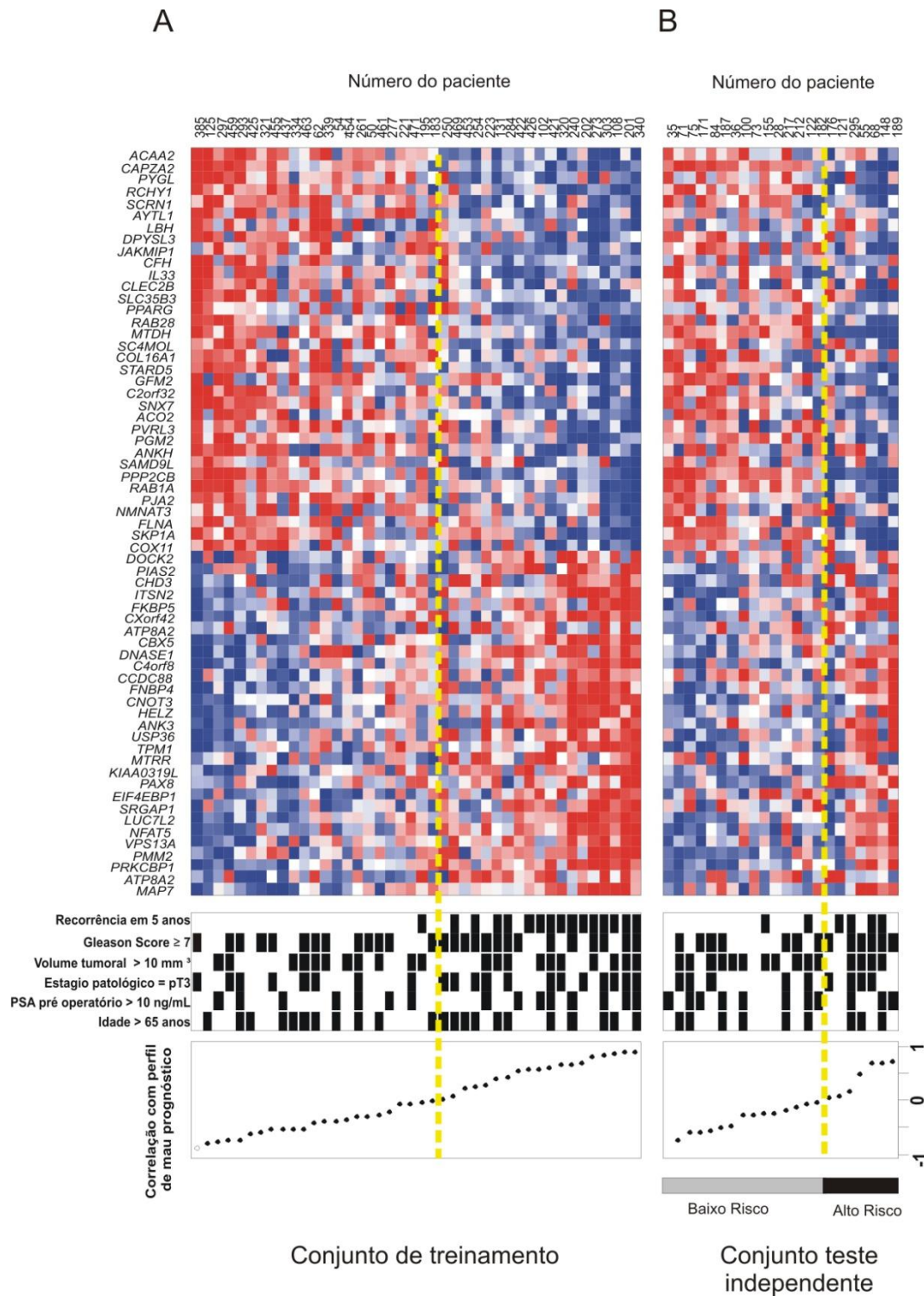


Figura 24: Perfil de expressão de genes codificadores de proteína pela estratégia reamostragem múltipla aleatória A: Genes diferencialmente expressos entre pacientes com recorrência e sem recorrência do conjunto de treinamento de 42 pacientes pela estratégia reamostragem múltipla aleatória. B: Conjunto teste independente composto por 22 amostras de pacientes, que foram classificadas aplicando-se o perfil de expressão de mau prognóstico de genes codificadores de proteína.

Tabela 11: Genes codificadores de proteína identificados pela estratégia de múltipla amostragem. *Genes também identifiados pela estratégia de SAM.

ID do probe	Símbolo do gene codificador	Descrição	Veze em que aparece como top 100 nas 500 permutações
Q1_P02080	ACAA2*	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	500
Q1_P17526	ACO2*	aconitase 2, mitochondrial	268
Q1_P07462	ANK3*	ankyrin 3, node of Ranvier (ankyrin G)	353
Q1_P04511	ANKH	ankylosis, progressive homolog (mouse)	268
Q1_P03774	ATP8A2*	ATPase, aminophospholipid transporter-like, Class I, type 8A, member 2	485
Q1_P04016	AYTL1	acyltransferase like 1	459
Q1_P03097	C2orf32	chromosome 2 open reading frame 32	394
Q1_P01168	C4orf8	chromosome 4 open reading frame 8	324
Q1_P01584	CAPZA2*	capping protein (actin filament) muscle Z-line, alpha 2	500
Q1_P02929	CBX5*	chromobox homolog 5 (HP1 alpha homolog, Drosophila)	279
Q1_P08602	CCDC88*	coiled-coil domain containing 88	333
Q1_P22023	CFH	complement factor H	365
Q1_P15898	CHD3	chromodomain helicase DNA binding protein 3	256
Q1_P14602	CLEC2B	C-type lectin domain family 2, member B	416
Q1_P16632	CNOT3*	CCR4-NOT transcription complex, subunit 3	364
Q1_P13341	COL16A1	collagen, type XVI, alpha 1	354
Q1_P03137	COX11	COX11 homolog, cytochrome c oxidase assembly protein (yeast)	260
Q1_P06530	CXorf42	chromosome X open reading frame 42	273
Q1_P15602	DNASE1	deoxyribonuclease I	265
Q1_P24593	DOCK2	dedicator of cytokinesis 2	266
Q1_P04615	DPYSL3*	dihydropyrimidinase-like 3	448
Q1_P05021	EIF4EBP1*	eukaryotic translation initiation factor 4E binding protein 1	422
Q1_P00163	FKBP5	FK506 binding protein 5	299
Q1_P06463	FLNA	filamin A, alpha (actin binding protein 280)	295
Q1_P22319	FNBP4*	formin binding protein 4	355
Q1_P18369	GFM2	G elongation factor, mitochondrial 2	360
Q1_P23357	HELZ*	helicase with zinc finger	359
Q1_P08252	IL33	interleukin 33	432
Q1_P03483	ITSN2*	intersectin 2	279
Q1_P06483	JAKMIP1	janus kinase and microtubule interacting protein 1	422
Q1_P21822	KIAA0319L*	KIAA0319-like	390
Q1_P05498	LBH*	limb bud and heart development homolog (mouse)	436
Q1_P19237	LUC7L2*	LUC7-like 2 (S. cerevisiae)	477
Q1_P24767	MAP7*	microtubule-associated protein 7	500
Q1_P05978	MTDH	metadherin	359
Q1_P01526	MTRR*	5-methyltetrahydrofolate-homocysteine methyltransferase reductase	365
Q1_P02195	NFAT5*	nuclear factor of activated T-cells 5, tonicity-responsive	476

ID do probe	Símbolo do gene codificador	Descrição	Veze em que aparece como top 100 nas 500 permutações
Q1_P16915	PAX8*	paired box gene 8	440
Q1_P04191	PGM2	phosphoglucomutase 2	329
Q1_P01345	PIAS2	protein inhibitor of activated STAT, 2	272
Q1_P08898	PJA2	praja 2, RING-H2 motif containing	279
Q1_P15624	PMM2*	phosphomannomutase 2	499
Q1_P05112	PPARG	peroxisome proliferator-activated receptor gamma	365
Q1_P19365	PPP2CB	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	297
Q1_P17299	PRKCBP1*	protein kinase C binding protein 1	498
Q1_P00624	PVRL3	poliovirus receptor-related 3	335
Q1_P06655	PYGL*	phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI)	500
Q1_P16819	RAB1A	RAB1A, member RAS oncogene family	291
Q1_P01467	RAB28	RAB28, member RAS oncogene family	441
Q1_P08889	RCHY1*	ring finger and CHY zinc finger domain containing 1	486
Q1_P06691	SAMD9L	sterile alpha motif domain containing 9-like	313
Q1_P08042	SC4MOL	sterol-C4-methyl oxidase-like	362
Q1_P08294	SCRN1*	secernin 1	487
Q1_P08326	SKP1A	S-phase kinase-associated protein 1A (p19A)	293
Q1_P18639	SLC35B3	solute carrier family 35, member B3	405
Q1_P03650	SNX7	sorting nexin 7	367
Q1_P04716	SRGAP1*	SLIT-ROBO Rho GTPase activating protein 1	452
Q1_P00425	STARD5	START domain containing 5	374
Q1_P00427	TPM1*	tropomyosin 1 (alpha)	317
Q1_P16199	USP36*	ubiquitin specific peptidase 36	357
Q1_P02749	VPS13A*	vacuolar protein sorting 13 homolog A (S. cerevisiae)	495

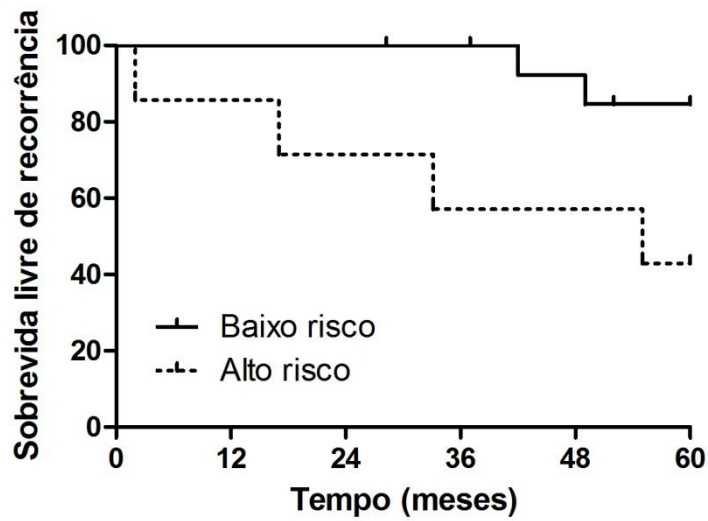


Figura 25: Análise de sobrevida para os pacientes no conjunto teste. Curvas de Kaplan-Meier para os grupos de alto e baixo risco de recorrência, definidos pela correlação de cada amostra com o perfil de mau prognóstico baseado na expressão dos genes codificadores de proteína diferencialmente expressos entre pacientes com recorrência e sem recorrência do conjunto de treinamento identificados pela estratégia de reamostragem múltipla aleatória.

5 Discussão

Utilizando a estratégia de múltipla amostragem treinamento-validação proposta por Michiels et al. (Michiels *et al.*, 2005), nós identificamos um conjunto de ncRNAs que é correlacionado com o status de recorrência do câncer de próstata em 5 anos de acompanhamento após a prostatectomia. Uma taxa de classificação errada abaixo de 0,5 foi obtida nas análises de reamostragem múltipla (0,424, 95 % CI = 0,434 a 0,414). O limite superior do intervalo de confiança encontrado é de 0,434, ou seja, menor que 0,5, para a taxa de classificação errada. Isto sugere uma habilidade significativamente melhor do classificador molecular que o esperado ao acaso (Michiels *et al.*, 2005). Esta análise mostra que um preditor robusto de recorrência bioquímica de PC pôde ser obtido com nosso oligoarranjo e nossa coorte de pacientes.

Nosso perfil preditor foi testado em um conjunto teste independente de 22 pacientes, mostrando uma classificação errada de 1 paciente dos 14 designados como pertencentes ao grupo de baixo risco (taxa de falso negativo de 7%). E uma classificação errada de 3 pacientes de 8 designados como pertencentes ao grupo de alto risco (taxa de falso positivo de 38%). Uma análise de Kaplan-Meier do conjunto independente mostrou um risco significativamente maior (23,4 vezes, com 95% CI = 3,6 a 151,2; teste *Log-rank* p-valor = 0,0009) de recorrência bioquímica para pacientes colocados no grupo de alto risco em oposição ao grupo de baixo risco, confirmando assim que este classificador molecular pode ser útil para identificar pacientes com alto risco de recorrência (Figura 20).

O perfil prognóstico baseado em 98 genes codificadores de proteína identificados pela análise estatística de SAM teve um desempenho pior do que o baseado no perfil de ncRNAs, mostrando uma diferença não significativa entre as curvas de alto e baixo risco ($p = 0,0588$ para os genes codificadores de proteína). A análise de sobrevida dos pacientes classificados como alto e baixo risco utilizando o perfil de 63 genes codificadores de proteína identificados

pela estratégia de múltipla amostragem demonstrou curvas estatisticamente distintas entre os dois grupos (*Log-rank test* $p = 0,0301$; *Hazard ratio* = 7,321, Intervalo de confiança 95% = 1,211 a 44,27) (Figura 25), mostrando que a estratégia de múltipla amostragem é mais adequada para a seleção de genes com poder prognóstico. O conjunto de dados indica que ambos transcritos codificadores e não codificadores identificados nas análises têm potencial como biomarcadores de recorrência para câncer de próstata.

É interessante notar que não foi observado um desempenho melhor ao se aplicar ambos os ncRNAs e os genes codificadores de proteína em um único perfil marcador (*Log-rank test* $p = 0,1669$; *Hazard ratio* = 3,648, 95% Intervalo de confiança = 0,6929 a 19,21), mostrando que existe um ótimo de genes no perfil marcador para se obter um fator prognóstico significativo.

A recorrência bioquímica após a prostatectomia radical é um evento documentado na clínica, e a chance de recorrência aumenta conforme a função de um número de fatores clássicos como estágio clínico avançado, Gleason Score, nível de PSA pré-operatório e estágio patológico (Han *et al.*, 2001). Foi mostrado que níveis de PSA acima de 17 ng/mL são bem correlacionados com a recorrência bioquímica, entretanto, para níveis de PSA entre 2 e 10 ng/mL não há correlação entre PSA pré-operativo e a recorrência bioquímica após a prostatectomia radical (Stamey, 2001). A esse respeito, é notável que nós não encontramos nenhuma correlação significativa entre os níveis de PSA acima de 10 ng/mL e a recorrência bioquímica em nossa coorte teste (Figura 19 painel do meio e Figura 21 A). Deve ser notado que metade de nosso grupo teste (11 de 22 = 50%), tiveram PSA pré-operatório entre 2 e 10 ng/mL e que aproximadamente metade dos pacientes restantes (5/11) tinham PSA pré-operatório entre 2 e 10 ng/mL (Tabela 4). É possível que a falha do PSA pré-operatório em prever a recorrência em nossa coorte possa ser devido a alta proporção de pacientes com níveis de PSA abaixo de 17 ng/mL. De fato, com o amplo uso de teste de PSA e exame retal é

esperado que a proporção de homens diagnosticados com PC inicial continuará a crescer. Neste cenário, é importante apontar que, apesar da falha no PSA pré operatório como fator de prognóstico em nosso conjunto de teste, o perfil de expressão prognóstico de ncRNAs foi capaz de estatisticamente prever o risco de recorrência (Figura 19, painel do meio e Figura 20).

O Gleason Score tem sido demonstrado ao longo dos anos como sendo um preditor poderoso de recorrência de câncer de próstata após a prostatectomia radical (Epstein *et al.*, 1996; Lerner *et al.*, 1996). A importância de um grau terciário de Gleason tem sido demonstrada apenas recentemente (Pan *et al.*, 2000), e uma revisão (Harnden *et al.*, 2007) que computou um taxa de risco de quatro trabalhos publicados mostrou que a presença de um grau terciário aumenta o risco de recorrência bioquímica após a prostatectomia radical cerca de 2,5 vezes. A frequência de um grau terciário de qualquer padrão no único trabalho publicado com uma coorte não selecionada foi de 48% (106 de 223) (van Oort *et al.*, 2005). Assim, é possível que a presença de um grau terciário nas amostras de nossa coorte contribua para a heterogeneidade e para a correlação não significativa entre o GS e o desfecho.

Dentre os genes codificadores de proteína diferencialmente expressos entre pacientes com recorrência e sem recorrência, quatro genes, *EIF4EBP1*, *PARP*, *HIPK2* e *GSK3B* já foram descritos como relacionados com PC. A regulação positiva do gene *EIF4EBP1* já foi associada com PC em tumores expressando *MYC* (Balakumaran *et al.*, 2009). Para o gene *PARP2*, um inibidor para a proteína produto do gene, chamado ABT-888, está em ensaio clínico de fase I para o tratamento de PC (Penning *et al.*, 2009). O aumento de expressão do gene *HIPK2* já foi associado com PC (Capra *et al.*, 2006). Para o gene *GSK3B*, um inibidor para a proteína produto do gene, chamado enzastaurina, está em ensaio clínico de fase II para o tratamento de câncer de próstata. Estes resultados sugerem que estes genes estão não só envolvidos na transformação tumoral como também na agressividade do tumor.

Entre os genes codificadores de proteínas encontrados como diferencialmente expressos em nosso trabalho, buscamos aqueles que já tivessem sido descritos como biomarcadores de recorrência de PC nos cinco trabalhos publicados sobre biomarcadores prognósticos em PC (Singh *et al.*, 2002; Glinsky *et al.*, 2004; Lapointe *et al.*, 2004; Yu *et al.*, 2004; Varambally *et al.*, 2005; Nanni *et al.*, 2006; Nakagawa *et al.*, 2008). O gene *Microtubule-associated protein 7 (MAP7)* já foi descrito como aumentado em PC em um trabalho (Nanni *et al.*, 2006). Já o gene *sel-1 suppressor of lin-12-like 3 (SEL1L3)* já foi descrito como biomarcador de recorrência em PC (Varambally *et al.*, 2005), embora tenha sido descrito como diminuído em pacientes com recorrência. Grande parte da variação entre os estudos pode ser explicada pelo uso de diferentes plataformas de arranjos, contendo sequências sondas diferentes e tipos diferentes de sondas (cDNA versus oligonucleotídeos), variação na preparação das amostras (microdissecção a laser ou aparagem crua), e protocolos de marcação (ex. com ou sem amplificação adicional de RNA), bem como o número limitado de amostras de pacientes de PC analisados na maioria dos estudos. É interessante notar que a comparação entre estes 5 trabalhos da literatura mostra muito pouca sobreposição de genes marcadores, isto é, somente 8 genes, cada um presente em apenas dois trabalhos, dos 5 da literatura (Tabela 12). Em sete dos oito casos o mesmo gene é identificado com variações opostas entre dois trabalhos da literatura, para o mesmo desfecho da doença do paciente (Tabela 12). Apenas o gene *IGFBP3*, presente nas assinaturas dos trabalhos de (Nakagawa *et al.*, 2008) e (Singh *et al.*, 2002), é coincidente em ambos trabalhos, apresentando a expressão diminuída em pacientes com recorrência (Tabela 12).

Tabela 12: Sobreposição de genes diferencialmente expressos entre pacientes com recorrência e sem recorrência nos 5 trabalhos da literatura e com o presente trabalho.

Símbolo do gene	(Singh <i>et al.</i> , 2002)	(Nanni <i>et al.</i> , 2006)	(Nakagawa <i>et al.</i> , 2008)	(Lapointe <i>et al.</i> , 2004)	(Glinsky <i>et al.</i> , 2004)	Presente trabalho
COL1A2	diminuído				aumentado	
GPR137B			diminuído		aumentado	
HOXC6	diminuído				aumentado	
IGFBP3	diminuído		diminuído			
MAP7		aumentado				aumentado
MYBPC1			aumentado		diminuído	
NIPA2		diminuído			aumentado	
PAGE4			aumentado		diminuído	
PTPRN2			aumentado		diminuído	
SEL1L3					diminuído	aumentado

Os 98 genes codificadores de proteína diferencialmente expressos entre amostras de pacientes com recorrência e sem recorrência com uma taxa de falsa descoberta de 5% (Tabela 8) foram usados para buscar redes gênicas estatisticamente alteradas. Os genes identificados foram mapeados em redes gênicas disponíveis no banco de dados do programa IPA e foram ordenados pelo *score*. O *score* mede a probabilidade de uma coleção de genes igual ou maior que o número presente em uma rede ser obtida ao acaso. Um *score* de 3 indica que existe uma chance de 10^{-3} (ou seja 0,001) de que os genes estejam presentes na rede ao acaso. Este *score* foi utilizado como critério para a identificação de redes gênicas estatisticamente alteradas (Tabela 13). As funções biológicas mais relevantes encontradas para as três redes gênicas mais enriquecidas foram *Cell Morphology*, *Cellular Function and Maintenance*, *Developmental Disorder* para a rede 1 (Figura 26) e *Cellular Development*, *Cellular Growth and Proliferation*, *Endocrine System Development and Function* para a rede 2 (Figura 27).

Tabela 13: Redes gênicas estatisticamente enriquecidas. ↑ = Aumentado em pacientes com recorrência. ↓ = Diminuído em pacientes com recorrência.

ID	Moléculas na rede	Score	Genes diferencialmente expressos	Funções mais relevantes
1	Akt, ↑ANK3, ↑CANX, ↑CBX5, ↓CCNDBP1, ↑CHEK1, Ck2, Creb, ↑CTBP2, Cyclin D, ↓DPYSL3, E2f, ↑EIF4EBP1, ↑FKBP5, FNBP4, ↑GSK3B, ↑HELZ, Histone h3, ↑HNRNPA2B1, ↑LBR, ↑LUC7L2, ↑MACF1, ↓MFAP1, ↑NASP, ↑NFATC2IP, NFkB (complex), PI3K, ↑PTMA, ↑PTPN2, RNA polymerase II, ↑SOX4, ↑SRRM2, ↑TCEA1, ↑TCOF1, ↑WWC1	54	26	<i>Cell Morphology, Cellular Function and Maintenance, Developmental Disorder</i>
2	↓ACAA2, ↓ACO2, APC, BAD, beta-estradiol, ↓CAPZA2, CAPZB, CDK4, CNTF, ↑CPNE7, DLG4, ↑GPR123, GRB2, HGS, HNRNPK, HTR2C, lfn, ↑ITSN2, KCNMA1, MIR124, MIR124-1 (includes EG:406907), MIR124-2 (includes EG:406908), MIR124-3 (includes EG:406909), NKX2-1, ↑NOP56, ↓OXCT1, PCBP1 (includes EG:5093), ↑PMM2, ↑PTTG1IP, ↑RALGPS1, ↓SCRN1, ↑TRIM25, UBE2N, UGP2, ↑USP48	26	15	<i>Cellular Development, Cellular Growth and Proliferation, Endocrine System Development and Function</i>
3	BAZ1B, ↑COL4A4 (includes EG:1286), COPS5, CRP, Cyclin D, ↑DDX46, ERK1/2, FHL1, FYN, ↑HIPK2, ITGB4, KLF5, KLK3, ↓LBH, ↑MATN4, MIF, ↑MON2, ↑NFAT5, Nfat (family), NKX2-1, ↑PARP2, PCNA, POLH, PRKAC, RARRES3, ↑SEL1L3, ↑SHPRH, ↑SMARCC1, TNF, TNFRSF1A, TP53, TP53INP1, ↑TPP2, VCL, ↑VPS13A	19	13	<i>Gene Expression, Cancer, Genetic Disorder</i>
4	Actin, AKAP13, BAT1, CEBPZ, CIR1, ↑CNOT3, CREBBP, EIF2C4, FHL1, ↑FRYL, HEY1, ↑HIPK2, HNF4A, MAML1, ↑MAP7, MFNG, MLXIPL, NLK, RBM15, RBPJ, retinoic acid, RORC, ↑RPL8, ↑SFERS2IP, SND1, SP100, SPEN (includes EG:23013), SWI-SNF, TIP60, ↑TJP1, T↑NRC6B, ↑TPM1, ↑USP36, WSB1, YWHAB	18	11	<i>Gene Expression, Skeletal and Muscular System Development and Function, Cellular Development</i>

5	↑AKAP1, ↑ANKS3, CCDC85B, COL18A1, ↑DKC1, DUSP1, ↑EPC2, GATA1, GFI1B, HES1, IL17A (includes EG:3605), KAT2A, KLK3, LGALS3, LMNA, MLXIPL, ↑MTRR, NKX2-1, ↑NOP58, ↑PAX8, PLAGL2, PTEN, ↓PYGL, ↓RCHY1, RORC, RPS27A (includes EG:78294), RPS6KA1, RUNX1, RUVBL2, SUZ12, TGFB1, TIP60, ↑TOMM70A, VEGFA, ↑ZNF821	17	11 <i>Cardiovascular System Development and Function, Cellular Development, Cell Cycle</i>
6	AR, ARRB2, ↑BAZ1A, BCL6, ↑BOLA2, Calmodulin, CAMK2A, ↑CCDC88B, CREBBP, cyclic AMP, ↑DDX27, EP300, HRAS, HTT, IGFBP3, INS1, KAT2B, L- triiodothyronine, MAPK8, MDM2, N- cor, NCOR1, NCOR2, Pka, PRKCB, SATB1, SMAD2, SREBF2, ↑SRGAP1, TBP, TRPV4, UBR4, YWHAG, Z↑MYND8	9	7 <i>Gene Expression, Cell Cycle, Infection Mechanism</i>

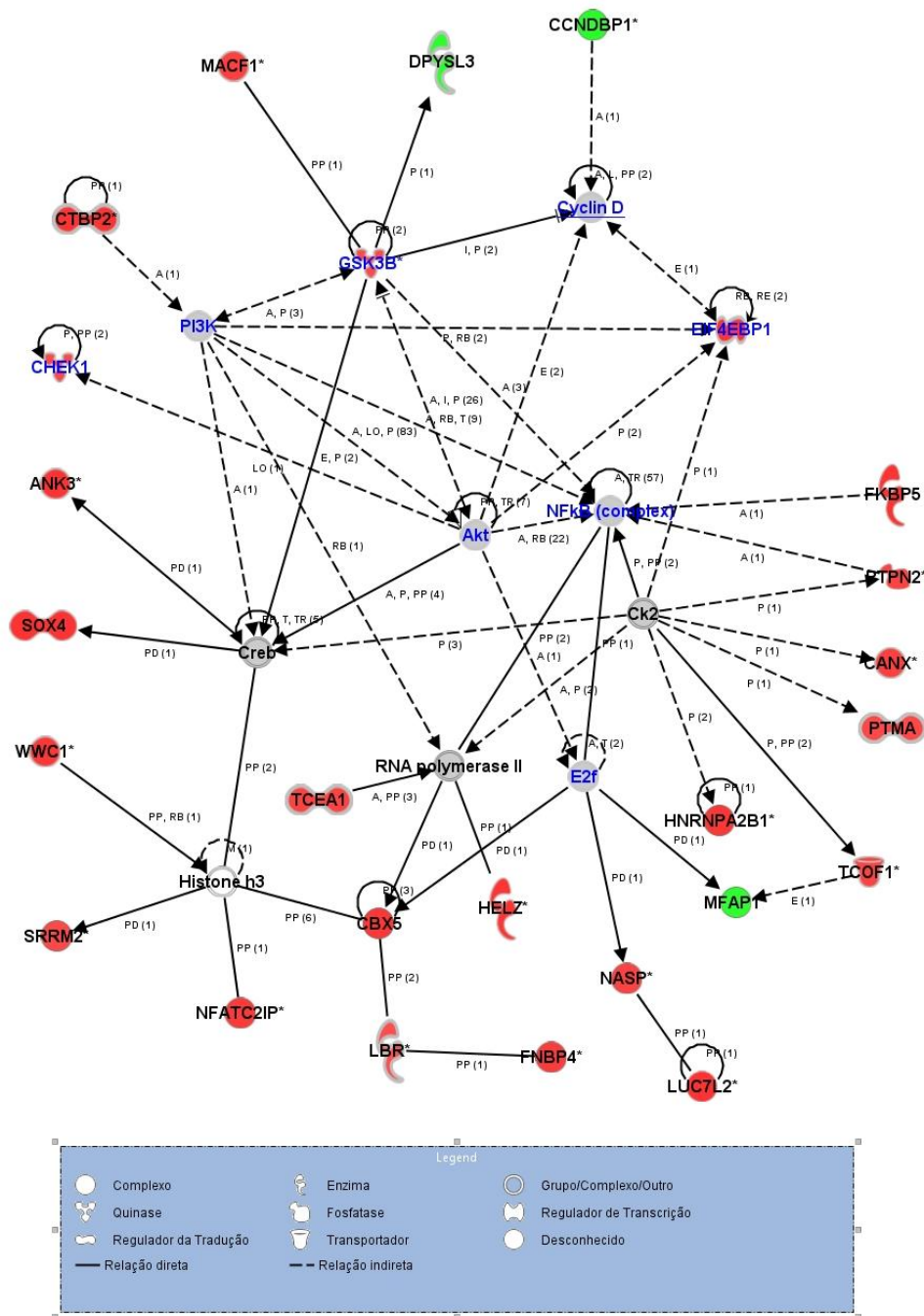


Figura 26: Rede gênica mais enriquecida com genes diferencialmente expressos entre amostras de pacientes com recorrência e sem recorrência. Genes aumentados em pacientes com recorrência estão em vermelho; genes diminuídos estão em verde; genes presentes no microarranjo sem alteração estão em cinza; genes ausentes no microarranjo estão em branco. Relações diretas estão marcadas em linhas contínuas; relações indiretas estão marcadas em linhas tracejadas; Próximo a cada relação estão letras indicando o tipo de relação entre as moléculas: A - Ativação; P - Fosforilação; PP - Ligação proteína-proteína; PD - Ligação proteína-DNA; E - Aumento de expressão; RB - Regulação de Ligação. Em parênteses estão os números de referências na literatura que indicam a relação envolvida. Genes descritos como correlacionados com câncer de próstata estão em azul. Ciclina D, descrita como relacionada a prognóstico de câncer de próstata, está sublinhado.

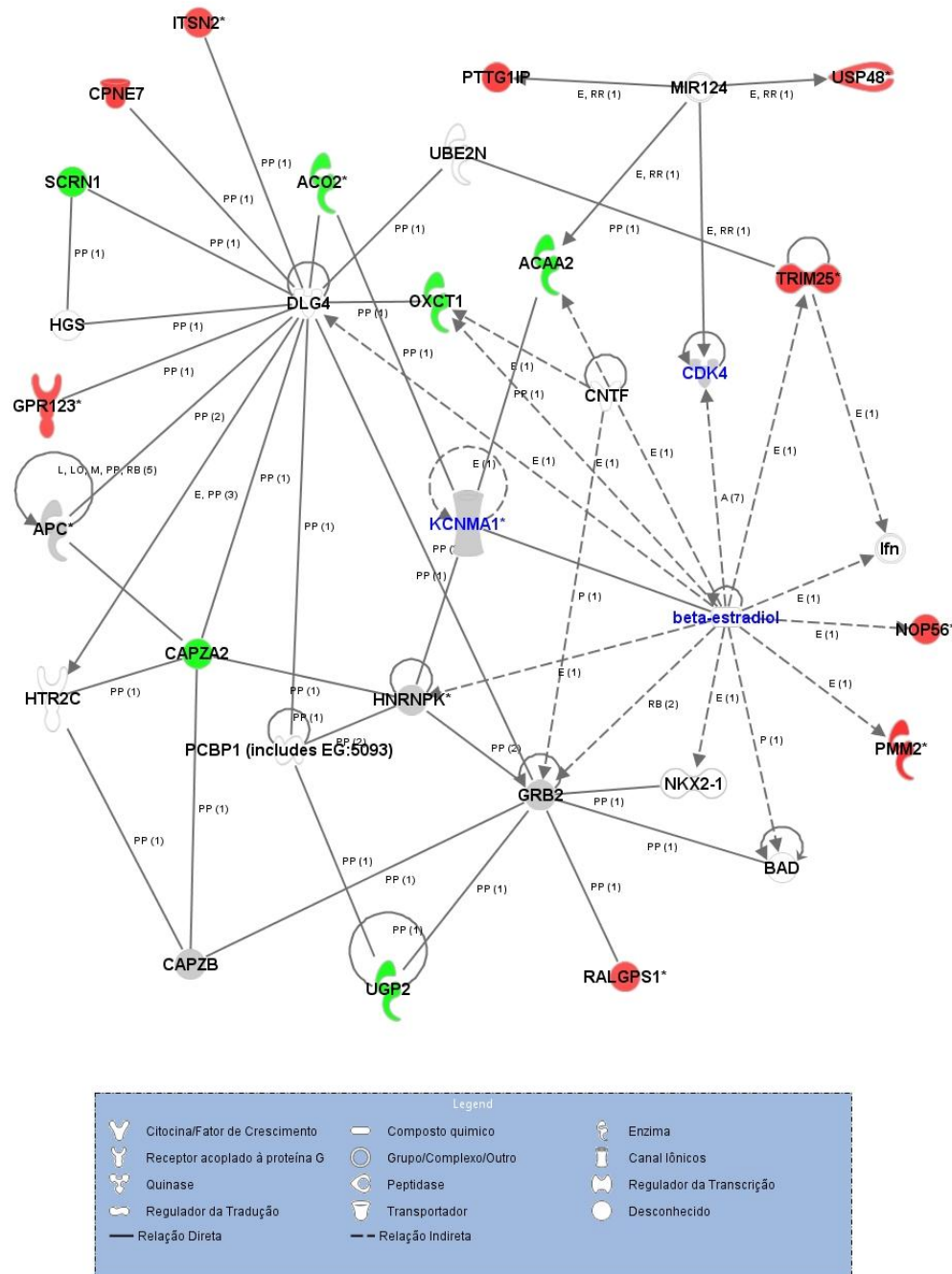


Figura 27: Segunda rede gênica mais enriquecida com genes diferencialmente expressos entre amostras de pacientes com recorrência e sem recorrência. Genes aumentados em pacientes com recorrência estão em vermelho; genes diminuídos estão em verde; genes presentes no microarranjo sem alteração estão em cinza; genes ausentes no microarranjo estão em branco. Relações diretas estão marcadas em linhas contínuas; relações indiretas estão marcadas em linhas tracejadas; Próximo a cada relação estão letras indicando o tipo de relação entre as moléculas: A - Ativação; P - Fosforilação; PP - Ligação proteína-proteína; PD - Ligação proteína-DNA; E - Aumento de expressão; RB - Regulação de Ligação. Em parênteses estão os números de referências na literatura que indicam a relação envolvida. Genes descritos como correlacionados com câncer de próstata estão em azul.

É interessante notar que genes já descritos como alterados em PC, que não foram encontrados como diferencialmente expressos, foram incluídos nas redes gênicas, pois possuem relações relevantes com genes encontrados como diferencialmente expressos. O gene *PIK3* interage com o gene *GSK3B*, encontrado como aumentado em pacientes com recorrência. Ambos fazem parte da via de sinalização do câncer de próstata (Figura 28).

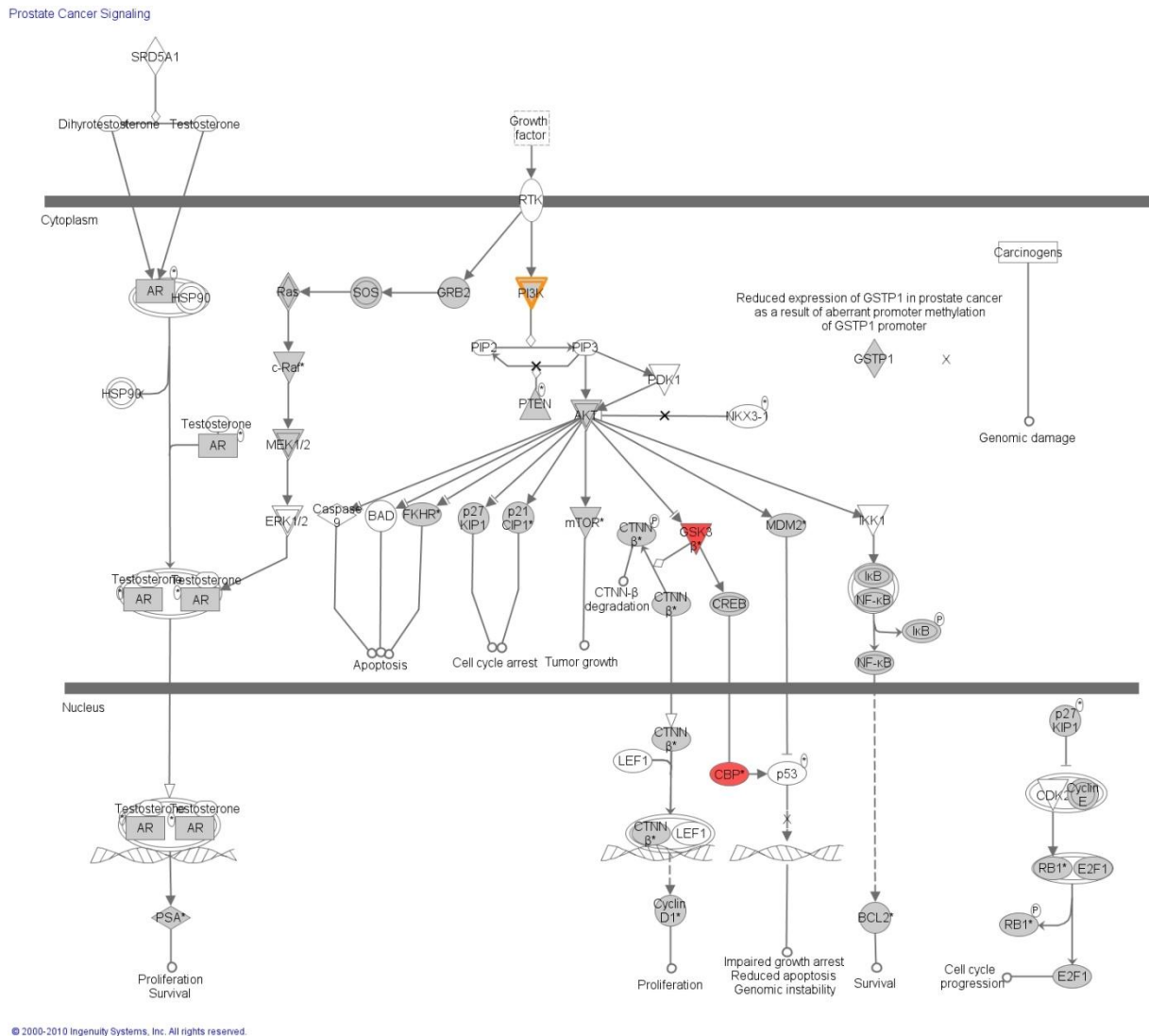


Figura 28: Via de sinalização do câncer de próstata. (Ingenuity Pathway Analysis). Genes em cinza não tiveram a expressão alterada. Genes em branco não estão presentes no array. GSK3 (vermelho) foi encontrado com aumentado em pacientes com recorrência (FDR = 5%). CBP (CREBP), também foi encontrado como diferencialmente expresso, embora com uma significância menor (FDR=10%).

Já os genes Ciclina D, *Akt*, *E2f* e *NFK-b*, embora não encontrados como diferencialmente expressos entre os pacientes com recorrência e sem recorrência, são pontos centrais de interação de vários genes diferencialmente expressos (Figura 26). O mesmo ocorre para os genes *CDK4* e *KCNMA1*, presentes na segunda rede gênica mais enriquecida (Figura 27). A alteração destes genes pode não estar na diferença de expressão deles, mas sim de modificações pós-transcricionais, como fosforilação e/ou degradação de proteína, relacionada ao desfecho no PC.

Entre os 51 ncRNAs presentes em nosso perfil prognóstico, quatro são transcritos a partir de regiões intrônicas de genes envolvidos em regulação da transcrição: *IGLI2*, *GTF3C1*, *INVS* e *CTNNB1* (Tabela 10). Estes ncRNAs podem afetar a função das vias controladas por esses genes reguladores de transcrição (Louro *et al.*, 2009; Mercer *et al.*, 2009), em analogia ao efeito do ncRNA intrônico longo do gene *CFTR* que atua sobre outros genes relacionados a função do gene codificador de proteína *CFTR* (Hill *et al.*, 2006). Isto pode levar a células tumorais mais agressivas e eventualmente refletir na recorrência do PC. A esse respeito o ncRNA do locus *CTNNB1* foi encontrado aqui correlacionado com recorrência enquanto o fator de transcrição *CTNNB1* é um membro da via de sinalização de PC e tem sido usado como biomarcador para diagnóstico do carcinoma de próstata (Huang *et al.*, 2007).

É notável que não houve sobreposição entre os *loci* genômicos dos 51 ncRNAs do perfil prognóstico e os *loci* dos 98 genes codificadores de proteína mais altamente correlacionados com a recorrência medidas em paralelo em nosso conjunto de 42 amostras, com exceção do *locus* do gene *ANK3*. Isto indica que a expressão de ncRNA não é uma desregulação generalizada em seu *locus* genômico, mas sim sugere uma regulação transcricional independente do ncRNA e dos mRNAs do mesmo *locus*. Além disso, a ausência de pares corregulados de ncRNAs e mRNAs do mesmo *locus* favorece a noção de

que qualquer efeito regulatório veiculado deve ocorrer em *trans*. Já para o locus *ANK3*, nós encontramos que ambos os transcritos senso e antissenso intrônicos estavam alterados nas amostras de pacientes com recorrência. Além disso, o gene codificador *ANK3* é altamente correlacionado com PC em nosso conjunto de amostras. É interessante notar que o gene codificador *ANK3* está presente em uma assinatura de 11 genes de PC primário relatada por Glinsky et al., que prediz a recorrência da doença em intervalo curto, metástase distante e morte após a terapia (Glinsky *et al.*, 2005). A correlação conhecida entre o gene codificador de proteína *ANK3* e a recorrência em PC pode ser uma consequência de uma regulação em cis, como descrito para outros ncRNAs longos (Sleutels *et al.*, 2002; Pauler *et al.*, 2007), ou neste caso uma alta atividade transcricional do *locus* genômico inteiro.

A observação de que os níveis de expressão de um número considerável de ncRNAs originados de regiões intrônicas do genoma são correlacionados com a recorrência de PC justifica uma investigação mais profunda das possíveis funções exercidas por essa classe de transcritos.

6 Conclusão

Nossas descobertas sugerem um papel potencial de ncRNAs longos na progressão de PC e apontam para os ncRNAs longos intrônicos como novos marcadores moleculares de recorrência de PC. A análise aqui apresentada ainda é limitada pelo número restrito de amostras no conjunto teste independente, e uma avaliação do classificador em uma coorte externa maior é necessária para determinar precisamente a extensão de sua relevância clínica e implicações para o gerenciamento dos pacientes.

Este perfil classificador de ncRNAs provavelmente será de grande utilidade como exame complementar para guiar a conduta pós-operatória no tratamento do PC. Antes de ser utilizado em larga escala na clínica, é necessário validar este perfil em um conjunto ainda maior de amostras de pacientes, em um projeto que deve se situar na interface da pesquisa básica com a pesquisa clínica aplicada.

Com a análise dos dados de microarranjo das 64 amostras descritas aqui foi possível gerar uma rotina de análise, pois todas as etapas de processamento das amostras, da normalização dos dados à busca de genes diferencialmente expressos foram automatizadas na plataforma R. Estas rotinas serão de grande valor para análises de amostras futuras. Estamos, no momento, empenhados na análise de um conjunto de 200 novas amostras de tumor de próstata de pacientes com longo seguimento, provenientes do Hospital A C Camargo, nosso parceiro no Instituto Nacional de Ciência e Tecnologia em Oncogenômica (INCiTO). Para a análise destas amostras, usaremos uma nova plataforma de microarranjo íntron-éxon, contendo 244 mil sondas (lâminas 244K). Este oligoarranjo foi fabricado sob encomenda pela Agilent e contém sondas para as fitas senso e antissenso de todos os ncRNAs das 72 mil regiões totalmente intrônicas e de todos os ncRNAs das 10 mil regiões parcialmente intrônicas com evidência de transcrição, baseado no banco público de ESTs do GenBank, além das sondas para todos os 30 mil genes codificadores de proteínas humanos confirmados

ou putativos encontrados na lâmina comercial Agilent. Colocamos também todas as sondas que estavam nas lâminas 44K com as quais foram gerados os resultados descritos na presente tese. Este estudo tem o objetivo de validar, em um conjunto mais amplo de pacientes, os resultados já encontrados, e identificar outros transcritos, codificadores e não codificadores de proteína, de interesse clínico.

7 Referências

- Ablin, R. J., P. Bronson, W. A. Soanes e E. Witebsky. Tissue- and species-specific antigens of normal human prostatic tissue. J Immunol, v.104, n.6, Jun, p.1329-39. 1970.
- Acs, A. C. S.-. Cancer Facts & Figures 2010. Atlanta: American Cancer Society. 2010
- Alter, O. e G. H. Golub. Singular value decomposition of genome-scale mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening. Proc Natl Acad Sci U S A, v.103, n.32, Aug 8, p.11828-33. 2006.
- Amling, C. L. Biochemical recurrence after localized treatment. Urol Clin North Am, v.33, n.2, May, p.147-59, v. 2006.
- Andreoiu, M. e L. Cheng. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. Hum Pathol, v.41, n.6, Jun, p.781-93. 2010.
- Ashida, S., H. Nakagawa, T. Katagiri, M. Furihata, M. Iizumi, Y. Anazawa, T. Tsunoda, R. Takata, K. Kasahara, T. Miki, T. Fujioka, T. Shuin e Y. Nakamura. 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, v.64, n.17, Sep 1, p.5963-72. 2004.
- Astro. Consensus statement: guidelines for PSA following radiation therapy. American Society for Therapeutic Radiology and Oncology Consensus Panel. Int J Radiat Oncol Biol Phys, v.37, n.5, Mar 15, p.1035-41. 1997.
- Aumuller, G. [Functional morphology of the prostate]. Urologe A, v.28, n.6, Nov, p.306-10. 1989.
- Bairati, I., F. Meyer, Y. Fradet e L. Moore. Dietary fat and advanced prostate cancer. J Urol, v.159, n.4, Apr, p.1271-5. 1998.
- Balakumaran, B. S., A. Porrello, D. S. Hsu, W. Glover, A. Foye, J. Y. Leung, B. A. Sullivan, W. C. Hahn, M. Loda e P. G. Febbo. MYC activity mitigates response to rapamycin in prostate cancer through eukaryotic initiation factor 4E-binding protein 1-mediated inhibition of autophagy. Cancer Res, v.69, n.19, Oct 1, p.7803-10. 2009.
- Barak, M., Y. Mecz, A. Lurie e N. Gruener. Evaluation of prostate-specific antigen as a marker for adenocarcinoma of the prostate. J Lab Clin Med, v.113, n.5, May, p.598-603. 1989.
- Barry, M. J. Clinical practice. Prostate-specific-antigen testing for early diagnosis of prostate cancer. N Engl J Med, v.344, n.18, May 3, p.1373-7. 2001.
- Benito, M., J. Parker, Q. Du, J. Wu, D. Xiang, C. M. Perou e J. S. Marron. Adjustment of systematic microarray data biases. Bioinformatics, v.20, n.1, Jan 1, p.105-14. 2004.
- Bertone, P., V. Stolc, T. E. Royce, J. S. Rozowsky, A. E. Urban, X. Zhu, J. L. Rinn, W. Tongprasit, M. Samanta, S. Weissman, M. Gerstein e M. Snyder. Global identification of

human transcribed sequences with genome tiling arrays. Science, v.306, n.5705, Dec 24, p.2242-6. 2004.

Best, C. J., I. M. Leiva, R. F. Chuaqui, J. W. Gillespie, P. H. Duray, M. Murgai, Y. Zhao, R. Simon, J. J. Kang, J. E. Green, D. G. Bostwick, W. M. Linehan e M. R. Emmert-Buck. Molecular differentiation of high- and moderate-grade human prostate cancer by cDNA microarray analysis. Diagn Mol Pathol, v.12, n.2, Jun, p.63-70. 2003.

Bickers, B. e C. Aukim-Hastie. New molecular biomarkers for the prognosis and management of prostate cancer--the post PSA era. Anticancer Res, v.29, n.8, Aug, p.3289-98. 2009.

Bill-Axelsson, A., L. Holmberg, M. Ruutu, M. Haggman, S. O. Andersson, S. Bratell, A. Spangberg, C. Busch, S. Nordling, H. Garmo, J. Palmgren, H. O. Adami, B. J. Norlen e J. E. Johansson. Radical prostatectomy versus watchful waiting in early prostate cancer. N Engl J Med, v.352, n.19, May 12, p.1977-84. 2005.

Birney, E., J. A. Stamatoyannopoulos, A. Dutta, R. Guigo, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, M. S. Kuehn, C. M. Taylor, S. Neph, C. M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. A. Greenbaum, R. M. Andrews, P. Flicek, P. J. Boyle, H. Cao, N. P. Carter, G. K. Clelland, S. Davis, N. Day, P. Dhami, S. C. Dillon, M. O. Dorschner, H. Fiegler, P. G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K. D. James, B. E. Johnson, E. M. Johnson, T. T. Frum, E. R. Rosenzweig, N. Karnani, K. Lee, G. C. Lefebvre, P. A. Navas, F. Neri, S. C. Parker, P. J. Sabo, R. Sandstrom, A. Shafer, D. Vetric, M. Weaver, S. Wilcox, M. Yu, F. S. Collins, J. Dekker, J. D. Lieb, T. D. Tullius, G. E. Crawford, S. Sunyaev, W. S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I. L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H. A. Hirsch, E. A. Sekinger, J. Lagarde, J. F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korb, O. Emanuelsson, J. S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M. C. Dickson, D. J. Thomas, M. T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K. G. Srinivasan, W. K. Sung, H. S. Ooi, K. P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M. L. Tress, A. Valencia, S. W. Choo, C. Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T. G. Clark, J. B. Brown, M. Ganesh, S. Patel, H. Tammana, J. Chrast, C. N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P. Bickel, J. S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R. M. Myers, J. Rogers, P. F. Stadler, T. M. Lowe, C. L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S. E. Antonarakis, Y. Fu, E. D. Green, U. Karaoz, A. Siepel, J. Taylor, L. A. Liefer, K. A. Wetterstrand, P. J. Good, E. A. Feingold, M. S. Guyer, G. M. Cooper, G. Asimenos, C. N. Dewey, M. Hou, S. Nikolaev, J. I. Montoya-Burgos, A. Loytynoja, S. Whelan, F. Pardi, T. Massingham, H. Huang, N. R. Zhang, I. Holmes, J. C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W. J. Kent, E. A. Stone, S. Batzoglou, N. Goldman, R. C. Hardison, D. Haussler, W. Miller, A. Sidow, N. D. Trinklein, Z. D. Zhang, L. Barrera, R. Stuart, D. C. King, A. Amez, S. Enroth, M. C. Bieda, J. Kim, A. A. Bhinge, N. Jiang, J. Liu, F. Yao, V. B. Vega, C. W. Lee, P. Ng, A. Shahab, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M. J. Oberley, D. Inman, M. A. Singer, T. A. Richmond, K. J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J. C. Fowler, P. Couttet, A. W. Bruce, O. M. Dovey, P. D. Ellis, C. F. Langford, D. A. Nix, G. Euskirchen, S. Hartman, A. E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T. H. Kim, K. Wang, C. Qu, G. Hon, R. Luna, C. K. Glass, M. G. Rosenfeld, S.

F. Aldred, S. J. Cooper, A. Halees, J. M. Lin, H. P. Shulha, X. Zhang, M. Xu, J. N. Haidar, Y. Yu, Y. Ruan, V. R. Iyer, R. D. Green, C. Wadelius, P. J. Farnham, B. Ren, R. A. Harte, A. S. Hinrichs, H. Trumbower, H. Clawson, J. Hillman-Jackson, A. S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R. M. Kuhn, D. Karolchik, L. Armengol, C. P. Bird, P. I. De Bakker, A. D. Kern, N. Lopez-Bigas, J. D. Martin, B. E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyra, I. B. Hallgrimsdottir, J. Huppert, M. C. Zody, G. R. Abecasis, X. Estivill, G. G. Bouffard, X. Guan, N. F. Hansen, J. R. Idol, V. V. Maduro, B. Maskeri, J. C. McDowell, M. Park, P. J. Thomas, A. C. Young, R. W. Blakesley, D. M. Muzny, E. Sodergren, D. A. Wheeler, K. C. Worley, H. Jiang, G. M. Weinstock, R. A. Gibbs, T. Graves, R. Fulton, E. R. Mardis, R. K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D. B. Jaffe, J. L. Chang, K. Lindblad-Toh, E. S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu e P. J. De Jong. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, v.447, n.7146, Jun 14, p.799-816. 2007.

Bismar, T. A., F. Demichelis, A. Riva, R. Kim, S. Varambally, L. He, J. Kutok, J. C. Aster, J. Tang, R. Kuefer, M. D. Hofer, P. G. Febbo, A. M. Chinnaiyan e M. A. Rubin. Defining aggressive prostate cancer using a 12-gene model. Neoplasia, v.8, n.1, Jan, p.59-68. 2006.

Bolstad, B. M., R. A. Irizarry, M. Astrand e T. P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, v.19, n.2, Jan 22, p.185-93. 2003.

Bostwick, D. G. Pathology of the prostate. New York: Churchill Livingstone. 1990. xiii, 250 p. p. (Contemporary issues in surgical pathology)

Bostwick, D. G. Grading prostate cancer. Am J Clin Pathol, v.102, n.4 Suppl 1, Oct, p.S38-56. 1994.

Branham, W. S., C. D. Melvin, T. Han, V. G. Desai, C. L. Moland, A. T. Scully e J. C. Fuscoe. Elimination of laboratory ozone leads to a dramatic improvement in the reproducibility of microarray gene expression measurements. BMC Biotechnol, v.7, p.8. 2007.

Buhmeida, A., S. Pyrhonen, M. Laato e Y. Collan. Prognostic factors in prostate cancer. Diagn Pathol, v.1, p.4. 2006.

Bussemakers, M. J., A. Van Bokhoven, G. W. Verhaegh, F. P. Smit, H. F. Karthaus, J. A. Schalken, F. M. Debruyne, N. Ru e W. B. Isaacs. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res, v.59, n.23, Dec 1, p.5975-9. 1999.

Calvano, S. E., W. Xiao, D. R. Richards, R. M. Felciano, H. V. Baker, R. J. Cho, R. O. Chen, B. H. Brownstein, J. P. Cobb, S. K. Tschoeke, C. Miller-Graziano, L. L. Moldawer, M. N. Mindrinos, R. W. Davis, R. G. Tompkins e S. F. Lowry. A network-based analysis of systemic inflammation in humans. Nature, v.437, n.7061, Oct 13, p.1032-7. 2005.

Capra, M., P. G. Nuciforo, S. Confalonieri, M. Quarto, M. Bianchi, M. Nebuloni, R. Boldorini, F. Pallotti, G. Viale, M. L. Gishizky, G. F. Draetta e P. P. Di Fiore. Frequent alterations in the expression of serine/threonine kinases in human cancers. Cancer Res, v.66, n.16, Aug 15, p.8147-54. 2006.

Cardiff, R. D. e A. D. Borowsky. Precancer: sequentially acquired or predetermined? Toxicol Pathol, v.38, n.1, p.171-9. 2010.

Chan, A. S., P. S. Thorner, J. A. Squire e M. Zielenska. Identification of a novel gene NCRMS on chromosome 12q21 with differential expression between rhabdomyosarcoma subtypes. Oncogene, v.21, n.19, May 2, p.3029-37. 2002.

Cheadle, C., M. P. Vawter, W. J. Freed e K. G. Becker. Analysis of Microarray Data Using Z Score Transformation. J Mol Diagn, v.5, n.2, p.73-81. 2003.

Chen, J., M. Sun, W. J. Kent, X. Huang, H. Xie, W. Wang, G. Zhou, R. Z. Shi e J. D. Rowley. Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res, v.32, n.16, p.4812-20. 2004.

Chung, L. W. The role of stromal-epithelial interaction in normal and malignant growth. Cancer Surv, v.23, p.33-42. 1995.

Cnattingius, S., F. Lundberg, S. Sandin, H. Gronberg e A. Iliadou. Birth characteristics and risk of prostate cancer: the contribution of genetic factors. Cancer Epidemiol Biomarkers Prev, v.18, n.9, Sep, p.2422-6. 2009.

Coen, J. J., C. S. Chung, W. U. Shipley e A. L. Zietman. Influence of follow-up bias on PSA failure after external beam radiotherapy for localized prostate cancer: results from a 10-year cohort analysis. Int J Radiat Oncol Biol Phys, v.57, n.3, Nov 1, p.621-8. 2003.

Cookson, M. S., G. Aus, A. L. Burnett, E. D. Canby-Hagino, A. V. D'amico, R. R. Dmochowski, D. T. Eton, J. D. Forman, S. L. Goldenberg, J. Hernandez, C. S. Higano, S. R. Kraus, J. W. Moul, C. Tangen, J. B. Thrasher e I. Thompson. Variation in the definition of biochemical recurrence in patients treated for localized prostate cancer: the American Urological Association Prostate Guidelines for Localized Prostate Cancer Update Panel report and recommendations for a standard in the reporting of surgical outcomes. J Urol, v.177, n.2, Feb, p.540-5. 2007.

Critz, F. A. A standard definition of disease freedom is needed for prostate cancer: undetectable prostate specific antigen compared with the American Society of Therapeutic Radiology and Oncology consensus definition. J Urol, v.167, n.3, Mar, p.1310-3. 2002.

Cross, A. J., U. Peters, V. A. Kirsh, G. L. Andriole, D. Reding, R. B. Hayes e R. Sinha. A prospective study of meat and meat mutagens and prostate cancer risk. Cancer Res, v.65, n.24, Dec 15, p.11779-84. 2005.

D'amico, A. V., A. Desjardin, A. Chung e M. H. Chen. Assessment of outcome prediction models for localized prostate cancer in patients managed with external beam radiation therapy. Semin Urol Oncol, v.16, n.3, Aug, p.153-9. 1998.

De Kok, J. B., G. W. Verhaegh, R. W. Roelofs, D. Hessels, L. A. Kiemeney, T. W. Aalders, D. W. Swinkels e J. A. Schalken. DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. Cancer Res, v.62, n.9, May 1, p.2695-8. 2002.

Demark-Wahnefried, W. e M. A. Moyad. Dietary intervention in the management of prostate cancer. Curr Opin Urol, v.17, n.3, May, p.168-74. 2007.

Denmeade, S. R. e J. T. Isaacs. A history of prostate cancer treatment. Nat Rev Cancer, v.2, n.5, May, p.389-96. 2002.

Dhanasekaran, S. M., T. R. Barrette, D. Ghosh, R. Shah, S. Varambally, K. Kurachi, K. J. Pienta, M. A. Rubin e A. M. Chinnaiyan. Delineation of prognostic biomarkers in prostate cancer. Nature, v.412, n.6849, Aug 23, p.822-6. 2001.

Di Sant'agnese, P. A. Neuroendocrine cells of the prostate and neuroendocrine differentiation in prostatic carcinoma: a review of morphologic aspects. Urology, v.51, n.5A Suppl, May, p.121-4. 1998.

Dias Neto, E., R. G. Correa, S. Verjovski-Almeida, M. R. Briones, M. A. Nagai, W. Da Silva, Jr., M. A. Zago, S. Bordin, F. F. Costa, G. H. Goldman, A. F. Carvalho, A. Matsukuma, G. S. Baia, D. H. Simpson, A. Brunstein, P. S. De Oliveira, P. Bucher, C. V. Jongeneel, M. J. O'hare, F. Soares, R. R. Brentani, L. F. Reis, S. J. De Souza e A. J. Simpson. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proc Natl Acad Sci U S A, v.97, n.7, Mar 28, p.3491-6. 2000.

Djavan, B., K. Kadesky, B. Klopukh, M. Marberger e C. G. Roehrborn. Gleason scores from prostate biopsies obtained with 18-gauge biopsy needles poorly predict Gleason scores of radical prostatectomy specimens. Eur Urol, v.33, n.3, p.261-70. 1998.

Do, J. H. e D.-K. Choi. Normalization of microarray data: single-labeled and dual-labeled arrays. Molecules and cells, v.22, n.3, p.254-61. 2006.

Dugan, J. A., D. G. Bostwick, R. P. Myers, J. Qian, E. J. Bergstralh e J. E. Oesterling. The definition and preoperative prediction of clinically insignificant prostate cancer. JAMA, v.275, n.4, Jan 24-31, p.288-94. 1996.

Eastham, J. A., M. W. Kattan e P. T. Scardino. Nomograms as predictive models. Semin Urol Oncol, v.20, n.2, May, p.108-15. 2002.

Eisen, M. B., P. T. Spellman, P. O. Brown e D. Botstein. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A, v.95, n.25, Dec 8, p.14863-8. 1998.

Epstein, J. I., A. W. Partin, J. Sauvageot e P. C. Walsh. Prediction of progression following radical prostatectomy. A multivariate analysis of 721 men with long-term follow-up. Am J Surg Pathol, v.20, n.3, Mar, p.286-92. 1996.

Epstein, J. I., P. C. Walsh, M. Carmichael e C. B. Brendler. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. JAMA, v.271, n.5, Feb 2, p.368-74. 1994.

Esquela-Kerscher, A. e F. J. Slack. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer, v.6, n.4, Apr, p.259-69. 2006.

- Eton, D. T. e S. J. Lepore. Prostate cancer and health-related quality of life: a review of the literature. Psychooncology, v.11, n.4, Jul-Aug, p.307-26. 2002.
- Fare, T. L., E. M. Coffey, H. Dai, Y. D. He, D. A. Kessler, K. A. Kilian, J. E. Koch, E. Leproust, M. J. Marton, M. R. Meyer, R. B. Stoughton, G. Y. Tokiwa e Y. Wang. Effects of atmospheric ozone on microarray data quality. Anal Chem, v.75, n.17, Sep 1, p.4672-5. 2003.
- Felton, J. S., M. G. Knize, N. H. Shen, P. R. Lewis, B. D. Andresen, J. Happe e F. T. Hatch. The isolation and identification of a new mutagen from fried ground beef: 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP). Carcinogenesis, v.7, n.7, Jul, p.1081-6. 1986.
- Ferro, M. A., D. Gillatt, M. O. Symes e P. J. Smith. High-dose intravenous estrogen therapy in advanced prostatic carcinoma. Use of serum prostate-specific antigen to monitor response. Urology, v.34, n.3, Sep, p.134-8. 1989.
- Fleming, C., J. H. Wasson, P. C. Albertsen, M. J. Barry e J. E. Wennberg. A decision analysis of alternative treatment strategies for clinically localized prostate cancer. Prostate Patient Outcomes Research Team. JAMA, v.269, n.20, May 26, p.2650-8. 1993.
- Gautier, L., L. Cope, B. M. Bolstad e R. A. Irizarry. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, v.20, n.3, Feb 12, p.307-15. 2004.
- Giovannucci, E., Y. Liu, E. A. Platz, M. J. Stampfer e W. C. Willett. Risk factors for prostate cancer incidence and progression in the health professionals follow-up study. Int J Cancer, v.121, n.7, Oct 1, p.1571-8. 2007.
- Gleason, D. F. Classification of prostatic carcinomas. Cancer Chemother Rep, v.50, n.3, Mar, p.125-8. 1966.
- Gleason, D. F. Histologic grading of prostate cancer: a perspective. Hum Pathol, v.23, n.3, Mar, p.273-9. 1992.
- Glinsky, G. V., O. Berezovska e A. B. Glinskii. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. J Clin Invest, v.115, n.6, Jun, p.1503-21. 2005.
- Glinsky, G. V., A. B. Glinskii, A. J. Stephenson, R. M. Hoffman e W. L. Gerald. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest, v.113, n.6, Mar, p.913-23. 2004.
- Goldstraw, M. A. Prostate cancer treatment. The case for radical prostatectomy. Ann R Coll Surg Engl, v.88, n.5, Sep, p.439-42. 2006.
- Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield e E. S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science, v.286, n.5439, Oct 15, p.531-7. 1999.
- Gopalkrishnan, R. V., D. C. Kang e P. B. Fisher. Molecular markers and determinants of prostate cancer metastasis. J Cell Physiol, v.189, n.3, Dec, p.245-56. 2001.

Greene, D. R., T. M. Wheeler, S. Egawa, R. P. Weaver e P. T. Scardino. Relationship between clinical stage and histological zone of origin in early prostate cancer: morphometric analysis. Br J Urol, v.68, n.5, Nov, p.499-509. 1991.

Gretzer, M. B., B. J. Trock, M. Han e P. C. Walsh. A critical analysis of the interpretation of biochemical failure in surgically treated patients using the American Society for Therapeutic Radiation and Oncology criteria. J Urol, v.168, n.4 Pt 1, Oct, p.1419-22. 2002.

Guinan, P., R. Bhatti e P. Ray. An evaluation of prostate specific antigen in prostatic cancer. J Urol, v.137, n.4, Apr, p.686-9. 1987.

Halvorsen, O. J., A. M. Oyan, T. H. Bo, S. Olsen, K. Rostad, S. A. Haukaas, A. M. Bakke, B. Marzolf, K. Dimitrov, L. Stordrange, B. Lin, I. Jonassen, L. Hood, L. A. Akslen e K. H. Kalland. Gene expression profiles in prostate cancer: association with patient subgroups and tumour differentiation. Int J Oncol, v.26, n.2, Feb, p.329-36. 2005.

Han, M., A. W. Partin, C. R. Pound, J. I. Epstein e P. C. Walsh. Long-term biochemical disease-free and cancer-specific survival following anatomic radical retropubic prostatectomy. The 15-year Johns Hopkins experience. Urol Clin North Am, v.28, n.3, Aug, p.555-65. 2001.

Hara, M., Y. Koyanagi, T. Inoue e T. Fukuyama. [Some physico-chemical characteristics of "-seminoprotein", an antigenic component specific for human seminal plasma. Forensic immunological study of body fluids and secretion. VII]. Nihon Hoigaku Zasshi, v.25, n.4, Jul, p.322-4. 1971.

Harnden, P., M. D. Shelley, B. Coles, J. Staffurth e M. D. Mason. Should the Gleason grading system for prostate cancer be modified to account for high-grade tertiary components? A systematic review and meta-analysis. Lancet Oncol, v.8, n.5, May, p.411-9. 2007.

Hayes, R. B., J. M. Liff, L. M. Pottern, R. S. Greenberg, J. B. Schoenberg, A. G. Schwartz, G. M. Swanson, D. T. Silverman, L. M. Brown, R. N. Hoover e Et Al. Prostate cancer risk in U.S. blacks and whites with a family history of cancer. Int J Cancer, v.60, n.3, Jan 27, p.361-4. 1995.

Hernandez, J. e I. M. Thompson. Prostate-specific antigen: a review of the validation of the most commonly used cancer biomarker. Cancer, v.101, n.5, Sep 1, p.894-904. 2004.

Hill, A. E., J. S. Hong, H. Wen, L. Teng, D. T. Mcpherson, S. A. Mcpherson, D. N. Levasseur e E. J. Sorscher. Micro-RNA-like effects of complete intronic sequences. Front Biosci, v.11, p.1998-2006. 2006.

Hofmann, R., A. Heidenreich e J. W. Moul. Prostate cancer : diagnosis and surgical treatment. Berlin ; New York: Springer. 2003. xii, 272 p. p.

Holland, J. F. Cancer medicine. Baltimore: Williams & Wilkins. 1997

Horwitz, E. M., H. D. Thames, D. A. Kuban, L. B. Levy, P. A. Kupelian, A. A. Martinez, J. M. Michalski, T. M. Pisansky, H. M. Sandler, W. U. Shipley, M. J. Zelefsky, G. E. Hanks e A. L. Zietman. Definitions of biochemical failure that best predict clinical failure in patients with prostate cancer treated with external beam radiation alone: a multi-institutional pooled analysis. J Urol, v.173, n.3, Mar, p.797-802. 2005.

Hsing, A. W., L. Tsao e S. S. Devesa. International trends and patterns of prostate cancer incidence and mortality. Int J Cancer, v.85, n.1, Jan 1, p.60-7. 2000.

Huang, D., G. P. Casale, J. Tian, N. K. Wehbi, N. A. Abrahams, Z. Kaleem, L. M. Smith, S. L. Johansson, J. E. Elkahwaji e G. P. Hemstreet, 3rd. Quantitative fluorescence imaging analysis for cancer biomarker discovery: application to beta-catenin in archived prostate specimens. Cancer Epidemiol Biomarkers Prev, v.16, n.7, Jul, p.1371-81. 2007.

Hudson, M. A., R. R. Bahnson e W. J. Catalona. Clinical use of prostate specific antigen in patients with prostate cancer. J Urol, v.142, n.4, Oct, p.1011-7. 1989.

Hughes, T. R., M. Mao, A. R. Jones, J. Burchard, M. J. Marton, K. W. Shannon, S. M. Lefkowitz, M. Ziman, J. M. Schelter, M. R. Meyer, S. Kobayashi, C. Davis, H. Dai, Y. D. He, S. B. Stephanians, G. Cavet, W. L. Walker, A. West, E. Coffey, D. D. Shoemaker, R. Stoughton, A. P. Blanchard, S. H. Friend e P. S. Linsley. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol, v.19, n.4, Apr, p.342-7. 2001.

Humphrey, P. A. Gleason grading and prognostic factors in carcinoma of the prostate. Mod Pathol, v.17, n.3, Mar, p.292-306. 2004.

Humphrey, P. A., D. W. Keetch, D. S. Smith, D. L. Shepherd e W. J. Catalona. Prospective characterization of pathological features of prostatic carcinomas detected via serum prostate specific antigen based screening. J Urol, v.155, n.3, Mar, p.816-20. 1996.

Inca, B. M. D. S. I. N. D. C.-. Estimativa 2010: incidência de câncer no Brasil Rio de Janeiro. 2009. 98 p.

Isaacs, J. T. Prostatic structure and function in relation to the etiology of prostatic cancer. Prostate, v.4, n.4, p.351-66. 1983.

Jemal, A., R. Siegel, E. Ward, Y. Hao, J. Xu e M. J. Thun. Cancer statistics, 2009. CA Cancer J Clin, v.59, n.4, Jul-Aug, p.225-49. 2009.

Jeronimo, C., R. Henrique, P. F. Campos, J. Oliveira, O. L. Caballero, C. Lopes e D. Sidransky. Endothelin B receptor gene hypermethylation in prostate adenocarcinoma. J Clin Pathol, v.56, n.1, Jan, p.52-5. 2003.

Jeronimo, C., R. Henrique, M. O. Hoque, E. Mambo, F. R. Ribeiro, G. Varzim, J. Oliveira, M. R. Teixeira, C. Lopes e D. Sidransky. A quantitative promoter methylation profile of prostate cancer. Clin Cancer Res, v.10, n.24, Dec 15, p.8472-8. 2004.

Ji, P., S. Diederichs, W. Wang, S. Boing, R. Metzger, P. M. Schneider, N. Tidow, B. Brandt, H. Buerger, E. Bulk, M. Thomas, W. E. Berdel, H. Serve e C. Muller-Tidow. MALAT-1, a

novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene, v.22, n.39, Sep 11, p.8031-41. 2003.

Johnson, W. E., C. Li e A. Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, v.8, n.1, Jan, p.118-27. 2007.

Kampa, D., J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt, H. Tammana e T. R. Gingeras. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res, v.14, n.3, Mar, p.331-42. 2004.

Kirsh, V. A., U. Peters, S. T. Mayne, A. F. Subar, N. Chatterjee, C. C. Johnson e R. B. Hayes. Prospective study of fruit and vegetable intake and risk of prostate cancer. J Natl Cancer Inst, v.99, n.15, Aug 1, p.1200-9. 2007.

Knight, J. F., C. J. Shepherd, S. Rizzo, D. Brewer, S. Jhavar, A. R. Dodson, C. S. Cooper, R. Eeles, A. Falconer, G. Kovacs, M. D. Garrett, A. R. Norman, J. Shipley e D. L. Hudson. TEAD1 and c-Cbl are novel prostate basal cell markers that correlate with poor clinical outcome in prostate cancer. Br J Cancer, v.99, n.11, Dec 2, p.1849-58. 2008.

Koide, T., P. A. Zaini, L. M. Moreira, R. Z. Vencio, A. Y. Matsukuma, A. M. Durham, D. C. Teixeira, H. El-Dorry, P. B. Monteiro, A. C. Da Silva, S. Verjovski-Almeida, A. M. Da Silva e S. L. Gomes. DNA microarray-based genome comparison of a pathogenic and a nonpathogenic strain of *Xylella fastidiosa* delineates genes important for bacterial virulence. J Bacteriol, v.186, n.16, Aug, p.5442-9. 2004.

Kristiansen, G., C. Pilarsky, C. Wissmann, S. Kaiser, T. Bruemmendorf, S. Roepcke, E. Dahl, B. Hinzmann, T. Specht, J. Pervan, C. Stephan, S. Loening, M. Dietel e A. Rosenthal. Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. J Pathol, v.205, n.3, Feb, p.359-76. 2005.

Kuban, D., H. Thames, L. Levy, E. Horwitz, P. Kupelian, A. Martinez, J. Michalski, T. Pisansky, H. Sandler, W. Shipley, M. Zelefsky e A. Zietman. Failure definition-dependent differences in outcome following radiation for localized prostate cancer: can one size fit all? Int J Radiat Oncol Biol Phys, v.61, n.2, Feb 1, p.409-14. 2005.

Lander, E. S. Array of hope. Nature genetics, v.21, n.1 Suppl, p.3-4. 1999.

Lange, P. H., C. J. Ercole, D. J. Lightner, E. E. Fraley e R. Vessella. The value of serum prostate specific antigen determinations before and after radical prostatectomy. J Urol, v.141, n.4, Apr, p.873-9. 1989.

Lapointe, J., C. Li, J. P. Higgins, M. Van De Rijn, E. Bair, K. Montgomery, M. Ferrari, L. Egevad, W. Rayford, U. Bergerheim, P. Ekman, A. M. Demarzo, R. Tibshirani, D. Botstein, P. O. Brown, J. D. Brooks e J. R. Pollack. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A, v.101, n.3, Jan 20, p.811-6. 2004.

Lapointe, J., S. Malhotra, J. P. Higgins, E. Bair, M. Thompson, K. Salari, C. P. Giacomini, M. Ferrari, K. Montgomery, R. Tibshirani, M. Van De Rijn, J. D. Brooks e J. R. Pollack. hCAP-D3 expression marks a prostate cancer subtype with favorable clinical behavior and androgen signaling signature. Am J Surg Pathol, v.32, n.2, Feb, p.205-9. 2008.

Latulippe, E., J. Satagopan, A. Smith, H. Scher, P. Scardino, V. Reuter e W. L. Gerald. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Res, v.62, n.15, Aug 1, p.4499-506. 2002.

Lerner, S. E., M. L. Blute, E. J. Bergstralh, D. G. Bostwick, J. T. Eickholt e H. Zincke. Analysis of risk factors for progression in patients with pathologically confined prostate cancers after radical retropubic prostatectomy. J Urol, v.156, n.1, Jul, p.137-43. 1996.

Li, T. S. e C. G. Beling. Isolation and characterization of two specific antigens of human seminal plasma. Fertil Steril, v.24, n.2, Feb, p.134-44. 1973.

Lichtenstein, P., N. V. Holm, P. K. Verkasalo, A. Iliadou, J. Kaprio, M. Koskenvuo, E. Pukkala, A. Skytthe e K. Hemminki. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med, v.343, n.2, Jul 13, p.78-85. 2000.

Lieberman, M., A. D. Marks e C. M. Smith. Marks' basic medical biochemistry : a clinical approach. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins. 2009. x, 1011 p. p.

Lilja, H. A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. J Clin Invest, v.76, n.5, Nov, p.1899-903. 1985.

Looijenga, L. H., A. J. Verkerk, N. De Groot, A. A. Hochberg e J. W. Oosterhuis. H19 in normal development and neoplasia. Mol Reprod Dev, v.46, n.3, Mar, p.419-39. 1997.

Louro, R., T. El-Jundi, H. I. Nakaya, E. M. Reis e S. Verjovski-Almeida. Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. Genomics, v.92, n.1, Jul, p.18-25. 2008.

Louro, R., H. I. Nakaya, P. P. Amaral, F. Festa, M. C. Sogayar, A. M. Da Silva, S. Verjovski-Almeida e E. M. Reis. Androgen responsive intronic non-coding RNAs. BMC Biol, v.5, p.4. 2007.

Louro, R., A. S. Smirnova e S. Verjovski-Almeida. Long intronic noncoding RNA transcription: expression noise or expression choice? Genomics, v.93, n.4, Apr, p.291-8. 2009.

Luo, J., D. J. Duggan, Y. Chen, J. Sauvageot, C. M. Ewing, M. L. Bittner, J. M. Trent e W. B. Isaacs. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. Cancer Res, v.61, n.12, Jun 15, p.4683-8. 2001.

Magee, J. A., T. Araki, S. Patil, T. Ehrig, L. True, P. A. Humphrey, W. J. Catalona, M. A. Watson e J. Milbrandt. Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res, v.61, n.15, Aug 1, p.5692-6. 2001.

Mansson, J., C. Bjorkelund e R. Hultborn. Symptom pattern and diagnostic work-up of malignancy at first symptom presentation as related to level of care. A retrospective study from the primary health care centre area of Kungsbacka, Sweden. Neoplasma, v.46, n.2, p.93-9. 1999.

Maruyama, R., S. Toyooka, K. O. Toyooka, A. K. Virmani, S. Zochbauer-Muller, A. J. Farinas, J. D. Minna, J. Mcconnell, E. P. Frenkel e A. F. Gazdar. Aberrant promoter methylation profile of prostate cancers and its relationship to clinicopathological features. Clin Cancer Res, v.8, n.2, Feb, p.514-9. 2002.

Mattick, J. S. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep, v.2, n.11, Nov, p.986-91. 2001.

Mcneal, J. E. Normal anatomy of the prostate and changes in benign prostatic hypertrophy and carcinoma. Semin Ultrasound CT MR, v.9, n.5, Oct, p.329-34. 1988a.

Mcneal, J. E. Normal histology of the prostate. Am J Surg Pathol, v.12, n.8, Aug, p.619-33. 1988b.

Mercer, T. R., M. E. Dinger e J. S. Mattick. Long non-coding RNAs: insights into functions. Nat Rev Genet, v.10, n.3, Mar, p.155-9. 2009.

Merrick, G. S., K. E. Wallner, W. M. Butler, R. W. Galbreath, Z. A. Allen, E. Adamovich e L. True. Brachytherapy in men aged < or = 54 years with clinically localized prostate cancer. BJU Int, v.98, n.2, Aug, p.324-8. 2006.

Michiels, S., S. Koscielny e C. Hill. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet, v.365, n.9458, Feb 5-11, p.488-92. 2005.

Miyake, H., M. Muramaki, T. Kurahashi, A. Takenaka e M. Fujisawa. Expression of potential molecular markers in prostate cancer: correlation with clinicopathological outcomes in patients undergoing radical prostatectomy. Urol Oncol, v.28, n.2, Mar-Apr, p.145-51. 2008.

Moul, J. W. Prostate specific antigen only progression of prostate cancer. J Urol, v.163, n.6, Jun, p.1632-42. 2000.

Nakagawa, T., T. M. Kollmeyer, B. W. Morlan, S. K. Anderson, E. J. Bergstralh, B. J. Davis, Y. W. Asmann, G. G. Klee, K. V. Ballman e R. B. Jenkins. A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. PLoS One, v.3, n.5, p.e2318. 2008.

Nakaya, H. I., P. P. Amaral, R. Louro, A. Lopes, A. A. Fachel, Y. B. Moreira, T. A. El-Jundi, A. M. Da Silva, E. M. Reis e S. Verjovski-Almeida. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. Genome Biol, v.8, n.3, p.R43. 2007.

Nanni, S., C. Priolo, A. Grasselli, M. D'eleto, R. Merola, F. Moretti, M. Gallucci, P. De Carli, S. Sentinelli, A. M. Cianciulli, M. Mottolose, P. Carlini, D. Arcelli, M. Helmer-

Citterich, C. Gaetano, M. Loda, A. Pontecorvi, S. Bacchetti, A. Sacchi e A. Farsetti. Epithelial-restricted gene profile of primary cultures from human prostate tumors: a molecular approach to predict clinical behavior of prostate cancer. Mol Cancer Res, v.4, n.2, Feb, p.79-92. 2006.

Norrish, A. E., L. R. Ferguson, M. G. Knize, J. S. Felton, S. J. Sharpe e R. T. Jackson. Heterocyclic amine content of cooked meat and risk of prostate cancer. J Natl Cancer Inst, v.91, n.23, Dec 1, p.2038-44. 1999.

O'donnell, K. A. e J. D. Boeke. Mighty Piwis defend the germline against genome intruders. Cell, v.129, n.1, Apr 6, p.37-44. 2007.

Oesterling, J. E., D. W. Chan, J. I. Epstein, A. W. Kimball, Jr., D. J. Bruzek, R. C. Rock, C. B. Brendler e P. C. Walsh. Prostate specific antigen in the preoperative and postoperative evaluation of localized prostatic cancer treated with radical prostatectomy. J Urol, v.139, n.4, Apr, p.766-72. 1988.

Ohuri, M., T. M. Wheeler, J. K. Dunn, T. A. Stamey e P. T. Scardino. The pathological features and prognosis of prostate cancer detectable with current diagnostic tests. J Urol, v.152, n.5 Pt 2, Nov, p.1714-20. 1994.

Okutsu, T., Y. Kuroiwa, F. Kagitani, M. Kai, K. Aisaka, O. Tsutsumi, Y. Kaneko, K. Yokomori, M. A. Surani, T. Kohda, T. Kaneko-Ishino e F. Ishino. Expression and imprinting status of human PEG8/IGF2AS, a paternally expressed antisense transcript from the IGF2 locus, in Wilms' tumors. J Biochem, v.127, n.3, Mar, p.475-83. 2000.

Pan, C. C., S. R. Potter, A. W. Partin e J. I. Epstein. The prognostic significance of tertiary Gleason patterns of higher grade in radical prostatectomy specimens: a proposal to modify the Gleason grading system. Am J Surg Pathol, v.24, n.4, Apr, p.563-9. 2000.

Papini-Terzi, F. S., F. R. Rocha, R. Z. Vencio, K. C. Oliveira, M. Felix Jde, R. Vicentini, S. Rocha Cde, A. C. Simoes, E. C. Ulian, S. M. Di Mauro, A. M. Da Silva, C. A. Pereira, M. Menossi e G. M. Souza. Transcription profiling of signal transduction-related genes in sugarcane tissues. DNA Res, v.12, n.1, Feb 28, p.27-38. 2005.

Parkin, D. M., F. Bray, J. Ferlay e P. Pisani. Global cancer statistics, 2002. CA Cancer J Clin, v.55, n.2, Mar-Apr, p.74-108. 2005.

Pauler, F. M., M. V. Koerner e D. P. Barlow. Silencing by imprinted noncoding RNAs: is transcription the answer? Trends Genet, v.23, n.6, Jun, p.284-92. 2007.

Peixoto, B. R., R. Z. Vencio, C. M. Egidio, L. Mota-Vieira, S. Verjovski-Almeida e E. M. Reis. Evaluation of reference-based two-color methods for measurement of gene expression ratios using spotted cDNA microarrays. BMC Genomics, v.7, p.35. 2006.

Penning, T. D., G. D. Zhu, V. B. Gandhi, J. Gong, X. Liu, Y. Shi, V. Klinghofer, E. F. Johnson, C. K. Donawho, D. J. Frost, V. Bontcheva-Diaz, J. J. Bouska, D. J. Osterling, A. M. Olson, K. C. Marsh, Y. Luo e V. L. Giranda. Discovery of the Poly(ADP-ribose) polymerase (PARP) inhibitor 2-[(R)-2-methylpyrrolidin-2-yl]-1H-benzimidazole-4-carboxamide (ABT-888) for the treatment of cancer. J Med Chem, v.52, n.2, Jan 22, p.514-23. 2009.

Perrone, E. E., C. Theoharis, N. R. Mucci, S. Hayasaka, J. M. Taylor, K. A. Cooney e M. A. Rubin. Tissue microarray assessment of prostate cancer tumor proliferation in African-American and white men. J Natl Cancer Inst, v.92, n.11, Jun 7, p.937-9. 2000.

Petrovics, G., W. Zhang, M. Makarem, J. P. Street, R. Connelly, L. Sun, I. A. Sesterhenn, V. Srikantan, J. W. Moul e S. Srivastava. Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. Oncogene, v.23, n.2, Jan 15, p.605-11. 2004.

Pibouin, L., J. Villaudy, D. Ferbus, M. Muleris, M. T. Prosperi, Y. Remvikos e G. Goubin. Cloning of the mRNA of overexpression in colon carcinoma-1: a sequence overexpressed in a subset of colon carcinomas. Cancer Genet Cytogenet, v.133, n.1, Feb, p.55-60. 2002.

Quackenbush, J. Microarray data normalization and transformation. Nat Genet, v.32 Suppl, Dec, p.496-501. 2002.

R_Development_Core_Team. R: A language and environment for statistical computing. 2.3. Vienna. Austria: R Foundation for Statistical Computing. 2006.

Reis, E. M., R. Louro, H. I. Nakaya e S. Verjovski-Almeida. As antisense RNA gets intronic. OMICs, v.9, n.1, Spring, p.2-12. 2005.

Reis, E. M., H. I. Nakaya, R. Louro, F. C. Canavez, A. V. Flatschart, G. T. Almeida, C. M. Egidio, A. C. Paquola, A. A. Machado, F. Festa, D. Yamamoto, R. Alvarenga, C. C. Da Silva, G. C. Brito, S. D. Simon, C. A. Moreira-Filho, K. R. Leite, L. H. Camara-Lopes, F. S. Campos, E. Gimba, G. M. Vignal, H. El-Dorry, M. C. Sogayar, M. A. Barcinski, A. M. Da Silva e S. Verjovski-Almeida. Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. Oncogene, v.23, n.39, Aug 26, p.6684-92. 2004.

Reis, F., N. R. Netto, Jr., J. A. Reinato, M. Thiel e E. Zani. The impact of prostatectomy and brachytherapy in patients with localized prostate cancer. Int Urol Nephrol, v.36, n.2, p.187-90. 2004.

Reissigl, A., J. Pointner, H. Strasser, O. Ennemoser, H. Klocker e G. Bartsch. Frequency and clinical significance of transition zone cancer in prostate cancer screening. Prostate, v.30, n.2, Feb 1, p.130-5. 1997.

Reynolds, M. A., K. Kastury, J. Groskopf, J. A. Schalken e H. Rittenhouse. Molecular markers for prostate cancer. Cancer Lett, v.249, n.1, Apr 28, p.5-13. 2007.

Rhodes, D. R., T. R. Barrette, M. A. Rubin, D. Ghosh e A. M. Chinnaiyan. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Res, v.62, n.15, Aug 1, p.4427-33. 2002.

Rinn, J. L., G. Euskirchen, P. Bertone, R. Martone, N. M. Luscombe, S. Hartman, P. M. Harrison, F. K. Nelson, P. Miller, M. Gerstein, S. Weissman e M. Snyder. The transcriptional activity of human Chromosome 22. Genes Dev, v.17, n.4, Feb 15, p.529-40. 2003.

Rosenthal, S. A. e H. M. Sandler. Treatment strategies for high-risk locally advanced prostate cancer. Nat Rev Urol, v.7, n.1, Jan, p.31-8. 2010.

Ross, P. L., P. T. Scardino e M. W. Kattan. A catalog of prostate cancer nomograms. J Urol, v.165, n.5, May, p.1562-8. 2001.

Ruddon, R. W. Cancer biology. Oxford ; New York: Oxford University Press. 2007. xiv, 530 p. p.

Scher, H. I., A. Sarkis, V. Reuter, D. Cohen, G. Netto, D. Petrylak, P. Lianes, Z. Fuks, J. Mendelsohn e C. Cordon-Cardo. Changing pattern of expression of the epidermal growth factor receptor and transforming growth factor alpha in the progression of prostatic neoplasms. Clin Cancer Res, v.1, n.5, May, p.545-50. 1995.

Scosyrev, E., J. Messing, K. Noyes, P. Veazie e E. Messing. Surveillance Epidemiology and End Results (SEER) program and population-based research in urologic oncology: An overview. Urol Oncol, Apr 2. 2010.

Sensabaugh, G. F. Isolation and characterization of a semen-specific protein from human seminal plasma: a potential new marker for semen identification. J Forensic Sci, v.23, n.1, Jan, p.106-15. 1978.

Shabalina, S. A., A. Y. Ogurtsov, V. A. Kondrashov e A. S. Kondrashov. Selective constraint in intergenic regions of human and mouse genomes. Trends Genet, v.17, n.7, Jul, p.373-6. 2001.

Shariat, S. F., J. A. Karam, V. Margulis e P. I. Karakiewicz. New blood-based biomarkers for the diagnosis, staging and prognosis of prostate cancer. BJU Int, v.101, n.6, Mar, p.675-83. 2008.

Shirai, T., M. Sano, S. Tamano, S. Takahashi, M. Hirose, M. Futakuchi, R. Hasegawa, K. Imaida, K. Matsumoto, K. Wakabayashi, T. Sugimura e N. Ito. The prostate: a target for carcinogenicity of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) derived from cooked foods. Cancer Res, v.57, n.2, Jan 15, p.195-8. 1997.

Siddiqui, E., F. H. Mumtaz e J. Gelister. Understanding prostate cancer. J R Soc Promot Health, v.124, n.5, Sep, p.219-21. 2004.

Sies, H. Oxidative stress: oxidants and antioxidants. Exp Physiol, v.82, n.2, Mar, p.291-5. 1997.

Singal, R., L. Ferdinand, I. M. Reis e J. J. Schlesselman. Methylation of multiple genes in prostate cancer and the relationship with clinicopathological features of disease. Oncol Rep, v.12, n.3, Sep, p.631-7. 2004.

Singh, D., P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub e W. R. Sellers. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell, v.1, n.2, Mar, p.203-9. 2002.

Sleutels, F., R. Zwart e D. P. Barlow. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature, v.415, n.6873, Feb 14, p.810-3. 2002.

Smith, D. S. e W. J. Catalona. The nature of prostate cancer detected through prostate specific antigen based screening. J Urol, v.152, n.5 Pt 2, Nov, p.1732-6. 1994.

Smith, R. A., V. Cokkinides e H. J. Eyre. American Cancer Society guidelines for the early detection of cancer, 2006. CA Cancer J Clin, v.56, n.1, Jan-Feb, p.11-25; quiz 49-50. 2006.

Sogani, P. C., A. Israel, P. H. Lieberman, M. L. Lesser e W. F. Whitmore, Jr. Gleason grading of prostate cancer: a predictor of survival. Urology, v.25, n.3, Mar, p.223-7. 1985.

Sorensen, K. D. e T. F. Orntoft. Discovery of prostate cancer biomarkers by microarray gene expression profiling. Expert Rev Mol Diagn, v.10, n.1, Jan, p.49-64. 2010.

Sorensen, K. D., P. J. Wild, A. Mortezavi, K. Adolf, N. Topping, S. Heeboll, B. P. Ulhøi, P. Ottosen, T. Sulser, T. Hermanns, H. Moch, M. Borre, T. F. Orntoft e L. Dyrskjøt. Genetic and epigenetic SLC18A2 silencing in prostate cancer is an independent adverse predictor of biochemical recurrence after radical prostatectomy. Clin Cancer Res, v.15, n.4, Feb 15, p.1400-10. 2009.

Srikantan, V., Z. Zou, G. Petrovics, L. Xu, M. Augustus, L. Davis, J. R. Livezey, T. Connell, I. A. Sesterhenn, K. Yoshino, G. S. Buzard, F. K. Mostofi, D. G. McLeod, J. W. Moul e S. Srivastava. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. Proc Natl Acad Sci U S A, v.97, n.22, Oct 24, p.12216-21. 2000.

Stamey, T. A. Preoperative serum prostate-specific antigen (PSA) below 10 microg/l predicts neither the presence of prostate cancer nor the rate of postoperative PSA failure. Clin Chem, v.47, n.4, Apr, p.631-4. 2001.

Stamey, T. A., F. S. Freiha, J. E. Mcneal, E. A. Redwine, A. S. Whittemore e H. P. Schmid. Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. Cancer, v.71, n.3 Suppl, Feb 1, p.933-8. 1993.

Stamey, T. A., N. Yang, A. R. Hay, J. E. Mcneal, F. S. Freiha e E. Redwine. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N Engl J Med, v.317, n.15, Oct 8, p.909-16. 1987.

Steinmetz, K. A. e J. D. Potter. Vegetables, fruit, and cancer. II. Mechanisms. Cancer Causes Control, v.2, n.6, Nov, p.427-42. 1991.

Szymanski, M., M. Z. Barciszewska, V. A. Erdmann e J. Barciszewski. A new frontier for molecular medicine: noncoding RNAs. Biochim Biophys Acta, v.1756, n.1, Sep 25, p.65-75. 2005.

Tamura, K., M. Furihata, T. Tsunoda, S. Ashida, R. Takata, W. Obara, H. Yoshioka, Y. Daigo, Y. Nasu, H. Kumon, H. Konaka, M. Namiki, K. Tozawa, K. Kohri, N. Tanji, M. Yokoyama, T. Shimazui, H. Akaza, Y. Mizutani, T. Miki, T. Fujioka, T. Shuin, Y. Nakamura e H. Nakagawa. Molecular features of hormone-refractory prostate cancer cells by genome-wide gene expression profiles. Cancer Res, v.67, n.11, Jun 1, p.5117-25. 2007.

Tang, D., J. J. Liu, A. Rundle, C. Neslund-Dudas, A. T. Savera, C. H. Bock, N. L. Nock, J. J. Yang e B. A. Rybicki. Grilled meat consumption and PhIP-DNA adducts in prostate carcinogenesis. Cancer Epidemiol Biomarkers Prev, v.16, n.4, Apr, p.803-8. 2007.

Tannenbaum, M. P. Urologic pathology : the prostate. Philadelphia: Lea & Febiger. 1977. xii, 419 p. p.

Tokumar, Y., S. V. Harden, D. I. Sun, K. Yamashita, J. I. Epstein e D. Sidransky. Optimal use of a panel of methylation markers with GSTP1 hypermethylation in the diagnosis of prostate adenocarcinoma. Clin Cancer Res, v.10, n.16, Aug 15, p.5518-22. 2004.

Tomlins, S. A., R. Mehra, D. R. Rhodes, X. Cao, L. Wang, S. M. Dhanasekaran, S. Kalyana-Sundaram, J. T. Wei, M. A. Rubin, K. J. Pienta, R. B. Shah e A. M. Chinnaiyan. Integrative molecular concept modeling of prostate cancer progression. Nat Genet, v.39, n.1, Jan, p.41-51. 2007.

Trapasso, J. G., J. B. Dekernion, R. B. Smith e F. Dorey. The incidence and significance of detectable levels of serum prostate specific antigen after radical prostatectomy. J Urol, v.152, n.5 Pt 2, Nov, p.1821-5. 1994.

True, L., I. Coleman, S. Hawley, C. Y. Huang, D. Gifford, R. Coleman, T. M. Beer, E. Gelmann, M. Datta, E. Mostaghel, B. Knudsen, P. Lange, R. Vessella, D. Lin, L. Hood e P. S. Nelson. A molecular correlate to the Gleason grading system for prostate adenocarcinoma. Proc Natl Acad Sci U S A, v.103, n.29, Jul 18, p.10991-6. 2006.

Tusher, V. G., R. Tibshirani e G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A, v.98, n.9, Apr 24, p.5116-21. 2001.

Ulaner, G. A., T. H. Vu, T. Li, J. F. Hu, X. M. Yao, Y. Yang, R. Gorlick, P. Meyers, J. Healey, M. Ladanyi e A. R. Hoffman. Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. Hum Mol Genet, v.12, n.5, Mar 1, p.535-49. 2003.

Van 't Veer, L. J., H. Dai, M. J. Van De Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. Van Der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards e S. H. Friend. Gene expression profiling predicts clinical outcome of breast cancer. Nature, v.415, n.6871, Jan 31, p.530-6. 2002.

Van Oort, I. M., B. M. Schout, L. A. Kiemeney, C. A. Hulsbergen e J. A. Witjes. Does the tertiary Gleason pattern influence the PSA progression-free interval after retropubic radical prostatectomy for organ-confined prostate cancer? Eur Urol, v.48, n.4, Oct, p.572-6. 2005.

Varambally, S., S. M. Dhanasekaran, M. Zhou, T. R. Barrette, C. Kumar-Sinha, M. G. Sanda, D. Ghosh, K. J. Pienta, R. G. Sewalt, A. P. Otte, M. A. Rubin e A. M. Chinnaiyan. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature, v.419, n.6907, Oct 10, p.624-9. 2002.

Varambally, S., J. Yu, B. Laxman, D. R. Rhodes, R. Mehra, S. A. Tomlins, R. B. Shah, U. Chandran, F. A. Monzon, M. J. Becich, J. T. Wei, K. J. Pienta, D. Ghosh, M. A. Rubin e A.

- M. Chinnaiyan. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. Cancer Cell, v.8, n.5, Nov, p.393-406. 2005.
- Vicini, F. A., L. Kestin e A. Martinez. Biochemical outcomes of treatment for prostate cancer. JAMA, v.284, n.22, Dec 13, p.2869. 2000.
- Wang, M. C., L. A. Valenzuela, G. P. Murphy e T. M. Chu. Purification of a human prostate specific antigen. Invest Urol, v.17, n.2, Sep, p.159-63. 1979.
- Ware, J. L. Growth factors and their receptors as determinants in the proliferation and metastasis of human prostate cancer. Cancer Metastasis Rev, v.12, n.3-4, Sep, p.287-301. 1993.
- Ware, J. L. Prostate cancer progression. Implications of histopathology. Am J Pathol, v.145, n.5, Nov, p.983-93. 1994.
- Wattenberg, L. W., A. B. Hanley, G. Barany, V. L. Sparnins, L. K. Lam e G. R. Fenwick. Inhibition of carcinogenesis by some minor dietary constituents. Princess Takamatsu Symp, v.16, p.193-203. 1985.
- Welsh, J. B., L. M. Sapinoso, A. I. Su, S. G. Kern, J. Wang-Rodriguez, C. A. Moskaluk, H. F. Frierson, Jr. e G. M. Hampton. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res, v.61, n.16, Aug 15, p.5974-8. 2001.
- West, D. W., M. L. Slattery, L. M. Robison, T. K. French e A. W. Mahoney. Adult dietary intake and prostate cancer risk in Utah: a case-control study with special emphasis on aggressive tumors. Cancer Causes Control, v.2, n.2, Mar, p.85-94. 1991.
- Wilt, T. J. e M. K. Brawer. The Prostate Cancer Intervention Versus Observation Trial: a randomized trial comparing radical prostatectomy versus expectant management for the treatment of clinically localized prostate cancer. J Urol, v.152, n.5 Pt 2, Nov, p.1910-4. 1994.
- Wolf, S., D. Mertens, C. Schaffner, C. Korz, H. Dohner, S. Stilgenbauer e P. Lichter. B-cell neoplasia associated gene with multiple splicing (BCMS): the candidate B-CLL gene on 13q14 comprises more than 560 kb covering all critical regions. Hum Mol Genet, v.10, n.12, Jun 1, p.1275-85. 2001.
- Woodson, K., J. Gillespie, J. Hanson, M. Emmert-Buck, J. M. Phillips, W. M. Linehan e J. A. Tangrea. Heterogeneous gene methylation patterns among pre-invasive and cancerous lesions of the prostate: a histopathologic study of whole mount prostate specimens. Prostate, v.60, n.1, Jun 15, p.25-31. 2004.
- Yang, I. V., E. Chen, J. P. Hasseman, W. Liang, B. C. Frank, S. Wang, V. Sharov, A. I. Saeed, J. White, J. Li, N. H. Lee, T. J. Yeatman e J. Quackenbush. Within the fold: assessing differential expression measures and reproducibility in microarray assays. Genome Biol, v.3, n.11, Oct 24, p.research0062. 2002.
- Yu, Y. P., D. Landsittel, L. Jing, J. Nelson, B. Ren, L. Liu, C. McDonald, R. Thomas, R. Dhir, S. Finkelstein, G. Michalopoulos, M. Becich e J. H. Luo. Gene expression alterations in

prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol, v.22, n.14, Jul 15, p.2790-9. 2004.

Zhang, B., Q. Wang e X. Pan. MicroRNAs and their regulatory roles in animals and plants. J Cell Physiol, v.210, n.2, Feb, p.279-89. 2007.

Zincke, H., J. E. Oesterling, M. L. Blute, E. J. Bergstralh, R. P. Myers e D. M. Barrett. Long-term (15 years) results after radical prostatectomy for clinically localized (stage T2c or lower) prostate cancer. J Urol, v.152, n.5 Pt 2, Nov, p.1850-7. 1994.

Lista de Anexos

- **Súmula curricular do autor**

Anexos eletrônicos

- **Trabalhos publicados**

- Baratti, M., **Y. Moreira**, F. Traina, F. Costa, S. Verjovski-Almeida e S. Olalla-Saad. Identification of protein-coding and non-coding RNA expression profiles in CD34+ and in stromal cells in refractory anemia with ringed sideroblasts. BMC Medical Genomics, v.3, n.1, p.30. 2010.
- Nakaya, H. I., P. P. Amaral, R. Louro, A. Lopes, A. A. Fachel, **Y. B. Moreira**, T. A. El-Jundi, A. M. Da Silva, E. M. Reis e S. Verjovski-Almeida. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. Genome Biol, v.8, n.3, p.R43. 2007.
- Secco, M., **Y. B. Moreira**, E. Zucconi, N. M. Vieira, T. Jazedje, A. R. Muotri, O. K. Okamoto, S. Verjovski-Almeida e M. Zatz. Gene expression profile of mesenchymal stem cells from paired umbilical cord units: cord is different from blood. Stem Cell Rev, v.5, n.4, Dec, p.387-401. 2009.

- **Trabalho submetido**

- **Yuri B. Moreira**, Camila M. Egidio, Miguel Srougi, Katia Moreira Leite, Luiz H. Camara-Lopes, Eduardo M. Reis and Sergio Verjovski-Almeida, An expression signature of long intronic non-coding RNAs predictive of prostate cancer biochemical recurrence after radical prostatectomy. *Submitted*

Súmula curricular do autor

Nome: Yuri José de Camargo Barros Moreira

Contato

E-mail: ymoreira@iq.usp.br

yurijmoreira@gmail.com

Formação

Doutorado em Bioquímica e Biologia Molecular, Universidade de São Paulo, São Paulo, 2006 – presente.

Bacharel em Biologia, Universidade de São Paulo, São Paulo - 2001-2005.

Publicações

Baratti, M., **Y. Moreira**, F. Traina, F. Costa, S. Verjovski-Almeida e S. Olalla-Saad. Identification of protein-coding and non-coding RNA expression profiles in CD34+ and in stromal cells in refractory anemia with ringed sideroblasts. *BMC Medical Genomics*, v.3, n.1, p.30. 2010.

Secco, M.*, **Y. B. Moreira***, E. Zucconi, N. M. Vieira, T. Jazedje, A. R. Muotri, O. K. Okamoto, S. Verjovski-Almeida e M. Zatz. Gene expression profile of mesenchymal stem cells from paired umbilical cord units: cord is different from blood. *Stem Cell Rev*, v.5, n.4, Dec, p.387-401. 2009.

*Secco M. e Moreira, Y.B. contribuíram igualmente neste trabalho.

Nakaya, H. I., P. P. Amaral, R. Louro, A. Lopes, A. A. Fachel, **Y. B. Moreira**, T. A. El-Jundi, A. M. Da Silva, E. M. Reis e S. Verjovski-

Almeida. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol*, v.8, n.3, p.R43. 2007.

Resumos publicados em anais de congressos

MOREIRA, Y. B.; NAKAYA, H. I.; EGIDIO, C. M.; VENANCIO T. M.;;
LEITE, K. R.M. CÂMARA-LOPES, L.H.; REIS, E, M.;;
VERJOVSKI-ALMEIDA, S. Identification of noncoding RNAs as marker candidates for prostate cancer recurrence prognosis. 100th American Association for Cancer Research Meeting, 2009.

MOREIRA, Y. B.; NAKAYA, H. I.; EGIDIO, C. M.; VENANCIO T. M.;;
LEITE, K. R.M. CÂMARA-LOPES, L.H.; REIS, E, M.;;
VERJOVSKI-ALMEIDA, S. Identification of an intronic transcription profile as a candidate marker for prostate cancer prognosis using intron-exon oligonucleotide microarray. Cold Spring Harbor Laboratory Meeting: Mechanisms of Eukaryotic Transcription, 2007.

MOREIRA, Y. B.; NAKAYA, H. I.; EGIDIO, C. M.; VENANCIO T. M.;;
LEITE, K. R.M. CÂMARA-LOPES, L.H.; REIS, E, M.;;
VERJOVSKI-ALMEIDA, S.. Identification of Intronic Marker Candidates for Prostate Cancer Prognosis Using Oligonucleotide Microarrays. XXXVI Reunião Anual da SBBq e X IUBMB Conference, 2007.

MOREIRA, Y. B.; NAKAYA, H. I.; EGIDIO, C. M.; VENANCIO T. M.;;

LEITE, K. R.M. CÂMARA-LOPES, L.H.; REIS, E, M.;
VERJOVSKI-ALMEIDA, S., Identification of Molecular Markers
of Prostate Cancer Recurrence Using Combined Intron-Exon
Oligonucleotide Microarrays. AACR Special Conference in Cancer
Research: The Role of Non-Coding RNAs in Cancer. October 31-
November 3., Cambridge, MA, EUA, 2007.

BARATTI, M.O.; **MOREIRA, Y. B.**; TRAINA, F.; BORGES, L.;
VERJOVSKI-ALMEIDA, S.; SAAD, S. T. O. , Intronic RNA
Expression in Stromal Cells of Patients With Myelodysplastic
Syndrome by RNA Microarray Analysis..3th Congress of the
European Hematology Association, 2008, Dinamarca.
Hematologica,. v. 93. p. 277-277. 2008.

BARATTI, M.O.; **MOREIRA, Y. B.**; TRAINA, F.; BORGES, L.; COSTA,
F. F.; VERJOVSKI-ALMEIDA, S.; SAAD, S. T. O. . RNAs
Expression Profile in Stromal Cells of Patients with Myelodysplastic
Syndrome. In: XI São Paulo Research Conferences - Medicina
Molecular e Farmacogenética, 2008, São Paulo. Applied, 2008.

BARATTI, M.O.; **MOREIRA, Y. B.**; TRAINA, F.; BORGES, L.; COSTA,
F. F.; VERJOVSKI-ALMEIDA, S.; SAAD, S. T. O. . Identification
of Intronic RNA Expression in CD34+ Cells of Patients with
Myelodysplastic Syndrome by RNA Microarray Analysis. In: VIII
São Paulo Research Conferences - Câncer: da Biologia Molecular ao
Tratamento, 2007, São Paulo. Applied Cancer Research,. v. 27. p.
111-112. 2007.

BARATTI, M.O.; **MOREIRA, Y. B.**; TRAINA, F.; BORGES, L.; COSTA,

F. F.; VERJOVSKI-ALMEIDA, S.; SAAD, S. T. O. . Expressão de RNA intrônico em células CD34+ de pacientes com Síndrome Mielodisplástica por Microarray. In: Congresso Brasileiro de Hematologia e Hemoterapia, 2007, São Paulo. Revista Brasileira de Hematologia e Hemoterapia. São Paulo,. v. 29. p. 207. 2007

BARATTI, M.O.; **MOREIRA, Y. B.**; TRAINA, F.; BORGES, L.; COSTA, Fernando Ferreira ; VERJOVSKI-ALMEIDA, S.; SAAD, Sara Teresinha Olalla . Identification of Intronic RNA Expression in CD34+ Cells of Patients with Myelodysplastic Syndrome by RNA Microarray Analysis.. In: The American Society of Hematology 49th Annual Meeting and Exposition, 2007, Atlanta. Blood (ASH Annual Meeting Abstracts), v. 110. 2007.

FACHEL, A. A.; TAHIRA, A. C.; **MOREIRA, Y.**; GIMBA, E.; Vignal, G.; CAMPOS, F.; LOURO, R.; REIS, E.; VERJOVSKI-ALMEIDA, S. Intronic noncoding RNAs correlated with carcinogenesis and survival in clear cell renal carcinoma identified with cDNA microarrays. In: 100th Annual Meeting AACR, 2009, Denver, CO, USA. 100th Annual Meeting AACR, 2009.

FACHEL, A. A.; TAHIRA, A. C.; **MOREIRA, Y.**; JUNDI, T. E.; GIMBA, E.; Vignal, G.; CAMPOS, F.; LOURO, R.; REIS, E.; VERJOVSKI-ALMEIDA, S. Intronic Noncoding RNAs Correlated with Carcinogenesis and Patient Survival in Clear Cell Renal Cell Carcinoma. In: 3rd EFS Functional Genomics Conference, 2008, Innsbruck, Àustria. 3rd EFS Functional Genomics Conference, 2008.

SILVEIRA R. A.; FACHEL, A. A; **MOREIRA, Y. B.**; DELAMAIN, M. T.; DE SOUZA, C. A.; COSTA, F. F; VERJOVSKI-ALMEIDA S.; AND PAGNANO, K. B.B. Gene Expression Profile in Responsive and Non-Responsive Chronic Myeloid Leukemia Patients Treated with Dasatinib. 51st ASH Annual Meeting and Exposition, 5-8 Dezembro, New Orleans EUA. 2009.

Atividades Extras

Conferencista no II Seminários Avançados em Tecnologia de Microarrays e Next Generation Sequencing – Hospital Israelita Albert Einstein, 18-19 de março de 2010.

Conferencista no I Seminários Avançados em Tecnologia de Microarrays– Hospital Israelita Albert Einstein, 12-13 de março de, 2009.