

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
Programa de Pós-Graduação Bioquímica**

CAMILA DE MOURA EGIDIO

**Análise da expressão de RNAs não-codificadores
intrônicos em tumores de mama**

São Paulo
Data do Depósito na SPG:
16/07/08

Livros Grátis

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

Milhares de livros grátis para download.

CAMILA DE MOURA EGIDIO

**Análise da expressão de RNAs não-codificadores
intrônicos em tumores de mama**

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

Orientador(a): Prof.. Dr. Eduardo Moraes Rego Reis
São Paulo
2008

Nome da Pós-Graduanda: Camila de Moura Egidio

Título da Dissertação de Tese de Doutorado: Análise da expressão de RNAs não-codificadores intrônicos em tumores de mama.

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

Aprovada em: _____

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Assinatura: _____

DEDICATÓRIA

À meu pai, sempre um grande exemplo de dedicação e esforço.

À minha mãe, exemplo de obstinação e persistência.

*À minha irmãzinha Nathália, a quem devo a oportunidade em minha vida
de almejar sempre acertar por inspirar exemplo a alguém.*

*Ao meu marido e grande companheiro Brian, de quem apoio e paciência
nunca faltou mesmo durante seu extenso período de “exílio” no Brasil.*

AGRADECIMENTOS

*“Um dia você aprende que o importante não é o que você tem na vida,
mas quem você tem na vida.*

E que bons amigos são a família que nos permitiram escolher.”

William Shakespeare.

Primeiramente, gostaria de agradecer ao meu orientador, Dr. Eduardo Reis, pela oportunidade de desenvolver este projeto como sua aluna. Durante todo o período do meu doutorado ele soube intercalar momentos de supervisão, participação e críticas com momentos de amizade e suporte. Muito obrigada.

Gostaria também de agradecer ao do Dr. Sergio Verjovski-Almeida a participação em meu período de doutoramento com críticas e sugestões e a permissão total para uso da estrutura de seu laboratório.

Todos que passam por um programa de doutorado na vida sabem que além de um orientador competente, é necessário também amigos habilidosos e inteligentes, que não só te ajudam durante o processo mas que também te inspiram a ser melhores. Nesse sentido, quero agradecer a amizade do Dr. Ricardo de Marco, que sempre esteve disponível para me ajudar em qualquer experimento ou apenas para um café e uma conversa; a Dr. Ana Cláudia Carrera, que nunca mediu esforços, minutos ou paciência para resolver um problema dentro do laboratório, incluindo os meus; ao Dr. Helder Nakaya, que além de sempre pacificar os momentos de desespero, ainda era capaz de propor novas idéias e soluções. Estes três doutores são além de grandes amigos, fontes de grande admiração.

Além destes três amigos doutores que me orientaram e apoiaram duante o doutorado, não posso deixar de mencionar três futuros doutores, amigos igualmente importantes em minha vida: Kátia, Yuri e Lauren. Com estes três doutores compartilhei conhecimento, risadas, cafés, estresses, ansios, experiências pessoais e até informações do Google (!) e aprendi que errar é humano já que compartilhamos também tantos segredos de insucesso do dia-a-dia do laboratório (como deixar a banda do gel cair ou “desenvolver” o melhor marcador de peso molecular). Muito obrigada pelo carinho e consideração dos três!

Ter uma irmã mais nova implica em te fazer almejar sempre ser uma pessoa melhor para servir de modelo a quem te observa no dia-a-dia. Neste aspecto, foi fundamental ter uma irmã mais nova na ciência como a Esther, que além de ser uma estudante muito dedicada é também muito inteligente em suas colocações. Não posso medir todo o conhecimento e ajuda que adquiri por tê-la ao meu lado em meus últimos dois anos de doutorado.

O laboratório do Dr. Sergio Verjovski-Almeida e do Dr. Eduardo Reis, combinados, formam uma grande família. Eu não poderia ter encontrado um ambiente mais agradável e amigo para os meus estudos. Ângela, Rodrigo, Giuliana, Ana Ayupe, Ana Tahira, Felipe, Vinícius, Glauber, Júlio e Thiago são alguns dos colegas de laboratório dos quais sempre guardarei boas lembranças, onde quer que esteja. Denise, Renato e Adriana, são além de amigos, pessoas que me ensinaram a crescer, em todos os aspectos. E como toda grande família tem sempre uma mãe, nós também temos o privilégio de contar com a Ana Paula., que não mede esforços para facilitar nossa vida dentro do laboratório. A todas estas pessoas, muito obrigada. Espero que eu tenha trazido tantas coisas boas às vidas de vocês como recebi.

A minha vida acadêmica começou antes do meu ingresso no grupo do Dr. Sergio Verjovski e Dr. Eduardo Reis. Durante o meu período de Iniciação Científica no laboratório da Dra. Aline da Silva tive a oportunidade de desenvolver uma pesquisa científica de qualidade com todo apoio e suporte desta professora. Após este período, pude sempre contar com sua amizade para me orientar em diversos momentos de decisões pertinentes à minha carreira científica.

Agradeço ao colaborador deste projeto, Dr. Bernardo Peixoto, que mesmo de longe nos foi sempre solícito, ajudando a resolver qualquer problema ou dúvida em relação às amostras de tumor de mama das pacientes do Hospital do Espírito Santo, Açores, Portugal.

Às amigas Daniela Munhoz e Larissa Fontes, amigas que nunca mediram esforços para me apoiar e confortar em momentos difíceis; amigas com as quais eu também sempre pude contar para os momentos de maior felicidade; amigas que valem por uma família inteira em São Paulo. Amo muito vocês!

À amiga e revisora ortográfica Regina Ambrogi, que sempre foi uma amiga que deu muito apoio e carinho em muitos momentos da minha vida, assim como os queridos amigos Laura Bugni, Camila Guindalini, Ricardo Godoi, Sávio Mourão, Sylvia Matsuda, Ana Cláudia Baldani e Álvaro Marciano. Nada disso poderia ter sido feito sem o apoio de cada um de vocês.

Aos meus tios Laércio, Kelly, Ângela e Amilton, por estarem sempre presentes em minha vida, sempre me apoiando e me dando muito carinho, amor e amizade.!

À minha vó Neusa, por estar sempre presente, torcendo por cada uma das pessoas desta família. Muito obrigada!

À querida amiga Carla, pela paciência e compreensão, pelo carinho e amizade. Pelo exemplo de mulher obstinada, independente e dedicada a um trabalho e família que ama. Muito obrigada!

Aos queridos “mom and dad”, John e Joan, de quem nunca faltou apoio e carinho. Muito obrigada por estar sempre presentes mesmo de tão longe.

À minha irmã Nathália, pelas cobranças dos meus pais de ser sempre um bom exemplo à você, já que por mais que errasse eu teria um “pingo de gente” me copiando sempre!

Obrigada por todas as conversas, as broncas, as brigas, os ensinamentos. Pela paciência, apoio e amizade incondicional. Te amo muito!

Aos meus pais pelo amor e apoio incondicional. Por acreditarem em mim até mesmo quando eu achava que não seria capaz e por me incentivarem ir mais longe. Por me fornecerem exemplos dignificantes de seres humanos e por me inspirarem grande admiração. Amo vocês.

Ao meu marido Brian, por estar sempre ao meu lado, me apoiando incondicionalmente com muito amor e carinho, suportando até mesmo a distância daqueles que ama por dois anos para estar comigo. Obrigada pelo depósito de confiança, pelo amor e pelo apoio emocional incondicional.

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

Egidio, C.M. **Análise da expressão de RNAs não-codificadores intrônicos em tumores de mama.** 2008. 200p. Dissertação de Tese de Doutorado – Programa de Pós-Graduação em Ciências (Bioquímica). Instituto de Química, Universidade de São Paulo, São Paulo.

O câncer de mama é o carcinoma que mais acomete mulheres no Brasil. Os tratamentos disponíveis são recomendados a partir da análise de fatores de prognóstico como a classificação pelo sistema TNM, tipo histológico, *status* de receptores hormonais e marcadores de proliferação tumoral. No entanto, a classificação dos tumores de mama é muito variável e o poder prognóstico dos marcadores tumorais atuais ainda é limitado, levando muitas pacientes à terapia adjuvante desnecessária. Portanto, novos métodos de prognóstico mais sensíveis são necessários para melhorar a tomada de decisão na clínica oncológica de pacientes com câncer de mama.

Do ponto de vista de ciência básica, as modificações transcricionais associadas à oncogênese e progressão do câncer de mama ainda são pouco conhecidas. Além da alteração na expressão de genes codificadores para proteínas, evidências recentes sugerem que RNAs não-codificadores (ncRNAs) podem ter um papel importante na transformação maligna.

Este projeto teve como principais objetivos: i) investigar a expressão de ncRNAs intrônicos em amostras de adenocarcinoma de mama e ii) identificar assinaturas de expressão gênica associadas a características anatomo-patológicas e clínicas de tumores de mama com potencial aplicação clínica. Para isso, foram comparados os perfis de expressão gênica de 58 amostras de tecido tumoral de mama, com seguimento clínico conhecido, utilizando uma plataforma de microarranjos de cDNA, enriquecida em ncRNAs provenientes de regiões intrônicas de genes humanos conhecidos.

Durante o projeto foram testadas diferentes metodologias para análise da expressão gênica utilizando microarranjos de cDNA com uma ou duas cores. O desenho experimental das hibridizações incluiu a co-hibridização de cada microarranjo com alvos fluorescentes representando o transcrito da amostra de tumor juntamente com um oligonucleotídeo referência complementar a uma região presente em todas as sondas de cDNA (*RefOligo*). Este desenho experimental permitiu a avaliação de duas abordagens de análise da expressão gênica: a primeira baseada nas intensidades diretas de cada transcrito (*One-Color*) e a segunda baseada em razões de expressão onde a intensidade de cada transcrito foi normalizada pelo oligonucleotídeo referência (*RefOligo*). A utilização direta das intensidades se mostrou mais reprodutível e sensível para a detecção de assinaturas de expressão correlacionadas com características das amostras de mama, e essa abordagem foi escolhida para as análises subsequentes.

Os dados provenientes dos experimentos de microarranjos revelaram níveis de expressão ubíquos dos transcritos intrônicos nas amostras analisadas, extendendo para o câncer de mama a relevância do estudo desta classe de ncRNAs. Além disso, foi identificada uma assinatura contendo 95 transcritos, correlacionada com o *status* de expressão do receptor de estrogênio (REr), dos quais cerca de 15% correspondem a ncRNAs. Utilizando apenas amostras com seguimento clínico superior a 4 anos, foi identificada uma assinatura com 113 transcritos, dos quais cerca de 30% são ncRNAs intrônicos, capaz de distinguir com 100% de acurácia pacientes que desenvolveram metástase daqueles que permaneceram livres da doença.

Além de contribuir com novos candidatos a marcadores de prognóstico no câncer de mama, este estudo aponta para a participação de ncRNAs intrônicos em complexas redes transcricionais, possivelmente modulando a expressão de genes codificadores para proteínas. A caracterização detalhada da função de ncRNAs com expressão

correlacionada a características fenotípicas e clínicas dos tumores de mama deverá fornecer novas informações sobre as bases moleculares da tumorigênese e progressão desta neoplasia.

Palavras-chave: câncer de mama, marcadores moleculares, RNAs não codificadores intrônicos, transcrição antisense, microarranjos de DNA, receptor de esteróides.

ABSTRACT

Egidio, C.M. **Gene expression analysis of intronic non-coding RNAs in breast tumors**. 2008. 200p. PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Breast carcinoma is the most frequently occurring cancer amongst women in Brazil. The treatments available for breast cancer are prescribed based on the results of prognostic factors, such as the TNM classification system, histological type, hormonal receptor *status* and tumoral markers for cell proliferation. Nevertheless, breast cancer classification can be variable and inconsistent, and the prognosis power of tumoral markers is still limited, resulting in many patients unnecessarily undergoing adjuvant therapy. Therefore, there is an urge for new prognosis methods that are more sensitive, as well as accurate, in order to improve treatment decisions for breast cancer patients.

From a basic science perspective, transcriptional modifications associated with oncogenesis and breast cancer progression are still poorly understood. Beyond alterations of the expression of protein-coding genes, recent evidences suggest that non-coding RNAs (ncRNAs) might have an important role in malignant transformation.

The main goals of this project are: i) to investigate the expression of intronic ncRNAs in breast cancer tissue and ii) to identify gene expression signatures correlated to anatomic-pathological and clinical characteristics of human breast tumors, with a potential clinical application. To achieve this, gene expression profiles of 58 breast tumor samples with clinical follow-up were compared using a microarray platform enriched in non-coding RNAs (ncRNAs) derived from intronic regions of known human genes.

During this project different gene expression methodologies were tested for the analysis of one- or two-color cDNA microarrays. The experimental design included the co-hybridization of the microarrays with fluorescent targets representing the tumor sample transcriptome with a reference oligonucleotide that is complementary to a

common region present in all cDNA probes (*RefOligo*). This experimental design permitted the evaluation of two gene expression analysis approaches: the first based on direct intensities of each transcript (*One-Color*) and the second based in expression ratios where the intensity of each transcript is normalized by the reference oligonucleotide (*RefOligo*). *One-Color* methodology has shown to provide a more reproducible and sensitive gene expression signatures correlated to the breast samples characteristics and, therefore, this approach was chosen for subsequent analysis.

The data provided by the microarray experiments revealed that ubiquitous expression of intronic ncRNAs was observed, confirming the relevance of investigating the role of this class of ncRNAs in breast cancer. Furthermore, a gene expression signature comprising 95 transcripts and correlated to the estrogen receptor *status* of breast tumor samples was identified, from which approximately 15% are ncRNAs. Using only samples from patients with known follow-up, a signature of 113 transcripts was identified, of which 30% are ncRNAs. This gene expression signature was able to distinguish with 100% accuracy patients that developed metastasis from those that remained disease-free up to 4 years after surgery.

Besides the contribution of new molecular prognostic markers for breast cancer, the present study indicates that intronic ncRNAs might play a role in complex transcriptional networks, possibly regulating the expression of protein-coding genes. The detailed characterization of the functional roles of ncRNAs, whose expression levels are correlated to fenotypical and clinical characteristics of breast tumors, is likely to provide new insights on the molecular basis of tumorigenesis and progression of this neoplasia.

Keywords: breast cancer, molecular markers, intronic non-coding RNAs, antisense transcription, DNA microarrays, steroid receptor.

LISTA DE ABREVIATURAS E SIGLAS

aa – Aminoácido

ACS – *American Cancer Society*

AJCC – *American Joint Committee on Cancer Staging*

aRNA – RNA amplificado, fita complementar

CAP – *College of American Pathologists*

CDI – Carcinoma Ductal Invasivo

cDNA – DNA complementar

DNA – Ácido desoxirribonucléico

DNase – Desoxirribonuclease

dNTP – Desoxirribonucleotídeo

HDES – Hospital do Divino Espírito Santo, Açores, Portugal

HER2 – Gene do receptor do fator de crescimento epidermal humano 2

HPRT1 – Gene da hipoxantina fosforibosil transferase

INCA – Instituto Nacional de Câncer

MIT - *Massachusetts Institute of Technology*.

NCCN – *National Comprehensive Cancer Network*

NCI – *National Cancer Institute*

ncRNA – RNA não-codificador de proteína

NIH – *National Institute of Health*

NKI – *National Cancer Institute*, Holanda

REr – Receptor de Estrogênio

REr+ – tumores de mama com expressão positiva para receptor de estrogênio

REr- – tumores de mama com ausência de expressão do receptor de estrogênio

RPr – Receptor de Progesterona

RPr+ – tumores de mama com expressão positiva para receptor de progesterona

RPr- – tumores de mama com expressão negativa para receptor de progesterona

RefOligo – Oligonucleotídeo sintético acoplado a uma molécula do fluoróforo Cy3

RefPool – *Pool* de amostras de RNA distintas utilizado como referência externa nos experimentos de microarranjo

RNA – Ácido Ribonucléico

RNAm – RNA mensageiro

rRNA – RNA ribossomal

RT-PCR – Transcrição reversa seguida de PCR

SAM – Ferramenta informática utilizada para a busca de transcritos diferencialmente expressos (*Significance Analysis of Microarrays*)

SNR – Razão sinal/ruído (*Signal-to-Noise*)

PCR – reação em cadeia da DNA Polimerase

pb – pares de base

SNPs – *Single Nucleotide Polymorphisms*

TRANSBIG – *Translational Research Network of the Breast International Group*

SUMÁRIO

1.	INTRODUÇÃO	20
1.1.	O câncer de mama.....	20
1.1.1.	Estatísticas populacionais.....	20
1.1.2.	Tipos histológicos de câncer de mama.....	21
1.1.3.	O câncer de mama hereditário.....	24
1.1.4.	Fatores de prognóstico clássicos no câncer de mama	26
1.1.5.	Tratamentos existentes	35
1.2.	Classificação molecular do câncer de mama utilizando microarranjos de DNA	37
1.3.	RNAs não-codificadores e câncer	47
2.	OBJETIVOS	55
2.1.	Objetivo Geral.....	55
2.2.	Objetivos Específicos.....	55
3.	MATERIAIS E MÉTODOS	57
3.1.	Coleta e armazenamento das amostras de tumor de mama.....	57
3.2.	Extração de RNA total de amostras de tumor de mama	57
3.3.	Amplificação linear do RNA mensageiro	59
3.4.	Acoplamento dos alvos de aRNA a moléculas fluorescentes	61
3.5.	Hibridização de alvos fluorescente com microarranjos de cDNA	62
3.6.	Captura da imagem das lâminas hibridizadas e extração dos dados.....	64
3.7.	Filtragem e normalização das intensidades de expressão gênica.....	64
3.8.	Identificação de assinaturas de expressão gênica correlacionadas com características clínicas e anatomo-patológicas.....	67
3.9.	Reações de RT-PCR fita específica	68
3.10.	PCR quantitativo em tempo real	72
3.11.	Anotação funcional dos transcritos identificados	74
4.	RESULTADOS.....	76
4.1.	Montagem de banco de amostras de tecido tumoral mamário e obtenção de RNA para experimentos de hibridização.....	76
4.2.	Hibridização de amostras de tumor de mama com o microarranjo de ~4.000 elementos.....	82
4.3.	Avaliação de uma metodologia para obtenção de medidas de expressão gênica utilizando um oligonucleotídeo referência (<i>RefOligo</i>).....	85
4.4.	Análise comparativa da metodologia <i>RefOligo</i> vs. <i>One-Color</i>	93
4.5.	Análise não supervisionada de um conjunto de 62 amostras de câncer de mama	96
4.6.	Análises supervisionadas utilizando informações clínicas e anatomo-patológicas	97
4.7.	Expressão de RNAs não-codificadores em amostras de tumor de mama.....	102
4.8.	Classificação dos tumores de acordo com <i>status</i> de receptor de estrogênio.....	103
4.9.	Classificação dos tumores de acordo com presença de metástase	112
4.10.	Caracterização de transcritos presentes na assinatura de metástase.....	125
5.	DISCUSSÃO	130
5.1.	Hibridização das amostras de tecido tumoral mamário	130
5.2.	Expressão de ncRNAs intrônicos em tumores de mama	132

5.3.	Avaliação de uma metodologia para obtenção de medidas de expressão gênica utilizando um oligonucleotídeo referência (<i>RefOligo</i>).....	133
5.4.	Análises não supervisionadas e supervisionadas utilizando características clínicas e anatomo-patológicas das amostras de câncer de mama	136
5.5.	Classificação dos tumores de acordo com <i>status</i> do receptor de estrogênio.....	139
5.6.	Classificação dos tumores de acordo com presença de metástase	151
5.7.	Caracterização de ncRNAs intrônicos correlacionados com metástase.....	164
6.	CONCLUSÕES	169
7.	REFERÊNCIAS.....	173
	ANEXOS	191
	Súmula Curricular da Autora	192
	ANEXOS ELETRÔNICOS	195

SUMÁRIO DE FIGURAS

Figura 1. Unidades alveolares dutais terminais da mama.....	22
Figura 2. Representação esquemática da hibridização do RefOligo com o microarranjo de cDNA.....	63
Figura 3. Fórmula para cálculo da razão de <i>SNR</i> para cada um dos transcritos analisados.	68
Figura 4. Eletroforese capilar de amostras de RNA total extraídas de amostras de tecido tumoral mamário	80
Figura 5. Imagem de um microarranjo de cDNA hibridizado com aRNA tumoral de mama (vermelho) e <i>RefOligo</i> (verde).	84
Figura 6. Representação esquemática do desenho experimental das hibridizações dos microarranjos, incluindo as hibridizações diretas e indiretas (<i>RefOligo</i> e <i>RefPool</i>).....	88
Figura 7. Gráficos de dispersão <i>M vs. A</i> para razões diretas e reconstruídas.....	89
Figura 8. Comparação das médias dos coeficientes de variação das réplicas intra- e inter-lides entre três metodologias, <i>RefOligo</i> , <i>RefPool</i> e <i>One-Color</i> em relação à hibridização direta.....	90
Figura 9. Variância das razões de expressão das quatro metodologias analisadas em faixas de intensidade de expressão.....	91
Figura 10. Correlação entre razões diretas e reconstruídas da comparação dos tecidos tumorais rim <i>vs.</i> mama.	92
Figura 11. Comparação dos conjuntos de transcritos identificados como expressos diferencialmente em cada uma das quatro metodologias utilizadas (hibridização direta, <i>RefOligo</i> , <i>RefPool</i> e <i>One-Color</i>).....	93
Figura 12. Mapa do agrupamento hierárquico não-supervisionado dos transcritos que apresentam maior variação da expressão entre 62 pacientes analisadas.....	97
Figura 13. Mapa do perfil de expressão de 62 amostras de tumor de mama de acordo com um grupo de 127 transcritos apontados como diferencialmente expressos entre tumores RPr+ <i>vs.</i> RPr-.	100
Figura 14. Mapa do perfil de expressão de 62 amostras de tumor de mama de acordo com um grupo de 63 transcritos apontados como diferencialmente expressos entre tumores REr+ <i>vs.</i> REr-.	101
Figura 15. Moléculas de RNA não codificadores provenientes de introns de genes conhecidos são ubiqüamente expressas em amostras de tumor de mama ductal invasivo.	103
Figura 16. Visualização do mapeamento genômico do transcrito intrônico do gene <i>OPCML</i> através da ferramenta BLAT da Universidade da Califórnia de Santa Cruz (genome.ucsc.edu).....	110
Figura 17. Mapa representativo do nível de expressão de 95 transcritos (linhas) indicados como diferencialmente expressos entre 58 tumores de mama REr+ <i>vs.</i> REr- (colunas).....	111
Figura 18. Visualização do mapeamento genômico dos transcritos intrônicos dos genes <i>MLLT3</i> (A), <i>NEBL</i> (B) e <i>ZMIZ1</i> (C).....	121
Figura 19. Mapa representativo do nível de expressão de 113 transcritos apontados como diferencialmente expressos (linhas) em 18 amostras de tumor de mama (colunas).....	123

Figura 20. Mapa representativo do nível de expressão de 32 transcritos intrônicos apontados como diferencialmente expressos (linhas) em 18 amostras de tumor de mama (colunas).....	124
Figura 21. Orientação da transcrição de seis mensagens intrônicas selecionadas do conjunto de 113 transcritos da assinatura de metástase em câncer de mama, determinado por RT-PCR fita específica	126
Figura 22. Níveis relativos de expressão de seis transcritos intrônicos (<i>CAB39</i> , <i>DNAJC3</i> , <i>GFRA1</i> , <i>ZBTB16</i> , <i>MLLT3</i> e <i>CNOT4</i>) e dois transcritos exônicos (<i>GATA3</i> e <i>JMJD2C</i>) em tumores de mama.....	128
Figura 23. Figura esquemática das relações diretas e indiretas dos genes relacionados a ciclo celular e câncer.....	148
Figura 24. Figura esquemática das relações diretas e indiretas dos genes relacionados a desenvolvimento celular, desenvolvimento e função dos sistemas linfático e imune e crescimento e proliferação celular.	161

SUMÁRIO DE TABELAS

Tabela 1. Classificação clínica do câncer de mama pelo sistema TNM.	28
Tabela 2. Estadiamento do câncer de mama em função das diversas combinações possíveis pelo sistema TNM.	29
Tabela 3. Seqüência dos <i>primers</i> utilizados nas reações de RT-PCR fita específica e PCR em tempo real.	70
Tabela 4. Características anatomo-patológicas do conjunto de amostras teciduais reunidos no banco de tecido tumoral mamário.	78
Tabela 5. Quadro de informações de rendimento total, qualidade e rendimento obtido da amplificação de amostras de RNA total extraídas de 74 amostras de tecido tumoral mamário.....	81
Tabela 6. Valores de correlação entre dados de lâminas de 5 amostras aleatórias analisadas por duas abordagens distintas, <i>RefOligo</i> e <i>One-Color</i>	95
Tabela 7. Conjunto de 95 transcritos expressos diferencialmente em tumores de mama REr+ <i>versus</i> tumores de mama REr-.....	104
Tabela 8. Relação dos transcritos da assinatura de REr identificados como diferencialmente expressos entre tumores de mama REr+ vs. REr-, em dados disponíveis na literatura.....	108
Tabela 9. Características anatomo-patológicas e moleculares de 18 amostras de tecido tumoral mamário que constaram da análise do nível de expressão dos transcritos do microarranjo de cDNA em tumores mamários metastáticos.	113
Tabela 10. Conjunto de 113 transcritos expressos diferencialmente em tumores de mama que desenvolveram metástase num período de acompanhamento de quatro anos (M+) <i>versus</i> tumores de mama sem metástase e/ou recidiva (M-).	115
Tabela 11. Relação dos transcritos da assinatura de metástase reportados como diferencialmente expressos em tumores de mama, em dados disponíveis na literatura.	119
Tabela 12. Comparação das razões do nível de expressão de transcritos avaliados através de PCR em tempo real e microarranjo, entre pacientes com e sem metástase... ..	129

1. INTRODUÇÃO

1.1. O câncer de mama

1.1.1. Estatísticas populacionais

Neste ano, Jemal e colaboradores apresentaram dados mostrando que apesar da redução e estabilização das taxas de incidência de todos os tipos de câncer na população americana desde o ano de 2004, o câncer ainda é responsável pelo maior número de mortes nos Estados Unidos, ficando à frente até mesmo de doenças cardíacas entre pessoas com menos de 85 anos de idade (Jemal et al., 2008). O número estimado de novos casos para 2008 é de cerca de um milhão e meio, sendo que um terço destes casos será fatal, resultando numa taxa de 25% das causas de morte nos Estados Unidos (Jemal et al., 2008).

No Brasil, cerca de 470 mil novos casos de câncer serão diagnosticados no ano de 2008, sendo que destes, 50 mil correspondem a câncer de mama, resultando num risco estimado de 51 casos a cada 100 mil mulheres (INCA, 2008). Nos Estados Unidos o câncer de mama é o câncer de maior incidência entre as mulheres com uma estimativa de 182 mil novos casos em 2008 e 40 mil mortes entre as mulheres, configurando como o segundo tipo de câncer mais fatal entre estas (Jemal et al., 2008) (INCA, 2008). Nos Estados Unidos a probabilidade de uma pessoa desenvolver qualquer tipo de câncer durante toda sua vida é de 45% e no caso de câncer de mama feminino é de 12%.

Na população mundial, a sobrevida média de pacientes com câncer de mama após cinco anos do diagnóstico da doença é de 61% (INCA, 2008), enquanto nos Estados Unidos é de 98% no caso de câncer de mama localizado. Se no momento do diagnóstico houver presença de metástase distante, essa taxa diminui para 27% (Jemal et al., 2008). Portanto, uma das formas de se diminuir o número de mortes no caso do

câncer de mama é o diagnóstico da doença em seu estágio inicial. No Brasil, as taxas de mortalidade por câncer de mama ainda são elevadas muito provavelmente por que a doença ainda é diagnosticada em estágios avançados (INCA, 2008).

Os métodos de diagnóstico do câncer de mama na rede de saúde pública no Brasil incluem o exame clínico anual das mamas para mulheres de 40 a 49 anos e o exame mamográfico pelo menos a cada dois anos para mulheres de 50 a 69 anos (INCA, 2008). Interessantemente, nos Estados Unidos, onde as taxas de incidência de câncer de mama vêm diminuindo progressivamente desde 2004, o auto-exame das mamas e o exame clínico com um profissional é recomendado para mulheres a partir dos 20 anos de idade, este último pelo menos a cada 3 anos (ACS, 2008). Já o exame mamográfico é recomendado a mulheres com mais de 40 anos, anualmente (ACS, 2008). Estas diferenças podem explicar por que as taxas de mortalidade por câncer de mama no Brasil ainda são tão altas apesar de ser considerado um câncer de relativamente bom prognóstico, se diagnosticado e tratado oportunamente, de acordo com o próprio Instituto Nacional de Câncer no Brasil.

1.1.2. Tipos histológicos de câncer de mama

As lesões malignas da mama mais comuns em humanos são originadas dos alvéolos secretórios de leite que estão agrupados em redes na região terminal da mama formando as unidades alveolares dutais terminais (Molyneux et al., 2007) (Figura 1). O crescimento celular neste caso, normalmente tem origem nos dutos ou lóbulos mamários.

Os carcinomas mamários podem se dividir em dois grandes grupos, dependendo de seu caráter invasivo. O câncer de mama *in situ* se refere ao câncer não-invasivo onde as células tumorais permanecem em seu local de origem, ficando contidas entre as

estruturas epiteliais mamárias, sem invadir a membrana basal ou o estroma do tecido mamário (Mallon et al., 2000). Existem vários tipos de câncer de mama não-invasivo, que se distinguem quanto às suas características histomorfológicas: hiperplasia sem atipia, hiperplasia ductal atípica, carcinoma ductal *in situ*, doença de Paget, carcinoma lobular *in situ* e hiperplasia lobular atípica (Mallon et al., 2000).

O carcinoma ductal é o mais freqüente câncer de mama *in situ*, correspondendo a 25% de todos os tipo de câncer de mama diagnosticados nos Estados Unidos (CAP, 2006). Destas pacientes, dois terços talvez nunca desenvolvam nenhum tipo de câncer invasivo (Mallon et al., 2000). Alguns especialistas consideram o carcinoma ductal *in situ* uma lesão pré-maligna pois embora seja o mais facilmente tratado, pode estar associado a outros tumores de mama invasivos na paciente (CAP, 2006).

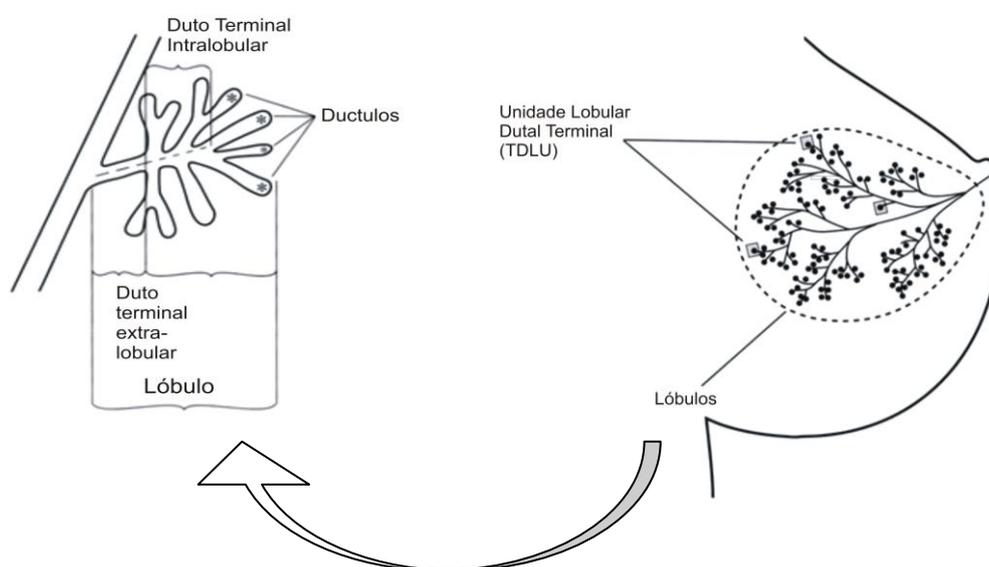


Figura 1. Unidades alveolares ductais terminais da mama. Este é o local onde o tumor se origina na maior parte dos tumores de mama (modificado de www.cancer.org – ACS).

Entre os tumores de mama invasivos, o carcinoma ductal invasivo (CDI) é o tipo histológico mais comum, sendo responsável por cerca de 65-85% dos casos de câncer de mama (CAP, 2006). Os carcinomas invasivos são caracterizados pela invasão celular dos

ductos ou lóbulos das glândulas mamárias, atingindo outros tecidos da mama. Podem ficar restritos ao local de origem ou migrar pelo corpo através da corrente sanguínea ou sistema linfático. O carcinoma lobular invasivo é o segundo carcinoma deste tipo mais comumente diagnosticado (cerca de 10-15% dos tumores de mama invasivos diagnosticados) e possui uma taxa de sobrevivência relativamente maior que o CDI (Mallon et al., 2000).

Outros tipos de carcinoma invasivo menos frequentes são: carcinoma tubular, com incidência de cerca de 10%, baixa ocorrência de metástases e longo período de sobrevivência livre da doença; carcinoma cribriforme, 3% dos casos de câncer de mama, com 100% de sobrevivência sem doença após 5 anos do diagnóstico; carcinoma mucinoso, com incidência de 1%, é mais comum em mulheres acima de 75 anos, apresentando uma taxa de sobrevivência sem doença após 5 anos de diagnóstico de 90%; carcinoma papilar, 1-2% dos casos de câncer de mama, apresenta processo de invasão celular muito semelhante ao carcinoma ductal invasivo mas com um prognóstico relativamente melhor; carcinoma secretório, forma muito rara de carcinoma mamário reportado em crianças, geralmente é associado a um prognóstico favorável; carcinoma apócrino, menos de 1% dos casos de câncer de mama. Existem ainda outras formas raras de câncer de mama invasivo (incidência < 1%), tais como o carcinoma adenóide cístico, carcinoma medular, mioepitelioma e carcinoma metaplástico (Mallon et al., 2000).

Os tumores filóides, responsáveis por menos de 1% dos casos de câncer de mama (Parfitt et al., 2004), correspondem ao único tipo de tumor de mama não-epitelial por conter, além de elementos epiteliais, elementos do estroma envolvidos na tumorigênese. Enquanto o componente epitelial destes tumores é totalmente benigno, o crescimento do estroma é o maior responsável por recorrências locais e metástases (Mallon et al., 2000).

1.1.3. O câncer de mama hereditário

O câncer de mama hereditário corresponde a cerca de 10% de todos os casos de câncer de mama (Lynch et al., 2008). A maior parte destes casos são de mutações associadas ao gene de alta penetrância *BRCA1* e *BRCA2*. Outros genes associados com a susceptibilidade do câncer de mama hereditário são *TP53*, *ATM*, *CHEK2*, *PTEN* e *LKB1/STK11*, sendo que a ocorrência de mutações nestes dois últimos genes são eventos mais raros na população (Bradbury e Olopade, 2007). O conjunto de todos estes genes não é responsável por mais que 25% dos casos de tumor de mama hereditários, apontando para a existência de vários genes de menor penetrância ainda desconhecidos (Bradbury e Olopade, 2007)

Os genes *BRCA1* e *BRCA2* estão envolvidos com reparo no DNA (Welsh et al., 2000) e mutações nestes genes estão associadas a um risco de 37-85% de uma mulher desenvolver câncer de mama durante sua vida e 15-40% no caso de câncer ovariano (Struewing et al., 1997).

Mutações no gene *TP53*, codificador para a proteína tumoral p53 estão associadas à síndrome Li-Fraumeni (Malkin et al., 1990), doença autossômica rara com ocorrência de diversos tipos de câncer, em especial, o câncer de mama em idades muito jovens (idade média no momento do diagnóstico de 36 anos) (Birch et al., 2001). A proteína p53 está envolvida com o controle do ciclo celular e apoptose e a ocorrência de mutações no seu gene é muito freqüente em diferentes tipos de tumor (Meulmeester e Jochemsen, 2008). Apesar de ter alta penetrância, a síndrome Li-Fraumeni e mutações no gene *TP53* são relativamente raras e são responsáveis por menos de 1% de todos os casos de câncer de mama (Sidransky et al., 1992).

A síndrome de Cowden é uma síndrome rara e autossômica e tem sido relacionada ao câncer de mama hereditário (Bradbury e Olopade, 2007). Mutações no

gene *PTEN*, um supressor de tumor, têm sido identificadas em famílias com esta síndrome (Liaw et al., 1997). O risco de uma mulher com esta síndrome desenvolver câncer de mama invasivo durante sua vida é de 25-50% (Uppal et al., 2007).

Mutações no gene *STK11/LKB1* foram identificadas e associadas com a síndrome Peutz-Jeghers (Lim et al., 2003). Esta síndrome rara é responsável por um pequeno número de casos de câncer de mama, no entanto, mulheres com esta síndrome apresentam um risco de desenvolver tumor de mama de 29-50% (Lim et al., 2003).

O gene *ATM* possui um importante papel no reconhecimento de dano e reparo ao DNA, além de atuar nas fases de transição G₁-S e G₂-M do ciclo celular (Dapic et al., 2005; Mavrou et al., 2008). Perda total da expressão da proteína ATM é causada por mutações homozigóticas na seqüência do gene, causando a síndrome ataxia-telangiectasia (Mavrou et al., 2008). A ataxia-telangiectasia é uma doença genética rara, autossômica recessiva, neurodegenerativa e hereditária e está associada à predisposição ao câncer, especialmente, ao câncer de mama (Khanna, 2000; Bradbury e Olopade, 2007). Apesar de mutações em heterozigose serem relativamente freqüentes na população no caso do gene *ATM* (1-2%), devido a sua baixa penetrância, estas mutações não são responsáveis por um grande número dos casos de câncer de mama hereditário (Dapic et al., 2005).

CHEK2 é um gene de susceptibilidade ao câncer de mama de baixa penetrância (Bradbury e Olopade, 2007). Este gene codifica para uma proteína quinase envolvida no ciclo celular e no reparo do DNA (Lee et al., 2000). Mutação neste gene confere às mulheres portadoras uma chance duas vezes maior de desenvolver câncer de mama e é responsável por cerca de 1% de todos os casos de câncer de mama (Meijers-Heijboer et al., 2002).

1.1.4. Fatores de prognóstico clássicos no câncer de mama

Cerca de 30% das pacientes com tumor de mama negativo para presença de células metastáticas nos nódulos linfáticos irão apresentar recorrência da doença nos primeiros 5 anos após o diagnóstico, enquanto as outras 70% irão permanecer livres da doença (Subramaniam e Isaacs, 2005). Deste fato é que se origina o grande dilema clínico no momento de se optar ou não pela terapia sistêmica adjuvante para as pacientes que se incluem neste caso. Portanto, os fatores de prognóstico têm um papel fundamental pois fornecem informações sobre a progressão da doença, indicando a melhor opção para a melhora da qualidade de vida da paciente doente.

A avaliação de parâmetros anatomo-patológicos é utilizada na rotina clínica para o prognóstico de cancer de mama. A informação sobre o *status* dos nódulos linfáticos axiais tem sido o fator de prognóstico utilizado de forma mais consistente na prática clínica como guia à terapia a ser utilizada (Rampaul et al., 2001). No entanto, embora seja útil para a estratificação de risco, a informação sobre o *status* dos nódulos linfáticos é dependente do tempo: quanto maior o crescimento do tumor, maior a probabilidade de infiltração de células metastáticas nos nódulos linfáticos. Dessa forma, a informação de infiltração de nódulos linfáticos isolada não é capaz de definir com total segurança a evolução da doença após o tratamento. (Rampaul et al., 2001)

Outro fator de prognóstico importante é o tamanho do tumor. Em um estudo com cerca de 13.000 mulheres com câncer de mama com *status* dos nódulos linfáticos negativo foi medida a taxa de sobrevida em um período de 5 anos de grupos de pacientes com tumores de diferentes tamanhos no diagnóstico (Carter et al., 1989). Cerca de 99% das pacientes com tumores menor que 1cm atingiram este período de sobrevida, enquanto apenas 89% das pacientes com tumores de 1 a 3 cm e 86% das pacientes com tumores de 3 a 5 cm conseguiram atingir este período de sobrevida. Juntamente com a

informação sobre a presença de metástases distantes ao tumor, este conjunto de informações forma o sistema TNM (tumor-nódulo-metástase - Tabela 1) e é o método de classificação mais utilizado e recomendado pelo “American Joint Committee on Cancer Staging” (AJCC) (Singletary et al., 2002). O sistema de estadiamento TNM inclui duas classificações principais, a clínica e a patológica (Singletary e Connolly, 2006). A classificação clínica é feita durante o momento do exame físico da paciente através do resultado de exames de mamografia, ultra-som e biópsias; a classificação patológica é baseada em evidências observadas no momento da cirurgia e na visualização do material tumoral ao microscópio (Singletary e Connolly, 2006). As informações combinadas sobre o tamanho, infiltração de nódulos linfáticos e presença de metástases distantes em outros órgãos compõem grupos de estadiamentos, nos quais os tumores que pertencem a um mesmo grupo são tratados da mesma maneira, além de possuírem uma taxa de sobrevida semelhante durante dez anos do diagnóstico da doença (NCCN, 2008) (Singletary e Connolly, 2006). Dessa forma, existem sete diferentes grupos de estadiamentos de tumores de mama (Tabela 2), que estão associados com o prognóstico da doença: cerca de 5 e 12% das pacientes com tumores de mama de estágio I e II morrem no período de dez anos após o diagnóstico da doença, comparado com cerca de 60 e 90% de mortalidade de pacientes com câncer de mama de estágio III e IV (NCCN, 2008) (Bland et al., 1998).

Tabela 1. Classificação clínica do câncer de mama pelo sistema TNM. Retirado de www.projetodiretrizes.org.br - Associação Médica Brasileira e Conselho Federal de Medicina.

T - Tumor	
TX	O tumor primário não pode ser avaliado
T0	Sem evidência de tumor primário
Tis	Carcinoma in situ: carcinoma intraductal ou carcinoma lobular in situ ou doença de Paget da papila sem tumor
T1	Tumor com 2 cm ou menos em sua maior dimensão
	T1a - tumor com 0,5 cm ou menos em sua maior dimensão
	T1b - tumor com mais de 0,5 cm e até 1 cm em sua maior dimensão
	T1c - tumor com mais de 1 cm e até 2 cm em sua maior dimensão
T2	Tumor com mais de 2 cm e até 5 cm em sua maior dimensão
T3	Tumor com mais de 5 cm em sua maior dimensão
T4	Tumor de qualquer tamanho, com extensão direta à parede torácica ou à pele
	T4a - extensão para parede torácica
	T4b - edema (incluindo peau d'orange) ou ulceração da pele da mama ou nódulos cutâneos satélites, confinados à mesma mama
	T4c - T4a e T4b associados
	T4d - carcinoma inflamatório
N - Linfonodais regionais	
NX	Os linfonodos regionais não podem ser avaliados (por ex. foram removidos previamente)
N0	Ausência de metástases nos linfonodos regionais
N1	Metástase em linfonodo(s) auxiliar(es) homolateral (is) móvel (is)
N2	Metástase nos linfonodos axilares homolaterais fixos uns aos outros ou a outras estruturas
N3	Metástase nos linfonodos da cadeia mamária interna homolateral
M - Metástases a distância	
MX	A presença de metástases a distância não pode ser avaliada
M0	Ausência de metástases a distância
M1	Metástases a distância (incluindo as metástases nos linfonodos supraclaviculares)

Tabela 2. Estadiamento do câncer de mama em função das diversas combinações possíveis pelo sistema TNM. Retirado de www.projetodiretrizes.org.br - Associação Médica Brasileira e Conselho Federal de Medicina.

Estadio 0	Tis	N0	M0
Estadio I	T1	N0	M0
Estadio IIa	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Estadio IIb	T2	N1	M0
	T3	N0	M0
Estadio IIIa	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1, N2	M0
Estadio IIIb	T4	qualquer N	M0
	qualquer T	N3	M0
Estadio IV	qualquer T	qualquer N	M1

Outro sistema de classificação do grau tumoral muito utilizado é o de Elston e Ellis (Elston e Ellis, 1991), modificado de Scarff-Bloom-Richardson (SBR) (Bloom e Richardson, 1957). Este sistema leva em consideração a porcentagem de formação de túbulos, o grau de pleiomorfismo nuclear e o número de células que se encontra em mitose. No entanto, este sistema apresentou alta variabilidade e falta de consistência metodológica na quantificação objetiva do grau tumoral. Portanto, a AJCC decidiu não incorporar este novo sistema na classificação tumoral TNM (Subramaniam e Isaacs, 2005).

Frente à grande quantidade de informação disponível na literatura sobre fatores de prognóstico em câncer de mama, o NIH (National Institute of Health, EUA) percebeu-se a necessidade de uma compilação e revisão sistemática sobre o uso dessas informações por profissionais da saúde. No ano de 2001, o NIH estimulou a elaboração por especialistas de um conjunto de critérios a ser utilizado por profissionais da saúde, no momento de tomar decisões sobre tratamentos para as pacientes com câncer de

mama. Os critérios do NIH para o tratamento do câncer de mama, no momento da seleção de pacientes para terapia adjuvante sistêmica e hormonal, levam em consideração quais agentes utilizar na quimioterapia, como avaliar os efeitos colaterais e a qualidade de vida da paciente (NIH Consensus Statement, (2001).

Outra tentativa de definição de critérios homogêneos para elegibilidade de pacientes com câncer de mama à terapia adjuvante, conhecida como critérios de St Gallen, determina que a primeira consideração a ser feita pelo profissional da saúde em uma paciente com câncer de mama é avaliar a presença de receptores hormonais no tecido tumoral (Goldhirsch et al., 2005). Em relação a esta informação, três categorias foram criadas, considerando a resposta à terapia endócrina, sendo elas: positiva, negativa e incerta (Goldhirsch et al., 2005). Cada uma destas categorias foram sub-divididas em três em relação ao risco de vida (baixo, intermediário e alto), de acordo com *status* da menopausa das pacientes (Goldhirsch et al., 2005). Além disso, ficou definido que a informação sobre infiltração de células metastáticas nos nódulos linfáticos axilares não constitui, automaticamente, alto risco de vida para a paciente (Goldhirsch et al., 2005).

Pacientes com câncer de mama com menos de 30/35 anos apresentam, em geral, um pior prognóstico da doença que aquelas pacientes com mais do que 40 anos. O *status* dos fatores de prognóstico dos tumores daquelas pacientes mais jovens normalmente vêm associados a resultados negativos como nódulos linfáticos positivos, REr-, super-expressão de *HER2* (Soerjomataram et al., 2008). Por outro lado, pacientes idosas com câncer de mama, com idade superior a 70 anos de idade apresentam uma taxa de mortalidade maior se comparado com as pacientes mais novas. Isto ocorre, muitas vezes, devido ao uso menos extensivo de tratamentos contra a doença, decorrente da pré-existência de outras doenças ou até mesmo pela idade avançada das pacientes (Soerjomataram et al., 2008).

O tipo histológico está associado a diferenças na biologia do tumor de mama, e fornece informações sobre o prognóstico da doença (Rampaul et al., 2001). Como já mencionado anteriormente, dos tipos de carcinomas invasivos, o tipo ductal invasivo é o que possui o pior prognóstico quando comparado aos outros. Metástases nos nódulos linfáticos foram encontradas em 35% das mulheres com este tipo tumoral *versus* 11% quando comparado aos outros tipos tumorais com melhor prognóstico (Wong et al., 2002).

Outra forma de se avaliar a agressividade do tumor é através da investigação da invasão vascular peritumoral, determinada por marcadores como hematoxilina e eosina em combinação com testes de imunohistoquímica que detectam estruturas vasculares (Subramaniam e Isaacs, 2005). A presença de invasão vascular peritumoral é um fator de prognóstico independente e aumenta as chances de recorrência do tumor em até 5 anos após o diagnóstico em 15% (Neville et al., 1992; Lauria et al., 1995).

Além de critérios anatomo-patológicos, informações relativas a um conjunto de marcadores moleculares são rotineiramente levadas em consideração para a avaliação prognóstica do câncer de mama. O *status* dos receptores de estrogênio e progesterona em tumores de mama são importantes fatores preditivos, pois funcionam como marcadores da resposta do tumor a terapias adjuvantes (Payne et al., 2008), além de prever a evolução da terapia hormonal a longo prazo (Soerjomataram et al., 2008). O tamoxifeno, droga utilizada em terapias endócrinas, é um modulador seletivo do receptor de estrogênio e inibe a ligação do deste receptor ao estradiol, interrompendo uma série de mecanismos que regulam a replicação e proliferação celular (Dinh et al., 2007). Desde 1970 esta droga tem sido prescrita a pacientes com câncer de mama, independente da informação de *status* do receptor de estrogênio no tecido tumoral (Dinh et al., 2007). Dessa forma, o tamoxifeno só era capaz de produzir resposta contra o tumor em 30%

dos casos (Dinh et al., 2007). Após 1990, esta taxa aumentou para 80% devido ao uso seletivo de tamoxifeno em pacientes com tumor de mama REr+ (Dinh et al., 2007). Assim, o *status* do receptor de estrogênio é, sozinho, o fator de prognóstico mais confiável disponível e, juntamente com a informação sobre o *status* do receptor de progesterona, podem determinar um baixo risco de mortalidade no caso de REr+/RPr+ quando comparados a tumores REr-/RPr- (Payne et al., 2008). O valor de predição, no entanto, não é tão poderoso e diminui ainda mais com o passar dos anos após o diagnóstico da doença (Costa et al., 2002).

O receptor do fator de crescimento epidermal humano 2 (Her2) é amplificado ou superexpresso em cerca de 20 a 30% dos tumores (Subramaniam e Isaacs, 2005). Em dois grupos de pacientes com nódulo-positivo, o primeiro com amplificação e super-expressão de Her2 e o segundo sem amplificação ou super-expressão deste gene, a taxa de sobrevivência após o diagnóstico do câncer de mama em 10 anos foi de, respectivamente, 50% e 65% (Pritchard et al., 2006). Tumores que superexpressam Her2 são mais prováveis de apresentar anormalidades na via da proteína tumoral p53, *status* de receptores hormonais negativos, infiltração de nódulos linfáticos e alto índice mitótico, todos marcadores de mau prognóstico em câncer de mama (Ferrero-Pous et al., 2000; Menard et al., 2002). No caso de pacientes com nódulo linfático negativo, a super-expressão de Her2 não afeta a sobrevivência destas a longo prazo de forma significativa (Ferrero-Pous et al., 2000; Menard et al., 2002). A avaliação do *status* de expressão de Her2 por imunohistoquímica tem sido muito valioso no momento da predição da resposta tumoral a tratamentos com o anticorpo trastuzumab e a certas terapias endócrinas e quimioterapia (Pritchard et al., 2006).

Nos últimos dez anos, surgiram novos fatores de prognóstico em câncer de mama baseados em marcadores moleculares, como o ativador plasminogênico da uroquinase

(uPA) (Duffy e Duggan, 2004). Esta proteína converte o plasminogênio em plasmina através de proteólise. A plasmina, por sua vez, ativa metaloproteinases da matriz que degradam a matriz extracelular e modulam a adesão celular, proliferação e migração. Ambos os genes *uPA* e seu inibidor fisiológico (inibidor da ativação do plasminogênio 1, *PAI-1*) foram encontrados superexpressos em vários tipos de câncer, especialmente de mama (Duffy e Duggan, 2004). Portanto, o gene *uPA* está altamente associado a tumores agressivos e à presença de metástases. Já foi mostrado que em um grupo de pacientes nódulo-negativo para metástases, existe uma diferença de 28% na taxa de sobrevida em 10 anos entre as pacientes com altos e baixos níveis de *uPA/PAI-1* (Look et al., 2002). Devido a este fato, a verificação dos níveis de *uPA/PAI-1* em pacientes com tumores de mama nódulo-negativo é o melhor fator de prognóstico utilizado atualmente (Hayes et al., 1996).

O papel de marcadores de proliferação tem sido muito estudado através de diferentes metodologias, como citometria de fluxo, fração da fase S, índice mitótico, incorporação de bromodeoxiuridina e testes de imunohistoquímica usando anticorpos direcionados a antígenos presentes durante a proliferação celular, como Ki-67 e PCNA (Subramaniam e Isaacs, 2005). A variabilidade das metodologias e sistemas utilizados torna difícil a comparação destes fatores de prognóstico aos existentes, e o uso como rotina na prática clínica. No entanto, já foi sugerido em vários estudos que a taxa de proliferação celular é um fator independente de predição da evolução da doença (Subramaniam e Isaacs, 2005).

Além da investigação de infiltração de células metastáticas nos nódulos linfáticos axiais da paciente, tem se mostrado que a investigação de infiltração de células metastáticas na medula óssea, através da técnica de aspiração, é um fator de prognóstico independente do risco de morte no câncer de mama (Braun et al., 2000). No entanto, este

fator ainda não foi incluído na prática clínica devido a diferenças na preparação celular, nos anticorpos utilizados e, em alguns casos, falta de informação sobre outros fatores de prognóstico clássicos (Subramaniam e Isaacs, 2005).

Historicamente, a seleção de terapias sistêmicas adjuvantes em câncer de mama dependia principalmente de características histológicas do tumor mamário (classificação TNM) (Dinh et al., 2007). Após a descoberta do receptor de estrogênio e Her2 como fatores de prognóstico, foi facilitada a decisão em relação a que terapias endócrinas recomendar à paciente, considerando o risco de vida associado à presença do tumor para o cálculo da relação de custo/benefício (Dinh et al., 2007). Atualmente, já foi possível evoluir muito em relação à caracterização molecular da doença de câncer de mama e foi descoberto que existem outros sub-tipos moleculares de tumor mamário, que não se beneficiam das terapias endócrinas como esperado quando utilizados fatores de prognóstico como o *status* de expressão dos receptores de estrogênio e do fator de crescimento epidermal 2. Contudo, observa-se uma significativa heterogeneidade quanto a evolução clínica de pacientes com câncer de mama com características anatomo-patológicas semelhantes e com mesmo padrão de expressão dos marcadores moleculares conhecidos, que se traduz em uma dificuldade de prever com 100% de acurácia a evolução clínica individual (Barginear et al., 2008). Este aspecto explica-se pela elevada variabilidade genética existente entre os diferentes indivíduos, e a penetrância parcial dos marcadores conhecidos. Portanto, são necessárias novas tecnologias capazes de analisar o perfil molecular dos vários tumores de mama, indicando assim novos sistemas de classificação para prover máximo benefício para as pacientes com a terapia endócrina escolhida.

1.1.5. Tratamentos existentes

Os diferentes tratamentos para o câncer de mama podem ser de quatro tipos: cirurgia, radioterapia, quimioterapia e terapia hormonal. A maior parte das pacientes com câncer de mama passam pelo processo cirúrgico de remoção do tecido tumoral, e inclusive alguns nódulos linfáticos axiais para investigação da presença de células metastáticas (NCI, 2008). A cirurgia da mama pode ser restrita à remoção do tecido tumoral juntamente com alguns nódulos linfáticos (lumpectomia e mastectomia parcial) ou pode abranger toda a mama e os nódulos linfáticos, incluindo o tecido muscular sob a mama (mastectomia total e radical) (NCI, 2008).

Após a cirurgia, a maioria das paciente serão indicadas à radioterapia, podendo, adicionalmente, serem também indicadas a uma terapia adjuvante, que pode ser quimioterapia ou hormonioterapia. A radioterapia consiste na exposição das células tumorais a raios-x altamente energéticos na tentativa de matá-las, contendo o crescimento tumoral (NCI, 2008). A quimioterapia utiliza drogas que interferem no ciclo celular, inibindo seu crescimento, ou ainda, agem de forma a matar as células tumorais (NCI, 2008). Radio e quimioterapias são tipicamente pouco seletivas para causar danos em células tumorais, e geralmente causam efeitos secundários indesejáveis que limitam o sucesso das estratégias.

As terapias hormonais são prescritas para mulheres com tumores responsivos a estrogênio e progesterona, ou ao menos progesterona, pois a resposta do receptor de progesterona depende da presença da via do receptor de estrogênio intacta no tumor (Newman e Singletary, 2007). O tamoxifeno tem sido a droga mais utilizada nos últimos 30 anos em terapias hormonais endócrinas em pacientes com tumor positivo para receptores hormonais. Como dito anteriormente, esta droga é um antagonista do estrogênio no tecido mamário, se ligando ao receptor hormonal e, conseqüentemente,

inibindo a ligação do estradiol ao receptor de estrogênio, inibindo a resposta intracelular de crescimento hormonal. O tamoxifeno também diminui as chances de recorrência do tumor em pacientes com câncer de mama responsivos à terapia endócrina em 3 a 50% dos casos (Newman e Singletary, 2007). No entanto, a droga tem efeito estrogênico no fígado, ossos e útero, aumentando as chances da paciente desenvolver outros tumores e ainda trombose venosa (Boughey et al., 2008). A terapia com o tamoxifeno pode ser complementada com o uso de inibidores de aromatase, que agem na inibição quase total da produção de estrogênio, bloqueando a conversão de precursores de estrogênio derivados da glândula adrenal pela enzima aromatase (Newman e Singletary, 2007). Os inibidores de aromatase têm a vantagem de não produzir nenhum efeito no tecido uterino. No entanto, sua desvantagem é o risco associado de sua terapia à ocorrência de osteoporose nas pacientes (Newman e Singletary, 2007). As pacientes com tumores de mama REr-/RPr- ou, ainda, resistentes à terapia endócrina, possuem um número menor de opções de tratamento. A quimioterapia, nestes casos, é o único tratamento efetivo.

Recentemente, tem-se percebido a necessidade de incorporar conhecimento molecular à clínica médica em câncer de mama para o desenvolvimento e sucesso de novas formas de terapias direcionadas às características moleculares da doença (Loi et al., 2004). Um exemplo bem sucedido e eficiente de terapia direcionada é o uso do anticorpo anti-Her2 (trastuzumab) em tumores de mama com super-expressão do gene *HER2*, normalmente associado a tumores mais agressivos e avançados (Perez, 2001). O anticorpo se liga a proteína Her2 e inibe a cascata de sinalização do receptor, com efeito inibitório sobre a proliferação celular. Esta droga é capaz de diminuir de 30 a 50% as chances de recorrência da doença em tumores de alto risco (Newman e Singletary, 2007). Cerca de 20% dos tumores de mama é positivo para o receptor Her2 (Whenham et al., 2008), no entanto, a maior parte dos tumores tratados apresentam progressão da

doença no período de um ano de tratamento (Bartsch et al., 2007). Portanto, compostos alternativos ao trastuzumab que agem sobre o receptor de Her2 ou ainda na sua via de sinalização são de extrema importância e já são utilizados na clínica médica, como o lapatinib (inibidor de tirosina quinase que age sobre os receptores Her2 e do fator de crescimento epidermal) (Bartsch et al., 2007).

1.2. Classificação molecular do câncer de mama utilizando microarranjos de DNA

Microarranjos de DNA consistem em coleções de sondas de DNA depositadas em um suporte sólido. São freqüentemente utilizados em experimentos de hibridização para se investigar a presença de DNA complementar (cDNA) na amostra que se deseja estudar e, assim, detectar o nível de expressão de genes e transcritos na amostra de interesse. As condições experimentais para o anelamento das fitas complementares de DNA foram definidas 10 anos após a elucidação da estrutura do DNA (Marmur e Doty, 1961). Em 1995 foi descrita pela primeira vez a utilização de microarranjos de DNA imobilizados para a obtenção simultânea de medidas de expressão de 45 genes em *A.taliana* (Schena et al., 1995). Evoluções tecnológicas permitem atualmente a construção de plataformas contendo milhares ou milhões de elementos de DNA, que têm sido largamente utilizadas para a comparação da expressão gênica global de tecidos diferentes, entre outros usos (Dieterich et al., 2008; Masri et al., 2008; McGraw et al., 2008).

As sondas de DNA podem ser sintetizadas na própria superfície sólida do microarranjo ou ainda ser ligadas covalentemente a ela. Após o primeiro grupo produzir *arrays* de oligonucleotídeos *in situ* (Maskos e Southern, 1992), os avanços na tecnologia e na química possibilitaram a confecção de microarranjos de oligonucleotídeos de alta

densidade sintetizados *in situ*, utilizando técnicas como a fotolitografia (Pease et al., 1994) e a impressão à jato de tinta (Hughes et al., 2001). Paralelamente, o desenvolvimento de técnicas de impressão (Shalon et al., 1996) permitiu o uso da robótica para a deposição de produtos de PCR, oligonucleotídeos sintéticos, clones de cDNA ou BAC nas lâminas de vidro.

Estas técnicas permitiram a construção de *arrays* de oligonucleotídeos altamente densos contendo 10^5 - 10^6 sondas em uma única lâmina de vidro (Gresham et al., 2008). Dessa forma, os microarranjos são capazes de interrogar o nível de expressão gênico de milhares de seqüências em um só experimento de hibridização e até mesmo de avaliar o perfil transcricional de genomas inteiros de organismos eucarióticos como *S. cerevisiae* (David et al., 2006) e *A. thaliana* (Clark et al., 2007). Adicionalmente, os microarranjos se tornaram uma ferramenta poderosa para medir os níveis de transcritos de RNA em tecidos humanos diferentes, tornando possível a análise comparativa dos padrões de expressão gênica em estados fisiológicos e patológicos distintos (Trevino et al., 2007).

Existem diversos desenhos experimentais possíveis para a análise de dados de expressão gênica utilizando microarranjos (Kerr e Churchill, 2001; Yang e Speed, 2002; Simon e Dobbin, 2003). Mais freqüentemente os microarranjos são hibridizados com uma única amostra experimental marcada fluorescentemente (Ex. *oligoarrays* da Affymetrix), ou co-hibridizados simultaneamente com duas amostras experimentais, marcadas com fluoróforos distintos (ex. Cy3 e Cy5) que permitem sua detecção individualmente. Um desenho experimental freqüente nos experimentos de 2 cores consiste na co-hibridização de cada amostra experimental com um mesmo RNA referência, tipicamente um *pool* de RNAs derivados das amostras ou de linhagens celulares (Alizadeh et al., 2000; Perou et al., 2000; van 't Veer et al., 2002). Esta estratégia melhora a comparação da expressão gênica entre amostras pois permite

corrigir diferenças existentes entre lâminas e lotes de lâminas, como a quantidade de DNA depositado e a morfologia dos pontos de DNA. No entanto, o uso de um RNA como referência exige grandes quantidades de material, o que pode ser uma limitação ao seu uso em projetos envolvendo um grande número de amostras a serem hibridizadas. Além disso, em estudos em que a coleção das amostras de RNA é prospectiva em relação ao começo dos experimentos de hibridização, como no caso deste projeto, esta metodologia também traz dificuldades uma vez que há a necessidade de garantir a existência do RNA referência ao longo de todo o projeto.

Já foram propostas diversas alternativas à utilização de um RNA referência, como o uso de DNA genômico (Kim et al., 2002), produtos de PCR representando a coleção de clones presentes no microarranjo (Gorreta et al., 2004), ou oligonucleotídeos complementares a um trecho comum presente nas sondas nos microarranjos (Dudley et al., 2002). A metodologia do oligonucleotídeo referência marcado fluorescentemente foi originalmente proposta por Church e colaboradores como uma forma de determinar razões de expressão gênica em experimentos com leveduras (Dudley et al., 2002). Como o oligonucleotídeo é complementar a todos os elementos no microarranjo, a fluorescência derivada desta molécula pode ser usada para avaliar a qualidade dos pontos de DNA na lâmina e facilitar o alinhamento da máscara de extração dos dados de intensidade pelo programa utilizado.

Devido à especificidade e sensibilidade da técnica de microarranjos, seu emprego se estende ainda mais, sendo usada para a detecção de SNPs (*Hap Map Project*) (Cutler et al., 2001), padrões aberrantes de metilações (Yan et al., 2001), alterações no número de cópias de genes (Pollack et al., 1999), *splicing* alternativo de RNA (Religio et al., 2005), detecção da presença de patógenos (Wang et al., 2002; Conejero-Goldberg et al.,

2005), comparação de genomas (Sidders et al., 2007; Ellison et al., 2008), entre outros usos.

Embora existam outras abordagens para a análise quantitativa da expressão gênica, tais como SAGE (Velculescu et al., 1995; Velculescu et al., 2000; Saha et al., 2002), MPSS (Brenner et al., 2000) e seqüenciamento de etiquetas de expressão (ESTs) (Adams et al., 1991), os custos associados a estas metodologias ainda é bastante elevado em comparação a hibridização com microarranjos de DNA (Wang, 2007). Por esta razão, estudos baseados na obtenção de perfis de expressão gênica individual de múltiplas amostras geralmente utilizam plataformas de microarranjos de DNA.

Dessa forma, os microarranjos de DNA têm sido extensivamente utilizados em estudos da expressão gênica de tecidos tumorais humanos (Virtanen e Woodgett, 2008). Esta técnica possibilita o estudo de padrões de expressão gênica e vias envolvidas com o câncer com potencial para (1) diagnóstico, prognóstico e predição de respostas a terapias (Ramaswamy et al., 2001; Dupuy e Simon, 2007), (2) descoberta de novos sub-tipos de câncer (Golub et al., 1999; Shedden et al., 2003; Lange et al., 2004) e (3) identificação de marcadores moleculares associados ao câncer e suas complexas interações (Shedden et al., 2003; Ransohoff, 2004).

Neste âmbito, o câncer de mama tem sido uma das doenças mais investigadas com o auxílio de microarranjos de DNA. A existência de uma elevada variabilidade genética entre os indivíduos limita a eficiência prognóstica dos marcadores moleculares conhecidos, como REr, RPr, Her2. Nos últimos 5 anos, grupos independentes têm conduzido estudos do perfil de expressão gênica com o objetivo de identificar assinaturas multi-gênicas capazes de fornecer uma classificação molecular mais robusta no câncer de mama (van 't Veer et al., 2002; Chang et al., 2005; Wang et al., 2005; Sotiriou et al., 2006; Naderi et al., 2007).

Inicialmente, a plataforma de microarranjos de cDNA foi utilizada para investigar a expressão de genes em tecido mamário normal, câncer de mama invasivo e metastático (Sgroi et al., 1999). Cerca de dois anos depois, esta mesma plataforma foi utilizada para investigar grupos de tumores de mama hereditários que possuíam mutações no gene *BRCA1* ou no gene *BRCA2*, pois as características histo-patológicas provenientes deste tumores são muito distintas (Hedenfalk et al., 2001). Assim, foram identificados perfis de expressão diferentes associados aos tumores de mama causados por mutações hereditárias distintas (Hedenfalk et al., 2001). Ainda no mesmo ano, outro grupo da Suécia decidiu explorar as diferenças fenotípicas associadas aos tumores de mama RER+ e RER- utilizando microarranjos de DNA, descobrindo distintos padrões de expressão entre os dois tipos tumorais (Gruvberger et al., 2001).

O primeiro grupo a encontrar um grupo de genes com padrão de expressão correlacionado com a ocorrência de metástase e levar estes resultados para a prática clínica foi um grupo de pesquisadores do *Netherlands Cancer Institute*, em Amsterdã (van 't Veer et al., 2002). Através de análise supervisionada utilizando 78 amostras de tumor de mama de pacientes com menos de 55 anos, com nódulos linfáticos negativo para infiltração de células metastáticas e que não passaram por nenhum tipo de terapia adjuvante, este grupo encontrou uma assinatura de prognóstico de 70 genes que é capaz de prever a ocorrência de metástase em 80% dos casos observados (van 't Veer et al., 2002). Um ano depois, este mesmo grupo validou este conjunto de 70 genes como fator de prognóstico em câncer de mama utilizando um conjunto de 295 amostras de tumores de mama de pacientes jovens com nódulo linfático negativo ou positivo para infiltração metastática e ainda pacientes que passaram ou não por terapias adjuvantes (van de Vijver et al., 2002). Ainda neste trabalho é mostrado que a assinatura de expressão dos 70 genes é o mais robusto fator de predição da evolução da doença em pacientes com

câncer de mama, independente de tratamentos adjuvantes utilizados, tamanho do tumor, grau histológico, idade e infiltração de células metastáticas nos nódulos linfáticos (van de Vijver et al., 2002). O conjunto de 70 genes constitui o primeiro kit de prognóstico molecular no câncer de mama aprovado pelo FDA (*Mamma print*).

Outro grupo de Rotterdam em colaboração com a empresa Veridex LLC (San Diego, USA), utilizou a mesma abordagem que o grupo NKI para chegar a um grupo de 76 genes, com poder discriminativo comparável à assinatura de 70 genes, capaz de prever a ocorrência de metástase em pacientes com câncer de mama de todas as idades, com nódulos negativos e que não receberam tratamento adjuvante (Wang et al., 2005). As duas principais diferenças entre estes dois estudos foi a plataforma de microarranjo utilizada (NKI, Agilent Technologies e Rotterdam, Affymetrix Technology) e o desenho experimental que, no caso do grupo Rotterdam, incluiu separar suas amostras em dois subgrupos de acordo com o *status* do receptor de estrogênio para construir sua assinatura. Assim como o grupo NKI, o grupo de Rotterdam utilizou um conjunto de 180 pacientes de múltiplas instituições que possuíam nódulo-negativo e que não passaram por tratamentos com terapias adjuvantes para validar sua assinatura de 76 genes (Foekens et al., 2006).

Estas duas assinaturas gênicas foram validadas pelo grupo internacional TRANSBIG, que conduziu uma validação independente em 302 pacientes provenientes de cinco diferentes centros europeus, utilizando plataformas de microarranjos e análises estatísticas de diferentes laboratórios (Buyse et al., 2006; Desmedt et al., 2007). Apesar das duas assinaturas gênicas apresentarem uma sobreposição de apenas três genes, as duas foram validadas nos mesmos conjuntos de pacientes.

Um grupo da Universidade de Stanford usou uma abordagem distinta para identificar uma assinatura gênica associada com a sobrevivência no câncer de mama.

Eles identificaram uma assinatura de expressão gênica associada com o mecanismo fisiológico de cicatrização que, de acordo com a hipótese levantada, estaria associada com a progressão do câncer e, portanto, com seu período de sobrevivência. A partir disto, eles identificaram um grupo de tumores de mama que expressava esta assinatura de cicatrização e perceberam que a evolução clínica deste grupo de tumores era pior que a do grupo que não expressava a assinatura de cicatrização (Chang et al., 2004). Este grupo mostrou que o classificador construído é capaz de melhorar a estratificação do risco de tumores iniciais quanto à recorrência do câncer (Chang et al., 2005).

Uma metodologia similar foi utilizada por Sotiriou e colaboradores na tentativa de encontrar padrões de expressão gênica associados ao grau histológico a fim de melhorar seu valor prognóstico (Sotiriou et al., 2006). Esta metodologia foi baseada no fato de que tumores com graus histológicos distintos possuem comportamentos clínicos também distintos (Elston e Ellis, 1991). No entanto, tumores com graus histológicos intermediários são a maior fonte de discrepância entre os patologistas, podendo apresentar fenótipos muito distintos (Loi et al., 2007). Dessa forma, foi construído um índice de expressão gênica do grau histológico de tumores de mama baseado na expressão gênica de um conjunto de 97 genes que foram encontrados expressos diferencialmente entre tumores com alto a baixo grau histológico (Sotiriou et al., 2006). Este conjunto de 97 genes possuía um enriquecimento na categoria funcional de proliferação celular. Interessantemente, os tumores de mama com graus histológicos intermediários não foram associados com nenhum padrão de expressão gênica distinto, mas aqueles que apresentaram má evolução da doença dentre estes tumores tiveram correlação com o padrão de expressão gênica correspondente a tumores com alto grau histológico. Resultados similares foram encontrados em conjuntos de pacientes de validação independentes e heterogêneos e em diferentes plataformas de microarranjos,

ênfatizando a reprodutibilidade desta assinatura e a relevância clínica dos dois sub-tipos identificados por esta assinatura de 96 genes associados a proliferação celular (Sotiriou et al., 2006).

A função biológica de proliferação celular também já foi aplicada em outros contextos, como para selecionar genes envolvidos com a biologia do câncer de mama para desenvolver um teste de prognóstico que associa uma taxa de recorrência da doença à amostra analisada (Oncotype DX, Genomic Health Inc). O teste, baseado no nível de expressão de 21 genes medido por RT-PCR, é capaz de prever a evolução clínica de pacientes com nódulos negativos e tratadas com tamoxifeno e já é usado na rotina clínica (Paik et al., 2004). Quando comparada a eficiência deste teste prognóstico, da assinatura gênica dos 96 genes (Sotiriou et al., 2006) e da assinatura de 76 genes associados à cicatrização (grande componente associado à proliferação celular) (Chang et al., 2004), os resultados encontrados foram muito similares, mostrando que o desempenho dos classificadores é comparável ao teste Oncotype DX já utilizado na clínica médica. Genes relacionados à proliferação celular estão presentes em uma porção significativa dos classificadores já desenvolvidos (Desmedt e Sotiriou, 2006), como nestas assinaturas discutidas acima.

Outras assinaturas de expressão gênica correlacionadas com a evolução clínica do câncer de mama já foram desenvolvidas baseadas, por exemplo, na diferença dos perfis de expressão gênica de tumores de mama selvagens e mutados para o gene *TP53* (Miller et al., 2005). Apesar desta assinatura com 32 genes não englobar genes alvos do *TP53* ou da via deste, ela apresentou um poder preditivo maior quanto à evolução dos pacientes com tumores de mama quando comparada à simples determinação da existência de mutação no gene *TP53* nos tumores através de seqüenciamento de DNA (Miller et al., 2005). A notar, esta assinatura também possui muitos genes associados à

proliferação celular, à semelhança de assinaturas identificadas em outros trabalhos de classificação molecular do câncer de mama (Ivshina et al., 2006; Sotiriou et al., 2006; Mosley e Keri, 2008).

A partir da hipótese de que a diversidade fenotípica dos tumores de mama deveria ser acompanhada de uma diversidade do perfil de expressão gênica correspondente que, por sua vez, poderia ser identificada por experimentos de microarranjos de DNA, um grupo da Universidade de Stanford encontrou 4 subtipos distintos de câncer de mama (Perou et al., 2000). Para tal, 65 amostras de tumores de mama, provenientes de 42 pacientes diferentes, foram hibridizadas contra um microarranjo de DNA representando cerca de 8 mil genes humanos (Perou et al., 2000). Inicialmente, as amostras foram subdivididas em dois grandes grupos, de acordo com o *status* do receptor de estrogênio. Padrões de expressão distintos de sub-grupos de tumores dentro destes dois grupos foram então identificados: grupo RER+, epitelial luminal; grupo RER-, perfil mama normal, perfil basal, ERBB2+ (Perou et al., 2000). Numa tentativa de validar a análise em um estudo mais amplo, um novo grupo de 85 amostras foi utilizado em experimentos de microarranjo para identificação dos sub-tipos moleculares encontrados previamente (Perou et al., 2000). Neste trabalho, foram identificados novos sub-grupos dentro do perfil de expressão do conjunto epitelial luminal, subdividindo este grupo em luminal A, B e C (Sorlie et al., 2001). Os diferentes sub-tipos foram investigados quanto ao tempo de sobrevida das pacientes e evolução clínica, tendo sido observado que pacientes com os sub-tipos tumorais ERBB2+ e basal apresentaram menor tempo de sobrevida e maior chance de recorrência da doença (Sorlie et al., 2001). Estes sub-tipos moleculares de tumores de mama foram confirmados em outros estudos (Sorlie et al., 2003) e continuam sendo referência quanto

à classificação dos tumores de mama em muitos outros estudos prospectivos (Finetti et al., 2008; Goncalves et al., 2008; Rakha et al., 2008; Smid et al., 2008).

Desta forma, muitos estudos já foram conduzidos em câncer de mama, desde a investigação fenotípica do *status* do receptor de estrogênio no tecido tumoral até a definição de perfis moleculares associados ao desenvolvimento de metástase na paciente. No entanto, estas assinaturas moleculares são distintas umas das outras, com quase nenhuma sobreposição nos genes que as compõem, e este aspecto não diminui o potencial prognóstico de cada uma delas em câncer de mama (Mosley e Keri, 2008). Por outro lado, apesar de algumas destas assinaturas moleculares já terem resultado em testes clínicos em fase de experimentação em conjuntos de pacientes independentes, muita discussão ainda é levantada sobre a capacidade preditiva e prognóstica destes no câncer de mama em relação a fatores e marcadores clássicos utilizados na clínica (Koscielny, 2008; Puztai, 2008). Este fato indica que esta área de investigação não foi esgotada, podendo ainda receber novas contribuições para o desenvolvimento de novas ferramentas de prognóstico ou terapias direcionadas, no câncer de mama (Garman et al., 2007).

Adicionalmente, as plataformas utilizadas até o momento são direcionadas para a identificação de genes codificadores para proteínas. Começa-se a abrir um novo campo de pesquisa no estudo da expressão de tecidos tumorais humanos envolvendo uma nova classe de moléculas que foi muito pouco explorada no contexto do câncer até o momento, que são os RNAs não-codificadores (ncRNAs) (Ji et al., 2003; Reis et al., 2004; Calin et al., 2007; Brito et al., 2008).

1.3. RNAs não-codificadores e câncer

O organismo humano apresenta um desenvolvimento muito mais complexo que um camundongo e, ainda assim, ambas espécies apresentam aproximadamente o mesmo número de genes (de 20 a 25 mil) e uma proporção comparável de seqüências codificadoras de proteínas em seus genomas (1,2% em humano e 1,3% em camundongo) (Frith et al., 2005). No entanto, notou-se uma grande diferença na quantidade de DNA não-codificador entre estes dois genomas, já que esta é estimada como sendo 47 vezes maior que a quantidade de DNA codificador no genoma humano, enquanto que em camundongos essa diferença é de apenas 43 vezes (Frith et al., 2005). A relevância desta diferença entre organismos mais e menos complexos só pôde ser compreendida com a descoberta de que a maior parte do genoma dos mamíferos é transcrita, dando origem a uma variedade de RNAs sem potencial codificador de proteínas uma ordem de magnitude maior que a prevista a partir do número de exons codificadores para proteína anotados no genoma humano (Kapranov et al., 2002; Carninci et al., 2005). Esta fração não-codificadora do transcrito humano consiste predominantemente de RNAs originados em regiões intrônicas e intergênicas do genoma (Cawley et al., 2004; Kampa et al., 2004; Kapranov et al., 2007).

Muitos destes estudos só foram possíveis graças ao avanço das técnicas de microarranjo com relação à quantidade de elementos depositados na lâmina, permitindo a construção de plataformas capazes de interrogar em um mesmo experimento a atividade transcricional em cromossomas completos, com alta resolução e sem viés para regiões codificadoras de proteínas (Cheng et al., 2005). Em uma análise do transcrito de dez cromossomos humanos utilizando a técnica de *tiling array* a uma resolução de cinco nucleotídeos, transcritos sem sinal de poliadenilação e sem anotação no genoma humano corresponderam à maior parte dos transcritos observados em oito linhagens

celulares diferentes (Cheng et al., 2005). Estes transcritos ainda possuíam expressão específica a determinadas linhagens celulares e, em geral, não apresentavam uma grande variação nos níveis de expressão (Cheng et al., 2005; Kapranov et al., 2005).

Ainda utilizando *tiling arrays* genômicos, um estudo recente gerou mapas transcricionais que revelam a existência de conjuntos de ncRNAs poliadenilados, que se acumulam preferencialmente na fração nuclear ou citoplasmática de células humanas e que potencialmente são processados em RNAs curtos (Kapranov et al., 2007). Contudo, os mecanismos de biogênese e as funções moleculares desses ncRNAs permanecem desconhecidos.

Nosso grupo tem caracterizado uma nova classe de ncRNAs longos, sem *splicing*, expressos em regiões intrônicas do genoma humano (Reis et al., 2004; Nakaya et al., 2007). Têm sido utilizados microarranjos de cDNA com 4 mil elementos construídos em nosso laboratório (Reis et al., 2004), e mais recentemente, têm sido utilizados *oligoarrays* intron-exon com 44k mil sondas sintetizados *in situ* sob encomenda (Nakaya et al., 2007). Os *oligoarrays* intron-exon contêm sondas para RNAs não-codificadores intrônicos e para o transcrito codificador de proteína correspondente, permitindo a avaliação simultânea do nível de expressão de transcritos exônicos e intrônicos originários do mesmo *locus* gênico (Nakaya et al., 2007). Utilizando esta plataforma, nosso grupo identificou conjuntos de transcritos intrônicos com expressão restrita a tecidos humanos provenientes de fígado, próstata e rim (Nakaya et al., 2007). A expressão de sub-conjuntos de transcritos intrônicos, expressos com orientação senso ou antisenso, apresenta-se correlacionada à expressão dos transcritos codificadores para proteína derivados dos mesmos *loci* gênicos, sugerindo que os RNAs intrônicos poderiam estar regulando a abundância ou o padrão de *splicing* destas seqüências codificadoras. Interessantemente, observou-se o enriquecimento de genes pertencentes a

categoria funcional de “Regulação da Transcrição” entre os genes com atividade transcricional intrônica antisense (Nakaya et al., 2007). Somado a outros estudos realizados em nosso laboratório (Reis et al., 2004; Louro et al., 2007; Brito et al., 2008), estes resultados apontam para um papel de ncRNAs intrônicos no controle da expressão gênica em tecidos humanos, e em especial, em doenças onde o descontrole da expressão gênica assume um componente relevante, como no câncer.

A conservação dos transcritos intrônicos em camundongos e humanos também foi investigada em nosso grupo. Utilizando-se um microarranjo de cDNA enriquecido em transcritos intrônicos, foram interrogados, separadamente, tecidos humanos e de camundongos provenientes do fígado, próstata e rim (Louro et al., 2008). A partir dos dados obtidos dos experimentos, foi possível identificar um conjunto de 198 transcritos em comum expressos nos diferentes tecidos humanos e de camundongos. Dentre estes, 22 mapeiam em regiões intrônicas com evidência de conservação evolucionária entre humanos e camundongos (Louro et al., 2008).

Com a exceção de pequenos RNAs envolvidos no processamento de RNAs no núcleo (snRNA, snoRNAs) ou na modulação da expressão gênica (miRNAs, piRNAs), ou ainda alguns ncRNAs longos envolvidos no silenciamento gênico (*XIST*, *HOTAIR*), as funções da maior parte dos diversos tipos de ncRNAs produzidos em células humanas permanecem desconhecidas (Mattick e Makunin, 2006; Pauler et al., 2007; Woo e Kingston, 2007).

Os microRNAs são RNAs não codificadores curtos, com 21 nucleotídeos, que possuem funções regulatórias através da repressão transducional e degradação de mRNAs (Amaral et al., 2008). MicroRNAs podem ser resultado do processamento de RNAs não codificadores longos, como mostrado por um estudo de He e colaboradores (He et al., 2008). Utilizando o conjunto de 34.030 RNAs não codificadores longos

expressos em camundongos, este grupo localizou 97 fortes candidatos a microRNAs. Deste grupo, um número igual a 52 RNAs não codificadores foram identificados como precursores de pré-miRNA, com evidência experimental obtida através de resultados de microarranjos (He et al., 2008).

Entre os poucos ncRNAs longos com função conhecida, encontram-se transcritos envolvidos em mecanismos de inativação do cromossomo X, *imprinting* e exclusão alélica (Yang e Kuroda, 2007) e regulação de formas variantes de *splicing* (Yan et al., 2005). O gene *SRA* (RNA ativador de receptor de esteróides) codifica um ncRNA longo que funciona como um co-ativador transcricional para certos receptores de hormônios esteróides (Lanz et al., 1999).

Foi mostrado em humanos que sítios dos fatores de transcrição Sp1, cMyc e p53 estão localizados não apenas na região terminal 5' de genes codificadores de proteínas mas também na região 3' de genes conhecidos e que estão significativamente correlacionados com ncRNAs intrônicos e intergênicos (Cawley et al., 2004). Este resultado sugere que possa existir co-regulação através dos mesmos fatores ambientais de pares de RNA codificadores para proteínas e não-codificadores (Cawley et al., 2004).

Em um ensaio com linhagens celulares expostas ao hormônio andrógeno, 39 ncRNAs intrônicos apresentaram níveis de expressão regulados pela presença do andrógeno (Louro et al., 2007). Os níveis de expressão de alguns destes transcritos intrônicos se mostraram correlacionados com os níveis de expressão dos genes codificadores para proteínas dos mesmos *loci* ou com formas alternativas de *splicing* das mesmas. Estes achados em nosso grupo indicaram que RNAs não-codificadores intrônicos podem ser regulados por sinais fisiológicos comuns, como hormônios (Louro et al., 2007).

As diversas linhas de evidência têm nos mostrado que a maior parte do genoma humano é transcrito em ncRNAs, e que, juntamente com as proteínas, estas moléculas são responsáveis pela execução da informação genética de diversos mecanismos celulares (Mattick e Makunin, 2006). A maior parte destas funções estão relacionadas a um papel importante exercido por ncRNAs na organização da expressão gênica, atuando na arquitetura da cromatina, memória epigenética, transcrição, *splicing*, editoração, tradução e *turnover* de RNA (Mattick e Makunin, 2006).

Embora os mecanismos moleculares através dos quais agem as várias classes de ncRNA ainda não tenham sido totalmente elucidados, as evidências apontam para um papel destas moléculas na modulação de características e processos fisiológicos complexos, incluindo doenças multi-fatoriais como o câncer (Mattick e Makunin, 2006). Dessa forma, torna-se extremamente relevante o estudo do papel destas moléculas em diversos tecidos tumorais, já que a expressão também é tecido específica (Mattick e Makunin, 2005). Assim, será possível se entender melhor a rede regulatória da qual estes ncRNAs participam, tornando possível a melhor compreensão da regulação da expressão gênica, da gênese ou progressão de estados patológicos como o câncer (Mattick e Makunin, 2006). Neste contexto, a classe de ncRNAs que têm mais sido investigada é a dos microRNAs. Um dos trabalhos pioneiros a mostrar o envolvimento destas moléculas com o câncer em grande escala, foi o desenvolvido por Calin e colaboradores em 2004, mostrando que metade dos microRNAs analisados por seu grupo (98 microRNAs) estavam localizados em regiões genômicas ou em sítios frágeis envolvidos em câncer, mostrando sua relevância na doença (Calin et al., 2004). Já em 2005, um grupo de Harvard e do MIT foi o primeiro a estudar o perfil de expressão gênico destas moléculas associado a diversos tipos de tumores humanos, observando expressão diferencial dos 217 microRNAs analisados entre todos os tipos tumorais estudados (Lu et al., 2005). Os

diferentes perfis de expressão deste conjunto de microRNAs foi capaz de classificar os tumores estudados de acordo com a origem da malignidade e identificar aqueles de pior prognóstico com maior acurácia quando comparado a perfis de expressão gênica definidos a partir de RNAs codificadores (Lu et al., 2005).

No contexto do câncer de mama, estudos utilizando a técnica de microarranjo foram capazes de identificar um grupo de microRNAs cujos níveis de expressão foram correlacionados com características bio-patológicas dos tumores, como: expressão dos receptores de estrogênio e progesterona, estágio tumoral, invasão vascular e índice de proliferação (Iorio et al., 2005). Além disto, o perfil de expressão dos microRNAs estudados foi também capaz de distinguir entre tecido mamário normal e tumoral (Iorio et al., 2005). Mais recentemente, estudos utilizando linhagens celulares metastáticas derivadas de câncer de mama humano foram capazes de identificar outro grupo de microRNAs que apresentavam baixa expressão nestas células (Tavazoie et al., 2008). Quando os níveis de expressão destes microRNAs foram restaurados nestas linhagens celulares, houve uma redução da disseminação metastática em relação ao local do tumor mamário primário (Tavazoie et al., 2008). Além disso, foram identificados dois microRNAs, miR-126 e miR-335, cujas expressões são perdidas em tumores de mama primários de pacientes que sofrem recorrência da doença. Portanto, a perda da expressão em tumores de mama de qualquer um destes dois microRNAs é associada com má evolução clínica da doença e metástase (Tavazoie et al., 2008).

Nosso grupo foi o primeiro a mostrar um conjunto de transcritos intrônicos longos expressos em tecido tumoral (Reis et al., 2004). Usando uma plataforma de cDNA com 4 mil elementos enriquecida em exons de genes relacionados com câncer na literatura, e fragmentos de transcritos mapeando em regiões intrônicas e intergênicas do genoma, detectamos níveis de expressão ubíquos de ncRNAs em uma coleção de 27

tumores de próstata (Reis et al., 2004). Interessantemente, foi observado que quando comparada à fração de RNAs codificadores (2%) e ncRNAs provenientes de regiões não anotadas do genoma (1%), uma fração maior dos ncRNAs intrônicos detectados nos tumores de próstata (7%) apresentou expressão correlacionada com a agressividade dos tumores analisados. (Reis et al., 2004). A análise molecular dos ncRNAs correlacionados com agressividade nos tumores de próstata revelou que seis dos doze ncRNAs com expressão mais correlacionada com o grau de diferenciação dos tumores de próstata apresentaram expressão antisense em relação ao gene codificador do mesmo *locus* gênico (Reis et al., 2004). A extensão destes transcritos não codificadores e antisense revelou transcritos longos variando de 0,6 kb até 1,1 kb. (Reis et al., 2004).

Resultados semelhantes, obtidos com o uso da mesma plataforma de microarranjos de DNA, foram observados no estudo da expressão de câncer renal em nosso grupo (Brito et al., 2008). Um conjunto de transcritos com baixa expressão em tecido de rim tumoral em relação ao normal foi identificado e, dentre estes, seis correspondem a ncRNAs provenientes de regiões intrônicas de genes conhecidos, indicando que estes transcritos possam um papel de supressores de tumores em tecido renal (Brito et al., 2008).

Como exemplo do envolvimento de ncRNAs longos em câncer, em 2007 foi descrito o potencial envolvimento do MALAT-1, um longo ncRNA com 7kb de tamanho, em câncer de fígado de camundongos (Lin et al., 2007). Esta molécula apresentou-se de 6 a 7 vezes mais expressa em tecido tumoral de fígado em relação ao tecido normal, funcionando como um marcador tumoral (Lin et al., 2007). Assim, baseado nestes achados, torna-se clara a necessidade de investigar o papel destas moléculas no contexto do câncer, explorando suas capacidades no prognóstico e classificação molecular desta doença.

Neste contexto, este projeto teve como interesse expandir as observações feitas em tumores de próstata e rim pelo nosso grupo em uma topologia distinta, o câncer de mama, avaliando a expressão de transcritos intrônicos em relação a aspectos relevantes à doença do câncer de mama como o *status* do receptor de estrogênio em tumores de mama e a ocorrência de metástase.

2. OBJETIVOS

2.1. Objetivo Geral

Investigar a expressão gênica em amostras de tumores de mama de pacientes com seguimento clínico conhecido, utilizando um microarranjo contendo sondas para genes codificadores de proteínas e RNAs não-codificadores para i) obter novas pistas sobre as bases moleculares da doença, incluindo alteração na expressão de RNAs não-codificadores e ii) identificar novos candidatos a marcadores de prognóstico no câncer de mama.

2.2. Objetivos Específicos

- 1) Montagem de um banco de tecidos frescos com amostras de tumores de mama para estudos de expressão gênica, provenientes de pacientes com acompanhamento clínico pré e pós-operatório.
- 2) Isolamento de RNA total, avaliação da qualidade e hibridização com microarranjos de cDNA com 4 mil elementos, enriquecidos em genes relevantes em câncer e ncRNAs intrônicos e intergênicos.
- 3) Análises dos dados de expressão gênica obtidos para avaliar o padrão de expressão de ncRNAs no câncer de mama
- 4) Análises supervisionadas para busca de assinaturas moleculares correlacionadas com características anatomo-patológicas e clínicas disponíveis.
- 5) Anotação genômica e funcional das assinaturas de expressão gênicas identificadas para identificação de vias moleculares relevantes na progressão tumoral.
- 6) Comparação dos resultados obtidos com estudos do perfil de expressão disponíveis na literatura e relevantes em câncer de mama.

7) Seleção de um grupo de transcritos para confirmação da expressão através de RT-qPCR e determinação da orientação de transcrição de ncRNAs intrônicos selecionados.

3. MATERIAIS E MÉTODOS

3.1. Coleta e armazenamento das amostras de tumor de mama

As amostras tumorais de mama utilizadas neste projeto foram recolhidas no Hospital HDES (Arquipélago de Açores, Portugal) no momento da cirurgia de mastectomia das pacientes com câncer de mama. As amostras foram excisadas do tecido tumoral e congeladas em nitrogênio líquido num período de até 20 minutos. Todos as pacientes assinaram termo de consentimento informado autorizando a utilização das amostras neste projeto de pesquisa, também aprovado pelo Comitê de Ética do HDES. Posteriormente, todas os fragmentos de tecido foram revisados e classificados por um mesmo patologista. Se necessário, as amostras foram macro-dissecadas de modo a garantir que pelo menos 80–90% do fragmento analisado correspondesse a tecido mamário tumoral.

O envio das amostras de Portugal para o nosso laboratório em São Paulo foi padronizado de forma a manter a integridade do RNA. Portanto, elas foram enviadas ainda na forma de tecido, embebidas no reagente *RNALater* e acondicionadas em gelo seco. Este reagente consegue manter a viabilidade das amostras por 7 dias (sob temperatura de 18 a 25°) ou até 4 semanas (sob temperatura de 2 a 8° C). No momento em que as amostras foram recebidas em nosso laboratório, elas foram estocadas no freezer -80°C, ainda submersas em *RNALater*. Este reagente só é descartado no momento da extração de RNA total do tecido tumoral mamário.

3.2. Extração de RNA total de amostras de tumor de mama

A extração de RNA total é feita através do método tiocianato de guanidina-fenol-clorofórmio. Para tal, utilizamos o reagente Trizol (Invitrogen) na relação de 1ml para

cada 50mg de tecido já macerado sob nitrogênio líquido. Após esta etapa, a amostra é homogeneizada e incubada por 5 minutos a temperatura ambiente. Divide-se então a amostra em alíquotas de 1ml por tubo, adiciona-se 200µl de clorofórmio e em seguida nova incubação a temperatura ambiente por 5 minutos. Segue-se uma centrifugação dos tubos a 11.000xg por 15 minutos a 2°C e a fase aquosa é separada. A ela são adicionados 500µl de isopropil seguido de incubação por 10 minutos a temperatura ambiente. Após nova centrifugação, o sobrenadante é então removido e descartado. 1ml de etanol 75% é adicionado ao tubo e segue-se nova centrifugação (7.000xg por 5 minutos a 2°C). Após remoção do sobrenadante, o precipitado é dissolvido em 30µl de água DEPC.

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 e por isso apresenta alta precisão quanto aos resultados. Como medida de qualidade quanto à degradação do RNA total utilizamos a razão das intensidades das bandas de RNA ribossomal 28S e 18S, sendo o valor de referência igual a 2, mas este valor pode variar de acordo com a topologia do tecido e organismo estudado. Outra forma de avaliar o sucesso da extração do RNA total do tecido tumoral mamário é a quantificação das razões de medida óptica realizadas a 260nm (pico de absorbância do RNA) e 280nm (pico de absorbância de proteínas), cujo valor de referência também é igual a 2, podendo variar de acordo com o organismo. Estas medidas de espectrofotometria são realizadas utilizando-se o espectrofotômetro NanoDrop (Thermo Scientific). Este equipamento produz medidas muito acuradas em relação a equipamentos de eletroforese comuns pois a distância do caminho ótico é muito reduzida. A partir deste equipamento também é possível obter medidas precisas da concentração das amostras de RNA total obtidas a partir da extração.

3.3. Amplificação linear do RNA mensageiro

A quantidade de RNA total obtida a partir de uma grande parte das amostras de tumores de mama se mostrou insuficiente para a obtenção de alvos fluorescentes e hibridização com microarranjos. Para contornar esta limitação, foi utilizado um protocolo de amplificação linear do RNA mensageiro (aRNA) descrito na literatura (Wang et al., 2000) e padronizado em nosso laboratório. Para obter aRNA suficiente para hibridizar 2 réplicas de cada amostra, foram realizados dois ciclos de amplificação com cada uma das 74 amostras de RNA total de mama:

1º Ciclo:

- Síntese da primeira fita de cDNA – 1µg de RNA total em 9µl de água é adicionado a 1µl de primer dT-T7 (0,5µg/µl) e incubado a 65° C por 5 minutos. A esta reação é adicionado então tampão 5X First Strand 1X (Invitrogen), 1µl de TS primer (0,5µg/µl), 2µl de 10mM dNTP, 1µl de RNAsin RNase Inhibitor (Promega), 2µl da enzima Superscript III (Invitrogen), completando 20µl de reação e incubado por 42°C por uma hora e meia.
- Síntese da segunda fita de cDNA – À reação anterior é adicionado 106µl de água, 15µl de tampão Advantage PCR (BD Biosciences), 3µl de 10mM dNTP, 3µl do mix da Advantage cDNA Polymerase (BD Biosciences), 1µl de RNase H (Invitrogen). A reação é incubada por 37°C por 5 minutos, 94°C por 2 minutos, 65°C por 1 minuto e 75°C por 30 minutos e logo depois interrompida com 7,5µl de 1M NaOH com 2mM EDTA a 65°C por 10 minutos.
- Purificação do cDNA fita dupla – À reação é adicionado 1µl de acrilamida linear (0,2µg/µl) e logo depois ela é transferida a tubos Phase Lock Gel (PLG - Eppendorf) contendo 160µl de Fenol:Clorofórmio:Álcool Isoamil. Os tubos são centrifugados por 6

minutos a 13.000 rpm. A fase aquosa é então transferida para tubos contendo 75µl de 7,5M de acetato de amônia e a estes é depois adicionado 1mL de etanol 100%. Segue-se centrifugação imediata por 20 minutos a 14.000 rpm. O precipitado então é lavado duas vezes com 500µl de etanol 100% (8 minutos a 14.000 rpm), seco e solubilizado em 8µl de água.

- Transcrição *In Vitro* – Ao cDNA dupla fita purificado é adicionado então 2µl de cada um dos nucleotídeos a 75mM, 2µl do tampão da reação e 2µl do mix da enzima T7 RNA polimerase (Ambion). A reação é incubada por 6 horas a 37°C.
- Purificação do aRNA – Neste procedimento, adicionamos ao tubo da reação de transcrição reversa 1mL de Trizol e 200µl de clorofórmio, misturando por inversão por 15 segundos. A reação é então incubada por 3 minutos à temperatura ambiente e centrifugada por 15 minutos a 13.000 rpm a 4°C. A fase aquosa é transferida a um novo tubo contendo 500µl de isopropanol e incubada por 10 minutos a temperatura ambiente. Segue-se nova centrifugação por 25 minutos a 13.000 rpm a 4°C. O precipitado é lavado duas vezes com 800µl de álcool 70% (centrífuga a 13.000 rpm por 10 minutos), seco e solubilizado em 20µl de água.

2º Ciclo:

- Síntese da primeira fita de cDNA – 1µg de aRNA em 8µl de água é adicionado a 1µl de *random hexamer* (2 µg/µl) e incubado a 65° C por 5 minutos. A esta reação é adicionado então tampão 5X First Strand 1X (Invitrogen), 2µl de dT-T7 primer (0,5µg/µl), 2µl de 10mM dNTP, 1µl de RNAsin RNase Inhibitor (Promega), 2µl da enzima Superscript III (Invitrogen), completando 20µl de reação e incubado por 42°C por uma hora e meia.
- Síntese da segunda fita de cDNA – Procedimentos idênticos à síntese da segunda fita no 1º ciclo.

- Purificação do cDNA fita dupla – Procedimentos idênticos à purificação do cDNA dupla fita no 1º ciclo.
- Transcrição *In Vitro* – Ao cDNA dupla fita purificado é adicionado então 4µl de cada um dos nucleotídeos A, G, e CTP a 75mM, 2µl do nucleotídeo UTP a 75mM, 3µl do nucleotídeo modificado 5-(3-amino-alil)-UTP a 50mM, 8µl de água, 4µl do tampão da reação e 4µl do mix da enzima T7 RNA polimerase (Ambion). A reação é incubada por 13 horas a 37°C.
- Purificação do aRNA e Tratamento com DNase – À reação é adicionado 4µl da DNase I da Ambion seguindo sua incubação a 37°C por 15 minutos. Seu volume é então completado para 100µl e segue-se o protocolo do *kit* da Qiagen *RNAeasy clean-up* para a purificação do aRNA.

3.4. Acoplamento dos alvos de aRNA a moléculas fluorescentes

Após a reação de amplificação, as moléculas de aRNA produzidas incorporaram análogos de amino-alil-UTP modificados. Dessa forma, a molécula de aRNA está pronta para ser acoplada a moléculas fluorescentes Cy5 através da esterificação destas à estrutura modificada do amino-alil-UTP:

- Reação de Acoplamento – Secar totalmente 5µg de aRNA em uma centrífuga à vácuo e solubilizá-lo em 4,5µl de tampão de acoplamento (0,1M tampão carbonato pH = 8,5-9,0). Adicionar 2,5µl de água e 4µl do corante monoreativo Cy5 (Amersham Biosciences). Misturar e incubar a reação no escuro por uma hora, misturando a cada 15 minutos.
- Reação de *Quenching* – Adicionar 6µl de hidroxilamina 4M e incubar por 15 minutos no escuro a temperatura ambiente.

- Purificação da sonda – Usar o *kit* da Qiagen *RNAeasy clean-up* para a purificação. Concentrar as amostras em uma centrífuga à vácuo para um volume final de até 50µl.

3.5. Hibridização de alvos fluorescente com microarranjos de cDNA

O microarranjo de cDNA utilizado nos experimentos de hibridização foi construído em nosso próprio laboratório em 2002, no começo do projeto CAGE-Câncer. Este microarranjo contém 3.840 fragmentos de cDNA que foram selecionados através de uma análise informática das 1.016.501 ESTs provenientes de diversos tipos de tumores humanos, geradas a partir da técnica ORESTES entre 2000-2001 durante o Projeto Genoma Humano do Câncer (HCGP - <http://www.ludwig.org.br/ORESTES>) (Dias-Neto et al., 2000; Camargo et al., 2004). O microarranjo contém também 30 controles positivos (*house keeping genes*) e negativos (regiões intergênicas de levedura e genes de bactéria e planta sem similaridade com transcritos humanos). Cerca de 60% (~2.200) das seqüências depositadas no microarranjo de cDNA correspondem a regiões exônicas de genes já conhecidos codificadores de proteínas relacionados com câncer na literatura, sendo que 250 destas já foram reportados como envolvidos na progressão e desenvolvimento de tumores de próstata. Aproximadamente 20% das seqüências contidas no microarranjo (~800) correspondem a regiões intrônicas de genes conhecidos, estando ainda presentes cerca de 250 fragmentos de cDNA que mapeiam em regiões intergênicas do genoma (Reis et al., 2004).

Os alvos fluorescentes de aRNA de cada um dos 74 tumores de mama marcados com Cy5 foram hibridizadas contra o microarranjo de cDNA, em duplicata. A hibridização foi feita ainda na presença de um oligonucleotídeo sintético de 23-mer acoplado a uma molécula de Cy3 em sua extremidade 5' (*RefOligo*) (Peixoto et al., 2006). Como todos os transcritos do microarranjo foram obtidos através da reação de

PCR utilizando *primers* universais do vetor pUC18, todos os cDNAs depositados no microarranjo possuem uma seqüência em comum, complementar a uma região do vetor pUC18. O *RefOligo* é complementar a esta região e por isso hibridiza especificamente a todos os pontos de cDNA presentes no microarranjo que foram provenientes de nossa biblioteca de cDNA (Figura 2).

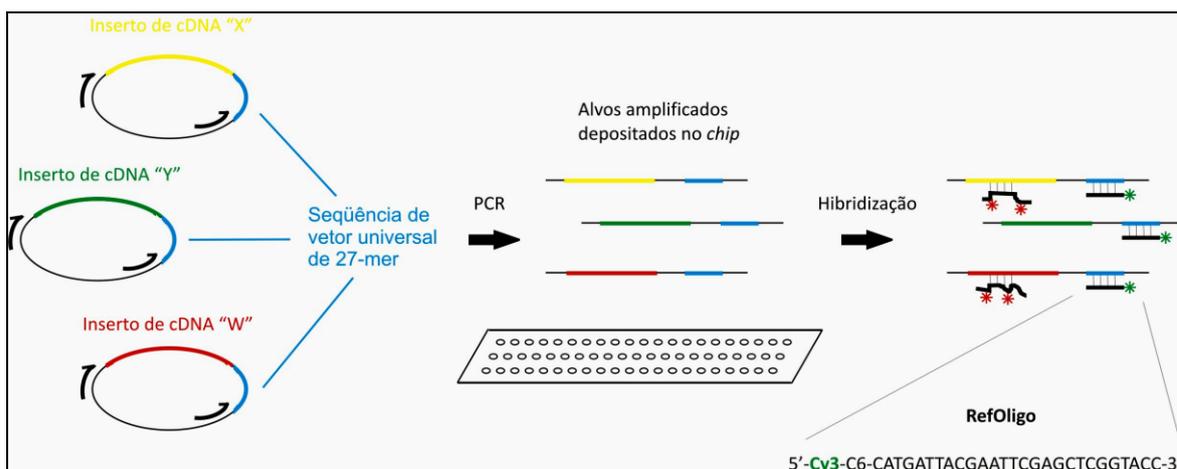


Figura 2. Representação esquemática da hibridização do *RefOligo* com o microarranjo de cDNA. Inseros de cDNA humanos que foram clonados em vetores pUC18 foram amplificados por PCR usando *primers* universais de forma que cada produto da reação de amplificação depositado no microarranjo contém uma seqüência comum na ponta 5' derivada do vetor. A molécula de *RefOligo* consiste em um oligonucleotídeo sintético de 27-mer acoplado ao fluoróforo Cy3 (marcado com asteriscos verdes) que é complementar à seqüência da ponta 5' do vetor, presente em cada uma das sondas de cDNA depositadas no microarranjo. Esta molécula de referência comum foi co-hibridizada com os microarranjos juntamente com os alvos produzidos a partir de RNA extraído de tecidos humanos e acoplados a uma molécula do fluoróforo Cy5 (marcado com asteriscos vermelhos).

A solução de hibridização é composta pelo tampão de hibridização (100µl de formamida 50%, 50µl de tampão de hibridização de microarranjo 4x versão 2 - Amersham Biosciences), o aRNA acoplado a Cy5 em um volume de 50µl e 1µl de *RefOligo* (100µM). A solução de hibridização é mantida a 92°C por dois minutos para desnaturação e transferido para gelo imediatamente. Segue-se centrifugação a 8.000xg por três minutos e aplicação no hibridizador automático (*Generation III Microarray Systems* - Amersham Biosciences), que fará a hibridização por 16 horas a 42°C. Todas

as 74 amostras do banco tumoral mamário foram hibridizadas no hibridizador automático.

Após a hibridização, segue-se a lavagem automática das lâminas ainda no hibridizador automático, que é feita com as soluções abaixo:

- Lavagem 1 – 1ml do tampão 1x SSC, 0,2% SDS por 3 minutos;
- Lavagem 2 – 1 ml do tampão 0,1x SSC, 0,2% SDS por 5 minutos;
- Lavagem 3 – jato com 2 ml do tampão 0,1x SSC;
- Lavagem 4 – jato com isopropanol para secagem das lâminas;
- Jato de ar.

3.6. Captura da imagem das lâminas hibridizadas e extração dos dados

Após a hibridização, a detecção dos alvos fluorescentes das lâminas é feita imediatamente para evitar o decaimento de sua fluorescência. As 148 lâminas hibridizadas contra as amostras de aRNA tumoral de mama e *RefOligo* tiveram sua imagem capturada utilizando-se o *scanner* que faz parte do conjunto de equipamentos *Generation III Microarray Systems* (Amersham Biosciences).

As imagens resultantes foram processadas utilizando-se o *software ArrayVision* (Imaging Research Inc.) que fornece intensidades totais de cada um dos pontos de cDNA depositados no microarranjo subtraídas do sinal de ruído local (intensidades medidas em uma moldura definida ao redor de cada ponto).

3.7. Filtragem e normalização das intensidades de expressão gênica

Para cada amostra de aRNA foram feitas 2 hibridizações (réplicas técnicas). Após a captura dos dados obtidos dos experimentos de hibridização, o primeiro passo é a filtragem dos dados de pontos de cDNA da lâmina que apresentam valores abaixo do sinal de ruído. Este sinal de ruído foi definido em nossos experimentos como sendo duas vezes o desvio padrão da média de intensidade de cerca de 130 pontos no microarranjo

onde foi depositado apenas solução de 50% DMSO. Optou-se por utilizar este critério por ser mais restritivo que a filtragem baseada no uso de cDNA de planta como controle negativo, resultando na filtragem de um número maior de elementos. Após a filtragem, segue-se a eliminação de valores espúrios de baixa qualidade utilizando-se o programa *ArrayStat* (v. 1.0 Imaging Research Inc.). O programa *ArrayStat* utiliza um modelo de distribuição de erro estimado a partir de todas as réplicas de todas as sondas analisadas para avaliar a reprodutibilidade das réplicas técnicas. Cada lâmina possui seu conjunto de pontos de cDNA depositado em duplicata. Portanto, para cada amostra existem 4 valores de intensidade para cada cDNA depositado na lâmina. Primeiro, o programa centraliza os dados de intensidade das 4 réplicas de cada sonda. Em seguida, são eliminados da análise os cDNAs que possuem mais do que uma das quatro réplicas com mais do que dois desvios absolutos da mediana (MADs). Em seguida foi calculada a média das sondas restantes.

Neste projeto foram testados dois tipos de normalização: pela média aparada e pelo método do quantil (Bolstad et al., 2003). O primeiro tipo de normalização consiste em calcular a média de cada um dos conjuntos de dados de cada amostra de aRNA, excluindo-se os valores mais altos e mais baixos. Neste caso, foram excluídos 40% dos dados no total. Esta média aparada é então utilizada para a normalização de cada um dos valores da lâmina.

O segundo método de normalização é mais eficaz contra efeitos de lotes de lâminas de microarranjo diferentes (Bolstad et al., 2003). Este método se baseia na ordenação decrescente dos valores de intensidade das diferentes lâminas (com uma etiqueta identificadora de cada transcrito) e cálculo da média de cada um dos pontos de cDNA. Estes valores médios irão substituir todos os valores de cada uma das lâminas e

através da etiqueta de identificação de cada transcrito os valores de cada um dos pontos de cDNA retornarão a sua posição original.

Paralelamente, além da utilização direta das intensidades de expressão provenientes do canal de RNA (desenho experimental de 1 cor, Cy5), foi testado o uso de razões onde a intensidade de expressão gênica de cada cDNA no microarranjo (Cy5) foram normalizados pelas intensidades do canal de um oligonucleotídeo referência marcado com Cy3 (*RefOligo*). O desenho experimental baseado na utilização de um oligonucleotídeo complementar a todos os pontos de cDNA no microarranjo como referência externa neste projeto, se baseou na ausência de amostras disponíveis de tecido mamário sadio para co-hibridização e normalização dos dados através da construção de razões de expressão, e pelo risco envolvido na utilização de um RNA referência que deveria estar disponível em grande quantidade e com qualidade adequada durante os 4 anos de duração do projeto. A metodologia de normalização de dados utilizando um oligonucleotídeo referência foi descrita em um estudo de expressão gênica em leveduras (Dudley et al., 2002). No início do projeto esta estratégia foi testada em nosso laboratório para a comparação da expressão gênica em tecidos humanos (Peixoto et al., 2006). Os resultados obtidos a partir da comparação de três tecidos humanos, demonstraram que a reprodutibilidade e acurácia das medidas de expressão gênica obtidas com uso de um *RefOligo* são comparáveis às obtidas utilizando um RNA referência e às medidas de expressão provenientes do canal de RNA (1 cor – *One-Color*) (Peixoto et al., 2006). A partir desses resultados, obtidos no início do projeto, o *RefOligo* foi utilizado como referência externa nos experimentos de hibridização das 78 amostras de tumores de mama, e a reprodutibilidade e acurácia desta metodologia foi comparada com as medidas de expressão gênica resultantes da utilização direta das intensidades de expressão provenientes do canal de RNA, como descrito no item 4.3.

3.8. Identificação de assinaturas de expressão gênica correlacionadas com características clínicas e anatomo-patológicas

Foram utilizadas duas abordagens distintas nas análises supervisionadas, onde as amostras foram agrupadas de acordo com características anatomo-patológicas ou clínicas, seguida da busca de genes diferencialmente expressos entre as classes. Uma delas consistiu no uso de uma ferramenta desenvolvida como um *plugin* do Microsoft Office Excel, específica para análise de dados de microarranjos de DNA, *SAM* – *Significance Analysis of Microarrays* (v. 1.0) (Tusher et al., 2001). A ferramenta *SAM* busca identificar genes diferencialmente expressos com significância estatística entre conjuntos de experimentos de microarranjo, através de permutações repetidas dos dados de expressão de cada transcrito, entre as amostras de dois ou mais grupos sendo comparados. Para cada gene, o valor real da correlação medida entre os grupos (valor observado) é comparada a um valor de correlação obtido através de permutação dos dados (valor esperado ao acaso) para checar o valor estatístico associado ao conjunto de transcritos encontrado como diferencialmente expresso entre duas condições analisadas. Se o valor de correlação observado excede o valor esperado, a diferença de expressão é considerada significativa. O programa também calcula uma taxa de falsa descoberta (*False Discovery Rate* – *FDR*), que expressa a porcentagem de falsos-positivos dentro do grupo de transcritos considerado diferencialmente expressos. A ferramenta permite ainda a seleção de conjuntos de transcritos usando limiares de *FDR* definidos pelo usuário.

A segunda forma utilizada para identificar transcritos expressos diferencialmente entre grupos de pacientes com características tumorais distintas, foi através da razão sinal-ruído (*Signal-to-Noise Ratio* – *SNR*). Este teste se baseia na comparação das médias de dois grupos distintos e verifica o quão distante elas estão entre si, levando em conta a variabilidade dos dados dentro dos grupos, utilizando a permutação dos dados

entre as classes para definir a significância estatística dos valores de *SNR* medidos (Golub et al., 1999) (Figura 3).

$\frac{\text{Signal}}{\text{Noise}} = \frac{\text{diferença entre as médias dos grupos}}{\text{variabilidade dos grupos}} = \frac{\bar{X}_a - \bar{X}_b}{\text{DP}(\bar{X}_a - \bar{X}_b)}$	$\bar{X}_a = \text{média grupo a}$ $\text{DP} = \text{desvio padrão}$
---	--

Figura 3. Fórmula para cálculo da razão de *SNR* para cada um dos transcritos analisados. Além do valor da razão de *SNR*, é determinado um valor *p* correspondente à razão calculada através da permutação dos dados.

Após a identificação de um conjunto de transcritos expressos diferencialmente, foram feitas análises de agrupamento hierárquico (*clustering*) das amostras utilizando-se o pacote *Functional genomics* do programa *SpotFire DecisionSite* (TIBCO) e como medida de similaridade foi utilizada a distância Euclidiana. Mapas de expressão (*Heat Maps*) dos dados dos transcritos expressos diferencialmente foram construídos utilizando-se o mesmo programa.

3.9. Reações de RT-PCR fita específica

O microarranjo utilizado contém sondas de cDNA dupla-fita, que não permitem determinar a orientação dos ncRNAs intrônicos identificados relativamente ao gene codificador de proteínas no mesmo *locus*. Para determinar a orientação dos transcritos de interesse a transcrição reversa foi feita utilizando *primers* complementares a uma possível orientação senso e antisenso do transcrito (Shendure e Church, 2002). Os *primers* foram desenhados utilizando-se o programa *Primer Express* (v. 2.0 - Applied Biosystems), visando um amplicon de tamanho entre 100-150pb e temperatura de *melting* igual a 60°C. O programa disponibiliza um documento contendo várias opções de pares de *primers* para cada um dos transcritos de interesse. O primeiro critério aplicado para a escolha de *primers* é a verificação de no máximo duas ou três bases fortes (C ou G) na ponta 3' de cada um deles. O segundo critério envolve a formação

indesejada de estruturas secundárias e dímeros de *primers*. Para a avaliação deste critério, foi utilizado o programa *Oligo Analyzer* (v. 1.1.2), ferramenta disponível na Internet e gratuita (<http://www.uku.fi/~kuulasma/OligoSoftware>), que determina propriedades dos pares de *primers* de interesse, como T_m, conteúdo de GC, dímeros de *primers*, formação de estruturas secundárias e compatibilidade *primer-primer*.

Outra ferramenta utilizada na análise dos *primers* e que foi desenvolvida pela Universidade da Califórnia, Santa Cruz, o PCR *in silico* (<http://genome.ucsc.edu/cgi-bin/hgPcr?org=Human&db=hg18&hgsid=84401276>), busca em seu banco de dados pela seqüência complementar ao par de *primers* fornecido pelo usuário. Dessa forma, pode-se avaliar a especificidade do produto da amplificação por PCR do seu transcrito, utilizando um par de *primers* específico. Assim, utilizando todos estes critérios, foram escolhidos os pares de *primers* mais eficientes e precisos para amplificação dos transcritos de interesse (Tabela 3).

A etapa de transcrição reversa foi feita com o *kit Superscript III First-Strand Synthesis SuperMix* (Invitrogen), partindo de 500ng de RNA total, a 55°C numa reação de volume final de 50µl, de acordo com o protocolo padrão do fabricante.

Tabela 3. Sequência dos *primers* utilizados nas reações de RT-PCR fita específica e PCR em tempo real.

Gene	Primers (senso e antisenso)	Tamanho (pb)	Tm (°C)	%GC
CAB39	TGGCCTTGCTACTAGGTTTTTGA	24	58	42
	GATGGTGAAGGCTAGAAGCCC	21	59	57
CNOT4	CCATCCTGTACTCCAAAGGTGTTAT	25	58	44
	TTGGCTTGTTTGACTTATCCTTAG	26	59	38
DNAJC3	TCACTGCCCTAGCTATTTGTGG	22	58	50
	ACTACAGAAGGCTGCCCTGG	20	58	60
GATA3	GTTAGAGCCCTGCTCGATGC	20	59	60
	CGCGTTTAGGCTTCATGATACTG	23	60	48
GFRA1	TCTGCTGGACATTTGTAAACATCC	24	59	42
	ACACATGTATTCCTCAGCCAGCT	23	59	48
GPHN	AACAGCAAGAATTTGAGATTTTTGAG	26	58	31
	TTCTCCACAGGCTTAAACAAGCC	23	58	48
JMJD2C	CCGAGCTTTCATGGCCAA	18	59	56
	CCCATGGCTCTCCACATCAC	20	60	60
MLLT3	GAGGCCACTGGCAAAACCTA	20	59	55
	GTTACTAAAGACCAGTGCTGCATTG	25	58	44
PTRH2	GCACTCACGTCCAATGGGA	19	59	58
	TAATGGGCCTGGAGTGAGGT	20	58	55
ZBTB16	CCCGCATCTTCTAAAGCATCTG	22	60	50
	GCCCCTGGGTCGTTGG	16	59	75
HPRT1	TGACACTGGCAAAACAATGCA	21	60	43
	GGTCCTTTTCACCAGCAAGCT	21	64	52

Para cada ncRNA, foram utilizados na reação *primers* específicos e complementares a fita senso ou antisenso do transcrito, em tubos separados. As reações de transcrição reversa fita-específica foram otimizadas individualmente para cada transcrito testado para garantir a ausência de transcrição a partir de *priming* formado por estruturas secundárias. Verificou-se que o *priming* inespecífico foi eliminado utilizando uma temperatura de desnaturação de 85°C por 5 minutos e em seguida transferindo as reações para 65°C por 5 minutos para evitar a re-formação de estruturas secundárias no RNA. As reações eram transferidas a 57°C por 5 minutos para adição de tampão e enzima de transcrição reversa e mantidas nesta temperatura por 50 minutos. Em paralelo, foram feitas reações sem a enzima transcriptase reversa para cada um dos pares de

primers para controle da contaminação com DNA genômico e reações sem *primers* para controle da formação de estruturas secundárias nas moléculas de RNA.

Para cada ncRNA investigado, os produtos das reações de transcrição reversa fita-específica foram utilizados como molde em reações de PCR, usando *primers* específicos. Foram realizados 40 ciclos de amplificação com uma temperatura de anelamento que variou entre 57°C e 60°C. Foi utilizada uma quantidade de cDNA inicial de aproximadamente 10ng, em uma reação de 50µl de volume final contendo 10µM de *primers*, 5U de DNA Polimerase, tampão de reação NH₄ 1X, 1,8mM de MgCl₂, 2mM de dNTPs. Em paralelo, foi feita a amplificação dos controles de transcrição sem DNA Polimerase (controle de ausência de contaminação genômica) e sem cada um dos pares de *primers* (controle de ausência de *priming* inespecífico na RT). Este último controle é essencial para garantir a confiabilidade da informação de orientação. A presença de produto de amplificação no controle sem *primer* na RT impede uma conclusão confiável sobre a orientação do transcrito analisado (Shendure e Church, 2002).

A análise dos produtos obtidos incluiu a separação em gel de 2% agarose corado com brometo de etídio e a observação de bandas nas reações realizadas sem a enzima DNA Polimerase (caso positivo, as amostras de RNA eram submetidas a novos tratamentos com DNase e novas reações de transcrição reversa) e nas reações realizadas na ausência de *primers*. Caso este último controle apresentasse resultado positivo, novas reações de PCR eram realizadas em condições mais estridentes, assim como no caso de bandas não específicas na reação teste. Em último caso, modificações nas condições das reações de transcrição reversa eram feitas para obtenção de produtos mais específicos.

3.10. PCR quantitativo em tempo real

A etapa de transcrição reversa foi feita com o *kit Superscript III First-Strand Synthesis SuperMix* (Invitrogen), partindo de 500ng de RNA total, a 55°C numa reação de volume final de 50µl, de acordo com o protocolo padrão do fabricante. Foi utilizado oligo dT como *primer*.

Os *primers* utilizados para as reações de PCR quantitativo em tempo real foram os mesmos utilizados nas reações de RT-PCR fita-específica (Tabela 3). Estes *primers* foram escolhidos de forma a apresentar um mínimo de formação de estruturas secundárias e dímeros de *primers*, conforme descrito no item 3.9 de Materiais e Métodos. A padronização da concentração saturante de *primers* a ser utilizada nas reações de PCR em tempo real foi feita para garantir a saturação do cDNA presente nas amostras ao mesmo tempo que a ausência de formação de dímeros dos *primers* utilizados. Assim, foram testadas concentrações crescentes dos *primers* para uma quantidade fixa de cDNA até que o ciclo de referência (Ct – número do ciclo em que a quantidade de alvo amplificada alcança um limiar fixo na faixa exponencial de amplificação) da reação se estabilizasse em um valor mínimo. Para cada concentração de *primer* testada, foi avaliada a ausência de formação de dímeros do *primer* através da análise de curvas de dissociação de cada reação. Após estes testes foi definida a concentração de 0,8µM para os transcritos avaliados.

Após determinada a concentração ótima de *primers*, foi feita a determinação da eficiência da amplificação de cada transcrito. Para o cálculo da eficiência da reação de PCR, foram feitas diluições de cDNA obtido da transcrição reversa de um pool de RNAs de diferentes tumores de mama utilizando-se 1, 5, 10, 15 e 30 ng de cDNA e *primers* na concentração de 0,8µM. A partir destes resultados, foi construída uma curva do log da quantidade de cDNA *versus* o Ct para o cálculo da eficiência da amplificação que é

definido a partir da inclinação da reta, dado pela equação $y = Ax + B$. A eficiência da amplificação de um dado par de *primers* (E) pode ser definida através da expressão matemática $E = 10^{[-1/A]}$, ou expressa em porcentagem, sendo $E(\%) = (E - 1) \times 100$ (Pfaffl, 2001). A partir deste dado de eficiência de amplificação de cada par de *primers*, é possível fazer a inferência da razão de expressão do transcrito de interesse em relação ao gene normalizador endógeno através da fórmula:

$$\text{Razão de Expressão} = \frac{(E_{\text{alvo}})^{\Delta C_{\text{t alvo}}}}{(E_{\text{ref}})^{\Delta C_{\text{t ref}}}}$$

onde o ΔC_{t} é o C_{t} da condição testada subtraído o C_{t} da condição controle para o transcrito alvo e o gene normalizador de referência (ex. tecido tratado e não tratado).

Para os amplicons em que a eficiência da PCR do gene alvo e do gene normalizador é próximo a 1, a fórmula então pode ser resumida a $2^{-\Delta C_{\text{t}}}$, que expressa a quantidade relativa de um gene alvo em relação ao gene referência, para cada amostra testada.

Os pares de *primers* do gene normalizador endógeno utilizados, hipoxantina fosforibosiltransferase 1 (*HPRT1*), foram sintetizados usando as mesmas seqüências testadas no trabalho de Vandesompele e colaboradores e descritas na tabela 3 (Vandesompele et al., 2002).

Todas as reações de amplificação em tempo real foram realizadas em 40 ciclos a 60°C, utilizando-se o reagente *SYBR Green PCR Master Mix*, o equipamento *7300 Real Time PCR System* (Applied Biosystems) e o gene endógeno normalizador *HPRT1*. Foram utilizados aproximadamente 5ng de cDNA para o início da reação e concentração final dos *primers* de 0,8µM. As reações de amplificação em tempo real foram ainda feitas em triplicata, e paralelamente, foram feitas reações sem cDNA molde para controlar contaminação com DNA amplificado ou genômico.

A significância estatística da diferença de expressão dos transcritos testados por qPCR entre grupos distintos de pacientes com câncer de mama foi medida utilizando o teste *t* de *Student*. Neste cálculo, foram considerados os valores médios de $2^{-\Delta Ct}$ dentro de cada grupo de pacientes (ex. pacientes com ou sem metástase) e o teste *t* (não-pareado, uni-caudal, homoscedástico) foi calculado utilizando-se estes grupo de valores para comparação.

3.11. Anotação funcional dos transcritos identificados

Os transcritos apontados como diferencialmente expressos nas análises supervisionadas foram anotados através de alinhamento na sequência genômica disponível no site Genome Browser (atualizado em março de 2006) da Universidade da Califórnia, Santa Cruz utilizando a ferramenta BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>). Os transcritos intrônicos foram analisados quanto à sua capacidade codificadora através da utilização de ferramentas disponíveis na Internet (*ESTScan2*, <http://www.ch.embnet.org/software/ESTScan2.html> e *ORF Finder*, <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>).

O potencial de produzir miRNAs e snoRNAs foi avaliado através do alinhamento das seqüências dos transcritos intrônicos com o banco de dados disponível na Internet (<http://microrna.sanger.ac.uk/sequences/> e <http://www-snoRNA.biotoul.fr/>).

Análises de enriquecimento da ontologia gênica dos grupos de transcritos apontados como expressos diferencialmente nas análises foram realizadas com as ferramentas disponíveis na Internet, eGOn e BinGO, ambas disponíveis no site: <http://www.geneontology.org/GO.tools.microarray.shtml>.

Outras 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

mama foram feitas utilizando-se a ferramenta *Ingenuity Pathways Analysis* (Ingenuity Systems) (Calvano et al., 2005) e *Oncomine Research Platform* (Compendia Bioscience) (Rhodes et al., 2007).

4. RESULTADOS

4.1. Montagem de banco de amostras de tecido tumoral mamário e obtenção de RNA para experimentos de hibridização

O objeto de pesquisa desta tese corresponde ao conjunto de amostras de tecido tumoral mamário de pacientes provenientes do HDES, Açores, Portugal. A previsão inicial era a montagem de um banco tumoral com cerca de 50 amostras. No entanto, já no segundo ano de recolhimento das amostras o banco tumoral contava com um número de 62 amostras clínicas de tecido tumoral, atingindo um número de 74 amostras em julho de 2007. Todas as amostras coletadas possuem uma descrição detalhada de suas características anatomo-patológicas além dos dados clínicos das pacientes (Tabela 4).

O tempo de seguimento médio das amostras é de 3,5 anos, sendo que todas as amostras provêm de pacientes com acompanhamento por um período mínimo de 1,5 anos (Tabela 4). A idade média apresentada pelo grupo de pacientes no momento do diagnóstico é de cerca de 61 anos, a mesma apresentada pelas pacientes com câncer de mama nos EUA durante o período de 2000-2004 (ACS, 2008). O tamanho médio do tumor, no entanto, é maior do que o da população de pacientes diagnosticadas com câncer de mama nos EUA no mesmo período: 2,7cm *versus* ≤ 2 cm (ACS, 2008). Neste conjunto de pacientes estudado, todos os tumores de mama são do tipo invasivo. A incidência de câncer de mama invasivo na população feminina dos EUA também é muito maior que o carcinoma *in situ*, cerca de três vezes mais freqüente. Estes últimos dois aspectos podem indicar uma menor prevenção da doença de câncer de mama em Portugal, quando comparado aos EUA já que os tumores são detectados com um tamanho maior e com um caráter invasivo mais freqüente. No entanto, a taxa de mortalidade das pacientes com câncer de mama incluídas neste projeto é de 25,6% a

mesma apresentada pela população de mulheres com câncer de mama nos EUA durante o período de 2000-2004 (ACS, 2007).

De acordo com Mallon e colaboradores, as taxas de incidência de tumores de mama invasivo do tipo ductal é de 65-85%; lobular, de 10-15%; medular, menos de 1%; mucinoso, 1% e papilar de 1-2% (Mallon et al., 2000). Em nosso conjunto de amostras de tecido tumoral mamário, encontramos taxas de incidências concordantes na maior parte dos tipos tumorais invasivos com ductal, 78%; lobular, 7%; medular, 1%; mucinoso, 4% e papilar, 3%.

O câncer de mama tem um pico de recorrência do tumor em dois anos após o começo do tratamento da doença e diminui com o passar dos anos, mas no entanto, as chances de recorrência nunca chegam a zero (Rugo, 2008). As recorrências podem ser de três tipos: local, quando restam células tumorais remanescentes da cirurgia de mastectomia do câncer primário e um novo câncer de mama se desenvolve, no mesmo exato local; regional, quando as células tumorais migram para os nódulos linfáticos axiais; e distante, quando as células tumorais migram pela corrente sanguínea ou vasos linfáticos, se instalando em outros tecidos do corpo, espalhando a doença (fonte: Imaginis). As recorrências mais comuns em câncer de mama são as distantes (metástases), independente do tempo decorrido do diagnóstico (Rugo, 2008). Esta informação vai de acordo com a obtida em nosso conjunto de 74 pacientes já que as metástases possuem incidência cerca de duas vezes maior que a recorrência local (16% *versus* 9%). No entanto, o número de metástases é baixo já que é reportado na literatura que até pacientes com baixo risco de desenvolver metástases, isto é, que possuam tumores de tamanhos pequenos e nódulos sem infiltração de células tumorais, apresentam um risco substancial de desenvolver metástases em 10-25% dos casos (Rugo, 2008).

Tabela 4. Características anatomo-patológicas do conjunto de amostras teciduais reunidos no banco de tecido tumoral mamário.

Idade		
30 a 50 anos	20	
51 a 60 anos	15	
61 a 70 anos	19	
71 a 87 anos	20	
Tamanho do tumor (mm)		
até 20	23	
de 21 a 30	18	
from 31 a 70	18	
n.d.	15	
Período de acompanhamento		
até 2 anos	18	
2.5 a 3 anos	7	
3.5 a 4 anos	24	
4.5 a 5 anos	25	
Tipo Histológico		
dutal invasivo	58	
dutal e lobular invasivo	2	
lobular invasivo	5	
medular	1	
mucinoso	3	
papilar invasivo	2	
mal diferenciado	3	
Infiltração dos nódulos linfáticos		
0	40	
1 a 5	22	
6 a 10	4	
11 a 15	8	
Grau Histológico		
I	7	
II	28	
III	22	
n.d.	17	
Estágio Tumoral		
I	21	
IIA - IIB	35	
IIIA - IIIB	9	
IIIC - IV	9	
Status do Paciente		
Metástase	doença em progressão	3
	morte	9
Reincidência	doença em progressão	1
	morte	1
Livre de metástase e/ou reincidência	livre da doença	25
	em tratamento	22
	morte	5
	n.d.	8
Status de ER		
0		22
+		2
++		2
+++		48
Status de PR		
0		30
+		7
++		2
+++		35

Todas as amostras coletadas foram enviadas ao nosso laboratório por via aérea, congeladas em gelo seco e acondicionadas em *RNALater*. Com este reagente as amostras de tecido tumoral são preservadas da degradação por até um mês a temperatura de 4°C. As amostras foram então processadas para extração de seu RNA total utilizando-se o método tiocianato de guanidina/fenol/clorofórmio (ver seção Material e Métodos).

Após este procedimento seguiu-se avaliação da quantidade e qualidade do RNA total obtido através de espectrofotometria (NanoDrop) e eletroforese capilar (Bioanalyzer 2100), respectivamente. A Figura 4 apresenta o resultado típico da análise por eletroforese capilar de RNA total isolado de amostras de tecido tumoral mamário processadas em nosso laboratório. A medida das razões de intensidade das bandas de RNA ribossomal 28S e 18S mede o nível de degradação na preparação de RNA. Como a banda de 28S apresenta o dobro do tamanho da banda de 18S, o valor ideal para a razão 28S/18S é 2, pois indica a ausência de degradação. No entanto, temos observado que tecidos provenientes de remoção cirúrgica e acondicionados de maneiras diversas nos hospitais (congelamento, blocos de parafina, etc.) raramente fornecem amostras de RNA total com razão 2. Cerca de 17% das amostras de RNA total obtidas do tecido tumoral mamário apresentaram razão mínima de 0,5, enquanto a maior parte das amostras (55%) apresentaram razão de 0,8 a 1,3. Apesar de 0,5 ser considerada uma razão baixa quando comparada ao valor de referência 2, nossa experiência tem nos mostrado que amostras de RNA total com esta razão mínima ainda permitem o prosseguimento dos experimentos de microarranjo, produzindo dados com qualidade comparável a de RNAs com razões mais próximas de 2, medido pela correlação entre réplicas e pelo número de transcritos detectados nas hibridizações e confirmado *a posteriori* (ver item 4.2). Desse modo, optamos por utilizar todos os RNAs com razão 28S/18S acima de 0,5.

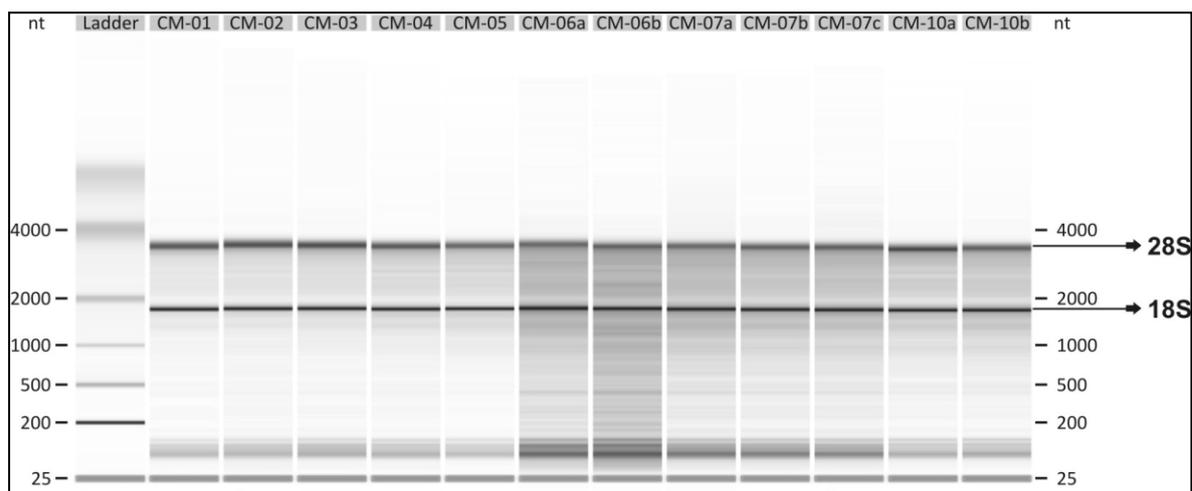


Figura 4. Eletroforese capilar de amostras de RNA total extraídas de amostras de tecido tumoral mamário. À direita estão indicados os picos das frações do RNA 28S e 18S (setas). Observar que as amostras CM-06a e CM-06b possuem um grau de degradação maior em relação às outras amostras. No entanto, a qualidade destas amostras de RNA ainda é satisfatória (razão 28S/18S > 1,0) para os experimentos de microarranjo.

A avaliação do grau de contaminação das frações de RNA com proteínas foi avaliada pela medida de absorção a 260nm (pico de absorbância do RNA) e 280nm (pico de absorbância de proteínas). A maior parte dos RNAs obtidos (84%) apresentou uma razão D.O. 260/280 igual ou maior que 1,8, o que indica a ausência de contaminação protéica significativa que poderia inibir reações enzimáticas e prejudicar a utilização das amostras nas etapas seguintes (Tabela 5).

A quantidade de RNA total obtida se mostrou bem variável: cerca de 30% das amostras apresentaram um rendimento menor que 10µg e a média e a mediana das quantidades de RNA total obtido das 74 amostras foram, respectivamente, 56 e 28µg (Tabela 5).

Tabela 5. Quadro de informações de rendimento total, qualidade e rendimento obtido da amplificação de amostras de RNA total extraídas de 74 amostras de tecido tumoral mamário.

Rendimento total (ug)	
até 5	11
de 6 a 10	10
de 11 a 20	10
de 21 a 40	12
de 41 a 80	15
> 80	16
média	55.94
mediana	28.35
Razão 260/280	
1.5 a 1.8	12
1.8 a 2.0	60
> 2.0	2
Razão RNAr 28S/18S	
0.5 a 0.7	25
0.8 a 1.0	20
1.1 a 1.3	20
1.4 a 2.2	9
Rendimento amplificação (ug)	
até 10	6
de 11 a 30	16
de 31 a 60	29
de 61 a 90	18
> 90	5
média	47.6
mediana	50.5

A restrição à quantidade de RNA nos estudos de expressão gênica de tecidos humanos, especialmente, é um problema enfrentado por vários laboratórios no mundo (Gomes et al., 2003; Stoyanova et al., 2004). Como alternativa, têm sido utilizado metodologias para amplificação linear do RNA mensageiro, de forma a obter material em quantidade suficiente para a realização de experimentos de hibridização de microarranjos (Van Gelder et al., 1990; Wang et al., 2000). Já foi demonstrado que a amplificação linear do RNA mensageiro em protocolos baseados em transcrição in vitro mediado pela T7 RNA polimerase mantém uma representação fiel do transcrito, demonstrado pela existência de alta correlação entre os resultados de microarranjo

obtidos a partir de RNA amplificado e não amplificado (Wang et al., 2000; Kaposi-Novak et al., 2004).

A quantidade mínima para a realização das reações de marcação de RNA total para produção de alvos fluorescentes para os experimentos de microarranjo é de 5µg; adicionando-se uma réplica técnica, 10µg. Portanto, 30% das amostras de tecido tumoral mamário não poderiam ser incluídas neste estudo. Para superar a limitação da pouca quantidade de RNA disponível decidimos realizar a amplificação de todas as amostras de RNA total disponíveis e com qualidade para as hibridizações. Dessa forma, fomos capazes de gerar quantidades suficientes de RNA total para todos os experimentos de hibridização subsequentes já que o protocolo utilizado (modificado da literatura) (Van Gelder et al., 1990) necessita de uma quantidade mínima inicial de RNA total igual a 1µg, quantidade disponível para todas as amostras. Assim, o protocolo de amplificação de RNA total gerou quantidades de aRNA maiores que 10µg para cerca de 93% das amostras, com uma média de 47,6µg e mediana 50,5µg para todas as 74 amostras (Tabela 5).

4.2. Hibridização de amostras de tumor de mama com o microarranjo de ~4.000 elementos

Cerca de 60% (~2.200) das seqüências depositadas no microarranjo de cDNA corresponde a regiões exônicas de genes já conhecidos e 250 destas têm papel já comprovado na progressão e desenvolvimento de tumores de próstata. Um número de 250 fragmentos de cDNA correspondem ainda a regiões intergênicas de genes conhecidos e, 20% das seqüências contidas no microarranjo (~800) correspondem a regiões intrônicas de genes conhecidos (Reis et al., 2004). Neste último aspecto reside a característica inovadora deste projeto, uma vez que grande parte dos estudos de

expressão gênica no câncer de mama são feitos com microarranjos comerciais que contém sondas para regiões codificadoras de proteínas. Também já foi demonstrado através de trabalhos de nosso grupo que o estudo dos padrões de expressão de ncRNAs intrônicos podem ter relevância para a identificação de marcadores moleculares no câncer de próstata e rim (Reis et al., 2004; Brito et al., 2008).

Após a amplificação das amostras de RNA total de tecido tumoral mamário foram feitos todos os experimentos de hibridização das 74 amostras de aRNA obtidas, cada uma com uma réplica técnica, totalizando 148 hibridizações. Todas as hibridizações foram feitas utilizando-se o aRNA do tecido tumoral marcado com o fluoróforo Cy5 e o oligonucleotídeo *RefOligo* marcado com Cy3 na mesma hibridização. O resultado típico de uma hibridização está apresentado na Figura 5.

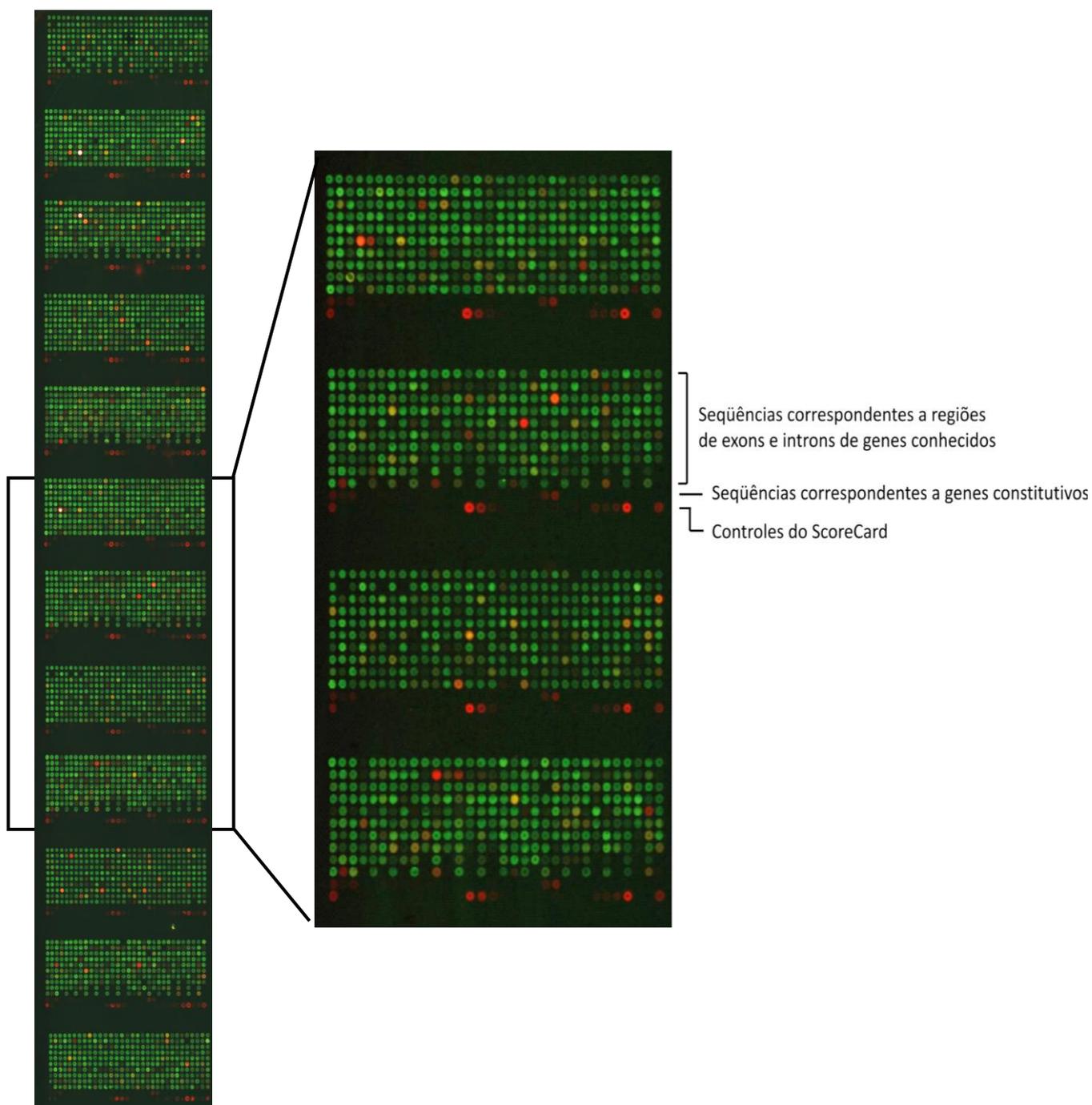


Figura 5. Imagem de um microarranjo de cDNA hibridizado com aRNA tumoral de mama (vermelho) e *RefOligo* (verde). Apenas um dos lados (réplica interna) do microarranjo está mostrado. No inserto à direita estão destacados quatro *subarrays*. Como esquematizado na figura, os transcritos correspondentes às seqüências de regiões exônicas e intrônicas de genes conhecidos estão distribuídos nas primeiras dez linhas de cada *subarray* e as duas últimas linhas correspondem a controles positivos e negativos do experimento. Observar a hibridização específica do oligonucleotídeo fluorescentemente marcado (*RefOligo*) em todos os pontos de cDNA das dez primeiras linhas de cada um dos *subarrays* e a ausência de sinal nos pontos de cDNA das duas últimas linhas, que são compostos de controles negativos e positivos que não possuem seqüência complementar ao *RefOligo*.

As 148 lâminas foram analisadas para avaliação da reprodutibilidade das medidas de expressão gênica obtidas, medida pela correlação média de todos os cDNAs do microarranjo, calculada entre réplicas dentro da mesma lâmina (intra-lâmina) ou entre lâminas distintas hibridizadas com a mesma amostra (inter-lâmina). Foi definido como critério mínimo de qualidade uma correlação intra-lâminas (entre os dois lados de uma mesma lâmina) de 0,90 e uma correlação inter-lâminas (entre duas réplicas técnicas) de 0,80. A correlação média e mediana inter-lâminas observadas nos experimentos foram de 0,83 e 0,84, respectivamente. Somente 4 microarranjos não atingiram os valores de correlação mínimos. Estas lâminas foram re-hibridizadas, de modo que todas as amostras possuíssem quatro réplicas, atendendo os critérios de qualidade definidos para o projeto.

Também foi avaliado se a qualidade do RNA total utilizado, medida pela razão 28S/18S, afetava a qualidade das medidas de expressão gênica obtidas. As correlações médias entre réplicas nas lâminas hibridizadas com RNAs com razão entre 0,5–0,7, 0,8–1,0, 1,1–1,3 e 1,4–2,2 foram respectivamente 0,97, 0,98, 0,97 e 0,97. Nos mesmos conjuntos de lâminas, o número médio de elementos detectados nas hibridizações após a filtragem e normalização foi de 3.340, 3.325, 3.164 e 3.185, respectivamente. Coletivamente estes resultados confirmam a ausência de diferenças significativas na qualidade dos experimentos de hibridização obtidos a partir de RNAs com razões de 28S/18S variando entre 0,5 e 2,2.

4.3. Avaliação de uma metodologia para obtenção de medidas de expressão gênica utilizando um oligonucleotídeo referência (*RefOligo*)

Microarranjos de cDNA são mais freqüentemente utilizados em experimentos de 2 cores, onde duas populações de alvos fluorescentes marcados com fluoróforos

distintos são hibridizadas competitivamente (Quackenbush, 2002; Yang e Speed, 2002). Após a extração dos dados obtém-se uma razão de expressão que reflete a abundância relativa de uma amostra teste em relação a uma amostra controle. Embora existam vários desenhos experimentais possíveis (Yang e Speed, 2002), em trabalhos de expressão gênica envolvendo um grande número de amostras experimentais, tipicamente utiliza-se como amostra controle uma mesma população de RNAs (RNA referência), contra a qual todas as amostras experimentais são comparadas (Kerr e Churchill, 2001; Yang e Speed, 2002). O RNA referência pode ser um *pool* de RNAs representando as amostras utilizadas no estudo (van 't Veer et al., 2002), ou provenientes de linhagens celulares (Alizadeh et al., 2000). Uma limitação do uso de um *pool* de RNA referência em estudos prospectivos como o deste trabalho, onde as amostras são coletadas ao longo do tempo, é a dificuldade de se obter um *pool* de RNA em quantidade e que mantenha a qualidade ao longo do projeto. A montagem de um *pool* de RNA tumoral para uso como referência externa aos experimentos de hibridização tem como desvantagem adicional, no caso deste projeto, a limitada quantidade das amostras de RNA envolvidas no estudo e o fato das amostras não estarem todas disponíveis no início do projeto quando o *pool* é montado. A utilização de RNA isolado de linhagens tumorais, embora permita a obtenção de quantidades maiores de RNA, tem a desvantagem de não se ter a garantia que a qualidade do RNA se manterá idêntica ao longo do projeto.

No caso de tumores de tecido mamário, o isolamento de tecido mamário não-tumoral adjacente é extremamente difícil, pois o tumor se desenvolve do tecido epitelial de glândulas mamárias, que apresenta volume muito pequeno no tecido mamário sadio. Esta limitação impede um desenho experimental onde cada amostra tumoral seja comparada contra tecido não-tumoral adjacente da mesma paciente.

Como alternativa, no início do projeto foi testada uma metodologia onde foi utilizada como referência um oligonucleotídeo sintético marcado fluorescentemente, complementar a todas as sondas presentes nos microarranjos (Figura 2) (Dudley et al., 2002). Foram conduzidos experimentos visando comparar a precisão e acurácia de medidas de expressão gênica sob a forma de intensidades (*One-Color*), ou com razões de expressão onde as intensidades fossem normalizadas utilizando uma referência externa (*RefOligo* ou *RefPool*)

Dessa forma, fizemos experimentos utilizando três topologias distintas de tecido tumoral (tumores de mama, próstata e rim), visando aumentar a heterogeneidade dos experimentos. Foram feitos quatro conjuntos de hibridizações com esse material (Figura 6): (i) no primeiro deles, foram hibridizados na mesma lâmina amostras de RNA provenientes de tecidos tumorais distintos, no que chamamos de experimentos diretos (Mama vs. Próstata, Mama vs. Rim, Próstata vs. Rim); (ii) a partir de quantidades equimolares de RNA destes três conjuntos de tumores foi construído um *pool* de RNA (*RefPool*) que foi utilizado na hibridização de um conjunto de lâminas, juntamente com cada uma das amostras de RNA provenientes dos três tipos tumorais (*RefPool* vs. Mama; Próstata; Rim); (iii) cada uma das amostras de RNA provenientes dos diferentes tecidos foi hibridizado na mesma lâmina com um oligonucleotídeo fluorescentemente marcado com Cy3 (*RefOligo*) e que é complementar a uma seqüência presente em cada um dos cDNAs depositados nas lâminas (*RefOligo* vs. Mama; Próstata; Rim); (vi) hibridizações *self-self*, onde cada um dos três tipos de RNA tumoral foi hibridizado numa mesma lâmina com ele mesmo, marcado com outro fluoróforo (*dye-swap*). Esse último conjunto de experimentos de hibridização foi feito para permitir a construção de gráficos de intensidade de sinal que permitissem delinear intervalos de confiança para definir conjuntos de transcritos expressos diferencialmente entre cada um dos tecidos estudados

com critério estatístico. Com os resultados em mãos, foram construídas razões de expressão para cada um dos tecidos a partir dos dados provenientes dos diferentes tipos de experimentos de hibridização, utilizando metodologias distintas. Dessa forma, obtivemos dois tipos de razão de expressão: diretas, provenientes das hibridizações diretas, e indiretas, provenientes das hibridizações indiretas, em que uma das referências é o denominador comum da equação. Além dos três tipos diferentes de razões geradas acima (razões diretas, indiretas usando *RefOligo* e indiretas usando *RefPool*), resolvemos também analisar a interferência da utilização de uma referência externa, usando dados de intensidade de sinal do canal do RNA de cada um dos três tecidos para a construção de razões (*One-Color*). Para tanto, utilizamos as medidas de intensidade do canal de RNA provenientes das hibridizações *RefOligo* vs. RNA (de cada um dos três tipos de tecidos tumorais, mama, próstata e rim). Esta escolha foi feita para assegurar que as medidas de intensidade do canal Cy3 do RNA não sofressem interferência de uma possível competição entre estes alvos acoplados a Cy3 com os alvos provenientes da referência externa *RefPool* acoplado a Cy5, apesar de já ter sido proposto que medidas de intensidade de *arrays* de oligonucleotídeos não são afetadas pela co-hibridização de alvos de RNA (t Hoen et al., 2004).

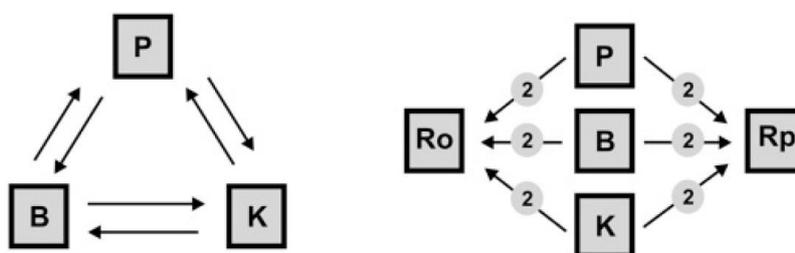


Figura 6. Representação esquemática do desenho experimental das hibridizações dos microarranjos, incluindo as hibridizações diretas e indiretas. Os quadrados representam alvos fluorescentes derivados de amostras de RNA (B, Mama; K, Rim; P, Próstata) ou de um dos dois tipos de referências externas, *RefPool* (Rp) ou *RefOligo* (Ro). Setas indicam alvos de RNA co-hibridizados com Cy3 e Cy5. O número de réplicas técnicas de cada experimento está indicado em pequenos círculos.

As razões de intensidade para uma das comparações Mama vs. Rim obtidas utilizando as quatro abordagens está mostrada na Figura 7 na forma de *plots* MxA. A partir desta figura pode-se perceber que as três razões reconstruídas (*RefOligo*, *RefPool* e *One-Color*) apresentam dados mais dispersos quando comparados às razões diretas. Resultados similares foram obtidos nas outras comparações entre tecidos. Uma hipótese para explicar esta diferença é o fato de que diferenças na incorporação do fluoróforo Cy3 e Cy5, ou até mesmo diferenças na intensidade de captação de sinal quando esses fluoróforos são excitados com luz UV no momento do *scanning*, são corrigidas pelo método do *dye-swap* no caso das hibridizações diretas. Nas hibridizações em que um dos canais é a referência externa, este método não pode ser aplicado e isto pode explicar a maior dispersão dos dados na figura 7.

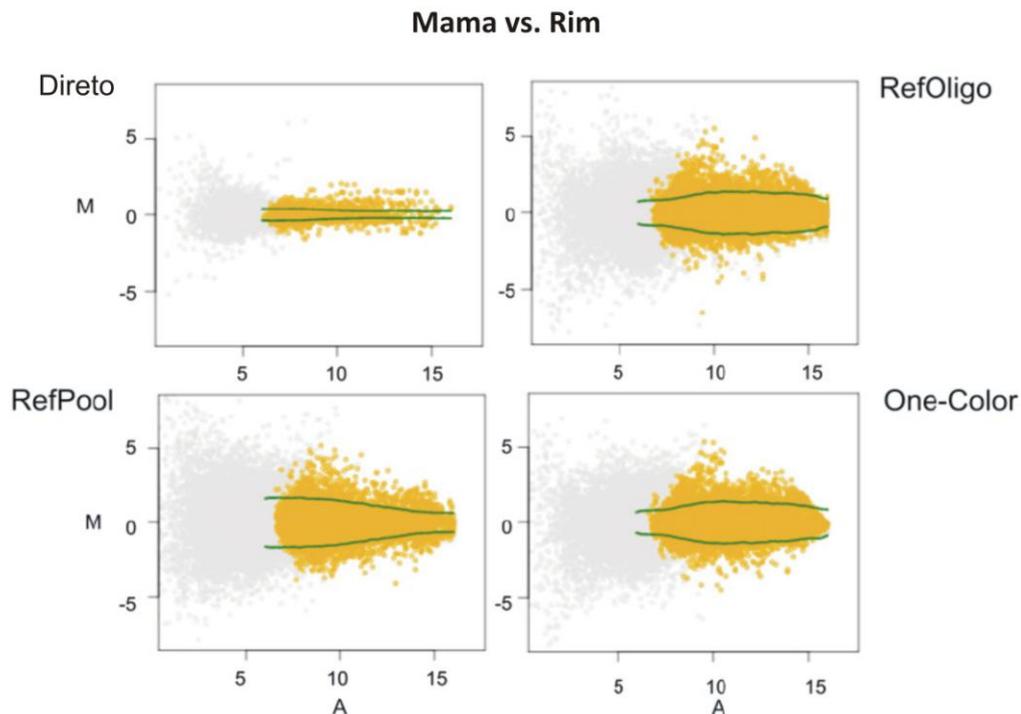


Figura 7. Gráficos de dispersão M vs. A para razões diretas e reconstruídas. Os pontos amarelos correspondem ao \log_2 da razão das intensidades (M) como uma função do \log_2 do produto das intensidades (A) para cada cDNA considerado válido do microarranjo (intensidade maior que 2 desvios-padrão em relação ao valor do ruído local). Os pontos cinzas correspondem aos cDNAs excluídos da análise. As linhas verdes apontam os limites do intervalo de confiança usado para definir o conjunto de transcritos considerados diferencialmente expressos.

As quatro metodologias distintas foram analisadas quanto a sua precisão através da comparação da variação das razões entre si, entre experimentos intra-slide (réplica interna de todos os cDNAs presentes em nosso microarranjo) e inter-slide. As razões diretas se mostraram mais reprodutíveis já que apresentaram coeficiente de variação quase duas vezes menor que os apresentados pelas razões indiretas (razões diretas, 23,6; *RefOligo*, 43,1; *RefPool*, 44,2; *One-Color*, 43,2 – Figura 8), dado já demonstrado anteriormente por Dudley *et al*, 2002. Por outro lado, o coeficiente de variação das razões reconstruídas através das hibridizações indiretas é comparável entre os três métodos testados (Figura 8).

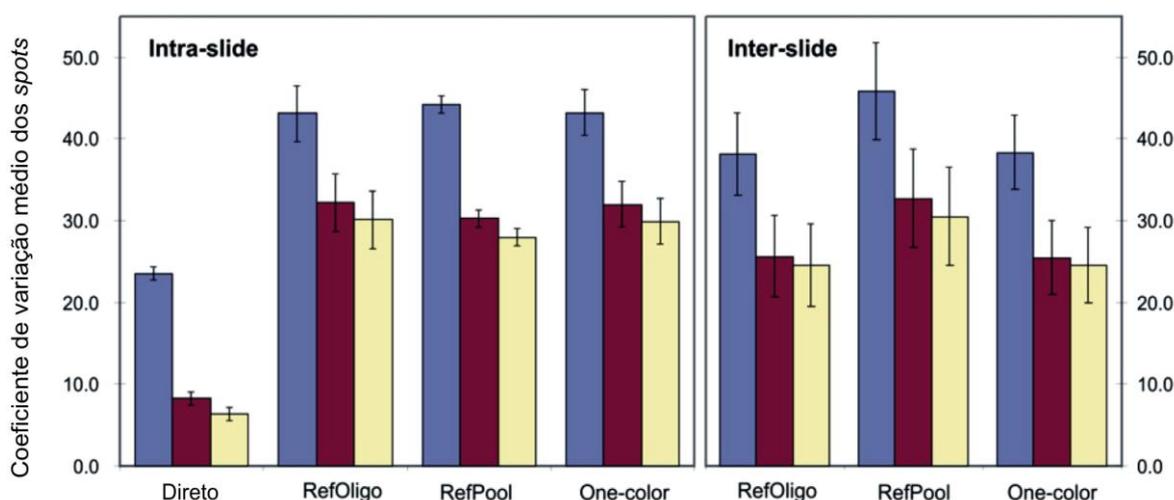


Figura 8. Comparação das médias dos coeficientes de variação das réplicas intra- e inter-lides entre três metodologias, *RefOligo*, *RefPool* e *One-Color* em relação à hibridização direta. Os coeficientes de variação foram calculados utilizando-se todas as razões (barras azuis) e razões calculadas a partir de *spots* com um ou dois desvios-padrão acima do valor de ruído local (respectivamente, barras vermelhas e amarelas). Barras de erros das comparações pareadas dos três tecidos (Próstata vs. Mama, Próstata vs. Rim ou Mama vs. Rim) estão representadas.

A precisão dos quatro métodos utilizados foi também avaliada através do cálculo da variância de todas as razões de expressão válidas medidas em faixas de intensidade de expressão (Figura 9). Da mesma forma, as medidas das razões reconstruídas são mais variáveis que as medidas das razões diretas, especialmente em intensidades de expressão

menores (Figura 9). No entanto, a variância dos dados obtidos por *RefOligo* e *RefPool* são muito similares a dos dados obtidos por *One-Color*. Estes resultados demonstram que as medidas feitas a partir do método *One-Color* são suficientes para reconstruir razões tão precisas quanto àquelas obtidas através de referência externas.

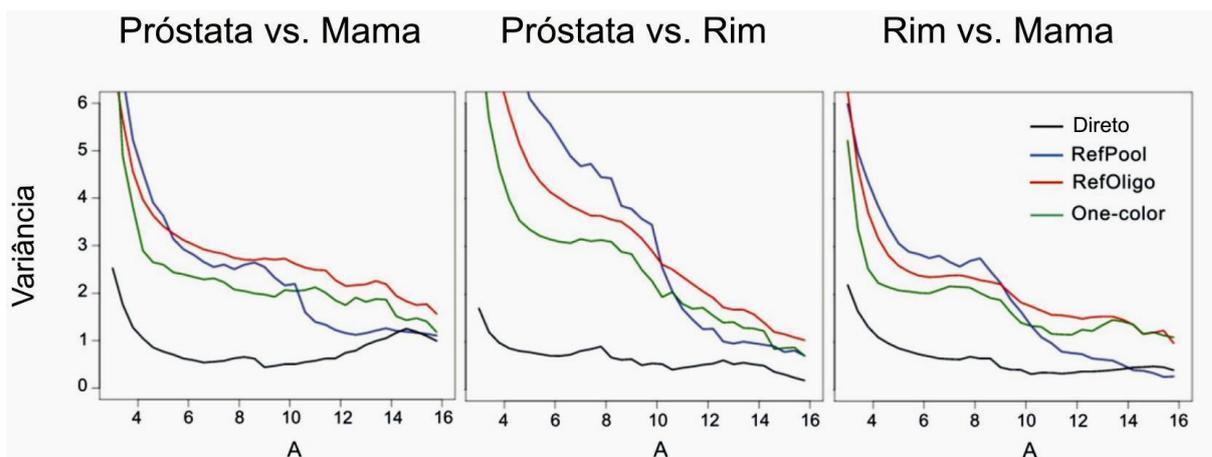


Figura 9. Variância das razões de expressão das quatro metodologias analisadas em faixas de intensidade de expressão. No eixo Y estão apresentados os valores de variância do \log_2 das razões da média de todas as réplicas técnicas, calculados usando-se janelas de intensidade com um espectro de 1,5 que se deslocam 0,4 pontos em direção ao \log_2 do produto da intensidade no eixo do X (A).

Outra forma de avaliar a reprodutibilidade das quatro metodologias analisadas foi através da comparação dos coeficientes de correlação calculados entre as razões diretas e cada uma das razões indiretas (Figura 10). Novamente, os resultados desta comparação não mostrou distinção entre nenhum dos três métodos (*RefOligo*, *RefPool* e *One-Color*) utilizados para o cálculo das razões indiretas quando comparado às razões diretas. Tanto as comparações entre os três métodos de referência externa entre si quanto cada um deles com as razões diretas apresentaram os mesmos resultados (Figura 10).

Em estudos que se utilizam a plataforma de microarranjos para se entender melhor a gênese e a progressão de uma doença como o câncer, é de extremo interesse delinear conjuntos de transcritos que apresentem uma diferença de expressão particular a

uma condição definida. Portanto, cada uma das metodologias analisadas foram testadas quanto à acurácia em determinar conjuntos de transcritos expressos diferencialmente em cada comparação tecidual.

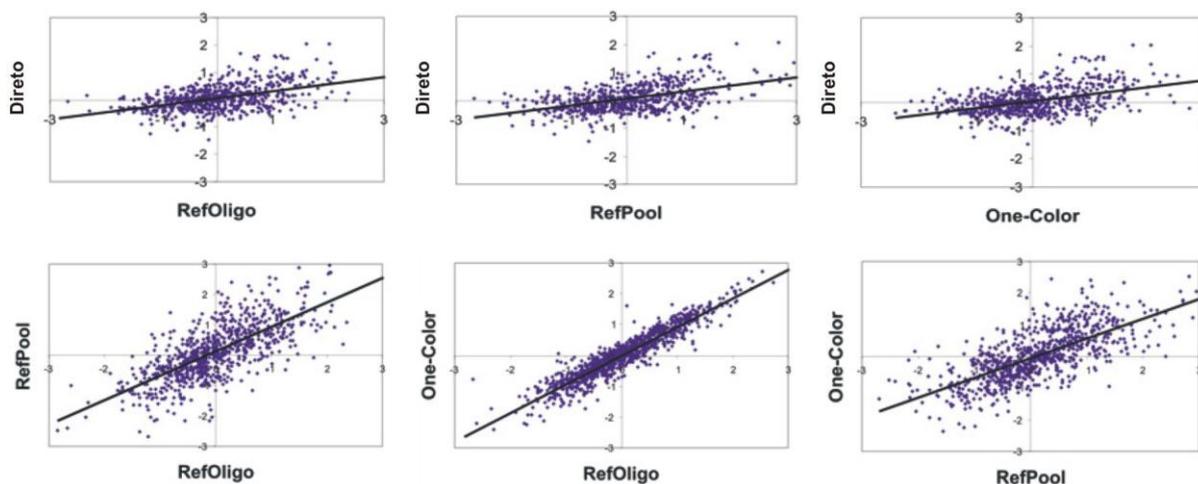


Figura 10. Correlação entre razões diretas e reconstruídas da comparação dos tecidos tumorais rim vs. mama. Somente razões de spots com valores de intensidade maiores que dois desvios-padrão acima do ruído médio da lâmina foram utilizados.

Transcritos expressos diferencialmente entre pares de tecidos das hibridizações diretas foram selecionados utilizando-se intervalos de confiança de 95% baseados nos resultados das hibridizações *self-self* (Vencio e Koide, 2005). Este conjunto de transcritos expressos diferencialmente foi considerado o conjunto padrão e serviu como base para comparação dos resultados obtidos das razões reconstruídas (Figura 11).

A acurácia em apontar transcritos expressos diferencialmente de cada uma das três metodologias indiretas foi muito similar visto que o número de transcritos concordantes de cada metodologia indireta com a direta foi em média de 58%, *RefOligo*; 63% *RefPool*; 60% *One-Color*. Da mesma forma, a fração de transcritos falso-positivos (transcritos presentes nas metodologias indiretas mas não no conjunto padrão da metodologia direta) apontados por cada uma das três metodologias indiretas é comparável: 42% em média, *RefOligo*; 37% em média, *RefPool*; 40% em média, *One-Color*. Dessa forma, o uso do *RefOligo* como uma referência externa em experimentos

de hibridização produz resultados tão precisos e acurados quanto o uso de uma referência comum como o *RefPool*. Além disso, razões de expressão gênica reconstruídas a partir de medidas de *One-Color* de experimentos de microarranjos diferentes são tão exatas e reprodutíveis quanto às obtidas através do uso de uma referência comum (*RefPool* ou *RefOligo*).

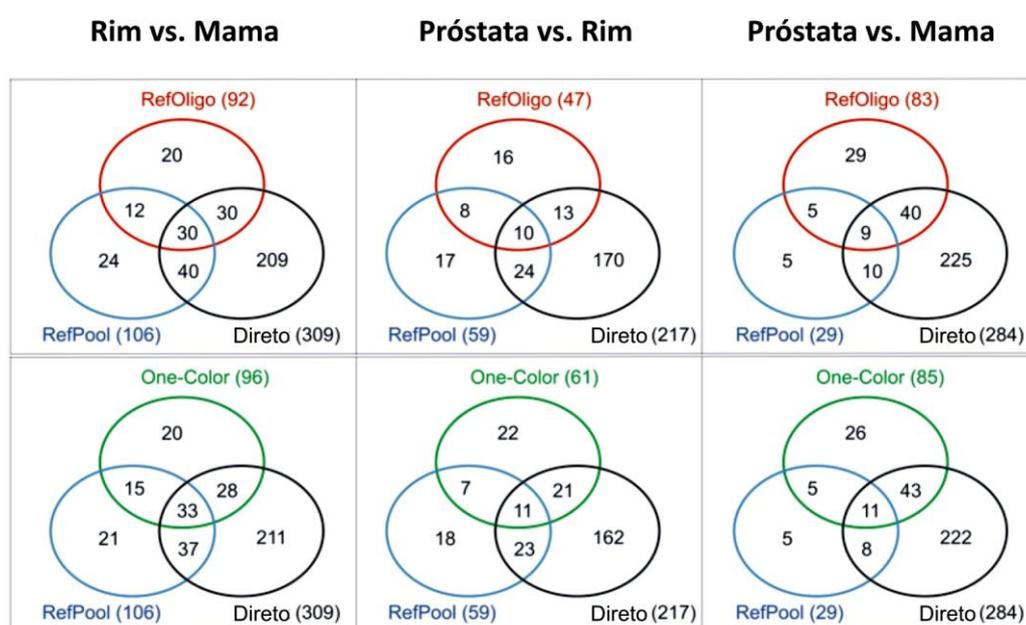


Figura 11. Comparação dos conjuntos de transcritos identificados como expressos diferencialmente em cada uma das quatro metodologias utilizadas (hibridização direta, *RefOligo*, *RefPool* e *One-Color*). O número total de transcritos identificados por cada metodologia está indicado entre parêntesis.

4.4. Análise comparativa da metodologia *RefOligo* vs. *One-Color*

Uma conclusão importante da análise descrita no item acima é que a utilização das intensidades de um único canal permite a identificação de genes diferencialmente expressos com acurácia e sensibilidade comparáveis às obtidas a partir de razões de expressão calculadas em relação à referência (Peixoto et al., 2006). Baseados nesta informação, o desenho experimental utilizado para a hibridização dos tecidos tumorais mamários baseou-se na hibridização de aRNA proveniente das amostras experimentais marcadas com Cy5 e um oligonucleotídeo sintético acoplado ao fluoróforo Cy3

(*RefOligo*) como referência externa. Dispondo agora de dados de expressão de um número significativamente maior de amostras, decidimos testar a reprodutibilidade e acurácia de medidas de expressão gênica baseadas em uma cor (*One-Color*) ou razões de expressão normalizadas através do *RefOligo*, utilizando agora os dados de 62 amostras de RNA tumoral de mama.

Os mesmos critérios de filtragem de *spots* fracos e normalização foram aplicados aos conjuntos de dados resultantes de ambos os canais de RNA e *RefOligo* das hibridizações (limiar de filtragem igual a média de intensidade dos *spots* vazios da lâmina + 2 desvios-padrão; normalização pela média aparada de 20% - ver seção Material e Métodos). Em média, foram filtrados 450 *spots* de cada paciente na análise por *RefOligo* e 165 *spots* por paciente na análise por *One-Color*. A filtragem na análise pelo *RefOligo* foi mais restritiva já que ela foi feita em cima de cada um dos conjuntos de dados separadamente (dados de RNA e dados do *RefOligo*) e as razões RNA/*RefOligo* foram calculadas a partir dos dados que ultrapassaram o limiar de filtragem (média de intensidade dos *spots* vazios + 2 desvios-padrão).

A robustez e reprodutibilidade dos dados de microarranjos podem ser avaliadas de acordo com o grau de correlação entre os dados obtidos de diferentes pacientes, já que é esperado que a maior parte dos transcritos não varie sua expressão (no caso de um mesmo organismo e tipo tecidual). A tabela 6 mostra um conjunto de dados de cinco amostras (dez microarranjos) escolhidas ao acaso para comparação do grau de correlação entre os dados obtidos por ambas as metodologias. Este conjunto de dados representa as análises feitas utilizando-se o conjunto total das 62 amostras de RNA de tumor mamário hibridizadas, já que a correlação de cerca de 40% das amostras analisadas por *RefOligo* ficou abaixo do limite mínimo que consideramos válido para um experimento de hibridização de qualidade (0,8). No caso das lâminas analisadas pela

metodologia de *One-Color*, 95% dos dados obtidos ultrapassam este limite de correlação.

Tabela 6. Valores de correlação entre dados de lâminas de 5 amostras aleatórias analisadas por duas abordagens distintas, *RefOligo* e *One-Color*. Observar que as correlações obtidas dos dados analisados pela metodologia de *One-Color* são maiores que os dados resultantes da análise usando o *RefOligo* como referência externa.

RefOligo (média = 0.81)					
	P13	P22	P32	P47	P56
P13	1,000				
P22	0,864	1,000			
P32	0,808	0,785	1,000		
P47	0,694	0,923	0,814	1,000	
P56	0,775	0,776	0,879	0,815	1,000
One-Color (média = 0.93)					
	P13	P22	P32	P47	P56
P13	1,000				
P22	0,960	1,000			
P32	0,929	0,896	1,000		
P47	0,928	0,919	0,909	1,000	
P56	0,934	0,912	0,920	0,962	1,000

Os dois conjuntos de dados analisados tanto pela metodologia de *One-Color* e *RefOligo* foram comparadas quanto à capacidade de apontar conjuntos de transcritos expressos diferencialmente entre pacientes com boa *versus* má evolução da doença em um período de 4,5 anos de acompanhamento. Os resultados obtidos pela metodologia de *One-Color* apontaram um conjunto de 320 transcritos expressos diferencialmente com uma taxa de falso-positivo de 0,27% (ferramenta *SAM* para Excel – ver seção Materiais e Métodos). No caso dos dados obtidos pela metodologia de *RefOligo*, 158 transcritos foram apontados como expressos diferencialmente com uma taxa de falso-positivo de cerca de 10%. Estes conjuntos de transcritos expressos diferencialmente foram então cruzados e apenas um transcrito apontado como diferencialmente expresso nas duas metodologias (*RefOligo* e *One-Color*) era coincidente. Dessa forma, como os dados normalizados pelo *RefOligo* apresentaram uma qualidade inferior quando comparados aos do *One-Color* e ainda apresentaram uma capacidade em apontar transcritos

expressos diferencialmente com menor acurácia, decidimos seguir com as análises do conjunto total de 74 amostras de tecido tumoral mamário utilizando a metodologia de *One-Color*.

4.5. Análise não supervisionada de um conjunto de 62 amostras de câncer de mama

Como as amostras foram coletadas ao longo de quatro anos, optamos por iniciar as análises após três anos de andamento do projeto, quando já haviam sido processadas e hibridizadas 62 amostras de tumor. Inicialmente, foi realizada uma análise não-supervisionada, baseada no agrupamento hierárquico das amostras unicamente em função do padrão de expressão dos genes selecionados. Para esta análise, foram selecionados os transcritos que apresentavam maior variação na expressão entre as amostras, dado pelo cálculo do desvio-padrão de cada transcrito dividido pela sua média. Um grupo de 396 transcritos com razão média de 1,55 foi selecionado e utilizado para a análise não supervisionada (Figura 12). Com este critério, procuramos enriquecer o conjunto em transcritos com alta variação da expressão em relação a média das amostras, que permitisse a identificação de sub-grupos de amostras baseados em características histológicas/clínicas.

O resultado obtido do agrupamento hierárquico sugere que há uma tendência de agrupamento das amostras de pacientes em função do *status* de expressão do receptor de estrogênio (Figura 12). Esta primeira análise confirma a presença de assinaturas gênicas correlacionadas com o *status* dos receptores de estrogênio em tumores de mama (Perou et al., 2000; Sorlie et al., 2001).

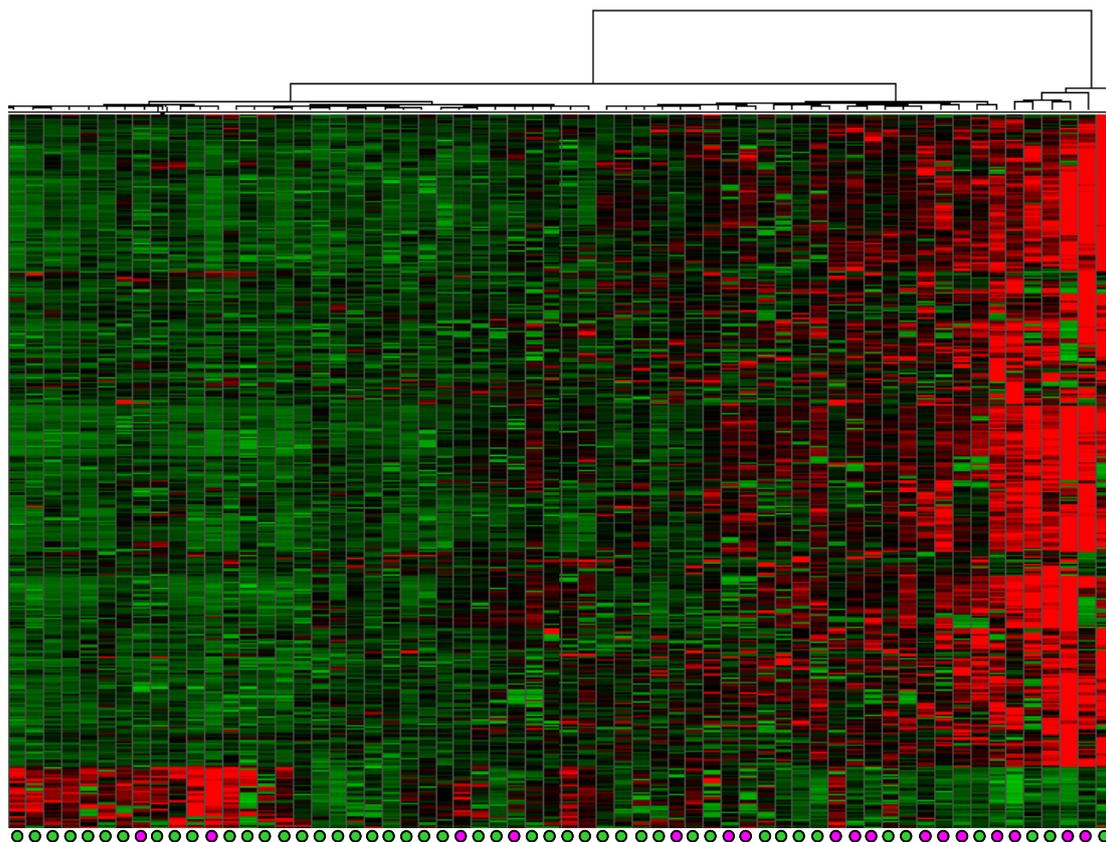


Figura 12. Mapa do agrupamento hierárquico não-supervisionado dos transcritos que apresentam maior variação da expressão entre 62 pacientes analisadas. Na horizontal estão representados cada um dos transcritos incluídos na análise (396 de 3.260 totais), enquanto cada coluna corresponde a uma amostra de tecido tumoral de mama proveniente de uma paciente. As cores verde, vermelha e preta no agrupamento correspondem, respectivamente, a sinais de intensidade total no mínimo duas vezes abaixo do desvio-padrão, duas vezes acima e intensidade total próxima de zero em relação ao desvio padrão. As cores verde e rosa no final do mapa correspondem, respectivamente, às amostras de pacientes com tumores REr+ e REr-. Para a construção deste agrupamento, foi utilizada como medida de similaridade a distância Euclidiana.

4.6. Análises supervisionadas utilizando informações clínicas e anatomo-patológicas

Uma das vantagens da colaboração que o nosso laboratório estabeleceu com o hospital HDES, em Portugal, no envio das amostras tumorais para estudo, é que estas amostras são provenientes de pacientes que tem um acompanhamento clínico permanente e atualizado, com as informações pertinentes aos tumores e pacientes, disponível para este estudo. Dessa forma, utilizamos as informações clínicas do conjunto inicial de 62 amostras de tumor de mama coletadas até 2006 para a realização de análises

supervisionadas e busca de perfis moleculares associados às características anatomo-patológicas dos tumores de mama. Apesar da disponibilidade de tecido tumoral mamário de 62 amostras, no momento da realização destas análises preliminares não havia ainda sido possível reunir as informações clínicas e anatomo-patológicas de todas as amostras. A atualização das características clínicas e patológicas foi feita quando completamos a coleta das amostras de tecido tumoral, completando um número de 74 amostras. Por essa razão nem todas as 62 amostras foram utilizadas nas análises descritas a seguir.

Inicialmente foi feita a busca de uma assinatura de expressão correlacionada a idade das pacientes. Grupos com amostras de pacientes com idade menor (16 amostras) e maior que 50 anos (43 amostras) foram comparados. Apesar da informação sobre a idade ser um fator de prognóstico, não foi possível identificar no nosso conjunto de dados um perfil de expressão com genes com expressão diferencial significativa entre os dois grupos de pacientes considerados.

Também não foi identificada um conjunto de genes capaz de discriminar amostras de pacientes com ou sem um componente hereditário (19 vs. 40 amostras, respectivamente). Foi considerado como possível câncer hereditário as pacientes que declararam a existência de doença na família. Contudo, as informações sobre o componente hereditário da doença não eram claras quanto ao grau de parentesco apresentado (a informação se resumia se existia presença de doença na família ou não). Esta falta de homogeneidade pode explicar a ausência de uma assinatura significativa.

No tocante ao tamanho do tumor, foi feita uma análise supervisionada com o objetivo de encontrar perfis de expressão de grupos de tumores menores (26 amostras) e maiores que 2cm (31 amostras). Da mesma forma como a informação de idade, estas características são muito amplas e gerais e possivelmente por isso não foi encontrado uma assinatura de expressão do tecido tumoral associado a elas.

A informação sobre a infiltração dos nódulos linfáticos é um importante fator de prognóstico na clínica médica do câncer de mama. No entanto, não foram encontrados assinaturas de expressão associadas a tumores que apresentavam maior quantidade de nódulos linfáticos infiltrados (35 amostras de pacientes sem infiltração *versus* 24 amostras de pacientes com infiltração dos nódulos linfáticos). A característica de estágio tumoral também foi utilizada em análises supervisionadas, mas não foi encontrado um perfil de expressão gênica capaz de distinguir amostras de estágio tumoral I (18 amostras) das que apresentavam estágio tumoral II (29 amostras) ou III (11 amostras). O mesmo resultado foi observado utilizando o grau histológico das amostras (5 amostras com grau histológico I, 29 com II e 20 com III).

A análise do perfil de expressão gênico de 37 amostras de tumor de mama RPr+ comparado ao perfil de expressão gênico de 25 amostras RPr- permitiu a identificação de uma assinatura molecular correlacionada com o *status* do receptor de progesterona contendo 127 transcritos (87 exônicos, 34 intrônicos e 6 intergênicos), com uma taxa de falso positivo de 3,7% e um valor *p* máximo de 0,01 (Figura 13). Este conjunto de transcritos foi capaz de agrupar as amostras RPr+ e RPr- baseado em seu nível de expressão com uma taxa de acerto de 76% (Figura 13).

Análises preliminares supervisionadas também foram realizadas partindo da informação do *status* do receptor de estrogênio nas 64 amostras de tumores de mama. Assim, o perfil de expressão de 45 amostras de tumor de mama REr+ foi comparado ao perfil de expressão de 17 amostras REr-, revelando uma assinatura molecular correlacionada com o *status* do receptor de estrogênio contendo 63 transcritos (47 exônicos, 14 intrônicos e 2 intergênicos), com uma taxa de falso positivo de 2,46% e um valor *p* máximo de 0,005 (Figura 14). Este conjunto de transcritos foi capaz de agrupar

as amostras REr+ e REr- baseado em seu nível de expressão com uma taxa de acerto de 85% (Figura 14).

Portanto, de todas as características anatomo-patológicas utilizadas nas análises supervisionadas, as únicas que produziram perfis de expressão reproduzíveis e consistentes foram as características relacionadas ao *status* de expressão do receptor de estrogênio e progesterona no tecido tumoral mamário.

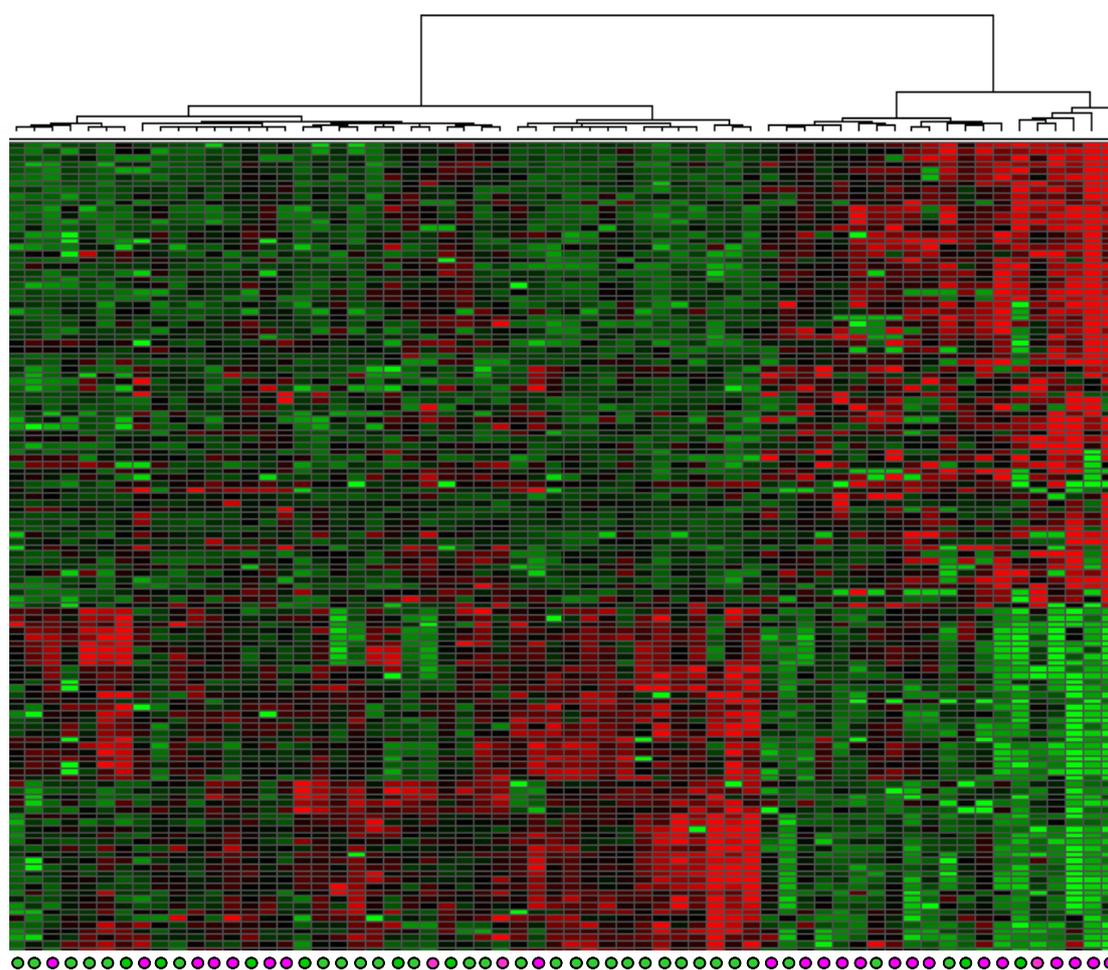


Figura 13. Mapa do perfil de expressão de 62 amostras de tumor de mama de acordo com um grupo de 127 transcritos apontados como diferencialmente expressos entre tumores RPr+ vs. RPr-. Na horizontal estão representados os 127 transcritos, enquanto cada coluna corresponde a cada uma das 62 amostras analisadas. Para cada linha, as intensidades de expressão entre os dois grupos foram normalizadas e coloridas em função do número de desvios-padrão de cada amostra em relação ao valor médio de intensidade de todas as amostras. No fim de cada coluna também está sinalizado, de acordo com as cores rosa ou verde, o grupo que cada amostra pertence, RPr- e RPr+ respectivamente. Foi utilizada a distância Euclidiana como medida de similaridade na análise de agrupamento hierárquico.

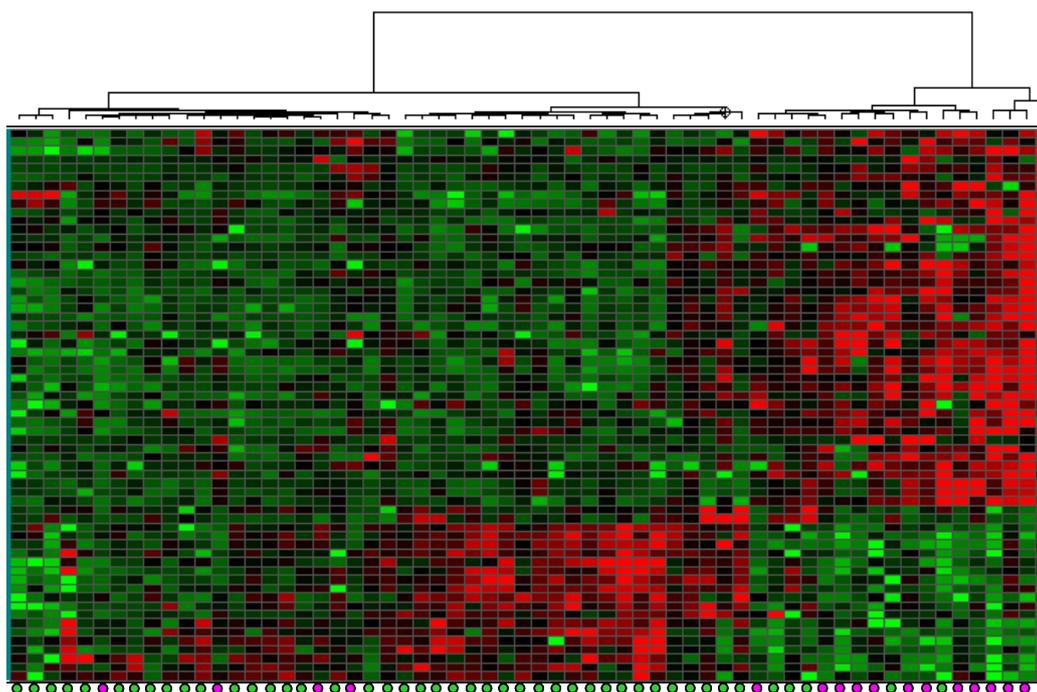


Figura 14. Mapa do perfil de expressão de 62 amostras de tumor de mama de acordo com um grupo de 63 transcritos apontados como diferencialmente expressos entre tumores RER+ vs.RER-. Na horizontal estão representados os 63 transcritos, enquanto cada coluna corresponde a cada uma das 62 amostras analisadas. Para cada linha, as intensidades de expressão entre os dois grupos foram normalizadas e coloridas em função do número de desvios-padrão de cada amostra em relação ao valor médio de intensidade de todas as amostras. No fim de cada coluna também está sinalizado, de acordo com as cores rosa ou verde, o grupo que cada amostra pertence, RER- e RER+ respectivamente. Foi utilizada a distância Euclidiana como medida de similaridade na análise de agrupamento hierárquico.

Após a conclusão da coleta de amostras, foram processadas e hibridizadas 74 amostras de adenocarcinoma de mama no total. A maior parte das 74 amostras de RNA de tecido tumoral mamário (78%) possui tipo histológico ductal invasivo (Tabela 4). Dezesesseis amostras correspondem a diversos tipos histológicos representados por poucas amostras. Apesar das proporções refletirem as da população de pacientes com câncer de mama em geral, a quantidade do tipo histológico ductal (58 amostras) é muito maior que a segunda maior classe de tipo histológico representada em nosso conjunto amostral (lobular, 5 amostras), o que limita a busca de marcadores específicos das demais histologias. Por essa razão, e também com o objetivo de diminuir o ruído das análises,

resolvemos limitar as análises subseqüentes apenas as 58 amostras com tipo histológico ductal invasivo.

4.7. Expressão de RNAs não-codificadores em amostras de tumor de mama

Os dados ds 58 amostras de carcinoma de mama ductal foram analisados para investigar o nível de expressão relativo dos transcritos codificadores de proteína e ncRNAs interrogados no microarranjo. Uma análise da distribuição cumulativa revelou a presença de frações comparáveis de transcritos intrônicos, intergênicos e exônicos considerados expressos nas 58 amostras de tumor de mama analisadas (Figura 15). O nível de expressão médio e a mediana destas três classes de transcritos são comparáveis; os valores arbitrários de intensidade (valores normalizados) observados são respectivamente de 14,3 e 6,9, para os exônicos, 14,5 e 6,3, para os intrônicos e 15,9 e 7,1, para os intergênicos. Além disso, cerca de 90% das moléculas alvo dos microarranjos de cada uma das três classes estão expressas em pelo menos 50 amostras, independente de suas características anatomo-patológicas (Figura 15). Este resultado mostra que RNAs intrônicos e intergênicos são expressos ubiqüamente em tecidos tumorais de mama, com uma abundância comparável a dos transcritos exônicos codificadores para proteínas..

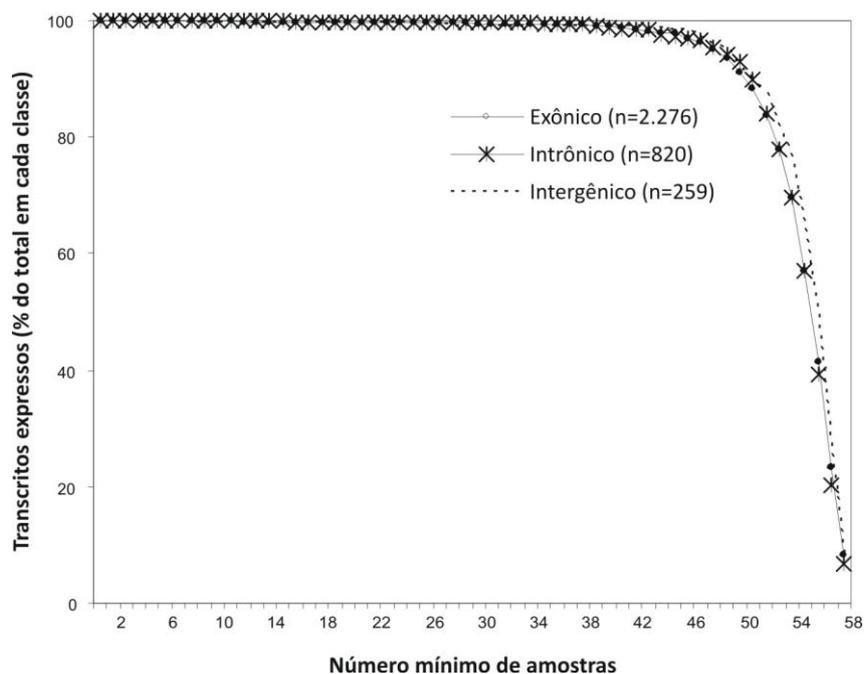


Figura 15. Moléculas de RNA não codificadores provenientes de introns de genes conhecidos são ubiqüamente expressas em amostras de tumor de mama ductal invasivo. Para cada amostra, a sonda do microarranjo foi considerada válida se ao menos 3 das 4 medidas de réplica técnica apresentassem intensidade de sinal mínima de 2 desvios-padrão acima do sinal médio de um conjunto de pontos vazios no microarranjo. A porcentagem total de sondas de cDNA no microarranjo (eixo Y) de cada uma das três diferentes categorias de sondas (exônico, intrônico e intergênico) é referente ao número mínimo de amostras em que suas expressões foram observadas (eixo X). O número total de sondas em cada uma das categorias está indicada em parêntesis.

4.8. Classificação dos tumores de acordo com *status* de receptor de estrogênio

A fim de identificar uma assinatura correlacionada com o *status* de expressão do receptor de estrogênio, foram comparados os perfis de expressão gênica de dois conjuntos de amostras: 36 tumores com positividade para receptor de estrogênio (REr+) e 22 tumores com negatividade para receptor de estrogênio (REr-).

Consideramos então como conjunto de transcritos válidos na análise aqueles presentes em 80% dos 58 pacientes, ou seja, 3.221 transcritos. Utilizando um programa desenvolvido pelo nosso grupo para o cálculo do *SNR*, 95 destes foram identificados como diferencialmente expressos entre os grupos de tumores REr+ e REr- com um valor $p \leq 0,002$ e 4.000 permutações dos dados (ver seção Materiais e Métodos). Dentre estes

transcritos, 76 correspondem a regiões exônicas de genes conhecidos, 14 mapeiam em regiões intrônicas e 5 em regiões intergênicas. (Tabela 7).

Tabela 7. Conjunto de 95 transcritos expressos diferencialmente em tumores de mama RER+ versus tumores de mama RER- (ordenados pela ordem alfabética do *locus* gênico), identificados por *SNR* com 4.000 permutações e limiar de significância de $p \leq 0,002$.

Gene ID	Tamanho do transcrito (pb)	Coordenada genômica	Locus gênico	Mapeamento exon/ intron	Anotação	Log ₂ (ER+/ER-)	valor p (ER+ vs, ER-)
29	438	chr17:954,074-954,473	ABR	Intron # 2	Active BCR-related gene	-0,43	0,0005
-	358	chr6:37,558,261-37,558,607	AI 096883	Região intergênica	EST: AI 096883	0,55	0,0005
203	544	chr9:129,670,106-129,674,917	AK1	Exon # 4, 5, 6, 7	Adenylate kinase 1	-0,44	0,0000
129138	361	chr22:36,557,390-36,557,746	ANKRD54	5' UTR	Ankyrin repeat domain 54	0,49	0,0010
290	291	chr15:88,136,393-88,136,713	ANPEP	Exon # 16, 17	Alanyl (membrane) aminopeptidase	-0,52	0,0010
324	512	chr5:112,207,490-112,207,994	APC	Exon # 16, 3' UTR	Adenomatosis polyposis coli	-0,99	0,0003
10095	545	chr7:98,823,746-98,828,294	ARPC1B	Exon # 4, 5, 6, 7, 8	Actin related protein 2/3 complex, subunit 1B, 41kDa	-0,56	0,0000
467	252	chr1:209,181,859-209,182,087	ATF3	3' UTR	Activating transcription factor 3	-0,48	0,0000
483	556	chr3:143,123,497-143,127,289	ATP1B3	Exon # 6, 7	ATPase, Na+/K+ transporting, beta 3 polypeptide	-0,54	0,0015
506	504	chr12:55,322,837-55,324,982	ATP5B	Exon # 3, 4, 5, 6	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide	-0,47	0,0003
-	445	chr11:121,465,359-121,465,782	AW 606760	Região Intergênica	EST: AW 606760	-0,30	0,0010
-	555	chr1:114,732,898-114,733,417	AW 902100	Região Intergênica	EST: AW 902100	0,62	0,0008
339862	335	chr3:18,120,451-18,120,786	BC 035826	Região intergênica	Hypothetical protein LOC339862	-0,40	0,0005
7862	622	chr3:9,759,698-9,760,956	BRPF1	Exon # 7, 8, 9	Bromodomain and PHD finger containing, 1	-0,44	0,0010
818	572	chr10:75,242,356-75,242,917	CAMK2G	3' UTR	Calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma	-0,32	0,0018
865	351	chr16:65,691,986-65,692,332	CBFB	3' UTR	Core-binding factor, beta subunit	-0,52	0,0018
595	356	chr11:69,177,768-69,178,113	CCND1	3' UTR	Cyclin D1	0,58	0,0010
965	242	chr1:116,880,108-116,880,341	CD58	Exon # 3	CD58 molecule	-0,35	0,0000
994	304	chr20:3,734,444-3,734,685	CDC25B	3' UTR	Cell division cycle 25 homolog B (S. pombe)	-0,68	0,0008
990	244	chr17:35,711,883-35,712,126	CDC6	3' UTR	Cell division cycle 6 homolog (S. cerevisiae)	-0,31	0,0008
83879	570	chr2:173,941,265-173,941,833	CDCA7	3' UTR	Cell division cycle associated 7	-0,49	0,0000
1000	541	chr18:23,786,173-23,819,046	CDH2	Exon # 11, 12, 13, 14	Cadherin 2, type 1, N-cadherin (neuronal)	-0,39	0,0000
80152	351	chr16:66,421,463-66,421,803	CENPT	Exon # 11, 12	Centromere protein T	-0,62	0,0008
1108	642	chr12:6,549,800-6,552,650	CHD4	Exon # 38, 39, 40, 3' UTR	Chromodomain helicase DNA binding protein 4	-0,40	0,0000
-	600	chr15:39,380,690-39,381,250	CK 902130	Região Intergênica	EST: CK 902130	-0,21	0,0015

4849	275	chr19:59,347,975-59,348,463	<i>CNOT3</i>	Exon # 15, 16	CCR4-NOT transcription complex, subunit 3	-0,35	0,0020
1312	270	chr22:18,330,262-18,331,270	<i>COMT</i>	Exon # 3, 4	Catechol-O-methyltransferase	-0,40	0,0008
10391	384	chr15:66,806,684-66,807,050	<i>CORO2B</i>	3' UTR	Coronin, actin binding protein, 2B	-0,49	0,0000
27132	474	chr16:88,173,288-88,173,741	<i>CPNE7</i>	Intron # 3	Copine VII	0,56	0,0000
1459	585	chr16:56,749,399-56,757,086	<i>CSNK2A2</i>	Exon # 10, 11, 3' UTR	Casein kinase 2, alpha prime polypeptide	-0,34	0,0015
1496	488	chr2:80,669,975-80,728,349	<i>CTNNA2</i>	Exon # 13, 14, 15, 16	Catenin (cadherin-associated protein), alpha 2	-0,32	0,0000
23191	626	chr15:20,554,379-20,554,999	<i>CYFIP1</i>	Exon # 27, 3' UTR	Cytoplasmic FMR1 interacting protein 1	-0,51	0,0003
9988	378	chr7:86,655,177-86,655,545	<i>DMTF1</i>	Exon # 13	Cyclin D binding myb-like transcription factor 1	-0,72	0,0000
23268	400	chr10:101,625,543-101,625,944	<i>DNMBP</i>	5' UTR	Dynamin binding protein	0,48	0,0000
10311	535	chr21:37,519,194-37,519,696	<i>DSCR3</i>	3' UTR	Down syndrome critical region gene 3	-0,46	0,0013
1869	419	chr20:31,727,925-31,728,315	<i>E2F1</i>	Exon # 7, 3' UTR	E2F transcription factor 1	0,41	0,0005
8668	402	chr1:32,464,356-32,466,767	<i>EIF3I</i>	Exon # 5, 6, 7	Eukaryotic translation initiation factor 3, subunit I	-0,38	0,0000
84034	420	chr18:2,881,078-2,881,474	<i>EMILIN2</i>	Exon # 3	Elastin microfibril interfacier 2	0,60	0,0008
84961	548	chr17:34,670,584-34,671,122	<i>FBXL20</i>	3' UTR	F-box and leucine-rich repeat protein 20	0,46	0,0008
201163	433	chr17:17,057,861-17,060,541	<i>FLCN</i>	Exon # 11, 12, 13, 14	Folliculin	0,72	0,0003
389337	432	chr5:148,992,153-148,992,564	<i>FLJ41603</i>	3' UTR	FLJ41603 protein	0,71	0,0020
3607	397	chr17:78,075,438-78,075,816	<i>FOXP2</i>	Intron # 1	Forkhead box K2	0,71	0,0020
2625	412	chr10:8,155,963-8,156,373	<i>GATA3</i>	3' UTR	GATA binding protein 3	1,49	0,0000
9843	407	chrX:65,403,310-65,403,704	<i>HEPH</i>	3' UTR	Hephaestin	-0,44	0,0020
221092	446	chr11:62,245,946-62,246,881	<i>HNRPUL2</i>	Exon # 5, 6, 7, 8	Heterogeneous nuclear ribonucleoprotein U-like 2	-0,36	0,0000
50809	375	chr1:20,980,342-20,980,697	<i>HP1BP3</i>	Intron # 12	Heterochromatin protein 1, binding protein 3	0,79	0,0005
3249	526	chr19:40,248,033-40,249,138	<i>HPN</i>	Exon # 11, 12, 13, 14, 3' UTR	Hepsin (transmembrane protease, serine 1)	-0,53	0,0013
3419	177	chr15:76,249,206-76,249,363	<i>IDH3A</i>	3' UTR	Isocitrate dehydrogenase 3 (NAD+) alpha	0,70	0,0013
3480	344	chr15:97,309,393-97,318,078	<i>IGF1R</i>	Exon # 16, 17, 18, 19	Insulin-like growth factor 1 receptor	-0,49	0,0000
55677	359	chr2:127,963,697-127,968,721	<i>IWS1</i>	Exon # 8, 9, 10, 11	IWS1 homolog (S. cerevisiae)	-0,31	0,0000
221955	494	chr7:6,467,304-6,467,788	<i>KDEL2</i>	3' UTR	Diacylglycerol lipase, beta	-0,65	0,0013
389072	363	chr2:208,574,282-208,574,634	<i>LOC389072</i>	Exon # 1, 5' UTR	Hypothetical protein LOC389072	-0,43	0,0003
10916	553	chrX:54,722,406-54,724,947	<i>MAGED2</i>	Exon # 7, 8, 9, 10, 11	Melanoma antigen family D, 2	0,44	0,0003
2011	399	chr11:63,429,057-63,433,218	<i>MARK2</i>	Exon # 12, 13, 14	MAP/microtubule affinity-regulating kinase 2	-0,35	0,0015
27430	130	chr5:162,878,072-162,878,199	<i>MAT2B</i>	3' UTR	Methionine adenosyltransferase II, beta	0,65	0,0000
55796	373	chrX:131,449,375-131,449,723	<i>MBNL3</i>	Intron # 1	Muscleblind-like 3 (Drosophila)	-0,69	0,0000
84365	659	chr2:122,201,579-122,209,474	<i>MKI67IP</i>	Exon # 2, 3, 4, 5, 6, 7	MKI67 (FHA domain) interacting nucleolar phosphoprotein	-0,37	0,0003
4318	489	chr20:44,075,707-44,078,389	<i>MMP9</i>	Exon # 10, 11, 12, 13	Matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	-0,59	0,0000

64419	447	chr3:9,705,656-9,714,551	<i>MTMR14</i>	Exon # 15, 16, 17	Myotubularin related protein 14	0,54	0,0003
9172	639	chr8:2,080,252-2,080,681	<i>MYOM2</i>	Exon # 32, 3' UTR	Myomesin (M-protein) 2, 165kDa	-0,59	0,0005
9111	354	chr2:151,835,560-151,840,325	<i>NMI</i>	Exon # 6, 7, 8	N-myc (and STAT) interactor	-0,63	0,0003
64710	369	chr1:203,952,418-203,952,782	<i>NUCKS1</i>	3' UTR	Nuclear casein kinase and cyclin-dependent	0,77	0,0015
4978	366	chr11:132,639,356-132,639,712	<i>OPCML</i>	Intron # 5	Opioid binding protein/cell adhesion molecule-like	-0,24	0,0003
26986	696	chr8:101,788,153-101,794,110	<i>PABPC1</i>	Exon # 6, 7, 8, 9, 10, 11	Poly(A) binding protein, cytoplasmic 1	-0,71	0,0005
11240	274	chr1:17,266,005-17,266,264	<i>PADI2</i>	3' UTR	Peptidyl arginine deiminase, type II	-1,20	0,0005
8566	432	chr21:43,991,509-43,991,920	<i>PDXK</i>	Intron # 5	Pyridoxal (pyridoxine, vitamin B6) kinase	-0,53	0,0000
51105	487	chr8:133,913,706-133,920,865	<i>PHF20L1</i>	Exon # 14, 15, 16, 17	PHD finger protein 20-like 1	1,01	0,0000
5256	371	chrX:18,821,585-18,822,445	<i>PHKA2</i>	Exon # 30, 31	Phosphorylase kinase, alpha 2 (liver)	-0,42	0,0010
9883	543	chr7:71,997,814-72,003,173	<i>POM121</i>	Intron # 2	Nuclear pore membrane protein 121	0,51	0,0005
5465	533	chr22:45,015,536-45,016,050	<i>PPARA</i>	3' UTR	Peroxisome proliferator-activated receptor alpha	-0,39	0,0000
81706	391	chr6:150,612,776-150,613,153	<i>PPP1R14C</i>	3' UTR	Protein phosphatase 1, regulatory (inhibitor) subunit 14C	-0,47	0,0010
5578	568	chr17:62,070,695-62,071,251	<i>PRKCA</i>	Intron # 3	Protein kinase C, alpha	0,56	0,0003
5719	675	chr11:237,408-242,646	<i>PSMD13</i>	Exon # 6, 7, 8, 9, 10, 11, 12, 3' UTR	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 13	-0,33	0,0005
53635	562	chr19:55,049,786-55,052,200	<i>PTOV1</i>	Exon # 4, 5, 6	Prostate tumor overexpressed gene 1	0,35	0,0003
5884	522	chr5:68,728,052-68,745,827	<i>RAD17</i>	Exon # 12, 13, 14, 15	RAD17 homolog (S. pombe)	0,71	0,0003
8737	563	chr6:3,050,875-3,056,139	<i>RIPK1</i>	Exon # 8, 9	Receptor (TNFRSF)-interacting serine-threonine kinase 1	0,48	0,0005
6382	533	chr2:20,265,355-20,265,868	<i>SDC1</i>	3' UTR	Syndecan 1	-0,67	0,0003
10806	435	chr1:241,537,735-241,560,483	<i>SDCCAG8</i>	Exon # 8, 9, 10	Serologically defined colon cancer antigen 8	-0,41	0,0000
9632	567	chr10:75,199,196-75,200,235	<i>SEC24C</i>	Exon # 16, 17, 18, 19	SEC24 related gene family, member C (S. cerevisiae)	0,54	0,0000
23176	491	chr5:132,112,961-132,114,876	<i>SEPT8</i>	5' UTR	Septin 8	0,80	0,0003
4087	503	chr18:43,621,900-43,625,788	<i>SMAD2</i>	Exon # 9, 10, 3' UTR	SMAD family member 2	-0,37	0,0010
10274	529	chr3:137,598,538-137,599,055	<i>STAG1</i>	Intron # 19	Stromal antigen 1	0,50	0,0018
6770	330	chr8:38,120,386-38,120,714	<i>STAR</i>	3' UTR	Steroidogenic acute regulatory protein	-0,39	0,0003
90627	335	chr13:32,802,644-32,802,961	<i>STARD13</i>	Intron # 4	StAR-related lipid transfer (START) domain containing 13	-0,26	0,0015
6612	672	chr21:45,050,625-45,051,291	<i>SUMO3</i>	3' UTR	SMT3 suppressor of mif two 3 homolog 3 (S. cerevisiae)	-0,36	0,0000
6857	447	chr12:78,351,291-78,351,736	<i>SYT1</i>	Intron # 6	Synaptotagmin I	0,50	0,0005
6938	250	chr15:55,366,737-55,366,981	<i>TCF12</i>	3' UTR	Transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)	-0,27	0,0000
55281	322	chr7:134,496,909-134,497,213	<i>TMEM140</i>	Intron # 1	Transmembrane protein 140	0,90	0,0003
28755	300	chr14:22,090,577-22,090,871	<i>TRAC</i>	3' UTR	T-cell receptor alpha chain C region.	-0,67	0,0020
7260	571	chr2:3,171,987-3,176,878	<i>TSSC1</i>	Exon # 7, 8, 9, 3' UTR	Tumor suppressing subtransferrable candidate 1	-0,48	0,0003
8078	353	chr12:6,842,677-6,843,616	<i>USP5</i>	Exon # 15, 16, 17	Ubiquitin specific peptidase 5 (isopeptidase T)	-0,25	0,0000

7403	280	chrX:44,855,734-44,856,003	<i>UTX</i>	3' UTR	Ubiquitously transcribed tetratricopeptide repeat, X chromosome	-0,55	0,0003
22884	390	chr10:1,140,549-1,140,717	<i>WDR37</i>	Intron # 9	WD repeat domain 37	-0,50	0,0000
56829	551	chr7:138,381,228-138,381,764	<i>ZC3HAV1</i>	3' UTR	Zinc finger CCCH-type, antiviral 1	-0,48	0,0010
7582	461	chr10:42,406,542-42,406,992	<i>ZNF33B</i>	3' UTR	Zinc finger protein 33B	-0,26	0,0015

Esta assinatura de estrogênio foi comparada a listas de genes identificadas em trabalhos similares utilizando plataformas de microarranjos comerciais, utilizando a ferramenta *Oncomine* (www.oncomine.org). Quarenta e um transcritos da assinatura de receptor de estrogênio mapeiam em genes encontrados em outros conjuntos de dados já publicados como diferencialmente expressos em amostras de tumor de mama quando feita a comparação de perfis de expressão gênica REr+ vs. REr- (Tabela 8). Destes 41 transcritos, 34 correspondem a transcritos exônicos, 6 são intrônicos e 1 é intergênico.

Os níveis de expressão dos transcritos da assinatura foram comparados aos dos encontrados em estudos já publicados comparando perfis de expressão gênica de tumores de mama em função do *status* de expressão do receptor de estrogênio (Chin et al., 2006; Desmedt et al., 2007; Ginestier et al., 2006; Minn et al., 2005; Richardson et al., 2006; Sotiriou et al., 2006; Turashvili et al., 2007; van de Vijver et al., 2002; Wang et al., 2005). Dos 34 transcritos exônicos que se sobrepõem às outras análises já descritas, 26 (76%) apresentam uma alteração na expressão entre amostras REr+ vs. REr- na mesma direção reportada na literatura. Apenas 8 apresentaram nível de expressão invertido ao do gene codificador reportado na literatura (Tabela 8). Dos 6 transcritos intrônicos, 4 apresentavam nível de expressão invertido ao do gene codificador do mesmo *locus* gênico e o único transcrito intergênico possui nível de expressão idêntico (Tabela 8).

Tabela 8. Relação dos transcritos da assinatura de RER identificados como diferencialmente expressos entre tumores de mama RER+ vs. RER-, em dados disponíveis na literatura. Os dados foram obtidos utilizando a ferramenta *Oncomine* (www.oncomine.org) e estão ordenados pela ordem alfabética do *locus* gênico.

Locus gênico	Mapeamento exon/intron	Log2 (ER+/ER-)	Concordância com dados de expressão reportados na literatura	Referência
AK1	Exon # 4, 5, 6, 7	-0.44	oposto	(Desmedt et al., 2007)
ARPC1B	Exon # 4, 5, 6, 7, 8	-0.56	igual	(van de Vijver et al., 2002)
ATP1B3	Exon # 6, 7	-0.54	igual	(van de Vijver et al., 2002)
AW 606760	Região Intergênica	-0.30	igual	(van de Vijver et al., 2002)
CBFB	3' UTR	-0.52	igual	(van de Vijver et al., 2002)
CCND1	3' UTR	0.58	igual	(Ginestier et al., 2006)
CD58	Exon # 3	-0.35	igual	(Wang et al., 2005)
CDC25B	3' UTR	-0.68	igual	(van de Vijver et al., 2002)
CDC6	3' UTR	-0.31	igual	(van de Vijver et al., 2002)
CDCA7	3' UTR	-0.49	igual	(van de Vijver et al., 2002)
CSNK2A2	Exon # 10, 11, 3' UTR	-0.34	igual	(Sotiriou et al., 2006)
DNMBP	5' UTR	0.48	igual	(van de Vijver et al., 2002)
E2F1	Exon # 7, 3' UTR	0.41	oposto	(van de Vijver et al., 2002)
EMLIN2	Exon # 3	0.60	oposto	(van de Vijver et al., 2002)
FLJ41603	3' UTR	0.71	igual	(van de Vijver et al., 2002)
FOXK2	Intron # 1	0.71	oposto	(Wang et al., 2005)
GATA3	3' UTR	1.49	igual	(van de Vijver et al., 2002)
HP1BP3	Intron # 12	0.79	oposto	(Turashvili et al., 2007)
HPN	Exon # 11, 12, 13, 14, 3' UTR	-0.53	oposto	(Sotiriou et al., 2006) (van de Vijver et al., 2002)
IGF1R	Exon # 16, 17, 18, 19	-0.49	oposto	(van de Vijver et al., 2002) (Desmedt et al., 2007)
KDELR2	3' UTR	-0.65	igual	(van de Vijver et al., 2002)
MAGED2	Exon # 7, 8, 9, 10, 11	0.44	igual	(van de Vijver et al., 2002)
MKI67IP	Exon # 2, 3, 4, 5, 6, 7	-0.37	igual	(Richardson et al., 2006)
MMP9	Exon # 10, 11, 12, 13	-0.59	igual	(van de Vijver et al., 2002)
NMI	Exon # 6, 7, 8	-0.63	igual	(van de Vijver et al., 2002)
PABPC1	Exon # 6, 7, 8, 9, 10, 11	-0.71	igual	(van de Vijver et al., 2002)
PADI2	3' UTR	-1.20	igual	(van de Vijver et al., 2002)
PDXK	Intron # 5	-0.53	igual	(van de Vijver et al., 2002)
PPARA	3' UTR	-0.39	igual	(van de Vijver et al., 2002)
PPP1R14C	3' UTR	-0.47	igual	(Richardson et al., 2006)
PRKCA	Intron # 3	0.56	oposto	(Chin et al., 2006)
PTOV1	Exon # 4, 5, 6	0.35	igual	(Minn et al., 2005)
RAD17	Exon # 12, 13, 14, 15	0.71	igual	(van de Vijver et al., 2002)
SDC1	3' UTR	-0.67	igual	(van de Vijver et al., 2002)
SEPT8	5' UTR	0.80	igual	(Sotiriou et al., 2006)
SMAD2	Exon # 9, 10, 3' UTR	-0.37	oposto	(van de Vijver et al., 2002)
STARD13	Intron # 4	-0.26	oposto	(van de Vijver et al., 2002)
SUMO3	3' UTR	-0.36	igual	(Richardson et al., 2006)
SYT1	Intron # 6	0.50	igual	(Richardson et al., 2006)
TCF12	3' UTR	-0.27	oposto	(Wang et al., 2005)
UTX	3' UTR	-0.55	oposto	(van de Vijver et al., 2002)

Os 14 transcritos intrônicos foram investigados quanto à capacidade de codificar proteínas (*ESTScan 2* e *ORF Finder*). Apenas 1 deles apresentou potencial para gerar

uma proteína com pelo menos 100 aminoácidos (*CPNE7*, 140aa), critério estabelecido para se considerar um gene como codificador para proteína (Mattick e Makunin, 2006).

Os transcritos selecionados também foram averiguados em relação à capacidade de produzir microRNAs (*miRBase Sequence Database*) e snoRNAs (*snoRNABase*) e nenhum deles apresentou alinhamento significativo com qualquer seqüência representada nesses bancos de dados. Quando analisados utilizando-se a ferramenta BLAT (*Genome Browser* da Universidade da Califórnia, Santa Cruz), 2 dos 14 transcritos intrônicos apresentaram sítio de ligação para fatores de transcrição na porção 5' (*OPCML* e *PRKCA*), podendo portanto apresentar alguma função de regulação da transcrição. Além disso, o transcrito do gene *OPCML* não só possui sítio de ligação para fator de transcrição próximo à sua ponta 5', mas também uma região genômica onde foi detectada sítio de ligação específico para o fator de transcrição c-Myc em células B humanas, podendo controlar a expressão deste ncRNA intrônico (Figura 16-A) (Zeller et al., 2006). O transcrito do gene *OPCML* em estudo se encontra em uma região em que foi mapeada a existência de um transcrito senso de pelo menos 200 mil pb, através do seqüenciamento de bibliotecas PET (*paired-end ditag*) a partir de linhagens celulares tratadas com 5-fluorouracil (droga utilizado no tratamento de câncer) (Ng et al., 2005; Choo et al., 2006) (Figura 16-B). Esta mesma região é ainda repleta de sítios de ligação para fatores de transcrição (Figura 16-B).

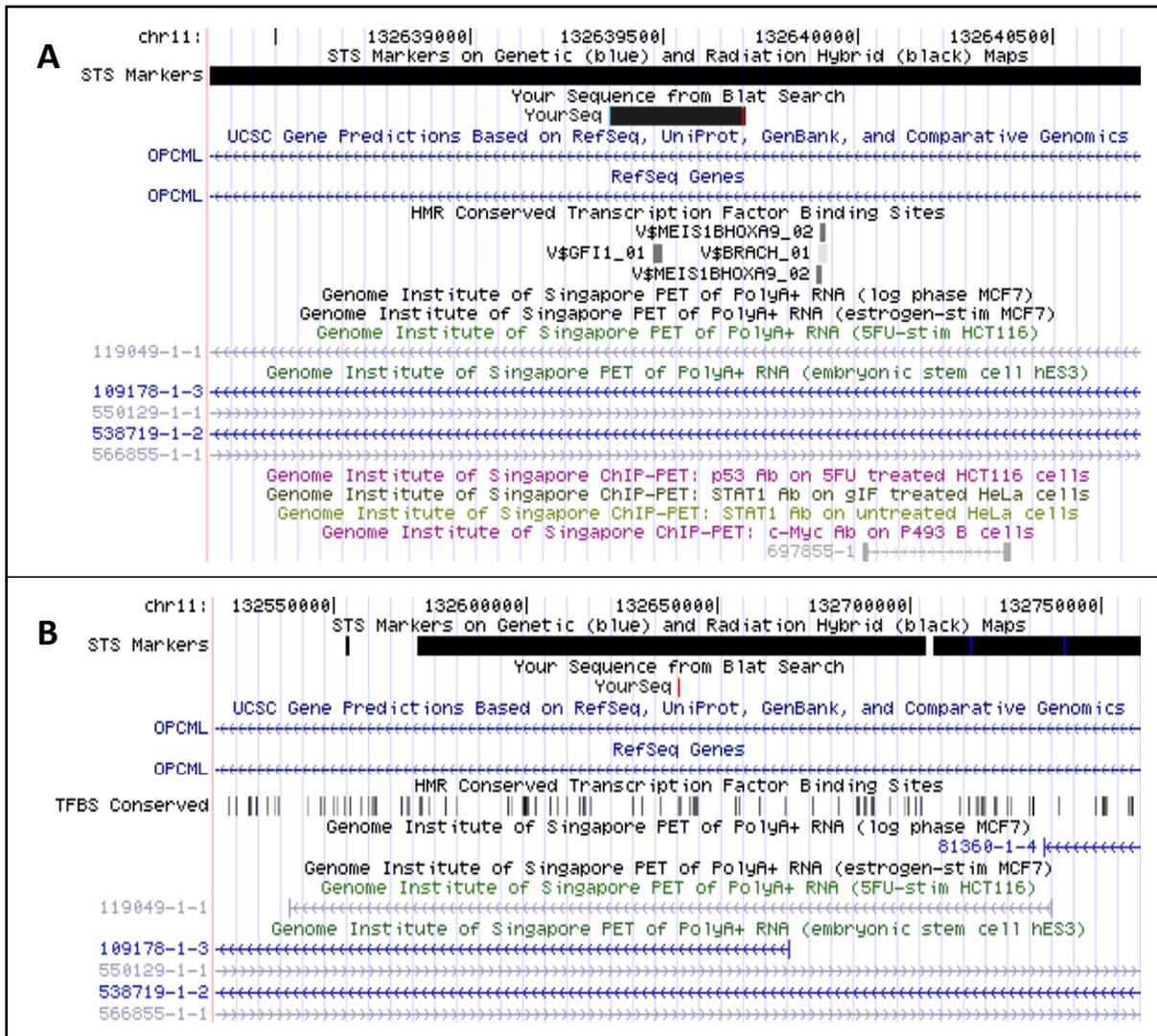


Figura 16. Visualização do mapeamento genômico do transcrito intrônico do gene *OPCML* através da ferramenta BLAT da Universidade da Califórnia de Santa Cruz (genome.ucsc.edu). Em A, observar sítios conservados de ligação de fatores de transcrição na região 5'da seqüência do transcrito e a região de ligação do fator de transcrição c-Myc detectado em células B humanas (última linha do painel). Em B foi feita a ampliação da região onde o transcrito de *OPCML* mapeia: existe um transcrito senso na região do gene *OPCML* (ver linha PET of PolyA+ RNA – 5FU-stim HCT116), de tamanho aproximado de 200 mil pb, na mesma região do transcrito de *OPCML* estudado, que também é rica em sítios conservados para ligação de fatores de transcrição.

Este grupo de 95 transcritos foi então utilizado para re-classificar as 58 amostras através de *clustering* hierárquico (Figura 17). De acordo com esta metodologia, o grupo de 95 transcritos expressos diferencialmente foi capaz de re-classificar as 58 amostras de câncer de mama com 90% de acerto; apenas 6 pacientes com tumores de mama REr+ foram agrupadas juntamente com pacientes com tumores REr- (Figura 17).

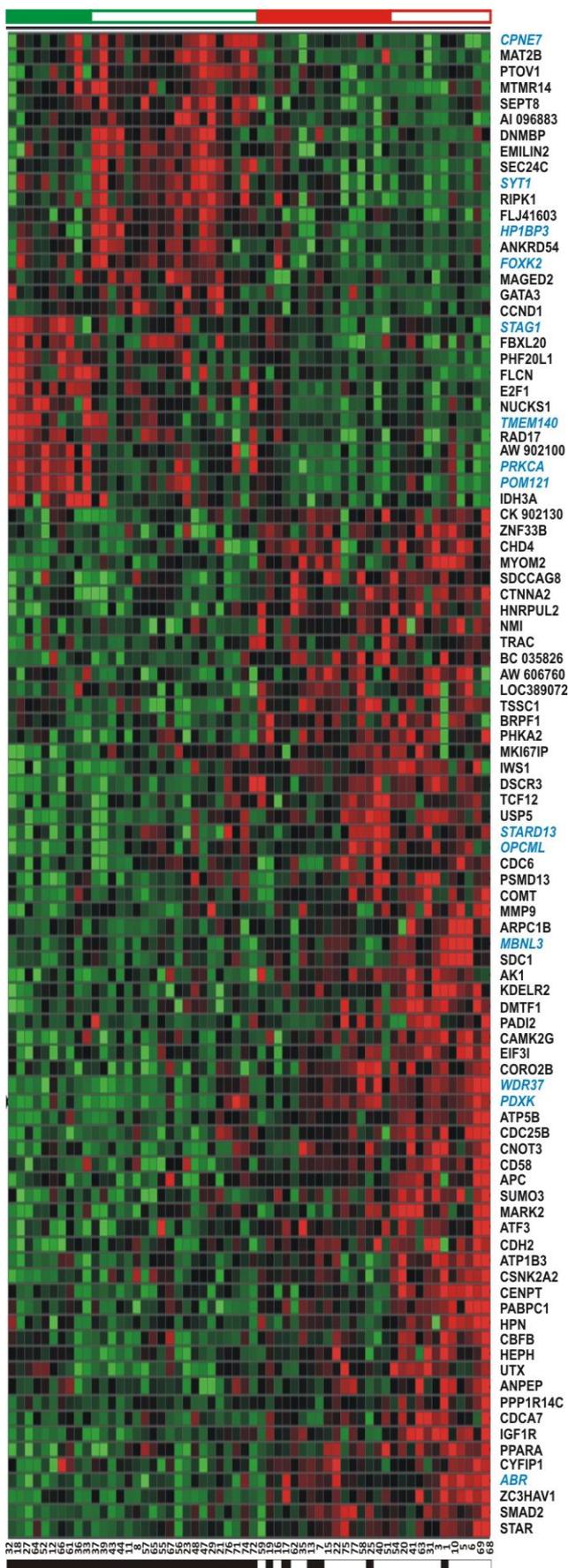


Figura 17. Mapa representativo do nível de expressão de 95 transcritos (linhas) indicados como diferencialmente expressos entre 58 tumores de mama REr+ vs. REr- (colunas). Perfil de expressão de transcritos selecionados como expressos diferencialmente através de análise supervisionada de duas classes de 58 amostras de tecido tumoral mamário (colunas) agrupadas de acordo com seu *status* de REr+ ou REr-. A assinatura baseada no perfil de expressão de 95 transcritos (linhas) foi capaz de re-classificar as amostras de tumor mamário (colunas) de acordo com seu *status* de REr com 90% de acurácia, utilizando como medida de similaridade na análise de agrupamento hierárquico a distância Euclidiana.

Barras pretas correspondem aos tumores REr+; barras brancas correspondem aos tumores REr-. Barras verdes e vermelhas, cheias e vazias correspondem a sub-grupos de pacientes agrupados dentro de cada um dos grupos de tumores REr+ e REr-. Na parte inferior da figura está indicado o número de identificação de cada paciente. Transcritos intrônicos foram identificados de acordo com o *RefSeq* do gene codificador em que mapeiam, onde a expressão ativa foi detectada (em azul).

4.9. Classificação dos tumores de acordo com presença de metástase

Além de interrogar o nível de expressão dos transcritos presentes no microarranjo de cDNA entre grupos de tumores com positividade ou negatividade para receptor de estrogênio, foram feitas também análises para investigar o nível de expressão destes transcritos em tecido mamário metastático. Para tal análise, restringimos as análises a um conjunto de tumores de mama do tipo ductal invasivo, livre de recidiva ou metástase durante um período de acompanhamento mínimo de quatro anos.

Para nos assegurarmos ainda mais em relação ao bom prognóstico deste tumores selecionados como livres de recidiva e/ou metástase, foram escolhidos para este grupo tumores com características moleculares sabidamente relacionadas à boa evolução da doença (REr+, RPr+, estágio tumoral I ou IIA, infiltração de nódulos linfáticos negativa) (Ciatto et al., 1990; Shi et al., 1997). Estes critérios limitaram o número de amostras incluídas na análise, determinando um grupo de oito amostras de tumores de mama com bom prognóstico da doença e livres de recidiva/metástase por um período mínimo de quatro anos e um grupo de dez amostras de tumores que desenvolveram metástase (Tabela 9).

Tabela 9. Características anatomo-patológicas e moleculares de 18 amostras de tecido tumoral mamário que constaram da análise do nível de expressão dos transcritos do microarranjo de cDNA em tumores mamários metastáticos.

Pacientes com metástase	Idade	Período de acompanhamento (anos)	Tamanho do tumor (mm)	Grau Histológico	Nódulos linfáticos infiltrados	Estágio tumoral	ER	PR	Recidiva local	Morte
3	57	5	40	III	0	IV	0	0	Sim	Julho/06
7	61	5	35	II	2	IIIB	+++	0	Sim	Setembro/06
15	38	5	28	II	3	IIB	0	0	Sim	Fevereiro/07
17	35	5	18	II	0	I	+++	+++	—	—
22	74	4,5	35	II	0	IIA	0	0	—	Outubro/05
35	71	4	18	III	0	I	0	0	—	Junho/05
38	66	4	40	II	15	IIIC	+++	+++	—	Mai/05
53	73	3	23	II	2	IIIA	+++	+++	—	Abril/05
55	33	3	20	III	0	I	+++	+	—	—
69	53	1,5	—	—	4	IIB	0	0	Sim	—
Pacientes sem metástase	Idade	Período de acompanhamento (anos)	Tamanho do tumor (mm)	Grau Histológico	Nódulos linfáticos infiltrados	Estágio tumoral	ER	PR	Recidiva (local)	Morte
11	66	5	20	II	0	I	++	0	—	—
12	66	5	15	III	0	I	+++	+++	—	—
18	78	4,5	11	I	0	I	+++	+++	—	—
23	38	4,5	18	III	0	I	+++	+++	—	—
27	72	4,5	17	III	0	I	+++	0	—	—
32	81	4	23	I	0	IIA	+++	+++	—	Abril/06
33	69	4	15	II	0	I	+++	+++	—	Setembro/05
39	69	4	12	II	3	IIA	+++	+++	—	—

Em função do número relativamente limitado de amostras de pacientes livres de metástase com seguimento clínico mínimo de 4 anos, delineamos uma estratégia de seleção aleatória de múltiplos grupos de treinamento e validação entre as amostras disponíveis para identificar uma assinatura de expressão gênica robusta correlacionada com a ocorrência de metástase. Esta estratégia foi adaptada de um procedimento descrito para avaliar o efeito da replicação em experimentos de classificação baseados em dados de expressão gênica de microarranjos (Pavlidis et al., 2003).

Sorteios de subconjuntos de igual número de amostras de cada um dos grupos de tumores de mama com e sem metástase foram realizados para definir dez grupos de treinamento e de validação (cada um com nove amostras de tumores de mama: cinco tumores com metástase e quatro livres de metástase). Cada um dos grupos de treinamento gerados (dez no total) foram então utilizados em análises supervisionadas com utilizando a ferramenta para Excel *SAM* a fim de encontrar transcritos expressos

diferencialmente entre grupos de tumores de mama que desenvolveram metástase e permaneceram livres de metástase. Dos dez grupos de treinamento gerados, dois não produziram resultados estatisticamente significativos, utilizando-se uma taxa de falso positivo máxima de 25%. Em média, os oito grupos de treinamento válidos apontaram 590 transcritos expressos diferencialmente entre tumores com e sem metástase de um total de 3.177 transcritos válidos. O conjunto selecionado a partir de cada grupo de treinamento foi utilizado para reclassificar as amostras do respectivo grupo de treinamento, de forma a testar a acurácia do classificador identificado. Quando cada uma destas listas de transcritos foi usada para re-classificar o grupo de validação associado ao de treinamento, a taxa de acerto apresentada em média foi de cerca de 78% (mín. 67% e máx. de 89%).

As listas de transcritos expressos diferencialmente foram então cruzadas para se encontrar um conjunto de transcritos mais robusto para reclassificar as amostras de tumores de mama. O conjunto mínimo capaz de reclassificar as amostras de tumor de mama com maior eficácia era composto de 113 transcritos presentes em seis das oito análises feitas com os grupos de treinamento. Dentre estes transcritos, 70 correspondem a regiões exônicas de genes conhecidos, 32 correspondem a regiões intrônicas e 12 a regiões intergênicas (Tabela 10). Os clones de cDNA representando o conjunto de 113 transcritos da assinatura de metástase foi re-seqüenciado para confirmação da seqüência depositada no microarranjo. Destes, foi possível obter seqüências válidas para 84 clones. Vinte e nove clones não puderam ser confirmados pois não foi possível recuperar colônias congeladas em glicerol. A confirmação destes clones ainda depende do re-seqüenciamento dos produtos de PCR presentes nas bibliotecas de cDNA utilizadas para a construção dos microarranjos. Dos 84 clones seqüenciados, 88% (74 clones) tiveram a identidade confirmada, 5 seqüências foram re-annotadas após o seqüenciamento e 5

seqüências não apresentaram qualidade suficiente para confirmação da identidade (Tabela 10).

Tabela 10. Conjunto de 113 transcritos expressos diferencialmente em tumores de mama que desenvolveram metástase num período de acompanhamento de quatro anos (M+) *versus* tumores de mama sem metástase e/ou recidiva (M-) (ordenados pela ordem alfabética do *locus* gênico). Este conjunto de transcritos foi identificado por *SAM* usando 1.000 permutações e limiar de significância de $FDR \leq 25\%$, em pelo menos seis de dez conjuntos de treinamento selecionados ao acaso entre as amostras disponíveis.

Gene ID	Tamanho do transcrito (pb)	Coordenada genômica	Locus gênico	Mapeamento exon/ intron	Anotação	Log ₂ (M+/M-)	valor q (M+ vs. M-)	Confirmação por seqüenciamento
53	276	chr11:47,218,055-47,218,298	ACP2	5' UTR, Exon # 1	Lysosomal acid phosphatase 2 precursor	-0,78	2,96	a confirmar
91	-	chr12:50,664,055-50,665,321	ACVR1B	Exon # 5,6	Activin A type IB receptor isoform c precursor	-0,97	0,00	re-anotada
8038	434	chr10:128,064,731-128,065,168	ADAM 12	Intron # 22	ADAM metalloproteinase domain 12 isoform 1	-0,99	1,88	confirmada
9411	596	chr1:94,412,669-94,416,113	ARHGAP29	Exon # 1, 2, 3	PTPL1-associated RhoGAP 1	-0,83	1,88	confirmada
8874	393	chr13:110,731,962-110,732,341	ARHGEF7	Intron # 15	Rho guanine nucleotide exchange factor (GEF) 7	-0,64	2,21	a confirmar
444	454	chr8:62,729,562-62,730,016	ASPH	Intron # 17	Aspartate beta-hydroxylase	-0,74	1,88	a confirmar
467	252	chr1:210,860,087-210,860,315	ATF3	3'UTR	Activating transcription factor 3	0,97	7,75	confirmada
79572	670	chr3:195,605,526-195,606,188	ATP13A3	5' UTR	ATPase type 13A3	-0,81	0,00	a confirmar
476	431	chr1:116,725,192-116,725,609	ATP1A1	Intron # 1	Na+/K+ -ATPase alpha 1 subunit isoform a	-1,79	0,00	confirmada
-	403	chr3:43,704,073-43,704,459	AW 749744	Região Intergênica	AW 749744	0,49	8,27	a confirmar
-	361	chrX:110,874,207-110,874,567	AW 902197	Região Intergênica	AW 902197	-0,84	0,00	confirmada
-	480	chr16:68,569,045-68,569,516	AW 935917	Região Intergênica	AW 935917	-1,33	0,00	confirmada
-	425	chr7:76,477,274-76,477,698	AW 937799	Região Intergênica	AW 937799	-0,72	1,88	a confirmar
-	439	chr22:30,441,121-30,441,551	AW 995449	Região Intergênica	AW 995449	-0,86	2,21	confirmada
-	488	chr8:62,577,326-62,577,811	BE 074326	Região Intergênica	BE 074326	-1,05	1,88	confirmada
-	324	chr5:132,191,548-132,191,864	BF 083158	Região Intergênica	BF 083158	-0,75	2,21	confirmada
-	323	chr2:158,256,798-158,257,085	BF 328120	Região Intergênica	BF 328120	-0,88	1,88	confirmada
-	300	chr1:179,294,839-179,295,071	BF 378192	Região Intergênica	BF 378192	-0,74	0,00	a confirmar
-	395	chr14:26,345,171-26,345,553	BF 920515	Região Intergênica	BF 920515	-1,30	0,00	confirmada
-	498	chr22:21,370,791-21,573,497	BU 942768	Região Intergênica	BU 942768	0,36	32,34	confirmada
388815	492	chr21:16,599,401-16,599,877	C21orf34	Intron # 3	Hypothetical protein LOC388815 isoform a	-1,39	0,00	confirmada
221545	443	chr6_qbl_hap2:1,865,625-1,867,447	C6orf136	Exon # 3, 4, 5, 6	Hypothetical protein LOC221545 isoform 2	-0,62	1,88	confirmada
51719	314	chr2:231,386,038-231,386,337	CAB39	Intron # 7	Calcium binding protein 39	0,64	7,75	confirmada
219771	325	chr10:35,898,382-35,898,613	CCNY	3' UTR	Cyclin Y	0,62	12,68	confirmada
22948	621	chr5:10,316,244-10,318,182	CCT5	Exon # 10, 11, 3' UTR	Chaperonin containing TCP1 subunit 5 (epsilon)	0,26	7,75	confirmada

8621	401	chr7:39,994,036-40,003,666	<i>CDC2L5</i>	Exon # 2, 3	Cell division cycle 2-like 5 isoform 2	0,77	8,89	confirmada
1005	293	chr18:61,661,979-61,662,256	<i>CDH7</i>	Exon # 7	Cadherin 7, type 2	-1,03	0,00	confirmada
4850	559	chr7:134,810,054-134,810,603	<i>CNOT4</i>	Intron # 11	CCR4-NOT transcription complex subunit 4	-0,90	1,88	confirmada
22837	425	chr2:165,257,003-165,259,202	<i>COBL1</i>	Exon # 2,3	COBL-like 1	-0,84	1,47	a confirmar
11151	396	chr16:30,105,968-30,106,846	<i>CORO1A</i>	Exon # 5, 6, 7, 8	Coronin, actin binding protein, 1A	1,08	7,75	confirmada
1459	585	chr16:56,749,399-56,757,086	<i>CSNK2A2</i>	5' UTR, Exon # 1, 2	Casein kinase 2 alpha prime polypeptide	0,65	7,75	confirmada
-	375	chr5:178,326,101-178,326,473	<i>CV 381527</i>	Região Intergênica	CV 381527	-1,03	1,47	confirmada
162989	378	chr19:47,409,548-47,409,919	<i>DEDD2</i>	Intron # 2	Death effector domain-containing DNA binding	-1,00	1,47	a confirmar
79665	433	chr17:55,018,294-55,031,573	<i>DHX40</i>	Exon # 11, 12, 13, 14	DEAH (Asp-Glu-Ala-His) box polypeptide 40	-1,16	0,00	a confirmar
9077	390	chr1:68,285,174-68,285,546	<i>DIRAS3</i>	Exon # 1	DIRAS family, GTP-binding RAS-like 3	-1,15	0,00	a confirmar
5611	345	chr13:95,230,043-95,230,370	<i>DNAJC3</i>	Intron # 7	DnaJ (Hsp40) homolog, subfamily C, member 3	0,67	8,89	confirmada
1786	392	chr19:10,107,931-10,109,658	<i>DNMT1</i>	Exon # 4, 5, 6	DNA (cytosine-5)-methyltransferase 1	0,67	8,89	confirmada
1946	452	chr5:106,750,837-106,791,110	<i>EFNA5</i>	Exon # 2, 3	Ephrin-A5	0,90	12,10	a confirmar
9941	503	chr3:38,523,391-38,540,800	<i>ENDOGL1</i>	Exon # 5, 6	Endonuclease G-like 1	-1,29	1,47	confirmada
2079	587	chr14:68,916,871-68,934,772	<i>ERH</i>	5' UTR, Exon # 1, 2, 3, 4	Enhancer of rudimentary homolog (<i>Drosophila</i>)	0,40	12,10	confirmada
10613	307	chr10:101,901,125-101,901,430	<i>ERLIN1</i>	5' UTR	ER lipid raft associated 1	-0,73	1,47	confirmada
2114	332	chr21:39,117,927-39,118,240	<i>ETS2</i>	3' UTR	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)	0,69	7,75	confirmada
259282	478	chr4:13,219,261-13,224,295	<i>FAM44A</i>	Exon # 17, 18, 19, 20	Family with sequence similarity 44, member A	0,48	7,75	confirmada
54751	347	chr1:15,961,753-15,962,100	<i>FBLIM1</i>	Intron # 2	Filamin-binding LIM protein-1 isoform a	-0,96	1,47	a confirmar
157574	301	chr8:28,360,558-28,360,857	<i>FBXO16</i>	Exon # 3	F-box protein 16	-0,88	1,88	confirmada
152273	352	chr3:14,935,326-14,939,067	<i>FGD5</i>	Exon # 10,11,12	FYVE RhoGEF and PH domain containing 5	0,71	8,89	confirmada
201163	433	chr17:17,057,861-17,060,541	<i>FLCN</i>	Exon # 1, 2, 3, 4	Folliculin isoform 1	-0,88	1,88	confirmada
2625	412	chr10:8,155,963-8,156,373	<i>GATA3</i>	Exon # 5, 3' UTR	GATA binding protein 3	-1,80	0,00	confirmada
2674	473	chr10:117,969,525-117,969,988	<i>GFRA1</i>	Intron # 6	GDNF family receptor alpha 1	0,58	8,27	confirmada
54865	457	chr1:154,831,166-154,831,580	<i>GPATCH4</i>	5' UTR	G patch domain containing 4 isoform 2	0,53	8,27	confirmada
10243	387	chr14:66,060,418-66,060,800	<i>GPHN</i>	Intron # 1	Gephyrin	0,28	12,68	confirmada
2821	504	chr19:39,568,710-39,569,214	<i>GPI</i>	Intron # 10	Glucose phosphate isomerase	0,61	7,75	a confirmar
9567	460	chr22:37,438,443-37,438,895	<i>GTPBP1</i>	Intron # 2	GTP binding protein 1	-0,86	1,88	a confirmar
3043	351	chr11:5,203,472-5,204,785	<i>HBB</i>	Exon # 1, 2, 3	Hemoglobin, beta	1,32	7,75	a confirmar
3419	177	chr15:76,249,206-76,249,363	<i>IDH3A</i>	3' UTR	Isocitrate dehydrogenase 3 (NAD+) alpha	-1,12	0,00	a confirmar
221895	300	chr7:28,126,211-28,126,507	<i>JAZF1</i>	Intron # 4	JAZF zinc finger 1	-1,12	0,00	confirmada
23081	622	chr9:6,974,281-6,976,841	<i>JMJD2C</i>	Exon # 9, 10	Jumonji domain containing 2C	-0,92	1,88	confirmada
3728	562	chr17:37,167,512-37,172,980	<i>JUP</i>	Exon # 4, 5, 6, 7	Junction plakoglobin	0,39	9,44	a confirmar
3936	625	chr13:45,599,058-45,599,662	<i>LCP1</i>	5' UTR	Lymphocyte cytosolic protein 1 (L-plastin)	-0,95	0,00	confirmada
114823	540	chr19:59,660,744-59,661,539	<i>LENG8</i>	Exon # 11, 12, 13, 14	Leukocyte receptor cluster (LRC) member 8	0,79	7,75	confirmada
3984	338	chr7:73,161,174-73,168,123	<i>LIMK1</i>	Exon # 9, 10, 11, 12	LIM domain kinase 1	1,41	7,75	a confirmar

11183	486	chr14:49,956,065-49,956,548	MAP4K5	5' UTR, Exon # 1	Mitogen-activated protein kinase kinase kinase	0,41	7,75	confirmada
9782	550	chr5:138,692,961-138,693,510	MATR3	3' UTR	Matrin 3	-1,02	1,88	confirmada
51573	312	chr16:19,427,687-19,427,850	MIR16	Intron # 3	Membrane interacting protein of RGS16	-1,26	0,00	confirmada
55904	634	chr7:104,535,423-104,538,209	MLL5	Exon # 20, 21, 22, 23	Myeloid/lymphoid or mixed-lineage leukemia 5	0,64	7,75	confirmada
4300	611	chr9:20,594,217-20,594,816	MLL3	Intron # 10	Myeloid/lymphoid or mixed-lineage leukemia	-1,42	0,00	confirmada
124540	470	chr17:52,863,641-52,864,102	MSI2	Intron # 3	Musashi 2 isoform a	0,57	8,27	confirmada
9961	399	chr16:29,759,153-29,760,761	MVP	Exon # 7, 8, 9	Major vault protein	0,42	9,44	confirmada
4644	374	chr15:50,394,513-50,394,878	MYO5A	Intron # 2	Myosin VA (heavy polypeptide 12 myoxin)	-2,07	2,21	a confirmar
10529	426	chr10:21,488,992-21,489,361	NEBL	Intron # 4	Nebulette	-0,94	0,00	a confirmar
4775	554	chr16:66,785,703-66,786,254	NFATC3	Intron # 9	Cytoplasmic nuclear factor of activated T-cells	-0,77	1,88	confirmada
11188	533	chr3:52,501,393-52,501,932	NISCH	Exon # 21, 3' UTR	Nischarin	0,68	8,89	confirmada
4893	521	chr1:115,052,107-115,058,049	NRAS	5' UTR, Exon # 1, 2, 3, 4	Neuroblastoma RAS viral (v-ras) oncogene homolog	-1,25	0,00	confirmada
8021	353	chr9:133,063,761-133,064,101	NUP214	Exon # 16	Nucleoporin 214kDa	-0,78	1,47	confirmada
5028	480	chr3:154,037,201-154,037,678	P2RY1	Exon # 1, 3' UTR	Purinergic receptor P2Y1	1,15	7,75	a confirmar
26986	696	chr8:101,788,153-101,794,110	PABPC1	Exon # 5, 6, 7, 8, 9, 10	Poly(A) binding protein cytoplasmic 1	0,87	8,89	a confirmar
445815	373	chr9:111,973,130-111,973,502	PALM2-AKAP2	Exon # 9	PALM2-AKAP2 protein isoform 2	0,54	10,63	confirmada
84105	-	chr5:134,290,725-134,290,880	PCBD2	Intron #2	Pterin-4 alpha-carbinolamine dehydratase 2	0,67	9,44	re-anotada
51105	487	chr8:133,913,706-133,920,865	PHF20L1	Exon # 15, 16, 17, 18	PHD finger protein 20-like 1	-1,19	0,00	confirmada
5256	371	chrX:18,821,585-18,822,445	PHKA2	Exon # 1, 2	Phosphorylase kinase alpha 2 (liver)	0,77	7,75	confirmada
57786	543	chr7:72,000,226-72,000,761	POM121	Intron # 3	Nuclear pore membrane protein 121	-1,10	0,00	confirmada
5527	557	chr14:101,462,525-101,463,078	PPP2R5C	3'UTR	Gamma isoform of regulatory subunit B56 protein	-0,67	4,25	a confirmar
5578	568	chr17:62,070,695-62,071,251	PRKCA	Intron # 4	Protein kinase C, alpha	-1,09	0,00	a confirmar
8073	529	chr1:32,146,931-32,154,162	PTP4A2	5' UTR, Exon # 1, 2, 3, 4	Protein tyrosine phosphatase type IVA member 2	-1,28	0,00	a confirmar
5782	464	chr7:77,103,035-77,107,029	PTPN12	Exon # 15, 16, 17, 18, 3' UTR	Protein tyrosine phosphatase non-receptor type	0,64	7,75	confirmada
51651	418	chr17:55,130,262-55,130,678	PTRH2	Intron # 1	BCL-2 inhibitor of transcription	-0,78	1,88	a confirmar
5879	433	chr7:6,406,282-6,408,632	RAC1	Exon 5, 6, 7	Ras-related C3 botulinum toxin substrate 1	0,81	8,89	confirmada
11030	478	chr8:30,542,657-30,543,120	RBPMS	Intron # 8	RNA-binding protein with multiple splicing	0,67	12,10	confirmada
6018	372	chr1:40,478,381-40,478,741	RLF	Exon # 8, 3' UTR	Rearranged L-myc fusion	-0,96	1,47	a confirmar
6093	407	chr18:16,789,132-16,801,036	ROCK1	Exon # 4, 5, 6	Rho-associated coiled-coil-containing protein kinase 1	0,59	8,89	confirmada
6132	-	chr8:145,985,876-145,986,756	RPL8	Exon # 5,6	Ribosomal Protein L8	1,18	7,75	re-anotada
201965	313	chr4:184,797,960-184,798,263	RWDD4A	5' UTR	RWD domain containing 4A	0,22	17,86	a confirmar
10806	435	chr1:241,537,735-241,560,483	SDCCAG8	Exon # 8, 9, 10	Serologically defined colon cancer antigen 8	0,51	9,44	confirmada
6464	408	chr1:153,202,640-153,203,032	SHC1	5' UTR, Exon # 1	Src homology 1 domain containing	-0,56	3,35	a confirmar
10274	529	chr3:137,598,538-137,599,055	STAG1	Intron # 11	Stromal antigen 1	-1,03	0,00	confirmada
261729	-	chr7:89,702,235-89,702,428	STEAP2	3' UTR	Six-transmembrane epithelial antigen of	0,33	12,10	re-anotada

					prostate 2			
171024	627	chr4:120,175,480-120,176,104	<i>SYNPO2</i>	Intron # 4	Synaptopodin 2	-0,78	1,47	a confirmar
6938	250	chr15:55,366,737-55,366,981	<i>TCF12</i>	Exon # 19	Transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)	0,26	13,52	a confirmar
27004	420	chr14:95,200,888-95,201,305	<i>TCL6</i>	5' UTR, Exon # 1	T-cell leukemia/lymphoma 6 isoform TCL6a1	0,64	8,27	confirmada
7011	198	chr14:19,906,427-19,906,856	<i>TEP1</i>	Exon # 1, 2	Telomerase-associated protein 1	-0,88	1,88	a confirmar
55281	322	chr7:134,496,909-134,497,213	<i>TMEM140</i>	Intron # 1	Transmembrane protein 140	-1,39	0,00	confirmada
55365	471	chr7:150,131,716-150,133,069	<i>TMEM176A</i>	Exon # 4, 5, 6, 3' UTR	Hepatocellular carcinoma-associated antigen 112	0,48	11,39	confirmada
8805	600	chr7:137,908,899-137,914,804	<i>TRIM24</i>	Exon # 12, 13, 14, 15	Tripartite motif-containing 24	0,26	12,68	a confirmar
7251	638	chr11:18,458,519-18,462,147	<i>TSG101</i>	5' UTR, Exon # 1, 2, 3	Tumor susceptibility gene 101	-0,91	0,00	a confirmar
10190	444	chr2:99,302,472-99,305,062	<i>TXNDC9</i>	5' UTR, Exon # 1	ATP binding protein associated with cell	-0,71	1,88	confirmada
8239	401	chrX:40,967,416-40,969,146	<i>USP9X</i>	Exon # 39, 40	Ubiquitin specific protease 9 X-linked isoform	-0,77	1,88	confirmada
54832	-	chr15:59,934,408-59,935,899	<i>VPS13C</i>	Exon # 44,45	Vacuolar protein sorting 13C protein isoform 1A	1,03	8,89	re-anotada
10406	333	chr20:43,531,824-43,542,055	<i>WFDC2</i>	Exon # 1, 2, 3	WAP four-disulfide core domain 2	-0,83	0,00	confirmada
55249	358	chr1:153,922,939-153,923,284	<i>YY1AP1</i>	Intron # 9	YY1 associated protein 1	0,50	8,89	confirmada
7704	385	chr11:113,479,313-113,479,676	<i>ZBTB16</i>	Intron # 2	Zinc finger and BTB domain containing 16	-0,82	3,35	confirmada
56829	551	chr7:138,381,228-138,381,764	<i>ZC3HAV1</i>	5' UTR	Zinc finger antiviral protein isoform 1	0,67	8,27	confirmada
84186	690	chr9:37,106,460-37,304,110	<i>ZCCHC7</i>	Exon # 1, 2, 3	Zinc finger CCHC domain containing 7	-1,41	2,21	confirmada
57178	473	chr10:80,659,218-80,659,691	<i>ZMIZ1</i>	Intron # 4	Retinoic acid induced 17	-0,85	1,88	a confirmar

A assinatura de metástase foi comparada a listas de genes identificadas em trabalhos similares utilizando plataformas de microarranjos comerciais (Desmedt et al., 2007; Radvanyi et al., 2005; Sorlie et al., 2003; van 't Veer et al., 2002), através da ferramenta *Oncomine* (www.oncomine.org). Oito transcritos da assinatura de metástase mapeiam em genes encontrados em outros conjuntos de dados já publicados como diferencialmente expressos em amostras de tumor de mama quando feita a comparação de perfis de expressão gênica de tumores com e sem metástase (Tabela 11). Destes oito transcritos, quatro correspondem a transcritos exônicos (*WFD2C*, *CCT5*, *JMJD2C*, *CORO1A*), três são intrônicos (*NFATC3*, *GFRA1*, *MYO5A*) e um é intergênico (*CXorf45*). Os níveis de expressão dos transcritos da assinatura foram comparados aos encontrados em estudos já publicados (Tabela 11). Dos quatro transcritos exônicos,

três (*WFD2C*, *CCT5*, *CORO1A*) apresentam níveis de expressão em outros estudos coincidentes com o que encontramos em nossas análises. O gene *WFD2C* se encontra com baixa expressão em tumores de pacientes com metástase no estudo de Radvanyi e colaboradores (Radvanyi et al., 2005). *CCT5* apresenta alta expressão em tumores de pacientes com metástase (Desmedt et al., 2007). Sorlie e colaboradores mostraram que *CORO1A* e *JMJD2C* também apresentam alta expressão em tumores com metástase (Sorlie et al., 2003). Este último gene foi o único representado por um transcrito em nossas análises que apresentou nível de expressão inverso em tecido tumoral mamário metastático. No mesmo estudo de Radvanyi e colaboradores (Radvanyi et al., 2005) foi também mostrado que o nível de expressão de *CXorf45* está diminuído em pacientes com tumor de mama metastático e este nível de expressão é coincidente ao que foi apresentado pelo transcrito em nosso microarranjo que representa este gene. Os três genes encontrados diferencialmente expressos em outros estudos que possuem representação neste microarranjo através dos transcritos intrônicos apresentaram expressão invertida em relação ao codificador medido nestes outros estudos (van 't Veer et al., 2002; Desmedt et al., 2007).

Tabela 11. Relação dos transcritos da assinatura de metástase reportados como diferencialmente expressos em tumores de mama, em dados disponíveis na literatura. Os dados foram obtidos utilizando a ferramenta *Oncomine* (www.oncomine.org) e estão ordenados pela ordem alfabética do *locus* gênico.

Locus gênico	Mapeamento exon/intron	Log2 (M+/M-)	Concordância com dados de expressão reportados na literatura	Referência
<i>CCT5</i>	Exon # 10,11, 3'UTR	0.26	igual	(Desmedt et al., 2007)
<i>CORO1A</i>	Exon # 5,6,7,8	1.08	igual	(Sorlie et al., 2003)
<i>CXorf45</i>	Região Intergênica	-0.84	igual	(Radvanyi et al., 2005)
<i>GFRA1</i>	Intron # 6	0.58	oposto	(van 't Veer et al., 2002)
<i>JMJD2C</i>	Exon # 9,10	-0.92	oposto	(Sorlie et al., 2003)
<i>MYO5A</i>	Intron # 2	-2.07	oposto	(Desmedt et al., 2007)
<i>NFATC3</i>	Intron # 9	-0.77	oposto	(van 't Veer et al., 2002)
<i>WFD2C</i>	Exon # 1,2,3	-0.83	igual	(Radvanyi et al., 2005)

Os 32 transcritos intrônicos foram investigados quanto à capacidade de codificar proteínas e nenhum deles apresentou potencial para produzir seqüências protéicas maiores que 100 aminoácidos. Estes transcritos também foram averiguados em relação à capacidade de produzir microRNAs e nenhum deles apresentou alinhamento significativo com qualquer seqüência dos bancos de dados.

Quando analisados utilizando-se a ferramenta BLAT (UCSC), 4 dos 32 transcritos intrônicos apresentaram sítios de ligação para fatores de transcrição na porção 5' de suas seqüências (*MLLT3*, *PRKCA*, *NEBL*, *ZMIZ1*), podendo portanto ser regulados por esses fatores, de forma semelhante a genes codificadores para proteínas (Figura 18).

Além da proximidade com sítios de ligação de fatores de transcrição, observou-se também que o o transcrito intrônico de *MLLT3* mapeia em uma região genômica identificada como sítio de ligação de histonas H3 contendo a lisina 27 trimetilada (H3K27me3) (Ng et al., 2005). (Figura 18, painel C). A modificação epigenética H3K27me3 é um importante regulador da transcrição gênica mediada pela atividade metiltransferase de uma das proteínas do grupo *Polycomb*, a *EZH2*, e está associada ao silenciamento gênico (Tiwari et al., 2008; Yu et al., 2008).

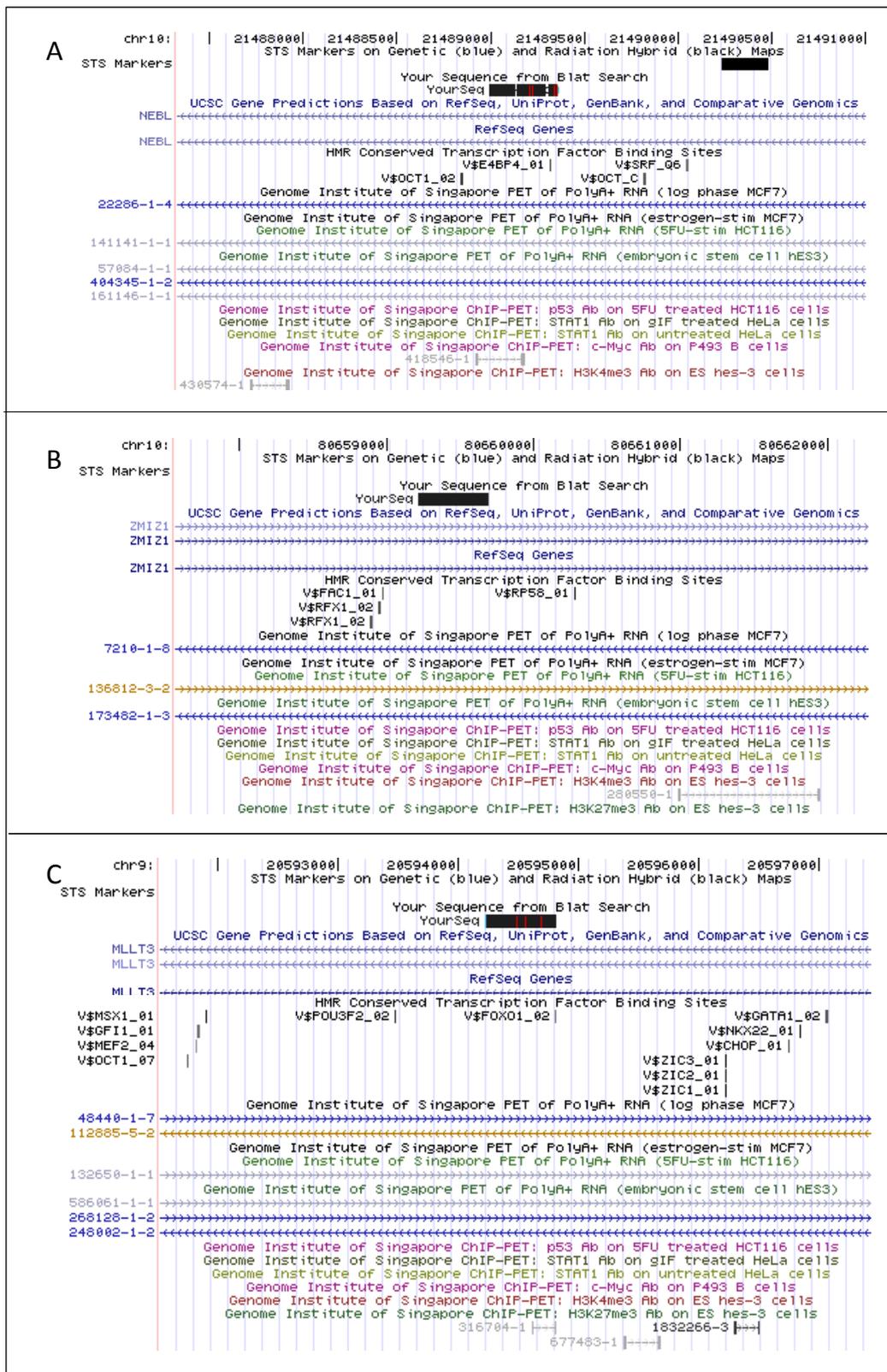


Figura 18. Visualização do mapeamento genômico dos transcritos intrônicos dos genes *MLLT3* (A), *NEBL* (B) e *ZMIZ1* (C). A seqüência parcial dos transcritos foi alinhada no genoma humano através da ferramenta BLAT (genome.ucsc.edu). Observar sítios conservados de ligação para fatores de transcrição na região 5' de cada uma das seqüências dos transcritos, além de sítios de ligação para o fator de transcrição c-Myc (A) e trimetilação da lisina 27 (C) de histonas próximas aos sítios de ligação para fatores de transcrição.

A partir destes 113 transcritos expressos diferencialmente, foi calculada uma assinatura de metástase, que corresponde à média de expressão dos 113 transcritos nas amostras de tumores do grupo metastático. Em seguida, foi calculada a correlação (r) de cada uma das 18 amostras incluídas nesta análise com esta assinatura de metástase para a re-classificação das pacientes (Figura 19). Do conjunto das 18 amostras incluídas nesta análise, somente uma proveniente de uma paciente que desenvolveu metástase (paciente 38) apresentou uma correlação negativa com a assinatura de metástase (Figura 19), tendo sido classificada de forma incorreta. Contudo, quando o limiar para classificação das amostras livres de metástase é ajustado para $r \geq -0,2$, todas as amostras passam a se classificar corretamente: todas as amostras que apresentam uma correlação com a assinatura de metástase maior que $-0,2$ são classificadas como livres de metástase e todas as amostras que apresentam uma correlação com a assinatura de metástase menor que $-0,2$ são classificadas como metastáticas (Figura 19). A determinação do limiar para classificação é arbitrária e uma mesma estratégia já foi utilizada anteriormente para se chegar a um conjunto de transcritos expressos diferencialmente entre grupos de tumores de mama com bom e mau prognóstico da doença (van 't Veer et al., 2002).

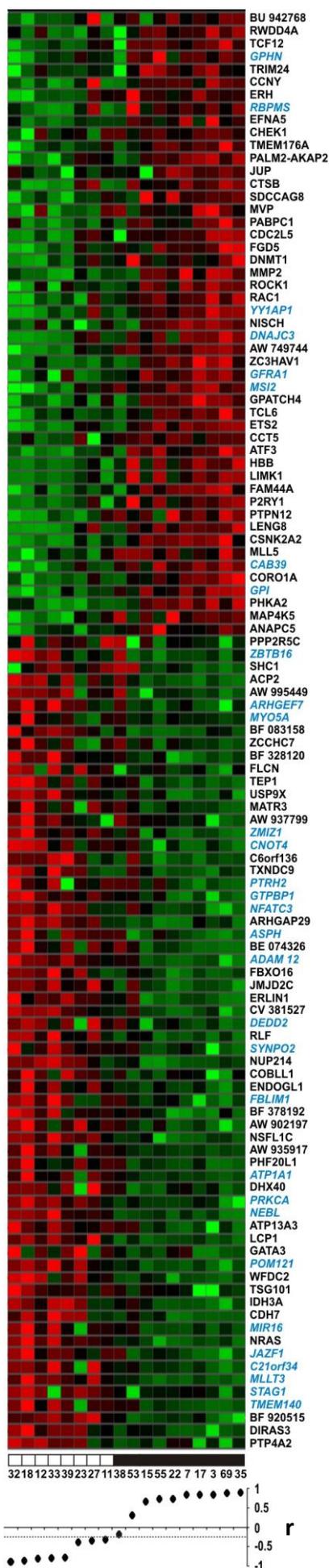


Figura 19. Mapa representativo do nível de expressão de 113 transcritos apontados como diferencialmente expressos (linhas) em 18 amostras de tumor de mama (colunas). Os nomes em azul correspondem a transcritos que mapeiam em regiões intrônicas de genes codificadores (de acordo com o *RefSeq*), onde a expressão ativa foi detectada. Os 113 transcritos foram ordenados de acordo com o valor da correlação entre as medidas de expressão nos grupos com e sem metástase. Na parte inferior da figura estão indicados os números de identificação de cada um dos pacientes.

As amostras (colunas) estão ordenadas de acordo com a correlação “r” de cada uma das amostras com o perfil de expressão dos 113 transcritos nas amostras com metástase. O painel inferior mostra um classificador gerado a partir das correlações de cada amostra com o perfil de mau prognóstico. Pacientes livres de metástase apresentam uma correlação $<-0,2$ com o classificador; pacientes que desenvolveram metástase apresentam uma correlação $>-0,2$ com o classificador. Pacientes livres de metástase, barras brancas; pacientes que desenvolveram metástase, barras pretas.

Com a finalidade de investigar o potencial de ncRNAs na classificação molecular de tumores de mama, o conjunto de 32 transcritos intrônicos que compõem a assinatura de metástase foi analisado independentemente para re-classificar as 18 amostras com e sem metástase. Esta assinatura intrônica foi capaz de re-classificar corretamente 17 das 18 amostras de tumor, a partir da correlação do nível de expressão dos 32 transcritos intrônicos, medidos em cada amostra, com o nível de expressão médio dos 32 transcritos nas amostras com metástase (Figura 20).

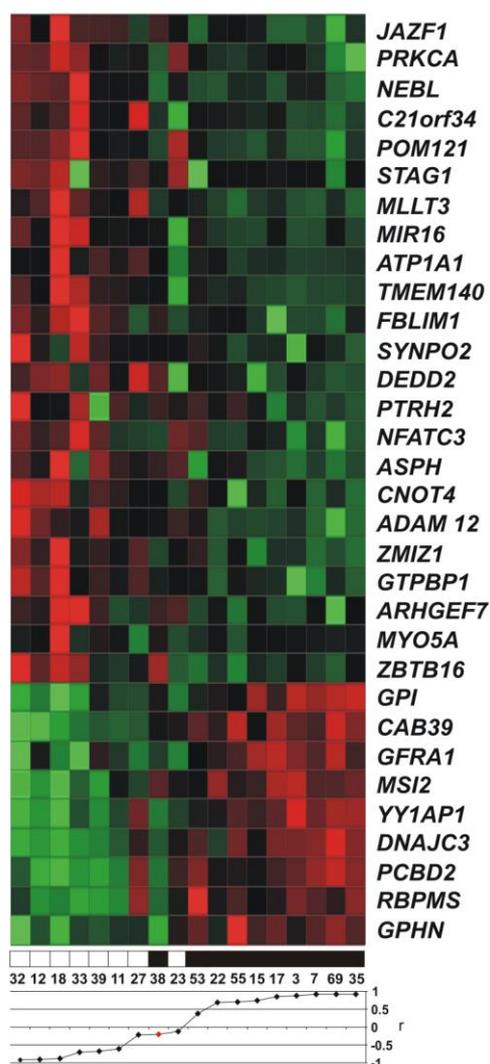


Figura 20. Mapa representativo do nível de expressão de 32 transcritos intrônicos apontados como diferencialmente expressos (linhas) em 18 amostras de tumor de mama (colunas). Os nomes indicados correspondem ao *RefSeq* onde os transcritos intrônicos com expressão ativa detectada foram mapeados. Os 32 transcritos foram ordenados de acordo com o valor da correlação entre as medidas de expressão nos grupos com e sem metástase. Na parte inferior da figura estão indicados os números de identificação de cada um dos pacientes. As amostras (colunas) estão ordenadas de acordo com a correlação “r” de cada uma das amostras com o perfil de expressão dos 32 transcritos nas amostras com metástase.

O painel inferior mostra um classificador gerado a partir das correlações de cada amostra com o perfil de metástase. A marcação vermelha no painel indica uma amostra com metástase que foi agrupada juntamente com as outras amostras sem metástase. Pacientes livres de metástase, barras brancas; pacientes que desenvolveram metástase, barras pretas.

O conjunto de 113 transcritos expressos diferencialmente entre tumores de pacientes com e sem metástase foi cruzado com o conjunto de 95 transcritos expressos diferencialmente entre tumores REr+ e REr- para checar possíveis transcritos em comum entre as duas assinaturas. Um grupo de quinze transcritos foi encontrado em ambas as assinaturas, sendo onze transcritos provenientes de regiões exônicas (*CSNK2A2*, *ATF3*, *SDCCAG8*, *TCF12*, *ZC3HAV1*, *PABPC1*, *PHF20L1*, *FLCN*, *GATA3*, *IDH3*, *PHKA2*) e quatro de regiões intrônicas de genes conhecidos (*PRKCA*, *POM121*, *STAG1*, *TMEM140*). Da mesma forma que os resultados encontrados no cruzamento de assinaturas já descritas na literatura com a assinatura de receptor de estrogênio encontrada, a assinatura de metástase também apresenta baixa sobreposição de genes em relação a outras listas disponíveis na literatura (cerca de 15% do total nas duas listas).

4.10. Caracterização de transcritos presentes na assinatura de metástase

Como o microarranjo de cDNA utilizado para hibridização das amostras de tumor de mama contém transcritos em fita dupla, não nos é possível afirmar de qual fita, senso ou antisenso, é proveniente o nível de expressão detectado. Em trabalho anterior, este microarranjo já foi utilizado para interrogar o nível de expressão de transcritos intrônicos envolvidos com a tumorigênese de câncer de próstata e seis transcritos intrônicos correlacionados com o grau de agressividade de tumores de próstata apresentaram expressão em ambas as fitas senso e antisenso (Reis et al., 2004).

Para investigar a orientação da transcrição em amostras de mama foram feitas reações de transcrição reversa das fitas específicas senso e antisenso de um conjunto de seis transcritos de interesse, selecionados da assinatura correlacionada com a ocorrência de metástase. Dos seis transcritos intrônicos avaliados, quatro apresentaram expressão apenas da fita antisenso (*GFRA1*, *DNAJC3*, *PTRH2*, *CNOT4*), um apresentou expressão

apenas na fita senso (*GPHN*) e um apresentou expressão com orientação senso e antisenso relativamente ao gene codificador no mesmo *locus* (*CAB39*) (Figura 21).

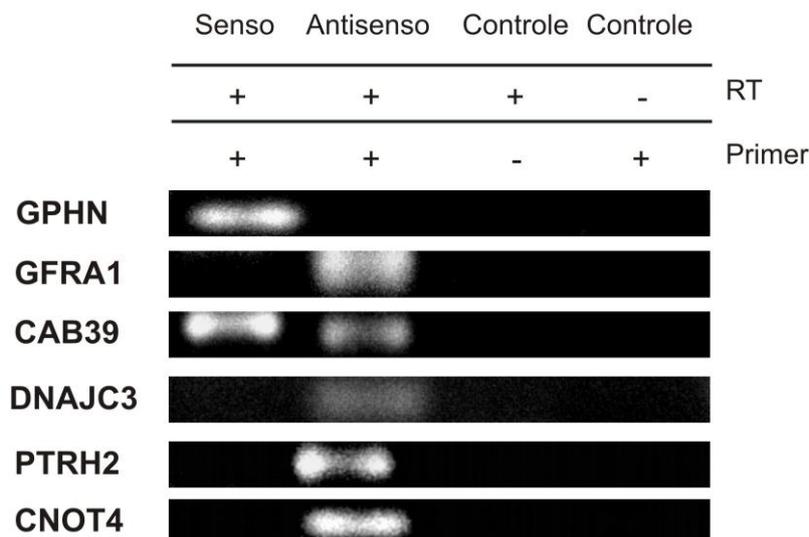


Figura 21. Orientação da transcrição de seis mensagens intrônicas selecionadas do conjunto de 113 transcritos da assinatura de metástase em câncer de mama, determinado por RT-PCR fita específica. Transcrição antisenso (*GFRA1*, *DNAJC3*, *PTRH2* e *CNOT4*), senso (*GPHN*) ou senso e antisenso simultâneas (*CAB39*) foram observadas. Controles para monitorar a ausência de contaminação com DNA genômico (controle - RT, + primer) ou a ausência de RNA proveniente de *self-priming* (controle + RT, - primer) estão indicados. Os nomes dos transcritos correspondem ao RefSeq de mRNAs codificadores para proteínas, onde os transcritos mapeiam na região intrônica, em que foi detectada transcrição ativa.

Alguns destes transcritos (*GFRA1*, *CAB39*, *DNAJC3* e *CNOT4*) foram avaliados por PCR em tempo real para verificar a alteração nos níveis de expressão em tumores livres de metástase relativamente a tumores que apresentaram metástase. No total, oito transcritos tiveram seu nível de expressão avaliado por qPCR (seis intrônicos: *CAB39*, *DNAJC3*, *GFRA1*, *ZBTB16* e *CNOT4*; dois exônicos: *GATA3* e *JMJD2C*). Nos experimentos de PCR em tempo real, *CAB39*, *DNAJC3*, *GFRA1* e *JMJD2C* apresentaram expressão aumentada e *ZBTB16*, *MLLT3*, *CNOT4* e *GATA3* apresentaram expressão diminuída nas pacientes com metástase (Figura 22). De um modo geral, observou-se uma tendência de confirmação da expressão medida nos experimentos de microarranjo (Figura 19 e Tabela 10). Apenas um dos transcritos avaliados (*JMJD2C*)

apresentou níveis de expressão entre os grupos de tumor de mama com e sem metástase discordantes dos medidos nos microarranjos: a maior parte das pacientes com metástase que tiveram expressão avaliada através de PCR em tempo real apresentou alta expressão de *JMJD2C* quando comparada ao grupo de pacientes livre de metástase (Figura 22). Aplicando um critério estatístico para avaliar a expressão diferencial entre grupos de pacientes com e sem metástase, do grupo de oito transcritos avaliados por reação de PCR em tempo real apenas dois transcritos apresentaram valores de significância com $p < 0,1$ em um teste t de Student: *DNAJC3* ($p < 0,1$) e *GATA3* ($p < 0,09$). A comparação dos níveis de expressão entre os transcritos que tiveram a expressão avaliada tanto por microarranjo e PCR em tempo real estão apresentadas na tabela 12 abaixo.

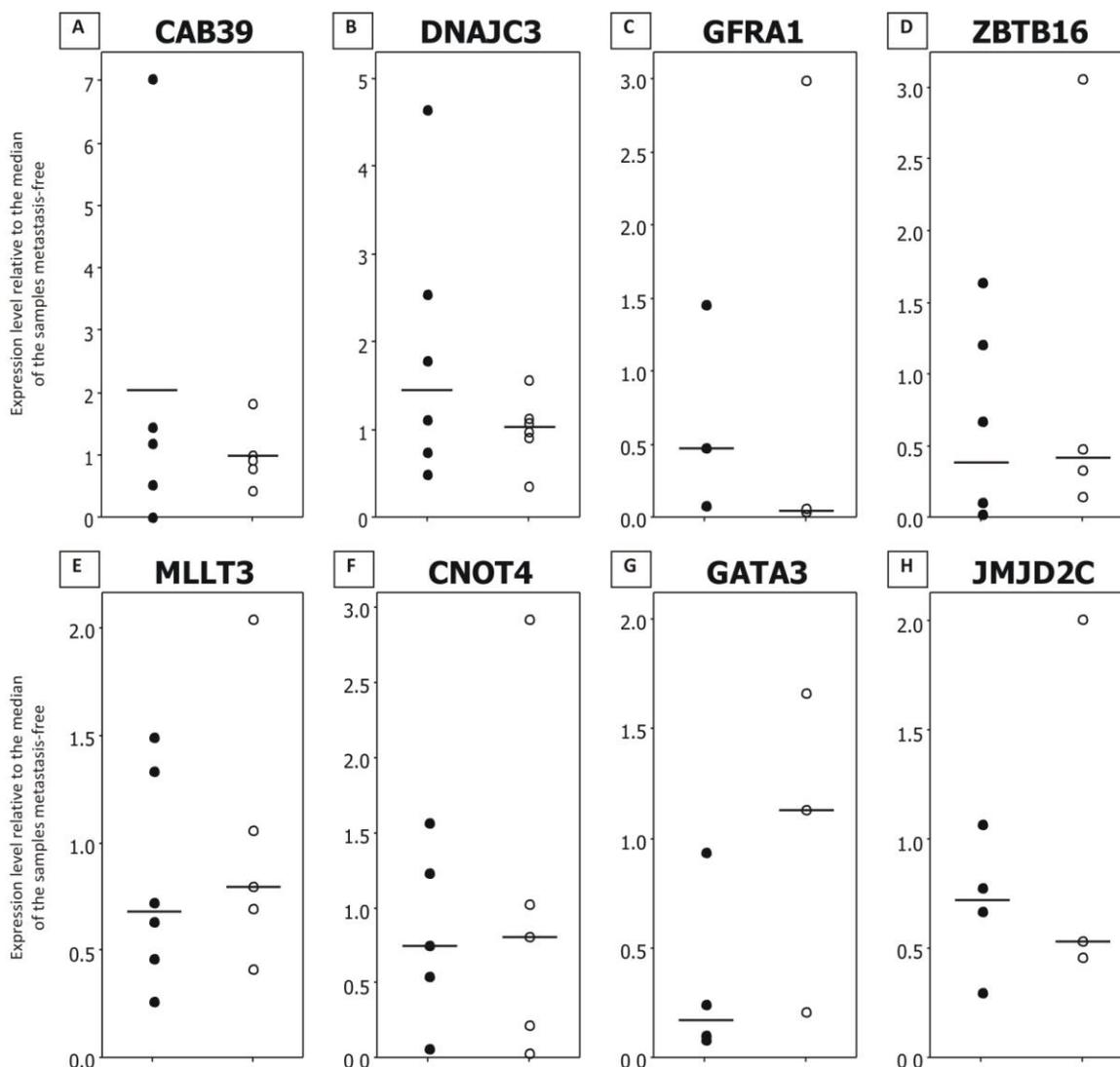


Figura 22. Níveis relativos de expressão de seis transcritos intrônicos (*CAB39*, *DNAJC3*, *GFRA1*, *ZBTB16*, *MLLT3* e *CNOT4*) e dois transcritos exônicos (*GATA3* e *JMJD2C*) em tumores de mama. Os níveis relativos dos transcritos foram medidos em relação à mediana das expressões obtidas das amostras livres de metástase. Cada transcrito teve sua expressão medida em um número de seis a doze pacientes (círculos cheios representam pacientes que desenvolveram metástase; círculos vazios representam pacientes livres de metástase). A diferença de expressão entre os dois grupos de amostras de tumor de mama é concordante em relação ao obtido por microarranjo, com exceção de *JMJD2C*, que apresentou expressão invertida à detectada no microarranjo. Para cada gene, os experimentos foram feitos em triplicata. A barra marca o valor da mediana de cada um dos grupos de dados. Para cada transcrito testado, foi utilizado o gene *HPRT1* (*Hypoxanthine phosphoribosyltransferase 1*) como referência para normalização entre as amostras de pacientes.

Tabela 12. Comparação das razões do nível de expressão de transcritos avaliados através de PCR em tempo real e microarranjo, entre pacientes com e sem metástase. *JMJD2C* apresentou expressão discordante nas duas metodologias.

Símbolo gênico	Log M+/M-		Nº amostras testadas no qPCR	
	Microarranjo	PCR em tempo real	Livre de metástase	Presença de metástase
<i>CAB39</i>	0.64	0.30	5	5
<i>CNOT4</i>	-0.90	-0.11	5	5
<i>DNAJC3</i>	0.67	0.50	6	6
<i>GATA3</i>	-1.80	-2.77	3	4
<i>GFRA1</i>	0.58	5.37	3	3
<i>JMJD2C</i>	-0.92	0.44	3	4
<i>MLLT3</i>	-0.23	-1.42	5	6
<i>ZBTB16</i>	-0.06	-0.82	4	5

5. DISCUSSÃO

5.1. Hibridização das amostras de tecido tumoral mamário

As amostras de tecido tumoral mamário que foram objeto de estudo deste projeto foram enviadas de Portugal ao Brasil, onde foram processadas e analisadas experimentalmente. À princípio, esta distância foi um obstáculo que tivemos que contornar através da padronização de formas diferentes de envio deste material, de maneira que o RNA fosse preservado ao máximo. A melhor forma encontrada foi o envio do RNA ainda sob a forma de tecido, conservado em *RNA Later*. No Brasil, em nosso laboratório, estas amostras foram processadas e analisadas quanto à qualidade e quantidade obtidas de RNA total.

Um dos métodos de avaliação da qualidade do RNA total obtido é a quantificação da razão dos valores de espectrofotometria medidos a 260 e 280nm. Uma razão próxima de 2 indica uma ótima eficiência de extração de RNA já que há pouca contaminação com proteínas. Das 74 amostras de RNA total obtidas, 80% apresentaram razão 260/280 entre 1,8 e 2,0 indicando que nosso protocolo de extração de RNA é eficaz para obtenção de RNA livre de proteínas.

Uma forma de se avaliar a degradação do RNA total é através da razão de intensidade das bandas de RNA ribossomal 28S e 18S. Novamente, o valor referência para esta razão é 2, que indica uma preparação de RNA com ausência de degradação. No entanto, amostras clínicas de fragmentos de tumores têm qualidade variável devido à presença de necrose e manipulação durante o procedimento cirúrgico, que frequentemente resultam na obtenção de RNA de pior qualidade. De fato, do conjunto total de 74 amostras de RNA total obtidas, a maior parte (90%) apresentou razão 28S/18S inferior a 1,3.

Nossa experiência no laboratório mostra que dados de expressão gênica obtidos com amostras clínicas com valores acima de 0,5 resultam em medidas reprodutíveis. Por essa razão optamos por hibridizar todas as amostras de RNA com razão 28S/18S superior a 0,5. A análise dos dados de expressão obtidos indicou que o número médio de transcritos detectados nas hibridizações com RNAs apresentando razão 28S/18S entre 0,5 e 1,3 são comparáveis aos detectados nas hibridizações com amostras apresentando razão 28S/18S maior que 1,4. Além disso, as correlações intra-lâminas de amostras com razão 28S/18S entre 0,5 e 1,3 apresentaram valores semelhantes às lâminas hibridizadas com amostras com razões maiores que 1,4.

Assim, utilizando as 74 amostras de tecido tumoral mamário, foi possível detectar 3.221 cDNAs com expressão válida em 46 pacientes e 265 em 58 pacientes, incluindo 56 transcritos intrônicos. A qualidade aceitável dos RNAs utilizados também foi constatada ao se avaliar as correlações do conjunto de 148 microarranjos obtidos a partir da hibridização destas amostras de tumor de mama: somente 4 lâminas não atingiram os critérios mínimos exigidos de qualidade (correlação inter-lâminas $\geq 0,8$ e correlação intra-lâminas $\geq 0,9$) e estes experimentos foram novamente realizados para reposição.

Quanto à quantidade de RNA total obtida, é sabido que amostras derivadas de remoção cirúrgica têm massa muito limitada e que esse é um problema enfrentado por muitos grupos no mundo (Gomes et al., 2003; Stoyanova et al., 2004). As amostras de tecido disponíveis para o projeto também eram limitadas e de um total de 74 amostras processadas, apenas cerca de 50 resultaram em uma quantidade de RNA total suficiente para os experimentos de hibridização (mínimo de 10 μ g). Com a aplicação de um método de amplificação linear do RNA mensageiro (Wang et al., 2000; Kaposi-Novak et al.,

2004), foi possível analisar o perfil de transcrição de todas as amostras disponíveis, viabilizando a execução do projeto.

5.2. Expressão de ncRNAs intrônicos em tumores de mama

Neste estudo observamos níveis de expressão ubíquos de ncRNAs intrônicos e intergênicos em amostras de tecido tumoral de mama, representando tipos histológicos distintos. Resultado semelhantes já haviam sido observados por nosso grupo utilizando tecidos tumorais de próstata (Reis et al., 2004) e rim (Brito et al., 2008).

Devido ao fato dos tumores de mama se desenvolverem de tecido epitelial de dutos e lóbulos mamários, a remoção cirúrgica de tecido não tumoral de mesma etiologia do tumor é muito difícil, impossibilitando nosso acesso a estas amostras. Dessa forma, esta limitação nos impossibilitou analisar o nível de expressão destas moléculas no tecido mamário normal. No entanto, a observação da expressão de ncRNAs em margens cirúrgicas histologicamente normais de próstata (Reis et al., 2004) e rim (Brito et al., 2008), sugere que a expressão de ncRNAs em tecido mamário não esteja associada somente ao tecido neoplásico. Serão necessários estudos adicionais comparando a expressão gênica em amostras de tecido normal mamário e adenocarcinoma da mama para avaliar alterações no padrão de expressão de ncRNAs associadas à transformação maligna.

Foi observado que em média, a abundância relativa dos ncRNAs intrônicos presentes no microarranjo é comparável a dos transcritos codificadores para proteínas. Resultados obtidos anteriormente por nosso grupo durante a análise de tumores de próstata e rim utilizando essa mesma plataforma de microarranjos de cDNA indicaram que a expressão de ncRNAs intrônicos é em média 20-40% inferior a de transcritos codificadores para proteína (Reis et al., 2004; Brito et al., 2008). A expressão em menor

nível de ncRNAs já foi também descrita em outros trabalhos da literatura (Kampa et al., 2004; Sasaki et al., 2007). Uma diferença no presente trabalho é que foi feita a amplificação linear do RNA mensageiro, posteriormente usado para a marcação de alvos fluorescentes. É possível que a estratégia de amplificação do mRNA empregada neste projeto tenha contribuído para aumentar a sensibilidade de detecção dos ncRNAs intrônicos, de forma que as diferenças de intensidade em relação aos transcritos codificadores sejam menos evidentes.

5.3. Avaliação de uma metodologia para obtenção de medidas de expressão gênica utilizando um oligonucleotídeo referência (*RefOligo*)

As amostras de tecido tumoral mamário enviadas pelo HDES não eram pareadas, ou seja, não possuíam o tecido mamário não-tumoral adjacente correspondente. Este aspecto não permitiu um desenho experimental onde cada amostra tumoral fosse comparada ao tecido normal correspondente através da co-hibridização no microarranjo. Uma alternativa seria a montagem de um *pool* de RNAs a ser utilizado como referência nos experimentos de hibridização competitiva com 2 cores, contra o qual todas as amostras tumorais seriam comparadas. A estratégia de utilização de um RNA referência apresenta dificuldades em projetos prospectivos onde as amostras experimentais são recolhidas ao longo do tempo. Nestes casos, não é fácil assegurar a disponibilidade de RNA referência em qualidade e quantidade suficientes durante toda a duração do projeto. Como alternativa, optamos testar um desenho experimental baseado na utilização de um oligonucleotídeo como referência externa aos experimentos de hibridização (Dudley et al., 2002).

Dados resultantes de três metodologias foram testados e comparados com a acurácia e eficiência de razões obtidas através de experimentos de hibridização diretos,

em que dois tecidos que se quer comparar são hibridizados em uma mesma lâmina. As três metodologias a serem testadas, incluíam: razões construídas a partir do uso de uma referência comum construída a partir de um conjunto de amostras de RNA, hibridizadas juntamente com o RNA do tecido de interesse à lâmina (*RefPool*); razões construídas a partir do uso de um oligonucleotídeo complementar a todos os pontos de cDNA no microarranjo como uma referência comum (*RefOligo*); razões construídas a partir de hibridizações distintas de dois microarranjos de dois tecidos que se quer comparar e que foram hibridizados, cada um deles, contra o *RefOligo* (*One-Color*).

Os resultados obtidos da análise dos conjuntos de dados provenientes das diferentes comparações dos 3 tecidos testados mostraram que o uso de hibridizações diretas produzem resultados mais acurados (menor variância dos dados) e eficientes em relação a indicar conjuntos de transcritos expressos diferencialmente. As razões obtidas através do uso do *RefOligo* e de *One-Color* apresentaram resultados comparáveis aos obtidos com o uso do *RefPool*. No entanto, o uso de uma metodologia como o *RefPool* traz dificuldades operacionais já que são necessárias grandes quantidades de amostras de RNA para a confecção de um *pool* com quantidade suficiente para todas as hibridizações. A coleção prospectiva de amostras em relação ao início dos experimentos de hibridização, como no nosso caso, traria uma dificuldade adicional, já que a quantidade de amostras no início do projeto era pequena, insuficiente para a montagem de um *pool* que seria utilizado em muitos outros experimentos de hibridização.

Em função dos resultados obtidos na etapa inicial do trabalho (Peixoto et al., 2006), optamos por um desenho experimental onde cada amostra de RNA tumoral de mama foi co-hibridizada com a mesma quantidade de *RefOligo*. Como o *RefOligo* é complementar a uma região do vetor presente na sonda, durante a hibridização não há competição entre o *RefOligo* e os alvos gerados a partir do RNA mensageiro. Este

aspecto confere flexibilidade durante a análise dos dados, pois também permite a obtenção de medidas de expressão gênica diretas (*One-Color*) ou indiretas, utilizando razões de expressão normalizadas pelo *RefOligo*.

Quando as duas metodologias, *RefOligo* e *One-Color*, foram comparadas utilizando-se as amostras de RNA tumoral de mama, 60% das razões obtidas a partir do *RefOligo* apresentavam correlação acima da mínima estabelecida para considerar o experimento como de boa qualidade (0,9 intra-lâmina e 0,8 inter-lâminas) enquanto 95% das medidas obtidas a partir de *One-Color* ultrapassaram estes mesmos limiares. Observou-se também uma maior reprodutibilidade das medidas a partir de uma cor em relação às razões normalizadas por *RefOligo*, medida pela correlação média do conjunto das lâminas hibridizadas.

O uso das medidas do *RefOligo* para normalização das intensidades implica em agregar a variabilidade de 2 medidas (RNA e *RefOligo*) na razão de expressão final. A menor correlação observada nas medidas com *RefOligo* obtidas com tumores de mama indica que em experimentos envolvendo um grande número de amostras, processadas ao longo do tempo, o uso do *RefOligo* introduz um fator de ruído nas medidas de expressão que é preponderante em relação aos benefícios obtidos com a correção local propiciada por este normalizador.

Outra evidência de que as medidas diretas de intensidade eram mais confiáveis foi obtida empiricamente, utilizando ambos conjuntos de dados (intensidades – *One-Color* e razões normalizadas por *RefOligo*) para identificar assinaturas correlacionadas com características clínicas associadas às amostras de mama. A metodologia de *One-Color* indicou 320 transcritos expressos diferencialmente entre pacientes com tumores de mama com boa *versus* má evolução da doença, com uma taxa de falso-positivo de 0,3%, enquanto a metodologia de *RefOligo* indicou apenas 158 transcritos com uma taxa

de falso-positivo igual a 10%. Dessa forma, apesar de ambas as metodologias apresentarem eficiência e acurácia muito semelhantes quando avaliadas em um conjunto restrito de amostras (Peixoto et al., 2006) a análise dos dados de um número maior de amostras clínicas, obtidos ao longo de um intervalo de tempo longo durante o qual foi executado o projeto, revelou que as medidas diretas baseadas em uma cor (*One-Color*) se mostraram mais precisas.

Em nossa análise dos perfis de tumores de mama, a metodologia baseada nas intensidades diretas (*One-Color*) se mostrou mais reprodutível e sensível para a identificação de assinaturas de expressão gênica quando comparada à abordagem de normalização por um oligonucleotídeo referência (*RefOligo*). Este resultado está de acordo com evidências da literatura de que classificadores gênicos baseados em medidas de intensidade podem apresentar uma maior acurácia em relação a classificadores baseados em razões normalizadas para uma referência (Attoor et al., 2004).

5.4. Análises não supervisionadas e supervisionadas utilizando características clínicas e anatomo-patológicas das amostras de câncer de mama

A análise não supervisionada dos dados de expressão gênica das amostras tumorais de mama mostrou que o *status* do receptor de estrogênio é uma característica clínica associada a perfis de expressão gênica muito distintos. A obtenção de perfis de expressão correlacionados com o *status* de estrogênio já foi reportada na literatura (Perou et al., 2000; Gruvberger et al., 2001; Ozdag et al., 2006) e esta característica foi a única em que se observou o agrupamento de amostras a partir dos perfis de expressão gênica na análise não-supervisionada.

Além de análises não supervisionadas, foram feitas análises supervisionadas buscando identificar assinaturas de expressão gênica correlacionadas a características

anatomo-patológicas ou clínicas associadas às amostras de tumor, tais como idade, tamanho do tumor, componente hereditário, estágio tumoral, grau histológico, infiltração dos nódulos tumorais, *status* de expressão de REr e RPr e presença/ausência de metástase.

A presença de tumores em mulheres entre 30 e 35 anos de idade ou ainda a presença de tumores maiores que 2cm em pacientes com câncer de mama, são indicativos do potencial agressivo da doença. No entanto, uma análise supervisionada agrupando as amostras de tumor em função da idade e tamanho do tumor, não revelou perfis de expressão correlacionados a estas características. Este resultado sugere que apesar destas informações serem fatores de prognóstico utilizados na clínica médica, elas não são tão informativas dos mecanismos moleculares envolvidos com as alterações de expressão gênica associadas à doença do câncer. Uma explicação alternativa é que as sondas representadas na plataforma de microarranjo utilizada não é capaz de capturar diferenças transcricionais existentes entre estes grupos de amostras.

Apesar da infiltração dos nódulos linfáticos ser uma informação relevante no momento do diagnóstico de tratamentos contra o câncer de mama e prognóstico da doença, esta não apresentou perfis de expressão correlacionados aos diversos graus de infiltração tumoral apresentados pelas pacientes. Este resultado também pode ser explicado pela ausência de uma assinatura molecular associada a esta característica clínica.

Outro aspecto a ser considerado é uma possível falta de homogeneidade na formação dos grupos de amostras para a análise de infiltração linfática. Não houve uma avaliação sistematizada do mesmo número de nódulos linfáticos em cada uma das pacientes incluídas no estudo. Portanto, não se tinha a informação do número de nódulos infiltrados por nódulo analisado para cada paciente, dificultando a criação de grupos de

pacientes homogêneos para a realização da análise. A nossa experiência em estudos de expressão gênica em amostras clínicas mostra que a dificuldade de sistematização da avaliação dos nódulos linfáticos axilares das pacientes não é restrita a este conjunto de dados proveniente do HDES, mas também está presente em outros projetos que utilizam conjuntos de amostras de hospitais diversos no Brasil.

As análises supervisionadas utilizando as informações sobre estágio tumoral de cada paciente também não produziram assinaturas de expressão gênica significativamente correlacionadas a estes estágios, refletindo a heterogeneidade das amostras de mama com mesma classificação patológica e as limitações de utilização do sistema de classificação TNM como fator de prognóstico em câncer de mama. Em especial, os estágios intermediários da classificação TNM consistem desafios ainda maiores no momento de utilizar esta classificação para a decisão sobre tratamentos a serem utilizados nas pacientes em questão, confirmando a necessidade de sistemas de classificação moleculares mais acurados e eficientes, capazes de estratificar melhor o risco das pacientes com câncer de mama.

A análise supervisionada utilizando as informações sobre o *status* do receptor de progesterona das amostras de tumor de mama resultou em um perfil de expressão gênica, capaz de agrupar as amostras em função do *status* RPr+ e RPr-. No entanto, o nível de acerto era mais baixo (78%) quando comparada à análise do *status* do receptor de estrogênio (85%), e este pareceu ser um fator importante que afetava o agrupamento das amostras quando utilizada a assinatura de expressão gênica correlacionada com o *status* do receptor de progesterona. Este resultado indica que o *status* de expressão do receptor de estrogênio é uma característica que tem uma influência maior que o *status* de expressão do receptor de progesterona na determinação do perfil transcricional do tecido tumoral de mama, resultando em grupos mais homogêneos. Esta observação está

de acordo com a descrição da literatura de que o *status* do receptor de estrogênio é um fator de prognóstico/predição mais robusto que o de progesterona (Kasami et al., 2008).

5.5. Classificação dos tumores de acordo com *status* do receptor de estrogênio

As análises iniciais realizadas com 62 amostras de mama, antes que houvessem sido coletadas e processadas a totalidade de amostras utilizadas no projeto, revelaram que o *status* do receptor de estrogênio é a característica molecular capaz de estratificar as amostras de modo mais robusto para a identificação de uma assinatura de expressão gênica. Este resultado é possivelmente explicado pela importância da sinalização através do receptor de estrogênio na definição do padrão transcricional do tecido mamário (Gruvberger et al., 2001).

Ao final do projeto foram processadas e gerados dados de expressão gênica de 74 amostras de adenocarcinoma de mama. A maior parte das amostras (78%) pertencem ao tipo histológico ductal invasivo, sendo as restantes 16 amostras representativas de quatro outros tipos histológicos de menor prevalência (lobular, mucinoso, medular e papilar), além de 3 amostras mal-diferenciadas. O número de amostras disponíveis dos tipos histológicos menos representados são insuficientes para permitir análises visando identificar sub-grupos clínicos. Para aumentar a homogeneidade do conjunto amostral, as análises subseqüentes concentraram-se nas 58 amostras de adenocarcinoma de mama com tipo histológico tipo ductal invasivo.

Uma vez que o *status* de expressão do receptor de estrogênio é um fator extensivamente utilizado para o prognóstico molecular do câncer de mama, investigamos em maior detalhe esta característica no conjunto de 58 amostras de tumores de mama do tipo ductal invasivo. Utilizando uma abordagem supervisionada, foi possível identificar uma assinatura de expressão gênica capaz de distinguir tumores

em função do *status* de expressão do receptor de estrogênio. O grupo de 95 transcritos identificados como expressos diferencialmente entre tumores de mama REr+ vs. REr- foi capaz de distinguir estes dois grupos de tumores com uma taxa de acerto de 90%. Das seis amostras de tumores REr+ que foram classificadas como REr-, quatro apresentaram má evolução da doença: duas pacientes (P.17 e P.7) apresentaram metástase, uma (P.19) apresentou recidiva da doença e uma (P.13) faleceu. Interessantemente, estas amostras foram agrupadas em um sub-grupo dentro do grupo de tumores REr- que contém cinco das oito pacientes presentes nesta análise que desenvolveram metástase durante os quatro anos de acompanhamento desta doença (Figura 17 – grupo indicado pela barra vermelha sólida). Este sub-agrupamento dentro do grupo de tumores REr- possui um perfil de expressão ligeiramente distinto dos outros tumores REr- deste conjunto (Figura 17 – grupos indicados pelas barras vermelhas vazia e sólida), podendo indicar um caráter mais agressivo destes tumores. Neste caso, o perfil de expressão associado à agressividade dos tumores foi mais relevante para agrupar as amostras REr+ com má evolução da doença que o perfil de expressão associado ao *status* do receptor de estrogênio. Existe uma alta correlação entre tumores REr- e a ocorrência de metástase e o tempo de sobrevivência de pacientes com tumores de mama (Lower et al., 2005). É possível que a perda do caráter responsivo do receptor de estrogênio nestas amostras de tumores de mama reflita um evento mais tardio na tumorigênese, decorrente de mudanças moleculares durante a evolução do câncer.

Estes sub-agrupamentos existentes dentro dos grupos de tumores REr+ e REr-, baseados em um perfil de expressão distinto, podem estar refletindo os diferentes sub-grupos moleculares definidos por Sorlie e colaboradores (Sorlie et al., 2001). Apesar do nosso microarranjo não possuir representantes dos genes utilizados para a classificação dos seis sub-grupos moleculares (tumores REr+: epitelial luminal A, B e C, tumores

RER-: mama normal, basal e ERBB2+), podemos especular que o conjunto de transcritos representados em nosso microarranjo seja capaz de detectar as diferenças fenotípicas celulares associadas a cada um destes sub-tipos moleculares, explicando o padrão de agrupamento dentro dos conjuntos de amostras de tecido tumoral RER+ e RER- de nosso conjunto amostral.

As outras duas pacientes com tumores RER+ que foram agrupadas com o grupo de tumores RER- (P.25 e P.1) não possuem atualmente evidência da doença de câncer de mama de acordo com o acompanhamento clínico destas, isto é, são pacientes que apresentaram boa evolução da doença. Em geral, a concordância dos resultados de imunohistoquímica em relação ao *status* do receptor de estrogênio entre laboratórios é alta, cerca de 90% de acordo com Badve e colaboradores (Badve et al., 2008). No entanto, apesar de erros do tipo falso-positivo (tumores RER- que são classificados como RER+ de acordo com teste de imunohistoquímica) serem menos comuns que os do tipo falso-negativo, eles também existem. Assim, podemos levantar a hipótese de que estas duas pacientes classificadas de acordo com a assinatura do *status* de estrogênio como RER- (P.25 e P.1), foram erroneamente identificadas como RER+ através de testes de imunohistoquímica.

A assinatura do *status* do receptor de estrogênio foi comparada a outras assinaturas de expressão gênica correlacionadas com o *status* deste receptor disponíveis na literatura. Esta comparação revelou que 41 transcritos da assinatura apresentada neste estudo mapeiam em genes encontrados como diferencialmente expressos entre amostras de câncer de mama RER+ e RER-, de acordo com outros estudos já publicados. *GATA3*, por exemplo, é um gene integral da via do receptor de estrogênio, apresentando co-expressão com este receptor e também com a proteína de *CCND1* (Wilson e Giguere, 2008). *GATA3* é um gene importante como marcador de prognóstico em câncer de

mama e sua baixa expressão está associada a tumores de mama REr- (Mehra et al., 2005). Ambos os transcritos exônicos que representam os genes *GATA3* e *CCND1* no nosso microarranjo foram também detectados com alta expressão nas amostras REr+ quando comparados com REr-, resultado semelhante encontrado em outros estudos de expressão gênica, realizando a mesma comparação entre tecidos tumorais mamários. Além disso, o alto nível de expressão de *GATA3* em nosso estudo foi também confirmado através de PCR em tempo real em amostras de tumores de mama REr+ comparado com amostras de tumores REr-. Portanto, os dados da literatura corroboram nossos achados, confirmando nossas análises e a assinatura correlacionada ao *status* do receptor de estrogênio encontrada.

O gene *MMP9* foi identificado em nossa análise com alta expressão em tumores de mama REr- em relação a REr+. Este resultado é coerente com o encontrado na literatura, já que tumores REr- são sabidamente mais agressivos que tumores de mama REr+ e o gene *MMP9* está envolvido com o caráter invasivo das células mamárias tumorais (Rizki et al., 2008). Da mesma forma que as informações da literatura sobre os genes *CCND1* e *GATA3* validam nossas análises de microarranjos, a informação sobre a expressão do gene *MMP9* em tecido tumoral mamário encontrada na literatura também corrobora os dados apresentados neste projeto.

Dentre os 95 transcritos da assinatura do *status* do receptor de estrogênio, 14 são provenientes de regiões intrônicas de genes conhecidos. Quando estes 14 transcritos intrônicos são excluídos do grupo e os dados são re-analisados por agrupamento hierárquico, não se observa mais o agrupamento das amostras em dois grandes grupos de tumores REr+ e REr- com expressão distinta. Dessa forma, podemos concluir que os transcritos intrônicos contribuem para uma melhor classificação das amostras em função de características fenotípicas dos tumores.

Muitos trabalhos vêm apontando para a atividade regulatória de moléculas de RNA não-codificadoras derivadas de introns regulando o gênero de plantas (Sugiyama et al., 2003; Dai et al., 2007; Hore et al., 2007), a estabilidade de moléculas de RNA em camundongos (Hirotsume et al., 2003) e o *imprinting* e compensação de dose do cromossomo X em humanos (Hore et al., 2007; Prasanth e Spector, 2007; Wutz, 2007). Nosso grupo também já identificou um conjunto de moléculas de longos RNAs não-codificadores com expressão altamente correlacionada com o grau de agressividade de tumores de próstata (Reis et al., 2004) e responsivo ao hormônio andrógeno (Louro et al., 2007). Esta é a primeira vez que um conjunto de moléculas de longos RNAs não-codificadores têm sua expressão correlacionada a características moleculares de tumores de estrogênio, o que abre novas perspectivas de estudo para avaliar o papel desta classe de transcritos na modulação da expressão de genes envolvidos na transdução intracelular da via de sinalização ativada em resposta ao estímulo por estrogênio.

Um dos transcritos detectados como diferencialmente expresso entre tumores de mama RER+ e RER- mapeia em uma região intrônica do gene *OPCML* e se encontra com alta expressão em tumores de mama RER-, de acordo com as análises apresentadas neste trabalho. Este transcrito ainda possui em sua ponta 5' sítios para ligação de fatores de transcrição e, em especial, sítio para ligação do fator de transcrição c-Myc, envolvido no mecanismo pró-mitogênico de células mamárias pois exerce uma regulação positiva sobre a via de sinalização WNT (Cappellen et al., 2007). Já está descrita a presença de sítios de ligação de fatores de transcrição de genes codificadores para proteínas, incluindo c-Myc, em regiões intrônicas e intergênicas do genoma, sugerindo que ncRNAs originados nessas regiões genômicas possam ser regulados pelos mesmos fatores que modulam a transcrição de genes codificadores para proteínas (Cawley et al., 2004). A observação de sítios de ligação de c-Myc na vizinhança de um transcrito

intrônico no locus *OPCML* sugere que a ligação do fator de transcrição possa modular diretamente os níveis de transcrição do ncRNA intrônico. Esta possibilidade é suportada pela identificação de promotores responsivos a estrogênio que contêm sítios de ligação do receptor de estrogênio alfa e c-Myc (Cheng et al., 2006). Foi observado que o estímulo com estrogênio facilitou a associação de RER, c-Myc e do co-ativador TRRAP na região de genes responsivos ao hormônio, resultando no remodelamento da cromatina e aumento da transcrição (Cheng et al., 2006).

Já foi mostrado que o gene codificador *OPCML* sofre uma modificação epigenética em tecido tumoral ovariano e linhagens tumorais de mama (metilação das ilhas CpG), estando o seu silenciamento gênico associado à presença de tumor neste tecido (Sellar et al., 2003; Teodoridis et al., 2005). Já foi descrito na literatura a participação de ncRNAs em complexos protéicos envolvendo proteínas *Polycomb*, responsáveis por rearranjos na cromatina através da desacetilação de histonas, modulando a expressão de genes (Akhtar et al., 2000; Imamura et al., 2004). Dessa forma, podemos postular que os níveis de expressão do transcrito intrônico de *OPCML* são finamente regulados por fatores de transcrição, podendo estar envolvido em mecanismos de silenciamento gênico em *cis* via participação do mecanismo de metilação. Este ncRNA estaria regulando os níveis de expressão do gene codificador *OPCML*, presente no mesmo locus gênico e descrito como supressor tumoral em câncer ovariano (Sellar et al., 2003; Teodoridis et al., 2005). Assim, já que a descrição do envolvimento deste gene em câncer ovariano foi confirmada em linhagem celular tumoral de mama, deve ser feita uma investigação do nível de expressão do gene codificador *OPCML* no tecido tumoral humano mamário e sua correlação com o nível de expressão do transcrito intrônico derivado do mesmo gene.

Um dos transcritos não-codificadores identificados como diferencialmente expresso entre tumores de mama RER+ e RER- mapeia em região intrônica do gene *PRKCA*. Além disso, este gene também foi encontrado diferencialmente expresso em outros estudos do *Oncomine* entre tumores de mama RER+ e RER- e ainda neste trabalho, foi encontrado seu transcrito intrônico diferencialmente expresso entre tumores de mama com e sem metástase, apresentando alta expressão em tumores de mama livres de metástase. O gene codificador *PRKCA* apresenta expressão suprimida em tecido tumoral mamário RER+, de acordo com informações da literatura (Assender et al., 2007). Além disso, o nível da proteína PKC-alfa no tecido tumoral mamário parece ser indicativo da resposta à terapia endócrina anti-estrogênio (Assender et al., 2007). Nossos resultados de microarranjos mostram que o nível de expressão do transcrito intrônico de *PRKCA* é alto em tecido tumoral mamário RER+ e baixo em tecido tumoral mamário metastático, indicando sua alta expressão em amostras de tumores de mama com bom prognóstico. Os dados da literatura associam baixos níveis da proteína codificada pelo gene *PRKCA* em tecido tumoral mamário RER+. Assim, estes resultados permitem levantar a hipótese de que o transcrito intrônico de *PRKCA* possa estar exercendo uma regulação negativa em *cis* sobre o transcrito codificador para proteína gerado no mesmo *locus* gênico, diminuindo sua expressão. Um efeito regulatório semelhante já foi observado associado a ncRNAs intrônicos regulados por andrógeno (Louro et al., 2007). Após confirmada a hipótese de regulação do gene *PRKCA* pelo ncRNA do mesmo *locus* gênico, a medição dos níveis do transcrito intrônico *PRKCA* pode se tornar um fator de predição da resposta de pacientes com câncer de mama à terapia endócrina.

Da mesma forma, os resultados dos experimentos de microarranjo apontaram alta expressão do transcrito intrônico de *STARD13* em tecido tumoral mamário RER-, enquanto na literatura o gene *STARD13* foi relatado como um supressor de tumor da

família Rho GTPase, apresentando baixa expressão em tumores metastáticos de diversas topologias, incluindo mama (Ullmannova e Popescu, 2006). Considerando que os tumores de mama REr- normalmente apresentam um pior prognóstico da doença, podemos comparar os níveis de expressão deste transcrito ao descrito na literatura para tumores avançados, pois este gene possui expressão suprimida em tumores metastáticos diversos. Dessa forma, assumindo que o gene codificador *STARD13* apresente baixa expressão em tumores de mama REr-, o transcrito intrônico deste gene pode estar modificando sua expressão através de uma regulação em *cis*, comprovando a importância de estudos da função destas moléculas de RNA não codificadoras no contexto do câncer.

O conjunto de 95 transcritos expressos diferencialmente entre tumores REr+ e REr- foi comparado ao conjunto total de 3.221 transcritos válidos do microarranjo para verificar o possível enriquecimento de alguma função molecular ou celular. Do conjunto de 95 transcritos, 74 foram considerados válidos na análise por interagir com outras moléculas e vias na base de dados da ferramenta *Ingenuity Pathways Analysis*. As vias associadas a este conjunto de 74 transcritos com alta significância incluíram vias relacionadas a câncer, morte celular e crescimento e proliferação celular (valor p mínimo de 10^{-34}). De 74 transcritos, 16 apresentaram enriquecimento na função celular e molecular de ciclo celular (valor p mínimo de $4,7 \times 10^{-02}$), dentre estes, 1 corresponde a um transcrito intrônico e 15 a exônicos. Outra categoria de função celular e molecular que apresentou enriquecimento neste conjunto de dados foi a função de expressão gênica (com 10 transcritos exônicos, valor p mínimo de $4,0 \times 10^{-02}$).

Já foi descrito na literatura que a função de proliferação celular parece ser o denominador comum da maioria das assinaturas de expressão gênica preditivas do prognóstico no câncer de mama, apresentando desempenho melhor até mesmo que a

informação do *status* do receptor de estrogênio (Desmedt e Sotiriou, 2006). O enriquecimento de nossa assinatura molecular associada ao *status* do receptor de estrogênio na função de proliferação celular valida nossos resultados, já que estamos conseguindo captar diferenças fisiológicas relevantes entre os dois conjuntos de amostras comparadas, de acordo com os dados da literatura (Desmedt e Sotiriou, 2006).

Outra via apontada como relevante nesta análise da expressão de grupos de tumores REr+ e REr- foi a de metabolismo de colesterol regulado por receptor hormonal. Esta via conta com 8 genes, sendo 1 deles (*Active BCR-related gene - ABR*) representado em nosso microarranjo por um transcrito que mapeia na região intrônica do gene e que se apresenta expresso diferencialmente entre tumores REr+ e REr-. Não existem muitas informações disponíveis sobre este gene na literatura mas sua expressão já foi detectada em câncer de pulmão em tecido humano (Bangur et al., 2002). Além disso, ele foi descrito a partir de sua homologia com o gene *BCR*, cuja translocação está presente em pacientes com leucemia mielóide crônica (Heisterkamp et al., 1989). Portanto, este gene é um candidato interessante para um estudo detalhado de sua função em tumores de mama, já que sua associação com a doença é inovadora na literatura.

O grupo de transcritos intrônicos da assinatura de REr também foi analisado em separado para verificar o enriquecimento das funções moleculares ou celulares específicas do conjunto de genes em que eles mapeiam, possivelmente envolvidas com o câncer. Do conjunto de 14 transcritos intrônicos que compõe o conjunto, 10 foram considerados válidos na análise por interagir com outras moléculas na base de dados da ferramenta *Ingenuity Pathways Analysis*. Deste conjunto de dez transcritos intrônicos, oito (*STAG1, ABR, PDXK, PKCA, STARD13, PRKCA, WDR37, SYTI*) apresentaram associação significativa com vias funcionais relacionadas a doenças metabólicas, ciclo celular e câncer (valor p igual a 10^{-19}) (Figura 23) e dentre estes, dois (*PRKCA,*

STARD13) apresentaram enriquecimento da função de crescimento e proliferação celular, com um valor p mínimo de $4,3 \times 10^{-02}$. Estes dois transcritos intrônicos com enriquecimento da função celular de crescimento e proliferação (*PRKCA*, *STARD13*) devem ser estudados em maior detalhe, pois já existem indicações na literatura da expressão diferencial dos genes codificadores presentes no mesmo *locus* gênico que estes ncRNAs, em tecido tumoral mamário (informações detalhadas nesta seção, nos parágrafos anteriores).

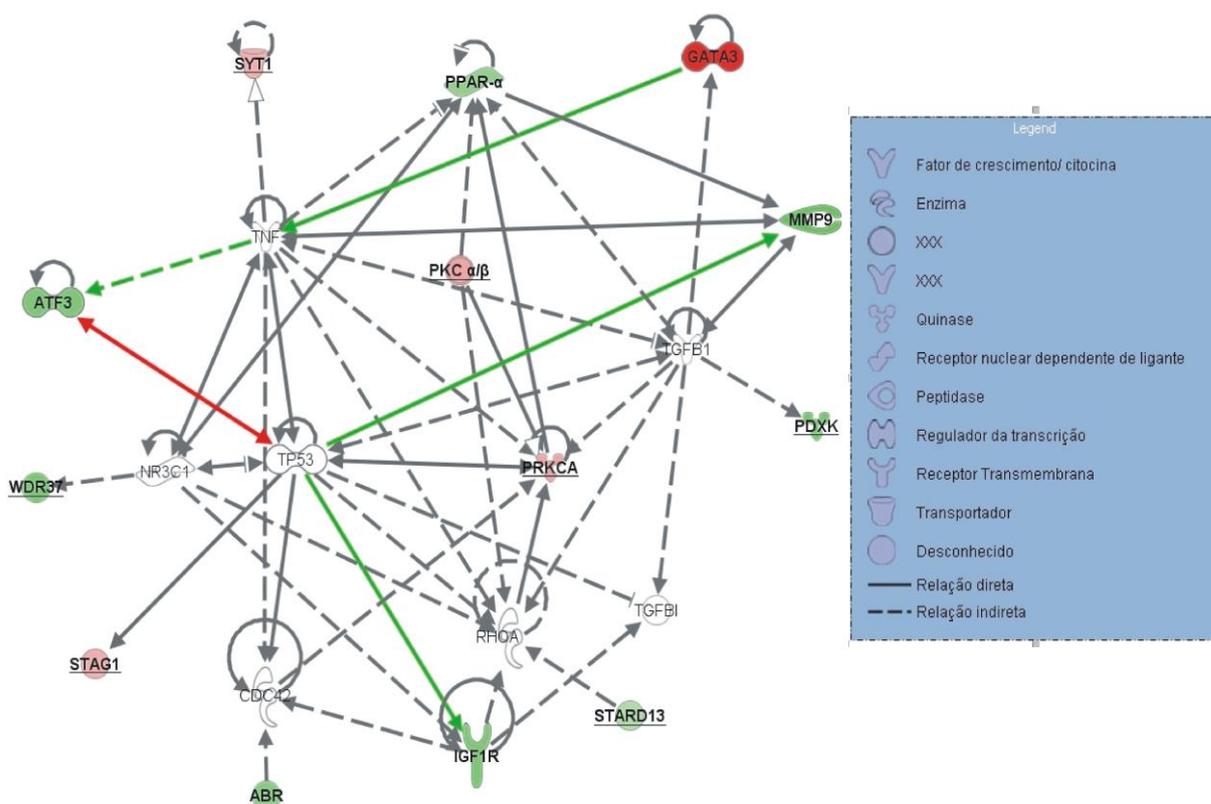


Figura 23. Figura esquemática das relações diretas e indiretas dos genes relacionados a ciclo celular e câncer. Sublinhados, estão identificados os transcritos intrônicos (seis de dez transcritos intrônicos no total), nomeados de acordo com o *RefSeq* codificador para proteína em que mapeiam. Moléculas brancas correspondem a genes que participam da rede mas não fazem parte do grupo de transcritos expressos diferencialmente entre tumores REr+ e REr-; moléculas coloridas correspondem aos genes representados no chip que são expressos diferencialmente entre tumores REr+ e REr- (cor verde indica baixa expressão e vermelha alta expressão em tumores REr+). Setas coloridas indicam inibição (verde) ou ativação (vermelha) da molécula relacionada.

A rede de associações apresentada na figura 23 não mostra apenas as relações entre os oito genes onde estão mapeados os transcritos intrônicos que compõem a

assinatura de RER+, presentes em vias celulares relacionadas a doenças metabólicas, ciclo celular e câncer, mas também alguns outros genes relacionados que estão representados em nosso microarranjo por transcritos exônicos. Estes transcritos também apareceram diferencialmente expressos na comparação do perfil de expressão gênica de tumores de mama RER+ vs. RER-.

O transcrito exônico *GATA-3* é um dos transcritos exônicos presente na rede da figura 23 e é altamente expresso em tumores de mama RER+, como já descrito na literatura em tecido tumoral mamário humano (Shen et al., 2005; Naderi et al., 2007). É sabido que ele interage diretamente com o fator de necrose tumoral (TNF) diminuindo sua expressão em camundongos (Ouyang et al., 1998). Considerando que esta regulação também esteja ocorrendo no conjunto de amostras de tumor de mama RER+ desta análise (o gene codificador para TNF não está representado no microarranjo), podemos supor que TNF apresenta baixa expressão em tumores RER+. Já se sabe que quando super-expresso o TNF é capaz de interagir de forma indireta com o transcrito exônico de *ATF3*, aumentando sua expressão nos tumores de mama RER- (Mallory et al., 2005; Masumoto et al., 2006). Dessa forma, é possível que nos tumores RER+, o aumento na expressão de *GATA-3* provoque uma redução de TNF que por sua vez regularia negativamente *ATF3*, o que de fato se observa.

Atf3 é um fator de transcrição que regula negativamente a expressão da proteína tumoral p53 (Kawauchi et al., 2002). A proteína p53 está envolvida com o mecanismo de tumorigênese e um dos tipos de câncer mais frequentemente associado a mutações no gene *P53* é o câncer de mama (Petitjean et al., 2007; Soerjomataram et al., 2008). Verificamos o nível de expressão deste gene nas amostras de tecido tumoral mamário através dos experimentos de microarranjo, mas não foi detectada expressão diferencial de *TP53* entre tumores RER+ e RER-, já que a razão de expressão deste gene é próximo

de zero na escala log (log da razão REr+/REr- é igual a -0,002). Uma explicação plausível para a baixa detecção deste transcrito em nossas amostras tumorais de mama é alguma característica da sonda depositada que não permite a detecção eficiente do transcrito. Portanto, não é possível determinar uma relação entre os níveis de expressão de *ATF3* e *TP53* utilizando nossos resultados dos experimentos de microarranjo.

Observou-se expressão diminuída do receptor do fator de crescimento *insulin-like 1 (IGF-1R)* nos tumores REr+ quando comparado a tumores de mama REr- (Figura 23). Os fatores de crescimento *insulin-like* fazem parte de uma família de polipeptídeos com papel importante no crescimento e diferenciação celular (Maor et al., 2006). O receptor IGF-1R medeia sua ação dentro das células, configurando como um mecanismo importante na iniciação do câncer de mama (Maor et al., 2006). Existem evidências na literatura de que a resposta da ligação do estrogênio ao seu receptor estimula a expressão e atividade dos componentes da via de *IGF-1R* (Maor et al., 2006). Sendo assim, a expressão de *IGF-1R* é alta em tecido tumoral mamário REr+ (Chong et al., 2006). No entanto, é reportada na literatura a alta expressão de *IGF-1R* em tecido tumoral mamário REr- como uma adaptação a baixos níveis de IGF-1 (Chong et al., 2006). Dessa forma, os níveis de expressão dos transcritos identificados na assinatura de estrogênio neste trabalho podem ser confirmados através dos dados experimentais relatados na literatura.

Também foi observado nos tumores REr+ o aumento da expressão do gene codificador para a metaloproteinase de matriz 9 (*MMP9*) (Figura 23). A expressão de *MMP9* é associado ao aumento da expressão da enzima responsável pela última etapa de produção de estrogênio em tecido tumoral mamário, as aromatases (Di et al., 2005). A super-expressão de aromatases no tecido tumoral causa uma alta concentração de estrogênio *in situ*, aumentando a expressão das metaloproteinases de matriz 2 e 9 (Di et al., 2005). O aumento da expressão destas metaloproteinases é um indicativo de pior

prognóstico e resistência à terapia com tamoxifeno em câncer de mama, onde os tumores são RER+ (Di et al., 2005). Estes dados da literatura também confirmam a validade desta assinatura de expressão gênica associada ao *status* do receptor de estrogênio em tecido tumoral mamário.

5.6. Classificação dos tumores de acordo com presença de metástase

A partir de análises de expressão entre dois grupos de tumores de mama de pacientes com e sem metástase foi possível identificar um conjunto de 113 transcritos expressos diferencialmente, capazes de reclassificar as amostras de acordo com a correlação com a assinatura de metástase com 100% de acerto. Este resultado foi possível devido ao critério rígido e cuidadoso na escolha dos tumores que compuseram a análise bem como à robustez da análise que incluiu a criação de grupos de treinamento e validação para encontrar e testar a assinatura de metástase (ver seção de Resultados 4.9).

Destes 113 transcritos identificados como expressos diferencialmente entre tumores de pacientes com e sem metástase, 32 (28%) são transcritos que mapeiam em regiões intrônicas de genes conhecidos. Estes dados corroboram resultados anteriores de nosso grupo usando a mesma plataforma de cDNA para a classificação de tumores de próstata (Reis et al., 2004) e rim (Brito et al., 2008), e reforçam a noção de que a inclusão de sondas para ncRNAs intrônicos em plataformas de microarranjos tem o potencial de aumentar a robustez das assinaturas de expressão identificadas para a classificação molecular de tumores.

Este conjunto de transcritos intrônicos presente na assinatura gênica associada com a presença/ausência de metástase nas pacientes apresenta uma alta associação a esta característica pois é capaz de re-classificar corretamente 17 das 18 amostras testadas em bom ou mau prognóstico, baseado em seu perfil de expressão, apresentando 95% de

acurácia. Uma hipótese para o motivo deste grupo de transcritos intrônicos não ter atingido 100% de acurácia como apresentado pelo conjunto total de 113 transcritos (utilizando como limiar para a classificação $r < -0,2$ para pacientes livres de metástase), seria o número reduzido de componentes desta assinatura de expressão gênica (32 transcritos intrônicos vs. 113 transcritos da assinatura de metástase), que representaria em menor grau a variabilidade existente entre as amostras tumorais de cada grupo. No entanto, podemos considerar que a utilização da assinatura de 32 ncRNAs intrônicos possui um desempenho comparável à assinatura de metástase composta por 113 transcritos na re-classificação dos tumores. Além disso, quando este conjunto de 32 transcritos intrônicos é removido da assinatura de 113 transcritos correlacionados à ocorrência de metástase, a discriminação entre as amostras dos grupos com presença/ausência é prejudicada, mostrando que os transcritos intrônicos melhoram o desempenho do classificador.

Foi encontrado na assinatura de metástase um ncRNA intrônico que mapeia no *locus* da ATPase de Na^+/K^+ (*ATP1A1*), cujo nível de expressão está relacionado à resposta de progestinas sintéticas (compostos utilizados em métodos contraceptivos orais, que simulam a resposta da progesterona no organismo) em linhagens tumorais mamárias (Bray et al., 2005). Observamos que o transcrito intrônico de *ATP1A1* possui baixa expressão em tecido tumoral mamário metastático quando comparado ao tecido tumoral mamário livre de metástase. A possibilidade de uma co-expressão do gene codificador e de um transcrito derivado de sua região intrônica no mesmo tipo tecidual sugere uma possível regulação do ncRNA sobre o gene codificador. Além disso, podemos levantar a hipótese de que a expressão de transcritos intrônicos pode estar associada a estímulos hormonais na célula.

Outros dois ncRNAs intrônicos presentes na assinatura correlacionada com metástase, mapeiam nos genes *ASPH* e *ARHGEF7*. Existem poucas informações na literatura sobre o envolvimento destes dois genes em câncer, mas eles estão envolvidos com o mecanismo de migração e motilidade celular em neuroblastoma (*ASPH*) e linhagem celular de câncer de cólon e adenocarcinomas (*ARHGEF7*) (Lahousse et al., 2006; Jones e Katan, 2007). A supressão da expressão de *ARHGEF7* nestas linhagens celulares através de siRNA inibe a migração celular (Jones e Katan, 2007). Assumindo que o caráter metastático das amostras de tumor de mama utilizadas neste projeto esteja associado à alta expressão destes genes no tecido tumoral, a baixa expressão dos ncRNAs de *ARHGEF7* e *ASPH* em tecido tumoral mamário metastático verificada nesta assinatura molecular de metástase, sugere novamente, uma possível regulação negativa destes ncRNAs sobre os genes codificadores dos respectivos *locus* gênicos. O envolvimento dos genes codificadores *ARHGEF7* e *ASPH* em câncer de mama deverá ser investigado em maiores detalhes, assim como sua possível regulação negativa, já que esta é a primeira vez que se estabelece essas associações com o câncer de mama.

O gene *CNOT4* possui representação no nosso microarranjo de DNA através de uma sonda de um transcrito proveniente de sua região intrônica e foi encontrado na assinatura de metástase com expressão aumentada em tumores livres de metástase. Foi confirmada a expressão diferencial do intrônico de *CNOT4* em tumores de mama com e sem metástase por PCR em tempo real. Experimentos de RT-fita específica revelaram que o intrônico de *CNOT4* é expresso com orientação antisenso em relação ao gene codificador para proteína existente no mesmo *locus* gênico. *CNOT4* é um homólogo de um gene descrito em levedura que participa do complexo protéico regulador da transcrição da RNA Polimerase II (Albert et al., 2000). Sua função ainda não foi descrita em tecido humano, mas a expressão diferencial de seu ncRNA pode indicar um possível

envolvimento de ambos RNAs codificadores e não codificadores na desregulação da transcrição que ocorre em tecidos tumorais humanos, sendo necessário estudos mais detalhados para a confirmação desta hipótese.

DNAJC3 é um gene que codifica uma proteína inibidora de Pkr (proteína quinase ativada por RNA dupla-fita) (Kim et al., 2000). O gene *PKR* consiste em um supressor tumoral em camundongos e sua expressão já foi identificada em linhagens tumorais e normais de células mamárias, mas a determinação de seu papel nestas células ainda é controversa (Kim et al., 2000). Na assinatura de metástase identificada neste trabalho, observamos o aumento da expressão do transcrito intrônico de *DNAJC3* nas amostras de tumor de mama que apresentaram metástase. Este resultado foi confirmado por qPCR e foi demonstrado que a orientação de transcrição é antisense em relação ao gene codificador para proteína através de RT-PCR fita-específica. O envolvimento dos RNAs codificadores e não codificadores de *DNAJC3* em câncer de mama deve ser estudado de forma mais sistemática assim como a relevância do gene *PKR* como supressor tumoral em adenocarcinoma de mama.

Quando ligado à molécula de estrogênio, o receptor de estrogênio alfa regula a expressão gênica através de sua ligação a seqüências de DNA palindrômicas, conhecidas como elementos responsivos ao estrogênio, na região promotora de genes regulados pelo estrogênio (Buluwela et al., 2005). Este mecanismo pode levar à iniciação do câncer através da resposta de sinalização de crescimento e proliferação celular nas células mamárias (Buluwela et al., 2005). Uma forma de inibir esta processo é a super-expressão da proteína de repressão transcricional *Zbtb16*, que age através do remodelamento da cromatina (Buluwela et al., 2005). O transcrito intrônico do gene *ZBTB16* foi encontrado diferencialmente expresso entre tumores de mama com e sem

metástase, apresentando baixa expressão em tumores metastáticos. Este ncRNA do gene *ZBTB16* pode estar participando do mecanismo de repressão transcricional de genes envolvidos com proliferação e crescimento celular em tecido mamário sadio.

A assinatura de 113 transcritos correlacionados com metástase foi comparada a listas de genes identificados como diferencialmente expressos entre tumores com e sem evidência de metástase em trabalhos utilizando plataformas de microarranjos comerciais (Desmedt et al., 2007; Radvanyi et al., 2005; Sorlie et al., 2003; van 't Veer et al., 2002), através da ferramenta *Oncomine* (www.oncomine.org). Dos oito transcritos da assinatura de metástase que mapeiam em genes encontrados em outros conjuntos de dados já publicados, quatro são exônicos (*WFDC2*, *CCT5*, *CORO1A*, *JMJD2C*), três mapeiam em regiões intrônicas (*MYO5A*, *GFRA1*, *NFATC3*) e um mapeia em uma região intergênica do genoma (*CXorf45*).

Três transcritos exônicos possuem nível de expressão em tumores de mama metastáticos semelhante ao dos genes do mesmo *locus* gênico encontrado em outros estudos, validando esta análise. O único transcrito que não apresentou nível de expressão coincidente em nossas análises de microarranjo em relação a outros estudos publicados foi *JMJD2C*. A expressão de *JMJD2C* está aumentada em tumores metastáticos (Sorlie, T., et al. 2003). Embora os resultados de microarranjo tenham apontado para uma diminuição da expressão de *JMJD2C* nos tumores metastáticos, os ensaios de PCR em tempo real apresentaram um resultado inverso, de acordo com os resultados descritos na literatura. Uma possível explicação para esta discrepância é que a sonda presente no microarranjo esteja detectando outras isoformas de *JMJD2C* com expressão distinta do transcrito medido no experimento de PCR em tempo real.

Existem três transcritos não-codificadores que mapeiam em regiões intrônicas de genes identificados como diferencialmente expressos em outros estudos comparando

tumores de mama com e sem metástase (*MYO5A*, *GFRA1*, *NFATC3*). O nível de expressão destes três ncRNAs intrônicos apresentam valores invertidos em relação aos genes codificadores identificados nos trabalhos (van 't Veer et al., 2002; Desmedt et al., 2007). Esta correlação invertida do nível de expressão de transcritos intrônicos em relação ao gene codificador de proteína do mesmo *locus* gênico pode estar associada a uma modulação em *cis* do gene codificador pelo transcrito intrônico associado, causando silenciamento da expressão gênica ou modificando o padrão de *splicing* e metilação do gene alvo (Reis et al., 2005).

O único transcrito intergênico que está presente como diferencialmente expresso em outros estudos de tumores metastáticos de mama é CXorf45, apresentando nível de expressão coincidente ao que foi apresentado em nosso microarranjo, confirmando também nossos resultados (Radvanyi et al., 2005).

Os quatro transcritos exônicos identificados no cruzamento com os dados disponíveis no *Oncomine* (*WFDC2*, *CCT5*, *CORO1A*, *JMJD2C*) também já foram estudados em maior detalhe em relação ao seu envolvimento com o câncer. Por exemplo, existem evidências na literatura do envolvimento do gene *WFDC2* com adenocarcinoma de pulmão (Bingle et al., 2006). Além disso, este gene é um marcador de câncer ovariano e a desregulação de sua expressão neste tecido está associada com o *splicing* alternativo (Bingle et al., 2002; Hellstrom et al., 2003). Este gene não havia sido implicado até o momento no câncer de mama, abrindo a perspectiva deste estudo contribuir não apenas com novos candidatos a marcadores de prognóstico em câncer de mama da classe de RNAs não codificadores, mas também de genes codificadores para proteínas.

O gene *CCT5* apresenta um papel importante no processo de desenvolvimento de resistência ao tratamento com docetaxel em pacientes com câncer de mama (Ooe et al.,

2007). *COROIA* está envolvido com o sistema imune, mais especificamente com processo de sobrevivência celular de células T (Mueller et al., 2008) e foi mostrado recentemente que o sistema imune humoral possui um papel importante na sobrevivência de pacientes com câncer de mama que não apresentam metástases (Schmidt et al., 2008). A amplificação do gene *JMJD2C* está presente na doença de leucemia mielóide aguda (Helias et al., 2008) e sua expressão é detectada em câncer de esôfago (Yang et al., 2001), pulmão (Italiano et al., 2006) e em células tronco não diferenciadas (Katoh e Katoh, 2007), justificando sua alta expressão em células metastáticas, encontrada em outros estudos. A presença destes genes em nossa assinatura correlacionada com a ocorrência de metástase confirma nossos resultados e aponta para um papel dos genes codificadores de proteína *COROIA* e *JMJD2C* na progressão dos tumores de mama.

Existem diversos grupos no mundo que investigam assinaturas moleculares capazes de permitir prognósticos mais eficientes no momento da decisão sobre terapias adjuvantes para pacientes com câncer de mama (van 't Veer et al., 2002; Chang et al., 2005; Wang et al., 2005; Crawford et al., 2007; Liu et al., 2007; Park et al., 2007; Welm et al., 2007). Estas assinaturas são extremamente diversas e apresentam uma sobreposição muito pequena entre os conjuntos de genes apontados como expressos diferencialmente nos diferentes estudos (Xu et al., 2008). Este fato pode ser explicado pelo uso de diferentes plataformas de microarranjo com diferentes conjuntos de sondas, métodos de normalização e até mesmo diferentes populações estudadas (Sotiriou e Piccart, 2007).

Neste trabalho também observamos uma baixa sobreposição entre a assinatura de metástase identificada e os conjuntos de genes identificados em outros estudos do perfil de expressão de câncer de mama (Desmedt et al., 2007; Radvanyi et al., 2005;

Sorlie et al., 2003; van 't Veer et al., 2002). Isto pode ser justificado em parte pelo limitado número de alvos de cDNA depositados em nosso microarranjo (4 mil elementos, sendo apenas cerca de 2 mil transcritos exônicos) quando comparado a outras plataformas comerciais utilizadas em estudos de expressão gênica, contendo de 20 a 40 mil elementos. Outra explicação para a identificação de genes distintos se relaciona ao uso de sondas que interrogam diferentes regiões dos transcritos, diferenças nos métodos de construção dos microarranjos, diferenças nos protocolos de marcação de alvos fluorescentes, entre outros.

Apesar da baixa sobreposição de genes, a maior parte dos classificadores desenvolvidos a partir do uso de microarranjos mostram um alto grau de concordância na predição da evolução da doença de grupos de pacientes independentes (van 't Veer et al., 2002; Wang et al., 2005). Uma análise do potencial de previsão da evolução clínica de assinaturas de expressão gênica relacionadas ao câncer de mama revelou que apesar do grau de sobreposição dos genes que participam destas assinaturas ser baixo, a concordância em relação à evolução da doença de câncer de mama entre as assinaturas é alta (Fan et al., 2006). Provavelmente, estas assinaturas são capazes de identificar os mesmos fenótipos biológicos apresentados pelo tumor apesar do diferente conjunto de genes envolvido em cada assinatura, assim como no caso da assinatura de metástase identificada neste trabalho.

Os três transcritos intrônicos resultantes do cruzamento com o *Oncomine* mapeiam em genes que ainda não foram descritos como relacionados ao câncer (*MYO5A*, *GFRA1*, *NFATC3*), mas que podem ter um papel na progressão do tumor. O gene *MYO5A*, por exemplo, possui uma importante função de distribuição das moléculas de RNA dentro da célula, disponibilizando-as para a tradução (Salerno et al., 2008). Esta função no cérebro de camundongos é afetada pela existência de diferentes isoformas da

proteína myo5a, causadas por um padrão específico de *splicing* destas moléculas neste tecido (Salerno et al., 2008). É possível que a expressão de um transcrito intrônico de *MYO5A* em tecido tumoral de mama pode afetar seu padrão de *splicing* neste tecido, comprometendo a expressão de outras proteínas envolvidas com o processo tumoral.

O gene *GFRA1* codifica um receptor da família GDNF, que é um co-receptor do receptor de tirosina quinase RET e ambos são super-expressos em tumores de mama invasivos RER+ (Esseghir et al., 2007). Esta super-expressão está associada ao aumento da proliferação e sobrevivência celular (Esseghir et al., 2007). O transcrito intrônico e antisenso (confirmado por PCR em tempo real) de *GFRA1* apresenta alta expressão em tumores de mama metastático e, portanto, pode estar atuando uma regulação em *cis* neste tecido, estabilizando a molécula do RNA codificador de *GFRA1*, exacerbando o processo de transformação maligna.

Da mesma forma pode estar atuando o transcrito intrônico de *NFATC3* sobre o RNA codificador proveniente do mesmo *locus* gênico, pois já existem evidências que este gene é um supressor de tumor em células mamárias de camundongos (Lee et al., 2005) e o nível de expressão de seu transcrito intrônico detectado neste estudo é baixo em células metastáticas mamárias. No entanto, em re-análises do *Oncomine* de estudos publicados na literatura, o nível de expressão do RNA codificador de *NFATC3* é alto em células metastáticas mamárias (van 't Veer, Dai et al. 2002), sendo necessário estudos mais detalhados para a confirmação do nível de expressão do gene codificador e do ncRNA intrônico de *NFATC3* em tecido tumoral mamário.

O grupo de 32 transcritos intrônicos expressos diferencialmente entre tumores com e sem metástase foi comparado ao conjunto total de 3.177 transcritos válidos do microarranjo para verificar possíveis enriquecimentos de alguma função molecular ou celular envolvidas com o câncer. Do conjunto de 32 transcritos intrônicos, 22 foram

considerados válidos na análise por interagir com outras moléculas e vias na base de dados da ferramenta *Ingenuity Pathways Analysis*. O resultado obtido relacionou seis dos 22 transcritos intrônicos associados à função celular de crescimento e proliferação (*MLLT3*, *ZBTB16*, *GFRA1*, *PRKCA*, *YYIAP1*, *MSI2*) e cinco transcritos foram associados à morte celular (*ATP1A1*, *MLLT3*, *PRKCA*, *ZBTB16*, *GFRA1*) (valor p mínimo de ambas as associações de 0,05). Além disso, o conjunto de 22 transcritos intrônicos apresentou alta associação a vias funcionais relacionadas à expressão gênica, ciclo celular e compromisso celular (esta última função envolve atrofia, dano, rompimento e inchaço celular) (valor p associado a essas vias de 10^{-3}), câncer, morte celular e doenças do tecido esquelético e muscular (valor p de 10^{-8}) e desenvolvimento celular, desenvolvimento e função dos sistemas linfático e imune e crescimento e proliferação celular (valor p de 10^{-32}). Destas 22 moléculas de ncRNAs, treze interagem de forma direta e indireta e esta rede de interação está apresentada na figura 25. Destes treze ncRNAs, três possuem sítio de ligação para fatores de transcrição na porção 5' de suas seqüências (*MLLT3*, *PRKCA*, *ZMIZ1*) e ainda o transcrito intrônico de *MLLT3* mapeia em uma região genômica identificada como sítio de ligação de histonas H3 contendo a lisina 27 trimetilada (H3K27me3). Estes dados apontam para um possível mecanismo de controle da expressão destes ncRNAs, da mesma natureza dos genes condutores para proteínas, indicando a relevância de seus níveis de expressão dentro da célula.

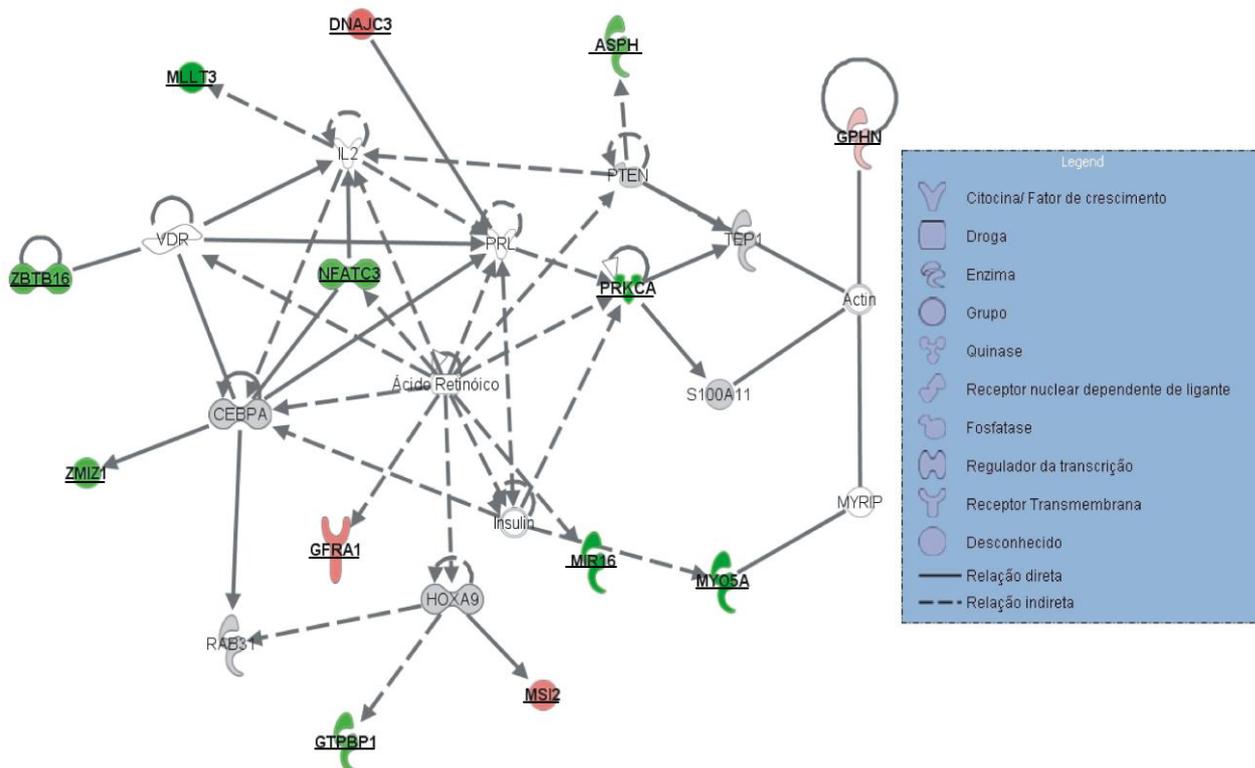


Figura 24. Figura esquemática das relações diretas e indiretas dos genes relacionados a desenvolvimento celular, desenvolvimento e função dos sistemas linfático e imune e crescimento e proliferação celular. Em negrito e sublinhado estão identificados os transcritos intrônicos (13 de 22 transcritos intrônicos no total), nomeados de acordo com o *RefSeq* codificador para proteína em que mapeiam. Moléculas brancas correspondem a genes que participam da rede mas não fazem parte do grupo de transcritos expressos diferencialmente entre tumores REr+ e REr-; moléculas cinzas correspondem a genes que participam da rede, estão representados no microarranjo de cDNA, mas não foram indicadas como expressas diferencialmente entre tumores de pacientes com e sem metástase; moléculas coloridas correspondem aos genes representados no chip que são expressos diferencialmente entre tumores REr+ e REr- (cor verde indica baixa expressão e vermelha alta expressão em tumores REr+).

Novamente, o enriquecimento deste grupo de transcritos intrônicos da assinatura de expressão gênica correlacionada à metástase na função de proliferação celular, valida o desempenho desta assinatura molecular como um fator de prognóstico em câncer de mama, comparável à informação do *status* do receptor de estrogênio, de acordo com a literatura (Desmedt e Sotiriou, 2006).

O transcrito intrônico de *NFATC3* faz parte da via funcional relacionada ao desenvolvimento celular, desenvolvimento e função dos sistemas linfático e imune e crescimento e proliferação celular e apresenta baixa expressão em tumores de pacientes

que desenvolveram metástase (Figura 25). Os processos biológicos associados a este gene (categorias de ontologia gênica do projeto *Gene Ontology* – www.geneontology.org) incluem regulação da transcrição dependente de DNA, regulação da transcrição através do promotor da RNA Polimerase II e resposta inflamatória. A proteína codificada pelo gene *NFATC3* é um fator de transcrição que se liga à região promotora de genes envolvidos com o processo de tumorigênese, como várias interleucinas (Humar et al., 2004), fator de necrose tumoral (Chen et al., 2003) e etc. modificando sua atividade transcricional. Especificamente, *Nfatc3* aumenta a transcrição da interleucina 2 (Rao et al., 1997), que tem o perfil de expressão muito alterado na presença de antígenos provenientes de tumores de mama quando comparado a outras infecções como vírus, tornando a resposta imune menos efetiva aos tumores de mama (Inokuma et al., 2007). Dessa forma, o transcrito intrônico de *NFATC3* deve exercer um papel regulatório sobre o fator de transcrição codificado pelo gene *NFATC3*. Este papel regulatório deve estar desregulado no tecido tumoral mamário, uma vez que já foi inclusive mostrado na literatura a expressão diferencial do gene codificador para proteína *NFATC3* entre tumores de mama com e sem metástase (análise de sobreposição da assinatura de metástase deste estudo com outros disponíveis na literatura através da ferramenta *Oncomine* – seção 4.9) (van 't Veer, Dai et al. 2002) e este gene já foi descrito como supressor de tumor em células tumorais mamárias de camundongos (Lee et al., 2005). Dessa forma e como já discutido anteriormente, este gene deve estar contribuindo para a progressão do câncer e constitui em um candidato interessante para estudos mais detalhados em câncer de mama.

Outros genes relacionados com o câncer de mama como o *PRKCA* (Tan et al., 2007), *GFRA1* (Naderi et al., 2007) e *MIR16* (Kobayashi et al., 2003), estão representados no microarranjo através de transcritos provenientes de RNAs não-

codificadores e fazem parte da via funcional relacionada ao desenvolvimento celular, desenvolvimento e função dos sistemas linfático e imune e crescimento e proliferação celular (Figura 25). O transcrito intrônico de *PRKCA* também compõe a assinatura de RER pois se apresenta diferencialmente expresso entre tumores de mama RER+ e RER-, mostrando que a assinatura de estrogênio também reflete, ainda que de forma limitada, as transformações associadas à evolução da doença em câncer de mama.

Aos RNAs não codificadores são atribuídas funções regulatórias e catalíticas controlando muitos processos fundamentais dentro das células através de três mecanismos básicos: catálise de reações biológicas, ligação e modulação da atividade protéica e pareamento com um ácido nucléico alvo (Goodrich e Kugel, 2006). Por exemplo, o transcrito intrônico de *NFATC3* pode estar exercendo uma destas funções sobre outros genes ou até mesmo sobre o gene codificador de *NFATC3*, pois já foi mostrado que este mesmo gene é regulado por um RNA não codificador longo (NRON RNA com 2-4 kilobases), que inibe sua ativação através da modificação de sua localização nuclear dentro da célula (Willingham et al., 2005). Além deste, existem já outros RNAs não codificadores descritos na literatura que atuam na regulação transcricional em algum nível, como o NRSE (*Neuron-restrictive silencer element*), RNA fita dupla de aproximadamente 20 nucleotídeos, que interage com a proteína NRSF/REST (*Neuron-restrictive silencing factor/RE-1-silencing*) convertendo-a em um fator de transcrição ativador ao invés de repressor, estimulando a diferenciação de células-tronco em neurônios (Kuwabara et al., 2004). Na classe dos RNAs não-codificadores longos, foi descrito o SRA (RNA ativador de receptor de esteróides), com cerca de 700 nucleotídeos, que funciona como um co-ativador transcricional para muitos receptores de hormônios esteróides (Lanz et al., 1999). Portanto, além de um mecanismo de regulação em *cis*, é plausível que um sub-conjunto de transcritos intrônicos

identificados neste trabalho exerçam regulação transcricional em *trans*, atuando sobre outros genes que participam da mesma via ou em vias distintas. Estes ncRNAs intrônicos poderiam agir através da formação de estruturas secundárias capazes de interagir com outros RNAs ou proteínas, ou alternativamente serem processados em pequenos RNAs com funções regulatórias. Estudos recentes utilizando *tiling arrays* indicam a existência de uma grande quantidade de pequenos RNAs complementares a regiões 5' e 3' de genes humanos, sendo uma fração destes originada de regiões intrônicas (Kapranov et al., 2007).

O envolvimento de microRNAs com câncer e, em especial, com o câncer de mama, tem sido mostrado de forma muito relevante na literatura para o processo de progressão da doença (Blenkiron et al., 2007; Zhang et al., 2007; Lehmann et al., 2008). No entanto, esta é a primeira vez que é mostrado um conjunto de RNAs longos, sem potencial codificador para proteínas, relacionado a vias de receptores hormonais e com a ocorrência de metástase. É possível que alguns destes ncRNAs desempenhem um papel na regulação da expressão gênica associada à tumorigênese ou progressão metastática no câncer de mama.

5.7. Caracterização de ncRNAs intrônicos correlacionados com metástase

A investigação da fita expressa dos transcritos intrônicos revelou que quatro dos seis transcritos testados (*GFRA1*, *DNAJC3*, *PTRH2*, *CNOT4*) possuem apenas expressão da fita antisenso e um (*CAB39*) possui expressão do par senso-antisenso.

A presença abundante de pares de transcritos senso-antisenso exônicos conservados evolutivamente já foi mostrado em plantas (Ma et al., 2006), camundongos (Kiyosawa et al., 2003) e também em humanos (Shendure e Church, 2002; Yelin et al., 2003; Chen et al., 2004), apontando para o fato de que este pode ser um mecanismo de

regulação pós-transcricional da regulação gênica em eucariotos (Silverman et al., 1992; Farrell e Lukens, 1995; Wutz et al., 1997). No entanto, existem poucos casos documentados mostrando evidências de funções regulatórias exercidas por ncRNAs intrônicos ou ainda por pares senso-antisense formados por estas moléculas (Reis et al., 2005).

Um dos casos de regulação por pares senso-antisense envolvendo ncRNAs é o descrito por Yan e colaboradores sobre o controle da apoptose celular exercido pelo gene *Fas* (Yan et al., 2005). Um transcrito intrônico não codificador denominado *Saf* de 1,5kb e expresso em diversos tecidos humanos e linhagens celulares, é derivado do intron 1 do gene *Fas* e está associado ao *splicing* alternativo deste gene, modificando a apoptose celular dos linfócitos T (Yan et al., 2005). Outro controle da regulação da expressão gênica exercido por pares senso-antisense de ncRNAs intrônicos envolve o impedimento físico da transcrição das duas fitas de DNA opostas, já que os transcritos senso e antisense se encontram no mesmo *locus* gênico (Prescott e Proudfoot, 2002).

Outras formas de regulação dos níveis de expressão gênicos por ncRNAs antisense envolvem a alteração do padrão de *splicing* do gene codificador em questão, que pode ser dado em *trans*. Esta regulação pode ocorrer através da ligação do transcrito intrônico antisense a elementos do *enhancer* intrônico do *splicing* ou silenciador, inibindo ou ativando a formação do spliceossomo, como ocorre com o receptor de hormônio da tireóide (Hastings et al., 2000). Além disso, é especulado que ncRNAs longos e antisense ao gene codificador do mesmo *locus* gênico podem modular a expressão gênica através da intervenção em várias etapas da transcrição ou processamento do RNAm, de forma semelhante à regulação por microRNAs (Bartel, 2004).

A maior parte dos transcritos intrônicos identificados nas assinaturas de metástase mapeiam em genes que estão envolvidos em vias relacionadas com o câncer. O estudo detalhado da possível função dos ncRNAs identificados na regulação da transcrição dos respectivos transcritos codificadores é altamente relevante e abrem a perspectiva de uma melhor compreensão do processo de progressão do tumor primário em tumor metastático.

Um sub-conjunto de transcritos identificados como diferencialmente expressos entre amostras de tumor provenientes de pacientes com ou sem metástase foi selecionado para confirmação do nível de expressão dos transcritos em experimentos de PCR em tempo real. Sete dos oito transcritos selecionados (*CAB39*, *DNAJC3*, *GFRA1*, *ZBTB16*, *MLLT3*, *CNOT4* e *GATA3*) apresentaram uma tendência de validação dos resultados de qPCR. Nas condições apresentadas, dois transcritos apresentaram uma diferença estatisticamente significativa: o exônico *GATA3*, $p < 0,09$ e o intrônico *DNAJC3*, $p < 0,1$.

Uma limitação para a validação de alguns transcritos de interesse por PCR em tempo real foi a pequena quantidade de RNA total disponível de amostras com tempo de seguimento longo (maior que quatro anos) e presença ou ausência de metástase. Das 18 amostras incluídas que atendiam os critérios de inclusão nas análises de microarranjo, apenas algumas dispunham de RNA total em quantidade suficiente para os experimentos de validação do nível de expressão por qPCR.

Como alternativa, foi testada sem sucesso a utilização de aRNA para a validação técnica dos dados de microarranjo. As reações de transcrição reversa utilizando-se aRNA não foram eficientes, produzindo resultados com alta variância tanto para transcritos exônicos como intrônicos. Uma hipótese para explicar este resultado é que durante a reação de amplificação do RNA total, são incorporadas uracilas com um

grupamento modificado para se acoplar o fluoróforo numa etapa posterior. Por isso decidimos utilizar o RNA total das amostras disponíveis para os experimentos de PCR em tempo real.

A existência de uma grande variabilidade individual no padrão transcricional entre amostras do mesmo tecido e tipo histológico, é sabidamente um fator que dificulta a identificação de genes com padrão de expressão semelhante em amostras de diferentes indivíduos. Para minimizar o efeito da variabilidade individual, utiliza-se freqüentemente as medidas de expressão gênica no tecido sadio, não-tumoral, do mesmo indivíduo para corrigir a variabilidade individual e tornar os perfis de expressão gênica dos tumores mais comparáveis. Neste caso, não nos era disponível o tecido sadio de cada uma das pacientes analisadas e por isso, as comparações dos níveis de expressão foram feitas entre pacientes diferentes, o que introduziu um nível maior de ruído experimental. Se houvesse, no entanto, quantidade de RNA disponível de todas as 18 amostras para os experimentos de PCR em tempo real, possivelmente a significância estatística da expressão diferencial entre os grupos com e sem metástase seria maior.

Será importante estender a validação da assinatura de metástase identificada a novos conjuntos de amostras tumorais com seguimento conhecido para avaliar o potencial preditivo deste conjunto de transcritos. A partir destes experimentos, espera-se que um conjunto mínimo de transcritos, possa ser otimizado em um teste baseado em reações de PCR tempo real, individualmente ou em reações multiplex, e utilizado como ferramenta complementar de prognóstico no câncer de mama.

Uma vez que RNAs não-codificadores compõem uma classe ampla de moléculas que possuem mecanismos muito diversos de regulação e em diferentes níveis (Goodrich e Kugel, 2006; Mattick e Makunin, 2006), antecipa-se que serão necessários estudos *in vitro* utilizando células em cultura para elucidar o mecanismo de ação dos ncRNAs

intrônicos identificados neste trabalho, assim como dos transcritos codificadores para proteína presentes nos *loci* correspondentes, para aprofundar o entendimento de seu papel na regulação da expressão gênica.

6. CONCLUSÕES

» Um estudo piloto indicou que metodologias de análise dos dados de microarranjo baseadas em razões de expressão a partir de uma referência externa/interna (*RefPool* ou *RefOligo*) ou de dados provenientes das intensidades diretas do canal de RNA (*One-Color*) produzem resultados comparáveis aos dados obtidos de hibridizações diretas de duas amostras que se deseja estudar em um mesmo microarranjo. No entanto, a utilização das medidas de intensidade (*One-Color*) se mostrou mais acurada e sensível em relação a razões de expressão baseadas em *RefOligo* quando ambas as metodologias foram aplicadas para a análise de dados de amostras clínicas numerosas, processadas ao longo de quatro anos do projeto.

» RNAs intrônicos não codificadores possuem níveis de expressão ubíquos em tecido tumoral mamário. Observou-se que os ncRNAs intrônicos e exônicos apresentaram nível de expressão comparáveis nas amostras de mama.

» Não foi possível identificar assinaturas de expressão gênica correlacionadas com idade, tamanho do tumor, presença de células metastáticas nos nódulos linfáticos, grau histológico e estágio tumoral, possivelmente refletindo a grande heterogeneidade entre as amostras que compõem os grupos baseados nestas características.

» Foi identificada uma assinatura de expressão gênica, contendo transcritos codificadores de proteínas e ncRNAs intrônicos, capaz de distinguir amostras de tecido tumoral mamário de acordo com o *status* de expressão do receptor de estrogênio,

apontando um conjunto de ncRNAs intrônicos possivelmente envolvidos na modulação da resposta intracelular a este estímulo hormonal.

» Foi identificada uma assinatura de expressão gênica, contendo transcritos codificadores de proteínas e ncRNAs intrônicos, capaz de distinguir com alta acurácia amostras de adenocarcinoma de mama com ou sem ocorrência de metástase. Este constitui um novo conjunto de genes marcadores no câncer de mama, com potencial aplicação na estratificação de risco de metástases à distância em pacientes com tumores do tipo ductal invasivo, que deverá ser ainda investigado em novos e independentes conjuntos de amostras.

» Uma assinatura contendo exclusivamente ncRNAs intrônicos apresentou 95% de acurácia na re-classificação de amostras de tumor de mama com ou sem ocorrência de metástase. O uso combinado de ncRNAs e transcritos codificadores para proteínas resultou em uma melhor acurácia de re-classificação de amostras em função da presença ou ausência de metástase, quando comparado a um classificador composto apenas por transcritos codificadores.

» Esta é a primeira vez que é mostrado um conjunto de RNAs longos, sem potencial codificador para proteínas, com expressão relacionada ao *status* de expressão do receptor de estrogênio e com a ocorrência de metástase em amostras de adenocarcinoma de mama do tipo ductal invasivo. Este achado levanta a hipótese de que alguns desses ncRNAs possam desempenhar um papel relevante na regulação da expressão gênica associada à tumorigênese ou progressão metastática no câncer de mama.

» Os níveis dos transcritos intrônicos não codificadores nas assinaturas de expressão gênica de REr e metástase se apresentam, muitas vezes, contrários ao relatado na literatura no mesmo tecido tumoral e análise, mas outras vezes se apresenta igual, sugerindo que a regulação gênica exercida por estas moléculas de ncRNAs deve se dar de modo complexo e diverso, possivelmente afetando em *cis* o padrão de *splicing* ou a abundância do gene codificador para proteína expresso no mesmo *locus*.

» As assinaturas de expressão gênica apresentadas neste trabalho indicaram transcritos intrônicos que mapeiam em genes que apresentam função relevante em câncer mais que ainda não foram descritos como envolvidos em tecido tumoral de mama (*OPCML*, *ASPH*, *MYO5A*, *CNOT4*). Dessa forma, os transcritos gerados em regiões codificadoras para proteína e não-codificadoras nestes *loci* gênicos devem ser estudados em maior detalhe para determinar uma possível correlação do nível de expressão, e um possível papel na transformação maligna/progressão do câncer de mama.

» Alguns transcritos identificados nas assinaturas de expressão gênica apresentadas neste projeto mapeiam em genes com função relevante descrita em câncer de mama (*PRKCA*, *STARD13*, *GFRA1*, *NFATC3*, *DNAJC3*, *ZBTB16*). Novos experimentos devem ser propostos para investigar o envolvimento dos ncRNAs derivados de regiões intrônicas destes genes na desregulação da função dos RNAs codificadores dos mesmos *loci* gênicos em tecido tumoral mamário.

» Foram identificados transcritos codificadores para proteínas que constituem novos candidatos a alvos moleculares no adenocarcinoma de mama ductal invasivo. Entre eles, destacam-se alguns genes já associados com câncer na literatura, porém nunca estudados

no câncer de mama: *WFDC2* (marcador de câncer ovariano), *CORO1A* (sistema imune – sobrevivência de células T) e *ABR* (câncer de pulmão).

7. REFERÊNCIAS

- Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno and et al. (1991). "Complementary DNA sequencing: expressed sequence tags and human genome project." Science 252(5013): 1651-6.
- Akhtar, A., D. Zink and P. B. Becker (2000). "Chromodomains are protein-RNA interaction modules." Nature 407(6802): 405-9.
- Albert, T. K., M. Lemaire, N. L. van Berkum, R. Gentz, M. A. Collart and H. T. Timmers (2000). "Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits." Nucleic Acids Res 28(3): 809-17.
- Alizadeh, A. A., M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown and L. M. Staudt (2000). "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling." Nature 403(6769): 503-11.
- Amaral, P. P., M. E. Dinger, T. R. Mercer and J. S. Mattick (2008). "The eukaryotic genome as an RNA machine." Science 319(5871): 1787-9.
- Assender, J. W., J. M. Gee, I. Lewis, I. O. Ellis, J. F. Robertson and R. I. Nicholson (2007). "Protein kinase C isoform expression as a predictor of disease outcome on endocrine therapy in breast cancer." J Clin Pathol 60(11): 1216-21.
- Attoor, S., E. R. Dougherty, Y. Chen, M. L. Bittner and J. M. Trent (2004). "Which is better for cDNA-microarray-based classification: ratios or direct intensities." Bioinformatics 20(16): 2513-20.
- Badve, S. S., F. L. Baehner, R. P. Gray, B. H. Childs, T. Maddala, M. L. Liu, S. C. Rowley, S. Shak, E. D. Perez, L. J. Shulman, S. Martino, N. E. Davidson, G. W. Sledge, L. J. Goldstein and J. A. Sparano (2008). "Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory." J Clin Oncol 26(15): 2473-81.
- Bangur, C. S., A. Switzer, L. Fan, M. J. Marton, M. R. Meyer and T. Wang (2002). "Identification of genes over-expressed in small cell lung carcinoma using suppression subtractive hybridization and cDNA microarray expression analysis." Oncogene 21(23): 3814-25.
- Barginear, M. F., T. Bradley, I. Shapira and D. R. Budman (2008). "Implications of applied research for prognosis and therapy of breast cancer." Crit Rev Oncol Hematol 65(3): 223-34.
- Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell 116(2): 281-97.
- Bartsch, R., C. Wenzel, C. C. Zielinski and G. G. Steger (2007). "HER-2-positive breast cancer: hope beyond trastuzumab." BioDrugs 21(2): 69-77.
- Bingle, L., S. S. Cross, A. S. High, W. A. Wallace, D. Rassl, G. Yuan, I. Hellstrom, M. A. Campos and C. D. Bingle (2006). "WFDC2 (HE4): a potential role in the innate immunity of the oral cavity and respiratory tract and the development of adenocarcinomas of the lung." Respir Res 7: 61.
- Bingle, L., V. Singleton and C. D. Bingle (2002). "The putative ovarian tumour marker gene HE4 (WFDC2), is expressed in normal tissues and undergoes complex alternative splicing to yield multiple protein isoforms." Oncogene 21(17): 2768-73.
- Birch, J. M., R. D. Alston, R. J. McNally, D. G. Evans, A. M. Kelsey, M. Harris, O. B. Eden and J. M. Varley (2001). "Relative frequency and morphology of cancers in carriers of germline TP53 mutations." Oncogene 20(34): 4621-8.

- Bland, K. I., H. R. Menck, C. E. Scott-Conner, M. Morrow, D. J. Winchester and D. P. Winchester (1998). "The National Cancer Data Base 10-year survey of breast carcinoma treatment at hospitals in the United States." *Cancer* 83(6): 1262-73.
- Blenkiron, C., L. D. Goldstein, N. P. Thorne, I. Spiteri, S. F. Chin, M. J. Dunning, N. L. Barbosa-Morais, A. E. Teschendorff, A. R. Green, I. O. Ellis, S. Tavare, C. Caldas and E. A. Miska (2007). "MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype." *Genome Biol* 8(10): R214.
- Bloom, H. J. and W. W. Richardson (1957). "Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years." *Br J Cancer* 11(3): 359-77.
- Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." *Bioinformatics* 19(2): 185-93.
- Boughey, J. C., A. U. Buzdar and K. K. Hunt (2008). "Recent advances in the hormonal treatment of breast cancer." *Curr Probl Surg* 45(1): 13-55.
- Bradbury, A. R. and O. I. Olopade (2007). "Genetic susceptibility to breast cancer." *Rev Endocr Metab Disord* 8(3): 255-67.
- Braun, S., K. Pantel, P. Muller, W. Janni, F. Hepp, C. R. Kentenich, S. Gastroph, A. Wischnik, T. Dimpfl, G. Kindermann, G. Riethmuller and G. Schlimok (2000). "Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer." *N Engl J Med* 342(8): 525-33.
- Bray, J. D., S. Jelinsky, R. Ghatge, J. A. Bray, C. Tunkey, K. Saraf, B. M. Jacobsen, J. K. Richer, E. L. Brown, R. C. Winneker, K. B. Horwitz and C. R. Lyttle (2005). "Quantitative analysis of gene regulation by seven clinically relevant progestins suggests a highly similar mechanism of action through progesterone receptors in T47D breast cancer cells." *J Steroid Biochem Mol Biol* 97(4): 328-41.
- Brenner, S., M. Johnson, J. Bridgham, G. Golda, D. H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S. R. Williams, K. Moon, T. Burcham, M. Pallas, R. B. DuBridge, J. Kirchner, K. Fearon, J. Mao and K. Corcoran (2000). "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays." *Nat Biotechnol* 18(6): 630-4.
- Brito, G. C., A. A. Fachel, A. L. Vettore, G. M. Vignal, E. R. Gimba, F. S. Campos, M. A. Barcinski, S. Verjovski-Almeida and E. M. Reis (2008). "Identification of protein-coding and intronic noncoding RNAs down-regulated in clear cell renal carcinoma." *Mol Carcinog*.
- Buluwela, L., J. Pike, D. Mazhar, T. Kamalati, S. M. Hart, R. Al-Jehani, H. Yahaya, N. Patel, N. Sarwar, D. A. Heathcote, O. Schwickerath, F. Phoenix, R. Hill, E. Aboagye, S. Shousha, J. Waxman, N. R. Lemoine, A. Zelent, R. C. Coombes and S. Ali (2005). "Inhibiting estrogen responses in breast cancer cells using a fusion protein encoding estrogen receptor-alpha and the transcriptional repressor PLZF." *Gene Ther* 12(5): 452-60.
- Buyse, M., S. Loi, L. van't Veer, G. Viale, M. Delorenzi, A. M. Glas, M. S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. Harris, J. Bogaerts, P. Therasse, A. Floore, M. Amakrane, F. Piette, E. Rutgers, C. Sotiriou, F. Cardoso and M. J. Piccart (2006). "Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer." *J Natl Cancer Inst* 98(17): 1183-92.
- Calin, G. A., C. G. Liu, M. Ferracin, T. Hyslop, R. Spizzo, C. Sevignani, M. Fabbri, A. Cimmino, E. J. Lee, S. E. Wojcik, M. Shimizu, E. Tili, S. Rossi, C. Taccioli, F. Pichiorri, X. Liu, S. Zupo, V. Herlea, L. Gramantieri, G. Lanza, H. Alder, L. Rassenti, S. Volinia, T. D. Schmittgen, T. J. Kipps, M. Negrini and C. M. Croce (2007). "Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas." *Cancer Cell* 12(3): 215-29.
- Calin, G. A., C. Sevignani, C. D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini and C. M. Croce (2004). "Human microRNA

- genes are frequently located at fragile sites and genomic regions involved in cancers." Proc. Natl. Acad. Sci. USA 101(9): 2999-3004.
- 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 and S. F. Lowry (2005). "A network-based analysis of systemic inflammation in humans." Nature 437(7061): 1032-7.
- Camargo, A. A., H. P. B. Samaia, E. Dias-Neto, D. F. Simao, I. A. Migotto, M. R. S. Briones, F. F. Costa, M. A. Nagai, S. Verjovski-Almeida, M. A. Zago, L. E. C. Andrade, H. Carrer, H. F. A. El-Dorry, E. M. Espreafico, A. Habr-Gama, D. Giannella-Neto, G. H. Goldman, A. Gruber, C. Hackel, E. T. Kimura, R. M. B. Maciel, S. K. N. Marie, E. A. L. Martins, M. P. Nobrega, M. L. Paco-Larson, M. Pardini, G. G. Pereira, J. B. Pesquero, V. Rodrigues, S. R. Rogatto, I. da Silva, M. C. Sogayar, M. D. Sonati, E. H. Tajara, S. R. Valentini, F. L. Alberto, M. E. J. Amaral, I. Aneas, L. A. T. Arnaldi, A. M. de Assis, M. H. Bengtson, N. A. Bergamo, V. Bombonato, M. E. R. de Camargo, R. A. Canevari, D. M. Carraro, J. M. Cerutti, M. L. C. Correa, R. F. R. Correa, M. C. R. Costa, C. Curcio, P. O. M. Hokama, A. J. S. Ferreira, G. K. Furuzawa, T. Gushiken, P. L. Ho, E. Kimura, J. E. Krieger, L. C. C. Leite, P. Majumder, M. Marins, E. R. Marques, A. S. A. Melo, M. Melo, C. A. Mestriner, E. C. Miracca, D. C. Miranda, A. Nascimento, F. G. Nobrega, T. P. B. Ojopi, J. R. C. Pandolfi, L. G. Pessoa, A. C. Prevedel, P. Rahal, C. A. Rainho, E. M. R. Reis, M. L. Ribeiro, N. da Ros, R. G. de Sa, M. M. Sales, S. C. Sant'anna, M. L. dos Santos, A. M. da Silva, N. P. da Silva, W. A. Silva, R. A. da Silveira, J. F. Sousa, D. Stecconi, F. Tsukumo, V. Valente, F. Soares, E. S. Moreira, D. N. Nunes, R. G. Correa, H. Zalberg, A. F. Carvalho, L. F. L. Reis, R. R. Brentani, A. J. G. Simpson and S. J. de Souza (2004). "The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome (vol 98, pg 12103, 2001)." Proceedings of the National Academy of Sciences of the United States of America 101(1): 414-414.
- Cappellen, D., T. Schlange, M. Bauer, F. Maurer and N. E. Hynes (2007). "Novel c-MYC target genes mediate differential effects on cell proliferation and migration." EMBO Rep 8(1): 70-6.
- Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V. B. Bajic, S. E. Brenner, S. Batalov, A. R. Forrest, M. Zavolan, M. J. Davis, L. G. Wilming, V. Aidinis, J. E. Allen, A. Ambesi-Impombato, R. Apweiler, R. N. Aturaliya, T. L. Bailey, M. Bansal, L. Baxter, K. W. Beisel, T. Bersano, H. Bono, A. M. Chalk, K. P. Chiu, V. Choudhary, A. Christoffels, D. R. Clutterbuck, M. L. Crowe, E. Dalla, B. P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C. F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T. R. Gingeras, T. Gojobori, R. E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T. K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasaki, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S. P. Krishnan, A. Kruger, S. K. Kummerfeld, I. V. Kurochkin, L. F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, F. Nori, O. Ohara, Y. Okazaki, V. Orlando, K. C. Pang, W. J. Pavan, G. Pavesi, G. Pesole, N. Petrovsky, S. Piazza, J. Reed, J. F. Reid, B. Z. Ring, M. Ringwald, B. Rost, Y. Ruan, S. L. Salzberg, A. Sandelin, C. Schneider, C. Schonbach, K. Sekiguchi, C. A. Semple, S. Seno, L. Sessa, Y. Sheng, Y. Shibata, H. Shimada, K. Shimada, D. Silva, B. Sinclair, S. Sperling, E. Stupka, K. Sugiura, R. Sultana, Y. Takenaka, K. Taki, K. Tammoja, S. L. Tan, S. Tang, M. S. Taylor, J. Tegner, S. A. Teichmann, H. R. Ueda, E. van Nimwegen, R. Verardo,

- C. L. Wei, K. Yagi, H. Yamanishi, E. Zabarovsky, S. Zhu, A. Zimmer, W. Hide, C. Bult, S. M. Grimmond, R. D. Teasdale, E. T. Liu, V. Brusica, J. Quackenbush, C. Wahlestedt, J. S. Mattick, D. A. Hume, C. Kai, D. Sasaki, Y. Tomaru, S. Fukuda, M. Kanamori-Katayama, M. Suzuki, J. Aoki, T. Arakawa, J. Iida, K. Imamura, M. Itoh, T. Kato, H. Kawaji, N. Kawagashira, T. Kawashima, M. Kojima, S. Kondo, H. Konno, K. Nakano, N. Ninomiya, T. Nishio, M. Okada, C. Plessy, K. Shibata, T. Shiraki, S. Suzuki, M. Tagami, K. Waki, A. Watahiki, Y. Okamura-Oho, H. Suzuki, J. Kawai and Y. Hayashizaki (2005). "The transcriptional landscape of the mammalian genome." Science 309(5740): 1559-63.
- Carter, C. L., C. Allen and D. E. Henson (1989). "Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases." Cancer 63(1): 181-7.
- Cawley, S., S. Bekiranov, H. H. Ng, P. Kapranov, E. A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A. J. Williams, R. Wheeler, B. Wong, J. Drenkow, M. Yamanaka, S. Patel, S. Brubaker, H. Tammana, G. Helt, K. Struhl and T. R. Gingeras (2004). "Unbiased Mapping of Transcription Factor Binding Sites along Human Chromosomes 21 and 22 Points to Widespread Regulation of Noncoding RNAs." Cell 116(4): 499-509.
- Chang, H. Y., D. S. Nuyten, J. B. Sneddon, T. Hastie, R. Tibshirani, T. Sorlie, H. Dai, Y. D. He, L. J. van't Veer, H. Bartelink, M. van de Rijn, P. O. Brown and M. J. van de Vijver (2005). "Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival." Proc Natl Acad Sci U S A 102(10): 3738-43.
- Chang, H. Y., J. B. Sneddon, A. A. Alizadeh, R. Sood, R. B. West, K. Montgomery, J. T. Chi, M. van de Rijn, D. Botstein and P. O. Brown (2004). "Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds." PLoS Biol 2(2): E7.
- Chen, J., Y. Amasaki, Y. Kamogawa, M. Nagoya, N. Arai, K. Arai and S. Miyatake (2003). "Role of NFATx (NFAT4/NFATc3) in expression of immunoregulatory genes in murine peripheral CD4+ T cells." J Immunol 170(6): 3109-17.
- Chen, J., M. Sun, W. J. Kent, X. Huang, H. Xie, W. Wang, G. Zhou, R. Z. Shi and J. D. Rowley (2004). "Over 20% of human transcripts might form sense-antisense pairs." Nucleic Acids Res 32(16): 4812-20.
- Cheng, A. S., V. X. Jin, M. Fan, L. T. Smith, S. Liyanarachchi, P. S. Yan, Y. W. Leu, M. W. Chan, C. Plass, K. P. Nephew, R. V. Davuluri and T. H. Huang (2006). "Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters." Mol Cell 21(3): 393-404.
- Cheng, J., P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt, V. Sementchenko, A. Piccolboni, S. Bekiranov, D. K. Bailey, M. Ganesh, S. Ghosh, I. Bell, D. S. Gerhard and T. R. Gingeras (2005). "Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution." Science 308(5725): 1149-54.
- Chin, K., S. DeVries, J. Fridlyand, P. T. Spellman, R. Roydasgupta, W. L. Kuo, A. Lapuk, R. M. Neve, Z. Qian, T. Ryder, F. Chen, H. Feiler, T. Tokuyasu, C. Kingsley, S. Dairkee, Z. Meng, K. Chew, D. Pinkel, A. Jain, B. M. Ljung, L. Esserman, D. G. Albertson, F. M. Waldman and J. W. Gray (2006). "Genomic and transcriptional aberrations linked to breast cancer pathophysiologies." Cancer Cell 10(6): 529-41.
- Chong, Y. M., K. Colston, W. G. Jiang, A. K. Sharma and K. Mokbel (2006). "The relationship between the insulin-like growth factor-1 system and the oestrogen metabolising enzymes in breast cancer tissue and its adjacent non-cancerous tissue." Breast Cancer Res Treat 99(3): 275-88.
- Choo, A., J. Padmanabhan, A. Chin, W. J. Fong and S. K. Oh (2006). "Immortalized feeders for the scale-up of human embryonic stem cells in feeder and feeder-free conditions." J Biotechnol 122(1): 130-41.

- Ciatto, S., S. Cecchini, G. Grazzini, A. Iossa, D. Bartoli and A. Rasponi (1990). "Tumor size and prognosis of breast cancer with negative axillary nodes." Neoplasma 37(2): 179-84.
- Clark, R. M., G. Schweikert, C. Toomajian, S. Ossowski, G. Zeller, P. Shinn, N. Warthmann, T. T. Hu, G. Fu, D. A. Hinds, H. Chen, K. A. Frazer, D. H. Huson, B. Scholkopf, M. Nordborg, G. Ratsch, J. R. Ecker and D. Weigel (2007). "Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*." Science 317(5836): 338-42.
- Conejero-Goldberg, C., E. Wang, C. Yi, T. E. Goldberg, L. Jones-Brando, F. M. Marincola, M. J. Webster and E. F. Torrey (2005). "Infectious pathogen detection arrays: viral detection in cell lines and postmortem brain tissue." Biotechniques 39(5): 741-51.
- Costa, S. D., S. Lange, K. Klinga, E. Merkle and M. Kaufmann (2002). "Factors influencing the prognostic role of oestrogen and progesterone receptor levels in breast cancer--results of the analysis of 670 patients with 11 years of follow-up." Eur J Cancer 38(10): 1329-34.
- Crawford, N. P., X. Qian, A. Ziogas, A. G. Papageorge, B. J. Boersma, R. C. Walker, L. Lukes, W. L. Rowe, J. Zhang, S. Ambs, D. R. Lowy, H. Anton-Culver and K. W. Hunter (2007). "Rrp1b, a new candidate susceptibility gene for breast cancer progression and metastasis." PLoS Genet 3(11): e214.
- Cutler, D. J., M. E. Zwick, M. M. Carrasquillo, C. T. Yohn, K. P. Tobin, C. Kashuk, D. J. Mathews, N. A. Shah, E. E. Eichler, J. A. Warrington and A. Chakravarti (2001). "High-throughput variation detection and genotyping using microarrays." Genome Res 11(11): 1913-25.
- Dai, S., T. Wang, X. Yan and S. Chen (2007). "Proteomics of pollen development and germination." J Proteome Res 6(12): 4556-63.
- Dapic, V., M. A. Carvalho and A. N. Monteiro (2005). "Breast cancer susceptibility and the DNA damage response." Cancer Control 12(2): 127-36.
- David, L., W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis and L. M. Steinmetz (2006). "A high-resolution map of transcription in the yeast genome." Proc Natl Acad Sci U S A 103(14): 5320-5.
- Desmedt, C., F. Piette, S. Loi, Y. Wang, F. Lallemand, B. Haibe-Kains, G. Viale, M. Delorenzi, Y. Zhang, M. S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. L. Harris, J. G. Klijn, J. A. Foekens, F. Cardoso, M. J. Piccart, M. Buyse and C. Sotiriou (2007). "Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series." Clin Cancer Res 13(11): 3207-14.
- Desmedt, C. and C. Sotiriou (2006). "Proliferation: the most prominent predictor of clinical outcome in breast cancer." Cell Cycle 5(19): 2198-202.
- Di, G. H., J. S. Lu, C. G. Song, H. C. Li, Z. Z. Shen and Z. M. Shao (2005). "Over expression of aromatase protein is highly related to MMPs levels in human breast carcinomas." J Exp Clin Cancer Res 24(4): 601-7.
- 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 and A. J. Simpson (2000). "Shotgun sequencing of the human transcriptome with ORF expressed sequence tags." Proc. Natl. Acad. Sci. USA 97(7): 3491-3496.
- Dieterich, C., D. Iglehart, E. Riccio, J. C. Mirsalis and H. H. Ng (2008). "Microarray Evaluation of the *Listeria monocytogenes* Infection and Amoxicillin Treatment in Mice." Int J Toxicol 27(3): 265-72.
- Dinh, P., C. Sotiriou and M. J. Piccart (2007). "The evolution of treatment strategies: aiming at the target." Breast 16 Suppl 2: S10-6.

- Dudley, A. M., J. Aach, M. A. Steffen and G. M. Church (2002). "Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range." Proc Natl Acad Sci U S A 99(11): 7554-9.
- Duffy, M. J. and C. Duggan (2004). "The urokinase plasminogen activator system: a rich source of tumour markers for the individualised management of patients with cancer." Clin Biochem 37(7): 541-8.
- Dupuy, A. and R. M. Simon (2007). "Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting." J Natl Cancer Inst 99(2): 147-57.
- Elenbaas, B., L. Spirio, F. Koerner, M. D. Fleming, D. B. Zimonjic, J. L. Donaher, N. C. Popescu, W. C. Hahn and R. A. Weinberg (2001). "Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells." Genes Dev 15(1): 50-65.
- Ellison, D. W., T. R. Clark, D. E. Sturdevant, K. Virtaneva, S. F. Porcella and T. Hackstadt (2008). "Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa." Infect Immun 76(2): 542-50.
- Elston, C. W. and I. O. Ellis (1991). "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up." Histopathology 19(5): 403-10.
- Esseghir, S., S. K. Todd, T. Hunt, R. Poulosom, I. Plaza-Menacho, J. S. Reis-Filho and C. M. Isacke (2007). "A role for glial cell derived neurotrophic factor induced expression by inflammatory cytokines and RET/GFR alpha 1 receptor up-regulation in breast cancer." Cancer Res 67(24): 11732-41.
- Fan, C., D. S. Oh, L. Wessels, B. Weigelt, D. S. Nuyten, A. B. Nobel, L. J. van't Veer and C. M. Perou (2006). "Concordance among gene-expression-based predictors for breast cancer." N Engl J Med 355(6): 560-9.
- Farrell, C. M. and L. N. Lukens (1995). "Naturally occurring antisense transcripts are present in chick embryo chondrocytes simultaneously with the down-regulation of the alpha 1 (I) collagen gene." J Biol Chem 270(7): 3400-8.
- Ferrero-Pous, M., K. Hacene, C. Bouchet, V. Le Doussal, M. Tubiana-Hulin and F. Spyrtos (2000). "Relationship between c-erbB-2 and other tumor characteristics in breast cancer prognosis." Clin Cancer Res 6(12): 4745-54.
- Finetti, P., N. Cervera, E. Charafe-Jauffret, C. Chabannon, C. Charpin, M. Chaffanet, J. Jacquemier, P. Viens, D. Birnbaum and F. Bertucci (2008). "Sixteen-kinase gene expression identifies luminal breast cancers with poor prognosis." Cancer Res 68(3): 767-76.
- Foekens, J. A., D. Atkins, Y. Zhang, F. C. Sweep, N. Harbeck, A. Paradiso, T. Cufer, A. M. Sieuwerts, D. Talantov, P. N. Span, V. C. Tjan-Heijnen, A. F. Zito, K. Specht, H. Hoefler, R. Golouh, F. Schittulli, M. Schmitt, L. V. Beex, J. G. Klijn and Y. Wang (2006). "Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer." J Clin Oncol 24(11): 1665-71.
- Frith, M. C., M. Pheasant and J. S. Mattick (2005). "The amazing complexity of the human transcriptome." Eur J Hum Genet 13(8): 894-7.
- Garman, K. S., J. R. Nevins and A. Potti (2007). "Genomic strategies for personalized cancer therapy." Hum Mol Genet 16 Spec No. 2: R226-32.
- Ginestier, C., N. Cervera, P. Finetti, S. Esteyries, B. Esterni, J. Adelaide, L. Xerri, P. Viens, J. Jacquemier, E. Charafe-Jauffret, M. Chaffanet, D. Birnbaum and F. Bertucci (2006). "Prognosis and gene expression profiling of 20q13-amplified breast cancers." Clin Cancer Res 12(15): 4533-44.
- Goldhirsch, A., J. H. Glick, R. D. Gelber, A. S. Coates, B. Thurlimann and H. J. Senn (2005). "Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005." Ann Oncol 16(10): 1569-83.
- 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 and E. S. Lander

- (1999). "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring." *Science* 286(5439): 531-7.
- Gomes, L. I., R. L. Silva, B. S. Stolf, E. B. Cristo, R. Hirata, F. A. Soares, L. F. Reis, E. J. Neves and A. F. Carvalho (2003). "Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray." *Anal Biochem* 321(2): 244-51.
- Goncalves, A., E. Charafe-Jauffret, F. Bertucci, S. Audebert, Y. Toiron, B. Esterni, F. Monville, C. Tarpin, J. Jacquemier, G. Houvenaeghel, C. Chabannon, J. M. Extra, P. Viens, J. P. Borg and D. Birnbaum (2008). "Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes." *Mol Cell Proteomics*.
- Goodrich, J. A. and J. F. Kugel (2006). "Non-coding-RNA regulators of RNA polymerase II transcription." *Nat Rev Mol Cell Biol* 7(8): 612-6.
- Gorreta, F., D. Barzaghi, A. J. VanMeter, V. Chandhoke and L. Del Giacco (2004). "Development of a new reference standard for microarray experiments." *Biotechniques* 36(6): 1002-9.
- Gresham, D., M. J. Dunham and D. Botstein (2008). "Comparing whole genomes using DNA microarrays." *Nat Rev Genet* 9(4): 291-302.
- Gruvberger, S., M. Ringner, Y. Chen, S. Panavally, L. H. Saal, A. Borg, M. Ferno, C. Peterson and P. S. Meltzer (2001). "Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns." *Cancer Res* 61(16): 5979-84.
- Hastings, M. L., H. A. Ingle, M. A. Lazar and S. H. Munroe (2000). "Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA." *J Biol Chem* 275(15): 11507-13.
- Hayes, D. F., R. C. Bast, C. E. Desch, H. Fritsche, Jr., N. E. Kemeny, J. M. Jessup, G. Y. Locker, J. S. Macdonald, R. G. Mennel, L. Norton, P. Ravdin, S. Taube and R. J. Winn (1996). "Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers." *J Natl Cancer Inst* 88(20): 1456-66.
- He, S., H. Su, C. Liu, G. Skogerbo, H. He, D. He, X. Zhu, T. Liu, Y. Zhao and R. Chen (2008). "MicroRNA-encoding long non-coding RNAs." *BMC Genomics* 9(1): 236.
- Hedenfalk, I., D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O. P. Kallioniemi, B. Wilfond, A. Borg, J. Trent, M. Raffeld, Z. Yakhini, A. Ben-Dor, E. Dougherty, J. Kononen, L. Bubendorf, W. Fehrle, S. Pittaluga, S. Gruvberger, N. Loman, O. Johannsson, H. Olsson and G. Sauter (2001). "Gene-expression profiles in hereditary breast cancer." *N Engl J Med* 344(8): 539-48.
- Heisterkamp, N., C. Morris and J. Groffen (1989). "ABR, an active BCR-related gene." *Nucleic Acids Res* 17(21): 8821-31.
- Helias, C., S. Struski, C. Gervais, V. Leymarie, L. Mauvieux, R. Herbrecht and M. Lessard (2008). "Polycythemia vera transforming to acute myeloid leukemia and complex abnormalities including 9p homogeneously staining region with amplification of MLLT3, JMJD2C, JAK2, and SMARCA2." *Cancer Genet Cytogenet* 180(1): 51-5.
- Hellstrom, I., J. Raycraft, M. Hayden-Ledbetter, J. A. Ledbetter, M. Schummer, M. McIntosh, C. Drescher, N. Urban and K. E. Hellstrom (2003). "The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma." *Cancer Res* 63(13): 3695-700.
- Hirotsune, S., N. Yoshida, A. Chen, L. Garrett, F. Sugiyama, S. Takahashi, K. Yagami, A. Wynshaw-Boris and A. Yoshiki (2003). "An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene." *Nature* 423(6935): 91-6.
- Hore, T. A., R. W. Rapkins and J. A. Graves (2007). "Construction and evolution of imprinted loci in mammals." *Trends Genet* 23(9): 440-8.
- 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. Stephanants, G. Cavet, W. L. Walker, A. West, E. Coffey, D. D. Shoemaker, R. Stoughton, A. P. Blanchard, S. H. Friend and P. S. Linsley (2001).

- "Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer." *Nat Biotechnol* 19(4): 342-7.
- Humar, M., S. E. Pischke, T. Loop, A. Hoetzel, R. Schmidt, C. Klaas, H. L. Pahl, K. K. Geiger and B. H. Pannen (2004). "Barbiturates directly inhibit the calmodulin/calcineurin complex: a novel mechanism of inhibition of nuclear factor of activated T cells." *Mol Pharmacol* 65(2): 350-61.
- Imamura, T., S. Yamamoto, J. Ohgane, N. Hattori, S. Tanaka and K. Shiota (2004). "Non-coding RNA directed DNA demethylation of Sphk1 CpG island." *Biochem Biophys Res Commun* 322(2): 593-600.
- Inokuma, M., C. dela Rosa, C. Schmitt, P. Haaland, J. Siebert, D. Petry, M. Tang, M. A. Suni, S. A. Ghanekar, D. Gladding, J. F. Dunne, V. C. Maino, M. L. Disis and H. T. Maecker (2007). "Functional T cell responses to tumor antigens in breast cancer patients have a distinct phenotype and cytokine signature." *J Immunol* 179(4): 2627-33.
- Iorio, M. V., M. Ferracin, C. G. Liu, A. Veronese, R. Spizzo, S. Sabbioni, E. Magri, M. Pedriali, M. Fabbri, M. Campiglio, S. Menard, J. P. Palazzo, A. Rosenberg, P. Musiani, S. Volinia, I. Nenci, G. A. Calin, P. Querzoli, M. Negrini and C. M. Croce (2005). "MicroRNA gene expression deregulation in human breast cancer." *Cancer Res* 65(16): 7065-70.
- Italiano, A., R. Attias, A. Aurias, G. Perot, F. Burel-Vandenbos, J. Otto, N. Venissac and F. Pedeutour (2006). "Molecular cytogenetic characterization of a metastatic lung sarcomatoid carcinoma: 9p23 neocentromere and 9p23-p24 amplification including JAK2 and JMJD2C." *Cancer Genet Cytogenet* 167(2): 122-30.
- Ivshina, A. V., J. George, O. Senko, B. Mow, T. C. Putti, J. Smeds, T. Lindahl, Y. Pawitan, P. Hall, H. Nordgren, J. E. Wong, E. T. Liu, J. Bergh, V. A. Kuznetsov and L. D. Miller (2006). "Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer." *Cancer Res* 66(21): 10292-301.
- Jemal, A., R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray and M. J. Thun (2008). "Cancer statistics, 2008." *CA Cancer J Clin* 58(2): 71-96.
- 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 and C. Muller-Tidow (2003). "MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer." *Oncogene* 22(39): 8031-41.
- Jones, N. P. and M. Katan (2007). "Role of phospholipase Cgamma1 in cell spreading requires association with a beta-Pix/GIT1-containing complex, leading to activation of Cdc42 and Rac1." *Mol Cell Biol* 27(16): 5790-805.
- Kampa, D., J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt, H. Tammana and T. R. Gingeras (2004). "Novel RNAs Identified From an In-Depth Analysis of the Transcriptome of Human Chromosomes 21 and 22." *Genome Res* 14(3): 331-342.
- Kaposi-Novak, P., J. S. Lee, A. Mikaelyan, V. Patel and S. S. Thorgeirsson (2004). "Oligonucleotide microarray analysis of aminoallyl-labeled cDNA targets from linear RNA amplification." *Biotechniques* 37(4): 580, 582-6, 588.
- Kapranov, P., S. E. Cawley, J. Drenkow, S. Bekiranov, R. L. Strausberg, S. P. Fodor and T. R. Gingeras (2002). "Large-scale transcriptional activity in chromosomes 21 and 22." *Science* 296(5569): 916-9.
- Kapranov, P., J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermuller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana and T. R. Gingeras (2007). "RNA maps reveal new RNA classes and a possible function for pervasive transcription." *Science* 316(5830): 1484-8.
- Kapranov, P., J. Drenkow, J. Cheng, J. Long, G. Helt, S. Dike and T. R. Gingeras (2005). "Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays." *Genome Res* 15(7): 987-97.

- Kasami, M., T. Uematsu, M. Honda, T. Yabuzaki, J. Sanuki, Y. Uchida and H. Sugimura (2008). "Comparison of estrogen receptor, progesterone receptor and Her-2 status in breast cancer pre- and post-neoadjuvant chemotherapy." Breast.
- Katoh, Y. and M. Katoh (2007). "Comparative integromics on JMJD2A, JMJD2B and JMJD2C: preferential expression of JMJD2C in undifferentiated ES cells." Int J Mol Med 20(2): 269-73.
- Kawauchi, J., C. Zhang, K. Nobori, Y. Hashimoto, M. T. Adachi, A. Noda, M. Sunamori and S. Kitajima (2002). "Transcriptional repressor activating transcription factor 3 protects human umbilical vein endothelial cells from tumor necrosis factor-alpha-induced apoptosis through down-regulation of p53 transcription." J Biol Chem 277(41): 39025-34.
- Kerr, M. K. and G. A. Churchill (2001). "Statistical design and the analysis of gene expression microarray data." Genet Res 77(2): 123-8.
- Khanna, K. K. (2000). "Cancer risk and the ATM gene: a continuing debate." J Natl Cancer Inst 92(10): 795-802.
- Kim, H., B. Zhao, E. C. Snestrud, B. J. Haas, C. D. Town and J. Quackenbush (2002). "Use of RNA and genomic DNA references for inferred comparisons in DNA microarray analyses." Biotechniques 33(4): 924-30.
- Kim, S. H., A. P. Forman, M. B. Mathews and S. Gunnery (2000). "Human breast cancer cells contain elevated levels and activity of the protein kinase, PKR." Oncogene 19(27): 3086-94.
- Kiyosawa, H., I. Yamanaka, N. Osato, S. Kondo and Y. Hayashizaki (2003). "Antisense transcripts with FANTOM2 clone set and their implications for gene regulation." Genome Res 13(6B): 1324-34.
- Kobayashi, S. D., J. M. Voyich, K. R. Braughton and F. R. DeLeo (2003). "Down-regulation of proinflammatory capacity during apoptosis in human polymorphonuclear leukocytes." J Immunol 170(6): 3357-68.
- Koscielny, S. (2008). "Critical review of microarray-based prognostic tests and trials in breast cancer." Curr Opin Obstet Gynecol 20(1): 47-50.
- Kuwabara, T., J. Hsieh, K. Nakashima, K. Taira and F. H. Gage (2004). "A small modulatory dsRNA specifies the fate of adult neural stem cells." Cell 116(6): 779-93.
- Lahousse, S. A., J. J. Carter, X. J. Xu, J. R. Wands and S. M. de la Monte (2006). "Differential growth factor regulation of aspartyl-(asparaginy)-beta-hydroxylase family genes in SH-Sy5y human neuroblastoma cells." BMC Cell Biol 7: 41.
- Lange, T., V. Roth, M. L. Braun and J. M. Buhmann (2004). "Stability-based validation of clustering solutions." Neural Comput 16(6): 1299-323.
- Lanz, R. B., N. J. McKenna, S. A. Onate, U. Albrecht, J. Wong, S. Y. Tsai, M. J. Tsai and B. W. O'Malley (1999). "A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex." Cell 97(1): 17-27.
- Lauria, R., F. Perrone, C. Carlomagno, M. De Laurentiis, A. Morabito, C. Gallo, E. Varriale, G. Pettinato, L. Panico, G. Petrella and et al. (1995). "The prognostic value of lymphatic and blood vessel invasion in operable breast cancer." Cancer 76(10): 1772-8.
- Lee, H., L. Chouinard, M. Bonin and R. N. Michel (2005). "NFATc3 deficiency may contribute to the development of mammary gland adenocarcinoma in aging female mice." Mol Carcinog 44(3): 219-22.
- Lee, J. S., K. M. Collins, A. L. Brown, C. H. Lee and J. H. Chung (2000). "hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response." Nature 404(6774): 201-4.
- Lehmann, U., B. Hasemeier, M. Christgen, M. Muller, D. Romermann, F. Langer and H. Kreipe (2008). "Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer." J Pathol 214(1): 17-24.
- Liaw, D., D. J. Marsh, J. Li, P. L. Dahia, S. I. Wang, Z. Zheng, S. Bose, K. M. Call, H. C. Tsou, M. Peacocke, C. Eng and R. Parsons (1997). "Germline mutations of the

- PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome." Nat Genet 16(1): 64-7.
- Lim, W., N. Hearle, B. Shah, V. Murday, S. V. Hodgson, A. Lucassen, D. Eccles, I. Talbot, K. Neale, A. G. Lim, J. O'Donohue, A. Donaldson, R. C. Macdonald, I. D. Young, M. H. Robinson, P. W. Lee, B. J. Stoodley, I. Tomlinson, D. Alderson, A. G. Holbrook, S. Vyas, E. T. Swarbrick, A. A. Lewis, R. K. Phillips and R. S. Houlston (2003). "Further observations on LKB1/STK11 status and cancer risk in Peutz-Jeghers syndrome." Br J Cancer 89(2): 308-13.
- Lin, R., S. Maeda, C. Liu, M. Karin and T. S. Edgington (2007). "A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas." Oncogene 26(6): 851-8.
- Liu, R., X. Wang, G. Y. Chen, P. Dalerba, A. Gurney, T. Hoey, G. Sherlock, J. Lewicki, K. Shedden and M. F. Clarke (2007). "The prognostic role of a gene signature from tumorigenic breast-cancer cells." N Engl J Med 356(3): 217-26.
- Loi, S., M. Buyse, C. Sotiriou and F. Cardoso (2004). "Challenges in breast cancer clinical trial design in the postgenomic era." Curr Opin Oncol 16(6): 536-41.
- Loi, S., B. Haibe-Kains, C. Desmedt, F. Lallemand, A. M. Tutt, C. Gillet, P. Ellis, A. Harris, J. Bergh, J. A. Foekens, J. G. Klijn, D. Larsimont, M. Buyse, G. Bontempi, M. Delorenzi, M. J. Piccart and C. Sotiriou (2007). "Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade." J Clin Oncol 25(10): 1239-46.
- Look, M. P., W. L. van Putten, M. J. Duffy, N. Harbeck, I. J. Christensen, C. Thomssen, R. Kates, F. Spyrtos, M. Ferno, S. Eppenberger-Castori, C. G. Sweep, K. Ulm, J. P. Peyrat, P. M. Martin, H. Magdelenat, N. Brunner, C. Duggan, B. W. Lisboa, P. O. Bendahl, V. Quillien, A. Daver, G. Ricolleau, M. E. Meijer-van Gelder, P. Manders, W. E. Fiets, M. A. Blankenstein, P. Broet, S. Romain, G. Daxenbichler, G. Windbichler, T. Cufer, S. Borstnar, W. Kueng, L. V. Beex, J. G. Klijn, N. O'Higgins, U. Eppenberger, F. Janicke, M. Schmitt and J. A. Foekens (2002). "Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients." J Natl Cancer Inst 94(2): 116-28.
- Louro, R., T. El-Jundi, H. I. Nakaya, E. M. Reis and S. Verjovski-Almeida (2008). "Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci." Genomics.
- Louro, R., H. I. Nakaya, P. P. Amaral, F. Festa, M. C. Sogayar, A. M. da Silva, S. Verjovski-Almeida and E. M. Reis (2007). "Androgen responsive intronic non-coding RNAs." BMC Biol 5(1): 4.
- Lower, E. E., E. L. Glass, D. A. Bradley, R. Blau and S. Heffelfinger (2005). "Impact of metastatic estrogen receptor and progesterone receptor status on survival." Breast Cancer Res Treat 90(1): 65-70.
- Lu, J., G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz and T. R. Golub (2005). "MicroRNA expression profiles classify human cancers." Nature 435(7043): 834-8.
- Lynch, H. T., E. Silva, C. Snyder and J. F. Lynch (2008). "Hereditary breast cancer: part I. Diagnosing hereditary breast cancer syndromes." Breast J 14(1): 3-13.
- Ma, J., D. J. Morrow, J. Fernandes and V. Walbot (2006). "Comparative profiling of the sense and antisense transcriptome of maize lines." Genome Biol 7(3): R22.
- Malkin, D., F. P. Li, L. C. Strong, J. F. Fraumeni, Jr., C. E. Nelson, D. H. Kim, J. Kassel, M. A. Gryka, F. Z. Bischoff, M. A. Tainsky and et al. (1990). "Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms." Science 250(4985): 1233-8.
- Mallon, E., P. Osin, N. Nasiri, I. Blain, B. Howard and B. Gusterson (2000). "The basic pathology of human breast cancer." J Mammary Gland Biol Neoplasia 5(2): 139-63.

- Mallory, J. C., G. Crudden, A. Oliva, C. Saunders, A. Stromberg and R. J. Craven (2005). "A novel group of genes regulates susceptibility to antineoplastic drugs in highly tumorigenic breast cancer cells." Mol Pharmacol 68(6): 1747-56.
- Maor, S., D. Mayer, R. I. Yarden, A. V. Lee, R. Sarfstein, H. Werner and M. Z. Papa (2006). "Estrogen receptor regulates insulin-like growth factor-I receptor gene expression in breast tumor cells: involvement of transcription factor Sp1." J Endocrinol 191(3): 605-12.
- Marmur, J. and P. Doty (1961). "Thermal renaturation of deoxyribonucleic acids." J Mol Biol 3: 585-94.
- Maskos, U. and E. M. Southern (1992). "Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ." Nucleic Acids Res 20(7): 1679-84.
- Masri, S., S. Phung, X. Wang, X. Wu, Y. C. Yuan, L. Wagman and S. Chen (2008). "Genome-wide analysis of aromatase inhibitor-resistant, tamoxifen-resistant, and long-term estrogen-deprived cells reveals a role for estrogen receptor." Cancer Res 68(12): 4910-8.
- Masumoto, J., K. Yang, S. Varambally, M. Hasegawa, S. A. Tomlins, S. Qiu, Y. Fujimoto, A. Kawasaki, S. J. Foster, Y. Horie, T. W. Mak, G. Nunez, A. M. Chinnaiyan, K. Fukase and N. Inohara (2006). "Nod1 acts as an intracellular receptor to stimulate chemokine production and neutrophil recruitment in vivo." J Exp Med 203(1): 203-13.
- Mattick, J. S. and I. V. Makunin (2005). "Small regulatory RNAs in mammals." Hum Mol Genet 14 Spec No 1: R121-32.
- Mattick, J. S. and I. V. Makunin (2006). "Non-coding RNA." Hum Mol Genet 15 Suppl 1: R17-29.
- Mavrou, A., G. T. Tsangaris, E. Roma and A. Kolialexi (2008). "The ATM gene and ataxia telangiectasia." Anticancer Res 28(1B): 401-5.
- McGraw, L. A., A. G. Clark and M. F. Wolfner (2008). "Post-mating gene expression profiles of female *Drosophila melanogaster* in response to time and to four male accessory gland proteins." Genetics.
- Mehra, R., S. Varambally, L. Ding, R. Shen, M. S. Sabel, D. Ghosh, A. M. Chinnaiyan and C. G. Kleer (2005). "Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis." Cancer Res 65(24): 11259-64.
- Meijers-Heijboer, H., A. van den Ouweland, J. Klijn, M. Wasielewski, A. de Snoo, R. Oldenburg, A. Hollestelle, M. Houben, E. Crepin, M. van Veghel-Plandsoen, F. Elstrodt, C. van Duijn, C. Bartels, C. Meijers, M. Schutte, L. McGuffog, D. Thompson, D. Easton, N. Sodha, S. Seal, R. Barfoot, J. Mangion, J. Chang-Claude, D. Eccles, R. Eeles, D. G. Evans, R. Houlston, V. Murday, S. Narod, T. Peretz, J. Peto, C. Phelan, H. X. Zhang, C. Szabo, P. Devilee, D. Goldgar, P. A. Futreal, K. L. Nathanson, B. Weber, N. Rahman and M. R. Stratton (2002). "Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations." Nat Genet 31(1): 55-9.
- Menard, S., A. Balsari, P. Casalini, E. Tagliabue, M. Campiglio, R. Bufalino and N. Cascinelli (2002). "HER-2-positive breast carcinomas as a particular subset with peculiar clinical behaviors." Clin Cancer Res 8(2): 520-5.
- Meulmeester, E. and A. G. Jochemsen (2008). "p53: a guide to apoptosis." Curr Cancer Drug Targets 8(2): 87-97.
- Miller, L. D., J. Smeds, J. George, V. B. Vega, L. Vergara, A. Ploner, Y. Pawitan, P. Hall, S. Klaar, E. T. Liu and J. Bergh (2005). "An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival." Proc Natl Acad Sci U S A 102(38): 13550-5.
- Minn, A. J., G. P. Gupta, P. M. Siegel, P. D. Bos, W. Shu, D. D. Giri, A. Viale, A. B. Olshen, W. L. Gerald and J. Massague (2005). "Genes that mediate breast cancer metastasis to lung." Nature 436(7050): 518-24.

- Molyneux, G., J. Regan and M. J. Smalley (2007). "Mammary stem cells and breast cancer." Cell Mol Life Sci 64(24): 3248-60.
- Mosley, J. D. and R. A. Keri (2008). "Cell cycle correlated genes dictate the prognostic power of breast cancer gene lists." BMC Med Genomics 1: 11.
- Mueller, P., J. Massner, R. Jayachandran, B. Combaluzier, I. Albrecht, J. Gatfield, C. Blum, R. Ceredig, H. R. Rodewald, A. G. Rolink and J. Pieters (2008). "Regulation of T cell survival through coronin-1-mediated generation of inositol-1,4,5-trisphosphate and calcium mobilization after T cell receptor triggering." Nat Immunol 9(4): 424-31.
- Naderi, A., A. E. Teschendorff, N. L. Barbosa-Morais, S. E. Pinder, A. R. Green, D. G. Powe, J. F. Robertson, S. Aparicio, I. O. Ellis, J. D. Brenton and C. Caldas (2007). "A gene-expression signature to predict survival in breast cancer across independent data sets." Oncogene 26(10): 1507-16.
- Naderi, A., A. E. Teschendorff, J. Beigel, M. Cariati, I. O. Ellis, J. D. Brenton and C. Caldas (2007). "BEX2 is overexpressed in a subset of primary breast cancers and mediates nerve growth factor/nuclear factor-kappaB inhibition of apoptosis in breast cancer cell lines." Cancer Res 67(14): 6725-36.
- 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 and S. Verjovski-Almeida (2007). "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 8(3): R43.
- Neville, A. M., R. Bettelheim, R. D. Gelber, J. Save-Soderbergh, B. W. Davis, R. Reed, J. Torhorst, R. Golouh, H. F. Peterson, K. N. Price and et al. (1992). "Factors predicting treatment responsiveness and prognosis in node-negative breast cancer. The International (Ludwig) Breast Cancer Study Group." J Clin Oncol 10(5): 696-705.
- Newman, L. A. and S. E. Singletary (2007). "Overview of adjuvant systemic therapy in early stage breast cancer." Surg Clin North Am 87(2): 499-509, xi.
- Ng, P., C. L. Wei, W. K. Sung, K. P. Chiu, L. Lipovich, C. C. Ang, S. Gupta, A. Shahab, A. Ridwan, C. H. Wong, E. T. Liu and Y. Ruan (2005). "Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation." Nat Methods 2(2): 105-11.
- Ooe, A., K. Kato and S. Noguchi (2007). "Possible involvement of CCT5, RGS3, and YKT6 genes up-regulated in p53-mutated tumors in resistance to docetaxel in human breast cancers." Breast Cancer Res Treat 101(3): 305-15.
- Ouyang, W., S. H. Ranganath, K. Weindel, D. Bhattacharya, T. L. Murphy, W. C. Sha and K. M. Murphy (1998). "Inhibition of Th1 development mediated by GATA-3 through an IL-4-independent mechanism." Immunity 9(5): 745-55.
- Ozdogan, H., A. E. Teschendorff, A. A. Ahmed, S. J. Hyland, C. Blenkiron, L. Bobrow, A. Veerakumarasivam, G. Burt, T. Subkhankulova, M. J. Arends, V. P. Collins, D. Bowtell, T. Kouzarides, J. D. Brenton and C. Caldas (2006). "Differential expression of selected histone modifier genes in human solid cancers." BMC Genomics 7: 90.
- Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant and N. Wolmark (2004). "A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer." N Engl J Med 351(27): 2817-26.
- Parfitt, J. R., C. Armstrong, F. O'Malley, J. Ross and A. B. Tuck (2004). "In-situ and invasive carcinoma within a phyllodes tumor associated with lymph node metastases." World J Surg Oncol 2: 46.
- Park, E. S., J. S. Lee, H. G. Woo, F. Zhan, J. H. Shih, J. D. Shaughnessy, Jr. and J. Frederic Mushinski (2007). "Heterologous tissue culture expression signature predicts human breast cancer prognosis." PLoS ONE 2(1): e145.

- Pauler, F. M., M. V. Koerner and D. P. Barlow (2007). "Silencing by imprinted noncoding RNAs: is transcription the answer?" Trends Genet 23(6): 284-92.
- Pavlidis, P., Q. Li and W. S. Noble (2003). "The effect of replication on gene expression microarray experiments." Bioinformatics 19(13): 1620-7.
- Payne, S. J., R. L. Bowen, J. L. Jones and C. A. Wells (2008). "Predictive markers in breast cancer--the present." Histopathology 52(1): 82-90.
- Pease, A. C., D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes and S. P. Fodor (1994). "Light-generated oligonucleotide arrays for rapid DNA sequence analysis." Proc Natl Acad Sci U S A 91(11): 5022-6.
- Peixoto, B. R., R. Z. Vencio, C. M. Egidio, L. Mota-Vieira, S. Verjovski-Almeida and E. M. Reis (2006). "Evaluation of reference-based two-color methods for measurement of gene expression ratios using spotted cDNA microarrays." BMC Genomics 7: 35.
- Perez, E. A. (2001). "The role of adjuvant monoclonal antibody therapy for breast cancer: rationale and new studies." Curr Oncol Rep 3(6): 516-22.
- Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein (2000). "Molecular portraits of human breast tumours." Nature 406(6797): 747-52.
- Petitjean, A., M. I. Achatz, A. L. Borresen-Dale, P. Hainaut and M. Olivier (2007). "TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes." Oncogene 26(15): 2157-65.
- Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res. 29(9): e45.
- Pollack, J. R., C. M. Perou, A. A. Alizadeh, M. B. Eisen, A. Pergamenschikov, C. F. Williams, S. S. Jeffrey, D. Botstein and P. O. Brown (1999). "Genome-wide analysis of DNA copy-number changes using cDNA microarrays." Nat Genet 23(1): 41-6.
- Prasanth, K. V. and D. L. Spector (2007). "Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum." Genes Dev 21(1): 11-42.
- Prescott, E. M. and N. J. Proudfoot (2002). "Transcriptional collision between convergent genes in budding yeast." Proc Natl Acad Sci U S A 99(13): 8796-801.
- Pritchard, K. I., L. E. Shepherd, F. P. O'Malley, I. L. Andrulis, D. Tu, V. H. Bramwell and M. N. Levine (2006). "HER2 and responsiveness of breast cancer to adjuvant chemotherapy." N Engl J Med 354(20): 2103-11.
- Pusztai, L. (2008). "Current status of prognostic profiling in breast cancer." Oncologist 13(4): 350-60.
- Quackenbush, J. (2002). "Microarray data normalization and transformation." Nat. Genet. 32 Suppl: 496-501.
- Radvanyi, L., D. Singh-Sandhu, S. Gallichan, C. Lovitt, A. Pedyczak, G. Mallo, K. Gish, K. Kwok, W. Hanna, J. Zubovits, J. Armes, D. Venter, J. Hakimi, J. Shortreed, M. Donovan, M. Parrington, P. Dunn, R. Oomen, J. Tartaglia and N. L. Berinstein (2005). "The gene associated with trichorhinophalangeal syndrome in humans is overexpressed in breast cancer." Proc Natl Acad Sci U S A 102(31): 11005-10.
- Rakha, E. A., M. E. El-Sayed, J. Reis-Filho and I. O. Ellis (2008). "Patho-biological aspects of basal-like breast cancer." Breast Cancer Res Treat.
- Ramaswamy, S., P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander and T. R. Golub (2001). "Multiclass cancer diagnosis using tumor gene expression signatures." Proc Natl Acad Sci U S A 98(26): 15149-54.
- Rampaul, R. S., S. E. Pinder, C. W. Elston and I. O. Ellis (2001). "Prognostic and predictive factors in primary breast cancer and their role in patient management: The Nottingham Breast Team." Eur J Surg Oncol 27(3): 229-38.
- Ransohoff, D. F. (2004). "Rules of evidence for cancer molecular-marker discovery and validation." Nat Rev Cancer 4(4): 309-14.
- Rao, A., C. Luo and P. G. Hogan (1997). "Transcription factors of the NFAT family: regulation and function." Annu Rev Immunol 15: 707-47.

- Reis, E. M., R. Louro, H. I. Nakaya and S. Verjovski-Almeida (2005). "As antisense RNA gets intronic." *Omics* 9(1): 2-12.
- 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 and S. Verjovski-Almeida (2004). "Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer." *Oncogene* 23(39): 6684-6692.
- Religio, A., C. Ben-Dov, M. Baum, M. Ruggiu, C. Gemund, V. Benes, R. B. Darnell and J. Valcarcel (2005). "Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells." *J Biol Chem* 280(6): 4779-84.
- Rhodes, D. R., S. Kalyana-Sundaram, V. Mahavisno, R. Varambally, J. Yu, B. B. Briggs, T. R. Barrette, M. J. Anstet, C. Kincead-Beal, P. Kulkarni, S. Varambally, D. Ghosh and A. M. Chinnaiyan (2007). "Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles." *Neoplasia* 9(2): 166-80.
- Richardson, A. L., Z. C. Wang, A. De Nicolo, X. Lu, M. Brown, A. Miron, X. Liao, J. D. Iglehart, D. M. Livingston and S. Ganesan (2006). "X chromosomal abnormalities in basal-like human breast cancer." *Cancer Cell* 9(2): 121-32.
- Rizki, A., V. M. Weaver, S. Y. Lee, G. I. Rozenberg, K. Chin, C. A. Myers, J. L. Bascom, J. D. Mott, J. R. Semeiks, L. R. Grate, I. S. Mian, A. D. Borowsky, R. A. Jensen, M. O. Idowu, F. Chen, D. J. Chen, O. W. Petersen, J. W. Gray and M. J. Bissell (2008). "A human breast cell model of preinvasive to invasive transition." *Cancer Res* 68(5): 1378-87.
- Rugo, H. S. (2008). "The importance of distant metastases in hormone-sensitive breast cancer." *Breast* 17 Suppl 1: S3-8.
- Saha, S., A. B. Sparks, C. Rago, V. Akmaev, C. J. Wang, B. Vogelstein, K. W. Kinzler and V. E. Velculescu (2002). "Using the transcriptome to annotate the genome." *Nat Biotechnol* 20(5): 508-12.
- Salerno, V. P., A. Calliari, D. W. Provance, Jr., J. R. Sotelo-Silveira, J. R. Sotelo and J. A. Mercer (2008). "Myosin-Va mediates RNA distribution in primary fibroblasts from multiple organs." *Cell Motil Cytoskeleton* 65(5): 422-33.
- Sasaki, Y. T., M. Sano, T. Ideue, T. Kin, K. Asai and T. Hirose (2007). "Identification and characterization of human non-coding RNAs with tissue-specific expression." *Biochem Biophys Res Commun* 357(4): 991-6.
- Schena, M., D. Shalon, R. W. Davis and P. O. Brown (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." *Science* 270(5235): 467-70.
- Schmidt, M., D. Bohm, C. von Torne, E. Steiner, A. Puhl, H. Pilch, H. A. Lehr, J. G. Hengstler, H. Kolbl and M. Gehrman (2008). "The humoral immune system has a key prognostic impact in node-negative breast cancer." *Cancer Res* 68(13): 5405-13.
- Sellar, G. C., K. P. Watt, G. J. Rabiasz, E. A. Stronach, L. Li, E. P. Miller, C. E. Massie, J. Miller, B. Contreras-Moreira, D. Scott, I. Brown, A. R. Williams, P. A. Bates, J. F. Smyth and H. Gabra (2003). "OPCML at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer." *Nat Genet* 34(3): 337-43.
- Sgroi, D. C., S. Teng, G. Robinson, R. LeVangie, J. R. Hudson, Jr. and A. G. Elkhouloun (1999). "In vivo gene expression profile analysis of human breast cancer progression." *Cancer Res* 59(22): 5656-61.
- Shalon, D., S. J. Smith and P. O. Brown (1996). "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization." *Genome Res* 6(7): 639-45.

- Shedden, K. A., J. M. Taylor, T. J. Giordano, R. Kuick, D. E. Misek, G. Rennert, D. R. Schwartz, S. B. Gruber, C. Logsdon, D. Simeone, S. L. Kardia, J. K. Greenon, K. R. Cho, D. G. Beer, E. R. Fearon and S. Hanash (2003). "Accurate molecular classification of human cancers based on gene expression using a simple classifier with a pathological tree-based framework." Am J Pathol 163(5): 1985-95.
- Shen, D., H. R. Chang, Z. Chen, J. He, V. Lonsberry, Y. Elshimali, D. Chia, D. Seligson, L. Goodglick, S. F. Nelson and J. A. Gornbein (2005). "Loss of annexin A1 expression in human breast cancer detected by multiple high-throughput analyses." Biochem Biophys Res Commun 326(1): 218-27.
- Shendure, J. and G. M. Church (2002). "Computational discovery of sense-antisense transcription in the human and mouse genomes." Genome Biol. 3(9): research0044
- Shi, Y., L. He, Y. Chen and J. Li (1997). "The correlations of clinicopathologic features, hormone receptor level with the prognosis of breast cancer: report of 308 cases in elderly women." Chin Med Sci J 12(1): 60-2.
- Sidders, B., M. Withers, S. L. Kendall, J. Bacon, S. J. Waddell, J. Hinds, P. Golby, F. Movahedzadeh, R. A. Cox, R. Frita, A. M. Ten Bokum, L. Wernisch and N. G. Stoker (2007). "Quantification of global transcription patterns in prokaryotes using spotted microarrays." Genome Biol 8(12): R265.
- Sidransky, D., T. Tokino, K. Helzlsouer, B. Zehnbauer, G. Rausch, B. Shelton, L. Prestigiacomo, B. Vogelstein and N. Davidson (1992). "Inherited p53 gene mutations in breast cancer." Cancer Res 52(10): 2984-6.
- Silverman, T. A., M. Noguchi and B. Safer (1992). "Role of sequences within the first intron in the regulation of expression of eukaryotic initiation factor 2 alpha." J Biol Chem 267(14): 9738-42.
- Simon, R. M. and K. Dobbin (2003). "Experimental design of DNA microarray experiments." Biotechniques Suppl: 16-21.
- Singletary, S. E., C. Allred, P. Ashley, L. W. Bassett, D. Berry, K. I. Bland, P. I. Borgen, G. Clark, S. B. Edge, D. F. Hayes, L. L. Hughes, R. V. Hutter, M. Morrow, D. L. Page, A. Recht, R. L. Theriault, A. Thor, D. L. Weaver, H. S. Wieand and F. L. Greene (2002). "Revision of the American Joint Committee on Cancer staging system for breast cancer." J Clin Oncol 20(17): 3628-36.
- Singletary, S. E. and J. L. Connolly (2006). "Breast cancer staging: working with the sixth edition of the AJCC Cancer Staging Manual." CA Cancer J Clin 56(1): 37-47; quiz 50-1.
- Smid, M., Y. Wang, Y. Zhang, A. M. Sieuwerts, J. Yu, J. G. Klijn, J. A. Foekens and J. W. Martens (2008). "Subtypes of breast cancer show preferential site of relapse." Cancer Res 68(9): 3108-14.
- Soerjomataram, I., M. W. Louwman, J. G. Ribot, J. A. Roukema and J. W. Coebergh (2008). "An overview of prognostic factors for long-term survivors of breast cancer." Breast Cancer Res Treat 107(3): 309-30.
- Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. Eystein Lonning and A. L. Borresen-Dale (2001). "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." Proc Natl Acad Sci U S A 98(19): 10869-74.
- Sorlie, T., R. Tibshirani, J. Parker, T. Hastie, J. S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, J. Demeter, C. M. Perou, P. E. Lonning, P. O. Brown, A. L. Borresen-Dale and D. Botstein (2003). "Repeated observation of breast tumor subtypes in independent gene expression data sets." Proc Natl Acad Sci U S A 100(14): 8418-23.
- Sotiriou, C. and M. J. Piccart (2007). "Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care?" Nat Rev Cancer 7(7): 545-53.
- Sotiriou, C., P. Wirapati, S. Loi, A. Harris, S. Fox, J. Smeds, H. Nordgren, P. Farmer, V. Praz, B. Haibe-Kains, C. Desmedt, D. Larsimont, F. Cardoso, H. Peterse, D.

- Nuyten, M. Buyse, M. J. Van de Vijver, J. Bergh, M. Piccart and M. Delorenzi (2006). "Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis." J Natl Cancer Inst 98(4): 262-72.
- Stoyanova, R., J. J. Upson, C. Patriotis, E. A. Ross, E. P. Henske, K. Datta, B. Boman, M. L. Clapper, A. G. Knudson and A. Bellacosa (2004). "Use of RNA amplification in the optimal characterization of global gene expression using cDNA microarrays." J Cell Physiol 201(3): 359-65.
- Struewing, J. P., P. Hartge, S. Wacholder, S. M. Baker, M. Berlin, M. McAdams, M. M. Timmerman, L. C. Brody and M. A. Tucker (1997). "The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews." N Engl J Med 336(20): 1401-8.
- Subramaniam, D. S. and C. Isaacs (2005). "Utilizing prognostic and predictive factors in breast cancer." Curr Treat Options Oncol 6(2): 147-59.
- Sugiyama, R., Y. Kazama, Y. Miyazawa, S. Matsunaga and S. Kawano (2003). "CCLS96.1, a member of a multicopy gene family, may encode a non-coding RNA preferentially transcribed in reproductive organs of *Silene latifolia*." DNA Res 10(5): 213-20.
- t Hoen, P. A., R. Turk, J. M. Boer, E. Sterrenburg, R. X. de Menezes, G. J. van Ommen and J. T. den Dunnen (2004). "Intensity-based analysis of two-colour microarrays enables efficient and flexible hybridization designs." Nucleic Acids Res 32(4): e41.
- Tan, J., X. Yang, L. Zhuang, X. Jiang, W. Chen, P. L. Lee, R. K. Karuturi, P. B. Tan, E. T. Liu and Q. Yu (2007). "Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells." Genes Dev 21(9): 1050-63.
- Tavazoie, S. F., C. Alarcon, T. Oskarsson, D. Padua, Q. Wang, P. D. Bos, W. L. Gerald and J. Massague (2008). "Endogenous human microRNAs that suppress breast cancer metastasis." Nature 451(7175): 147-52.
- Teodoridis, J. M., J. Hall, S. Marsh, H. D. Kannall, C. Smyth, J. Curto, N. Siddiqui, H. Gabra, H. L. McLeod, G. Strathdee and R. Brown (2005). "CpG island methylation of DNA damage response genes in advanced ovarian cancer." Cancer Res 65(19): 8961-7.
- Tiwari, V. K., L. Cope, K. M. McGarvey, J. E. Ohm and S. B. Baylin (2008). "A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations." Genome Res.
- Trevino, V., F. Falciani and H. A. Barrera-Saldana (2007). "DNA microarrays: a powerful genomic tool for biomedical and clinical research." Mol Med 13(9-10): 527-41.
- Turashvili, G., J. Bouchal, K. Baumforth, W. Wei, M. Dziechciarkova, J. Ehrmann, J. Klein, E. Fridman, J. Skarda, J. Srovnal, M. Hajduch, P. Murray and Z. Kolar (2007). "Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis." BMC Cancer 7: 55.
- Tusher, V. G., R. Tibshirani and G. Chu (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A 98(9): 5116-21.
- Ullmannova, V. and N. C. Popescu (2006). "Expression profile of the tumor suppressor genes DLC-1 and DLC-2 in solid tumors." Int J Oncol 29(5): 1127-32.
- Uppal, S., D. Mistry and A. P. Coatesworth (2007). "Cowden disease: a review." Int J Clin Pract 61(4): 645-52.
- 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 and S. H. Friend (2002). "Gene expression profiling predicts clinical outcome of breast cancer." Nature 415(6871): 530-536.
- van de Vijver, M. J., Y. D. He, L. J. van't Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E.

- T. Rutgers, S. H. Friend and R. Bernardts (2002). "A gene-expression signature as a predictor of survival in breast cancer." N Engl J Med 347(25): 1999-2009.
- Van Gelder, R. N., M. E. von Zastrow, A. Yool, W. C. Dement, J. D. Barchas and J. H. Eberwine (1990). "Amplified RNA synthesized from limited quantities of heterogeneous cDNA." Proc. Natl. Acad. Sci. USA 87(5): 1663-1667.
- Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe and F. Speleman (2002). "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." Genome Biol 3(7): RESEARCH0034.
- Velculescu, V. E., B. Vogelstein and K. W. Kinzler (2000). "Analysing uncharted transcriptomes with SAGE." Trends Genet 16(10): 423-5.
- Velculescu, V. E., L. Zhang, B. Vogelstein and K. W. Kinzler (1995). "Serial analysis of gene expression." Science 270(5235): 484-7.
- Vencio, R. Z. and T. Koide (2005). "HTself: self-self based statistical test for low replication microarray studies." DNA Res 12(3): 211-4.
- Virtanen, C. and J. Woodgett (2008). "Clinical uses of microarrays in cancer research." Methods Mol Med 141: 87-113.
- Wang, D., L. Coscoy, M. Zylberberg, P. C. Avila, H. A. Boushey, D. Ganem and J. L. DeRisi (2002). "Microarray-based detection and genotyping of viral pathogens." Proc Natl Acad Sci U S A 99(24): 15687-92.
- Wang, E., L. D. Miller, G. A. Ohnmacht, E. T. Liu and F. M. Marincola (2000). "High-fidelity mRNA amplification for gene profiling." Nat Biotechnol 18(4): 457-9.
- Wang, S. M. (2007). "Understanding SAGE data." Trends Genet 23(1): 42-50.
- Wang, Y., J. G. Klijn, Y. Zhang, A. M. Sieuwerts, M. P. Look, F. Yang, D. Talantov, M. Timmermans, M. E. Meijer-van Gelder, J. Yu, T. Jatkoe, E. M. Berns, D. Atkins and J. A. Foekens (2005). "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer." Lancet 365(9460): 671-9.
- Welsh, P. L., K. N. Owens and M. C. King (2000). "Insights into the functions of BRCA1 and BRCA2." Trends Genet 16(2): 69-74.
- Welm, A. L., J. B. Sneddon, C. Taylor, D. S. Nuyten, M. J. van de Vijver, B. H. Hasegawa and J. M. Bishop (2007). "The macrophage-stimulating protein pathway promotes metastasis in a mouse model for breast cancer and predicts poor prognosis in humans." Proc Natl Acad Sci U S A 104(18): 7570-5.
- Whenham, N., V. D'Hondt and M. J. Piccart (2008). "HER2-positive breast cancer: from trastuzumab to innovatory anti-HER2 strategies." Clin Breast Cancer 8(1): 38-49.
- Willingham, A. T., A. P. Orth, S. Batalov, E. C. Peters, B. G. Wen, P. Aza-Blanc, J. B. Hogenesch and P. G. Schultz (2005). "A strategy for probing the function of noncoding RNAs finds a repressor of NFAT." Science 309(5740): 1570-3.
- Wilson, B. J. and V. Giguere (2008). "Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway." Mol Cancer 7: 49.
- Wong, S. L., C. Chao, M. J. Edwards, D. J. Carlson, A. Laidley, R. D. Noyes, T. McGlothlin, P. B. Ley, T. Tuttle, M. Schadt, R. Pennington, M. Legenza, J. Morgan and K. M. McMasters (2002). "Frequency of sentinel lymph node metastases in patients with favorable breast cancer histologic subtypes." Am J Surg 184(6): 492-8; discussion 498.
- Woo, C. J. and R. E. Kingston (2007). "HOTAIR lifts noncoding RNAs to new levels." Cell 129(7): 1257-9.
- Wutz, A. (2007). "Xist function: bridging chromatin and stem cells." Trends Genet 23(9): 457-64.
- Wutz, A., O. W. Smrzka, N. Schweifer, K. Schellander, E. F. Wagner and D. P. Barlow (1997). "Imprinted expression of the Igf2r gene depends on an intronic CpG island." Nature 389(6652): 745-9.
- Xu, L., A. C. Tan, R. L. Winslow and D. Geman (2008). "Merging microarray data from separate breast cancer studies provides a robust prognostic test." BMC Bioinformatics 9: 125.

- Yan, M. D., C. C. Hong, G. M. Lai, A. L. Cheng, Y. W. Lin and S. E. Chuang (2005). "Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas." Hum Mol Genet 14(11): 1465-74.
- Yan, P. S., C. M. Chen, H. Shi, F. Rahmatpanah, S. H. Wei, C. W. Caldwell and T. H. Huang (2001). "Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays." Cancer Res 61(23): 8375-80.
- Yang, P. K. and M. I. Kuroda (2007). "Noncoding RNAs and intranuclear positioning in monoallelic gene expression." Cell 128(4): 777-86.
- Yang, Y. H. and T. Speed (2002). "Design issues for cDNA microarray experiments." Nat Rev Genet 3(8): 579-88.
- Yang, Z. Q., I. Imoto, A. Pimkhaokham, Y. Shimada, K. Sasaki, M. Oka and J. Inazawa (2001). "A novel amplicon at 9p23 - 24 in squamous cell carcinoma of the esophagus that lies proximal to GASC1 and harbors NFIB." Jpn J Cancer Res 92(4): 423-8.
- Yelin, R., D. Dahary, R. Sorek, E. Y. Levanon, O. Goldstein, A. Shoshan, A. Diber, S. Biton, Y. Tamir, R. Khosravi, S. Nemzer, E. Pinner, S. Walach, J. Bernstein, K. Savitsky and G. Rotman (2003). "Widespread occurrence of antisense transcription in the human genome." Nat Biotechnol 21(4): 379-86.
- Yu, H., S. Zhu, B. Zhou, H. Xue and J. D. Han (2008). "Inferring causal relationships among different histone modifications and gene expression." Genome Res.
- Zeller, K. I., X. Zhao, C. W. Lee, K. P. Chiu, F. Yao, J. T. Yustein, H. S. Ooi, Y. L. Orlov, A. Shahab, H. C. Yong, Y. Fu, Z. Weng, V. A. Kuznetsov, W. K. Sung, Y. Ruan, C. V. Dang and C. L. Wei (2006). "Global mapping of c-Myc binding sites and target gene networks in human B cells." Proc Natl Acad Sci U S A 103(47): 17834-9.
- Zhang, B., X. Pan, G. P. Cobb and T. A. Anderson (2007). "microRNAs as oncogenes and tumor suppressors." Dev Biol 302(1): 1-12.

ANEXOS

- SÚMULA CURRICULAR DA AUTORA

Súmula Curricular da Autora

Camila de Moura Egidio

Contato

E-mail: c_egidio@iq.usp.br

camila.egidio@gmail.com

Formação

Doutorado em Bioquímica e Biologia Molecular,
Universidade de São Paulo, São Paulo, 2003 – presente.

Bacharel em Biologia, Universidade de São Paulo, São Paulo - 2002.

Financiamentos e Prêmios

“Scholar-in-Training Grant”, da Fundação Avon - AACR Oncogenomics 2007, Phoenix, AZ, USA – Fevereiro de 2008.

Bolsa de pesquisa da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) durante o período de doutorado para desenvolver pesquisa com tumores de mama humanos no laboratório do Dr. Eduardo Reis, 2004 - presente.

Bolsa de pesquisa da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) para desenvolver pesquisa com tumores de próstata humanos no laboratório da Dr. Aline da Silva, 2000 - 2002.

Publicações

Reis EM, Nakaya HI, Louro R, Canavez FC, Flatschart AV, Almeida GT, **Egidio CM**, Paquola AC, Machado AA, Festa F, Yamamoto D, Alvarenga R, da Silva CC, Brito GC, Simon SD, Moreira-Filho CA, Leite KR, Camara-Lopes LH, Campos FS, Gimba E, Vignal GM, El-Dorry H, Sogayar MC, Barcinski MA, da Silva AM, Verjovski-Almeida S.

- Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene*. 2004, 23(39):6684-92.

Peixoto BR, Vêncio RZN, **Egidio CM**, Mota-Vieira L, Verjovski-Almeida S, Reis EM

- Evaluation of reference-based two-color methods for measurement of gene expression ratios using spotted cDNA microarrays.

BMC Genomics, 2006, 7:35.

Camila M Egidio, Esther Durães, Helder I Nakaya, André Albergaria, Vitor Carneiro, Sergio Verjovski-Almeida,

Trabalhos Apresentados em Congressos

Bernardo R. Peixoto, Eduardo M. Reis
 - A breast cancer outcome classifier based on an expression signature enriched in non-coding RNAs
Manuscrito em preparação.

Camila M. Egidio, Esther O. Durães, Bernardo R. Peixoto, Sergio Verjovski-Almeida, Eduardo M. Reis.
 - Molecular Classification of Breast Tumors Using an Intronic cDNA Microarray.
Oncogenomics, Phoenix, AZ, EUA – 2007.

Camila M. Egidio, Esther O. Durães, Bernardo R. Peixoto, Sergio Verjovski-Almeida, Eduardo M. Reis.
 - Análise da expressão gênica de tumores de mama utilizando um microarray de cDNA enriquecido em transcritos intrônicos.
SBG, Foz do Iguaçu, PR - 2006.

Egidio, C., Vencio, R.Z.N., Mota-Vieira, L, Verjovski-Almeida, S, Peixoto, BR e Reis, EM.
 - Indirect measurement of differential gene expression using an oligonucleotide reference sample.
SBBQ, Águas de Lindóia, SP - 2005.

Camila Egidio, Ricardo Z.N. Vêncio, Luiza Mota-Vieira, Sergio Verjovski-Almeida, Bernardo R. Peixoto e Eduardo M. Reis.
 - Reconstruction of microarray expression intensities using a common oligonucleotide reference.
ICOBICOB I, Angra dos Reis, RJ - 2004.

Atividades Extras

Orientação da aluna de graduação Esther Durães durante seu período de Iniciação Científica no laboratório do Dr. Eduardo Reis e sob a sua supervisão, fevereiro de 2006 - presente.

Desenvolvimento e aplicação do curso “Análise da expressão gênica utilizando microarrays de DNA”, ministrado a alunos de pós-graduação com duração de 2 semanas e sob a supervisão do Dr. Eduardo Reis – julho de 2007.

Monitoria na matéria de Biologia Molecular para a Faculdade de Farmácia da USP, sob a supervisão da Dra. Aline Maria da Silva e Dra. Suely Lopes Gomes - 2006.

Desenvolvimento e aplicação do curso “Bioquímica dos Anexos Epidérmicos”, ministrado a alunos de graduação com duração de 1 semana e sob a supervisão do Dr. Bayardo Torres – janeiro de 2005.

Monitoria na matéria de Química de Biomoléculas para a Faculdade de Nutrição da USP, sob a supervisão da Dra. Marisa Medeiros e Dr. Felipe Chambergo - 2005.

Monitoria na matéria de Bioquímica para a Faculdade de Medicina da USP, sob a supervisão do Dr. Frederico Gueiros Filho e Dra. Adelaide Faljoni-Alário - 2004.

ANEXOS ELETRÔNICOS

- TRABALHOS PUBLICADOS:

1) Reis EM, Nakaya HI, Louro R, Canavez FC, Flatschart AV, Almeida GT, **Egidio CM**, Paquola AC, Machado AA, Festa F, Yamamoto D, Alvarenga R, da Silva CC, Brito GC, Simon SD, Moreira-Filho CA, Leite KR, Camara-Lopes LH, Campos FS, Gimba E, Vignal GM, El-Dorry H, Sogayar MC, Barcinski MA, da Silva AM, Verjovski-Almeida S.

- Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer.

Oncogene. 2004 Aug 26;23(39):6684-92.

2) Bernardo R Peixoto, Ricardo ZN Vêncio, **Camila M Egidio**, Luisa Mota-Vieira, Sergio Verjovski-Almeida, Eduardo M Reis

- Evaluation of reference-based two-color methods for measurement of gene expression ratios using spotted cDNA microarrays.

BMC Genomics. 2006 Feb 24;7:35.

3) **Camila M Egidio**, Esther Durães, André Albergaria, Teresa Eloy, Vitor Carneiro, Sergio Verjovski-Almeida, Bernardo R. Peixoto, Eduardo M. Reis

- A breast cancer outcome classifier based on an expression signature enriched in non-coding RNAs.

Manuscrito em preparação

Livros Grátis

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

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

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

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

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