

**IDENTIFICAÇÃO DE GENES SUPOSTAMENTE
ENVOLVIDOS COM O PROCESSO DE
INVASÃO TUMORAL EM CÂNCER DE MAMA ATRAVÉS
DA TÉCNICA RASH**

PAULO HENRIQUE BALDAN PINEDA

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

Área de concentração: Oncologia

Orientador: Dra Dirce Maria Carraro

Co-Orientador: Dra. Anamaria Aranha Camargo

São Paulo

2007

FICHA CATALOGRÁFICA

**Preparada pela Biblioteca do Centro de Tratamento e Pesquisa
Hospital do Câncer A.C. Camargo**

Baldan-Pineda, PH.

**Identificação de genes supostamente envolvidos com o
processo de invasão tumoral em câncer de mama através da
técnica RaSH** / Paulo Henrique Baldan-Pineda -- São Paulo, 2007.

125p.

Dissertação (Mestrado)-Fundação Antônio Prudente.

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

Orientador: Dirce Maria Carraro

Descritores: 1. CÂNCER DA MAMA/genética. 2. GENES DE CÂNCER
3. EXPRESSÃO GÊNICA. 4. CARCINOMA DUCTAL DA MAMA/genética
5. BIBLIOTECA DE CDNA.

DEDICATÓRIA

*Dedico este esforço a minha mãe,
Maria Elza Baldan, por sempre ter
sido meu exemplo maior de
profissional, de garra, humildade e
paciência... Meu alicerce!*

AGRADECIMENTOS

A Dra. **Dirce Maria Carraro** pela oportunidade e confiança, por todos os ensinamentos, pela infindável paciência comigo, e pelos “puxões de orelha” que muito me ensinaram a ser mais objetivo e racional. Agradeço também pelo exemplo de profissional e pela orientação sempre presente!

A Dra. **Anamaria A. Camargo** pela co-orientação, pelo suporte quanto ao sequenciamento e validação por RT-PCR em tempo real, e pelo exemplo de carisma.

Ao Dr. **Wilson A. Silva Junior**, chefe do Laboratório de Genética Molecular e Bioinformática da Faculdade de Medicina da Universidade de São Paulo - *campus* Ribeirão Preto, pela oportunidade de apreender a metodologia de construção de bibliotecas subtrativas de cDNA RaSH.

Ao Conselho Nacional de Pesquisa (CNPq) pelo auxílio concedido.

Ao Instituto Ludwig de Pesquisa sobre o Câncer pela oportunidade e por toda a infra-estrutura que proporcionou a realização deste trabalho.

Ao Hospital do Câncer A.C. Camargo por toda a infra-estrutura que proporcionou a realização deste trabalho.

As pacientes que consentiram doação de suas amostras tumorais para o Banco de Tumores do Centro de Tratamento e Pesquisa do Hospital do Câncer A.C. Camargo, e ao **Dr. Fernando Augusto Soares** pela coordenação desse Banco de Tumores, possibilitando que este estudo fosse realizado.

A todos os colegas de trabalho do **Laboratório de Biologia Molecular do Centro de Pesquisa do Hospital A.C. Camargo**, que por toda a ajuda a mim prestada e todas as conversas proveitosas, acabaram por deixar de serem colegas para se tornarem verdadeiros amigos.

Ao Departamento de Anatomia Patológica do Hospital A.C. Camargo, em nome do **Dr. Fernando Augusto Soares**, em especial às patologistas **Cíntia Osório** e **Isabela Cunha**, pela microdissecção a laser das amostras utilizadas neste estudo.

A todos do Laboratório de Bioinformática do Hospital do Câncer, em nome da Dra. **Helena P. Brentani**, pela utilização das ferramentas disponíveis para análises dos dados gerados com os experimentos de Microarray.

A **Nadia Castro** pela imensa ajuda em muitas etapas deste trabalho, como a ajuda com a metodologia de amplificação de mRNA, pelas amostras gentilmente cedidas, por todas as conversas proveitosas, e pela amizade que cultivamos ao longo deste trabalho.

A **Anemari Dinarte**, do Laboratório de Genética Molecular e Bioinformática da Faculdade de Medicina da Universidade de São Paulo - *campus* Ribeirão Preto, pela ajuda na metodologia de construção das bibliotecas subtrativas RaSH e pela amizade.

A **Daniel Pinheiro**, do Laboratório de Genética Molecular e Bioinformática da Faculdade de Medicina da Universidade de São Paulo - *campus* Ribeirão Preto, pelo constante suporte com a bioinformática do projeto *RaSH-BDC*.

A **Wilson Malagó Junior** pela ajuda na organização dos dados gerados pelas bibliotecas RaSH, pelas palavras duras e sinceras que me direcionaram para o caminho correto e objetivo do trabalho, e pela amizade.

A **Elisa Napolitano Ferreira** pela grande ajuda com a análise das sequências *no-match*, organização dos dados, submissão das sequências RaSH ao dbEST, e pela amizade.

A **Murilo Geraldo** por compartilhar comigo toda sua experiência com a metodologia RaSH, por todas as conversas proveitosas e dicas recebidas. Pela sincera amizade.

A **Mariana Maschietto** pelo *tour* de apresentação a todos os colaboradores e instalações do Instituto Ludwig logo que cheguei, e pela ajuda incondicional em todos os momentos que precisei.

A **Maria Cristina Rangel** por todas as conversas proveitosas, dicas e consolos nos momentos difíceis... por toda nossa amizade.

A **Louise Mota** pelos ensinamentos de amplificação de mRNA, pela disposição em sempre ajudar, pelo alegre convívio e pelo exemplo de garra!

A **Waleska Martins** pela marcação e hibridização das amostras na etapa de Microarray e pela amizade.

A todos os amigos do Laboratório de Biologia Molecular do Centro de Tratamento e Pesquisa do Hospital A.C. Camargo: **Aderbal, Adriana, Elen, Elisa, Fábio, Louise, Maria Cristina, Mariana, Mev, Nadia, Thiago e Waleska** por todo o convívio, aprendizado, ajuda e pela amizade.

A Pós-Graduação da Fundação Antônio Prudente pelo excelente trabalho, em nome do **Dr. Luis F. L. Reis**, pelo incentivo à pesquisa e apoio ao ensino. E também agradeço a **Anamaria Kuninari** e **Luciana Pitombeira** por toda a ajuda prestada.

A toda a equipe da biblioteca do Centro de Tratamento e Pesquisa do Hospital A.C. Camargo, em nome da senhora **Suely**, por toda a ajuda e serviços prestados.

Ao quarteto de grandes amigos **André, Letícia, Luciana e Rafael** que sempre torceram por mim. E aos grandes amigos da UFSCar **Flavio e Vitor** pelo incentivo.

A minha mãe **Maria Elza Baldan** pelo seu esforço de uma vida inteira para me educar. Pelo exemplo de trabalho, garra e paciência. Por ser minha amiga e confidente (...) por tudo!

A toda minha família que sempre me apoiou. Por sempre estarem do meu lado!

Ao meu primo **Rodrigo Martins** pelo suporte que se propôs a dar logo que cheguei a São Paulo.

Ao meu primo **Cristiano Baldan** pela constante ajuda e apoio em tudo. Pelas conversas, por ouvir meus desabafos, pelo incentivo e pelo convívio na saudosa Avenida Odila!

Ao amigo **Gustavo Ladeira** pelo convívio, pelas conversas e pelo ombro amigo!

A todos os colegas do Instituto Ludwig pela amizade e convívio sempre saudável. E a todos aqueles que me ajudaram direta e indiretamente neste trabalho!

E por fim, mas não menos importante (pelo contrário), agradeço ao meu amor **Fernanda Grande**, que sempre me ajudou nos momentos mais difíceis. Agradeço por me ouvir, apoiar, incentivar e ajudar. Ajuda sem a qual eu não teria conseguido! Obrigado por ser meu anjo da guarda!!!

E agradeço a **Deus** por tudo, principalmente pela saúde concedida, a qual foi fundamental para que eu conseguisse chegar até aqui!

RESUMO

Baldan-Pineda PH. **Identificação de genes supostamente envolvidos com o processo de invasão tumoral em câncer de mama através da técnica RaSH.** São Paulo; 2007. [Dissertação de Mestrado-Fundação Antônio Prudente]

O câncer de mama apresenta uma alta frequência na população feminina brasileira e mundial, resultando num grande número de mortes, sendo que o processo de metástase é sua principal causa, a qual corresponde a cerca de 90% de todas as mortes. Assim, a identificação de genes envolvidos nas primeiras etapas da instalação desse processo pode contribuir para sua melhor compreensão, sendo que esses podem ser testados como marcadores moleculares de invasão e de prognóstico. Neste contexto, este trabalho teve como objetivo identificar genes diferencialmente expressos entre carcinoma ductal *in situ* e invasivo de mama, de uma única amostra tumoral, sendo genes potenciais para participarem da primeira etapa desse processo que é a invasão tumoral. Para esse propósito foram confeccionadas bibliotecas subtrativas de cDNA (RaSH - *Rapid Subtraction Hybridization*) de células tumorais de carcinoma ductal *in situ* e invasivo de mama, que foram microdissecadas a laser de uma mesma amostra. Esse delineamento experimental foi adotado a fim de minimizar a variabilidade genética entre as amostras de diferentes pacientes, aumentando a chance de identificar genes realmente envolvidos com a invasão tumoral. Após alterações no protocolo RaSH, para adaptá-lo à amostras de RNA que foram submetidas ao procedimento de amplificação, foi possível a confecção de 4 bibliotecas subtrativas (duas IDC-*tester* e duas DCIS-*tester*) a partir de dois casos pareados. Foram identificados 443 transcritos específicos dos 2 componentes (DCIS e IDC), dos quais somente 7 apresentaram o mesmo comportamento nos dois casos. Estes transcritos foram selecionados para confirmação de expressão diferencial através de RT_PCR em tempo real. Destes, 5 genes apresentaram dados confiáveis em termos de reprodutibilidade experimental. Dentre os quais, os genes EIF1 e

FN1 foram confirmados como 1,53 e 3,1 vezes mais expressos no componente IDC, respectivamente, e, o gene METTL3 confirmado como 3,25 vezes mais expresso no componente DCIS. Para uma validação em larga escala, utilizando um grupo independente de amostras, foi construída uma plataforma de microarray na qual foram imobilizados 414 transcritos específicos e hibridizados com RNAs amplificados de células microdissecadas a laser de um grupo de 6 casos pareados de carcinoma ductal *in situ* e invasivo de mama. Uma análise comparativa entre os 2 grupos de amostras -DCIS e IDC- dos 6 casos, identificou 5 transcritos diferencialmente expressos, que confirmaram os dados de RaSH. Dentre esses o gene TRIP6, identificado como *up*-regulado no componente IDC, auxilia no recrutamento de proteínas que participam do complexo de adesão focal sugerindo o envolvimento desse gene no processo de invasão tumoral no tecido adjacente em carcinoma ductal de mama. Uma análise individual em cada um dos 6 casos pareados (amostras DCIS e IDC de um mesmo caso) revelou 114 transcritos diferencialmente expressos. Vários desses genes têm sido associados com tumores mamários, tais como CTGF, ITGB3BP e THBS2, confirmados como *up*-regulados em IDC; e ANKRD30A, ANXA1, SFRP1, APLP2, PER2 e TFF1, confirmados como *down*-regulados em IDC. Outros genes que não apresentam relatos sobre algum envolvimento com o processo de invasão em carcinoma mamário, também foram confirmados como diferencialmente expressos entre DCIS e IDC, como por exemplo, o gene RDBP e a *EST* CN414308. Esse estudo identificou genes candidatos que podem apresentar um papel chave no mecanismo de invasão tumoral de carcinoma mamário sendo potenciais candidatos a serem testados como marcadores de invasão e prognóstico.

SUMMARY

Baldan-Pineda PH. **[Identification of genes supposedly involved with tumor invasion of breast cancer by RaSH methodology]**. São Paulo; 2007. [Dissertação de Mestrado-Fundação Antônio Prudente]

The breast cancer occurs at high frequency in the Brazilian and world-wide female population, resulting in a great number of deaths. Metastasis is its main cause, which corresponds to about 90% of all deaths. Thus, the identification of genes involved in early stages of this process may contribute for its better understanding, and these genes may be tested as molecular markers of invasion and prognosis. In this context, the aim of this study was the identification of differentially expressed genes between ductal carcinoma *in situ* (DCIS) and invasive (IDC) of the breast, from the same tumor sample, being potential genes to participate in the first step of this process, which is tumor invasion. For this purpose, subtractive libraries of cDNA were confectioned (RaSH - *Rapid Subtraction Hybridization*) of tumoral cells from ductal carcinoma *in situ* and invasive of the breast, which were laser microdissected from the same sample. This experimental design was adopted to minimize genetic variability from samples of different patients, increasing the chance of identification of really tumor invasion process-related genes. After few alterations in RaSH original protocol for using amplified RNA, it was possible to generate 4 subtractive libraries (two IDC-tester and two DCIS-tester) from two cases of matched samples. Four hundred and forty three specific genes were identified from both components (DCIS and IDC). Only 7 had similar behavior between the two cases. These selected genes were submitted to validation of differential expression by real time RT_PCR and 5 presented reproducible experimental data. From those, EIF1 and FN1 were confirmed 1,53 and 3,1 up-regulated in IDC, respectively and METTL3 was confirmed 3,25 up-regulated in DCIS. To aim a large scale differential expression confirmation using an independent group of samples a customized microarray platform was designed in which 414 specific genes were immobilized. These genes were hybridized

with amplified RNA from laser microdissected cells from a group of 6 matched DCIS-IDC samples. Comparative analyses between the 2 group of samples - DCIS and IDC- from 6 cases identified 5 differentially expressed genes, in agreement with RaSH results. Among these, the TRIP6 gene, identified as up-regulated in IDC, helps in protein recruitment to focal adhesion complex formation suggesting its involvement in tumoral invasion to adjacent tissues in breast ductal carcinoma. An individual analysis in each 6 matched DCIS-IDC cases (DCIS and IDC samples from the same case) identified 114 differentially expressed genes that confirmed RaSH results. Several of these genes have been associated to breast cancer, like CTGF, ITGB3BP and THBS2, all confirmed as up-regulated in IDC; and ANKRD30A, ANXA1, SFRP1, APLP2, PER2 e TFF1, all confirmed as down-regulated in IDC. Other genes that do not present any note about its involvement to invasion process in breast ductal carcinoma were also confirmed by this analysis, for example RDBP and the EST CN414308. Thus, this study identified candidate genes that may present key role in tumor invasion mechanism in breast ductal carcinoma being potential candidates to be tested as molecular markers of invasion and/or prognosis.

LISTA DE FIGURAS

Figura 1	Microdisseccção a laser de DCIS e IDC de uma mesma amostra tumoral	29
Figura 2	Síntese de cDNA e amplificação por transcrição <i>in vitro</i> de mRNA	35
Figura 3	Metodologia de amplificação de mRNA por <i>random primer</i>	37
Figura 4	Metodologia de amplificação de mRNA por <i>template switch</i>	38
Figura 5	Ilustração esquemática da metodologia RaSH de construção de biblioteca subtrativa de cDNA	45
Figura 6	<i>MM plot</i> da amostra 13-DCIS	60
Figura 7	Representação da fluorescência de todos os genes de uma lâmina	61
Figura 8	Gráficos representativos da normalização por <i>Lowess</i>	62
Figura 9	Eficiência da ligação dos adaptadores XDPN12 e XDPN14 às moléculas de cDNA dos componentes DCIS e IDC do caso 1	71
Figura 10	Resultado da análise do <i>pipeline</i> <i>GEAP-BDC</i> das sequências submetidas das 4 bibliotecas	72
Figura 11	Representação de parte da comparação das duas bibliotecas construídas com as amostras do primeiro paciente	75
Figura 12	Representação de parte da comparação das duas bibliotecas construídas com as amostras do segundo paciente	76
Figura 13	Genes comuns entre os mesmos componentes tumorais dos dois casos	78
Figura 14	Análise do <i>software</i> <i>geNorm</i> sobre o papel dos genes BCR, GAPDH e HPRT1 como normalizadores neste conjunto de amostras	83
Figura 15	Clusterização hierárquica. Dendrograma das lâminas controle e <i>swap</i> das 12 amostras tumorais (6 amostras DCIS e 6 amostras IDC)	91

LISTA DE TABELAS

Tabela 1	Genes selecionados, as seqüências dos <i>primers</i> que amplificam tais genes e o tamanho do fragmento amplificado (<i>Amplicon</i>)	49
Tabela 2	Genes normalizadores com suas seqüências e tamanho do fragmento amplificado (<i>Amplicon</i>)	50
Tabela 3	Casos pareados de carcinoma ductal <i>in situ</i> e invasivo de mama utilizadas na hibridização com os transcritos específicos de DCIS e IDC identificados por RaSH	55
Tabela 4	Características específicas de cada fluoróforo utilizado	58
Tabela 5	Casos previamente selecionados como material de estudo	64
Tabela 6	Características clínicas das pacientes, e, anatomopatológicas e imunohistoquímicas das amostras selecionadas para estudo	66
Tabela 7	Quantificação de aRNA após 1 <i>round</i> de amplificação por transcrição <i>in vitro</i>	67
Tabela 8	Quantificação dos fragmentos de DNA amplificados por PCR	71
Tabela 9	Número de seqüências relativas as categorias <i>RefSeq</i> , <i>Unigene</i> e <i>ESTs</i> , de acordo com cada biblioteca RaSH, e o percentual de seqüências destas categorias por biblioteca	74
Tabela 10	Número de genes classificados pelas categorias <i>RefSeq</i> , <i>Unigene</i> e <i>EST</i> , de acordo com cada biblioteca	74
Tabela 11A	Número de genes específicos da cada componente tumoral, DCIS e IDC do caso 1, classificados pelas categorias <i>RefSeq</i> , <i>Unigene</i> e <i>EST</i>	77
Tabela 11B	Número de genes específicos da cada componente tumoral, DCIS e IDC do caso 2, classificados pelas categorias <i>RefSeq</i> , <i>Unigene</i> e <i>EST</i>	77
Tabela 12	Genes selecionados para validação por RT_PCR em tempo real. Genes específicos de cada componente tumoral e comuns entre os casos 1 e 2.	79
Tabela 13	Padronização da quantidade de cDNA e das concentrações de <i>primers</i> para reação de RT_PCR em tempo real	81

Tabela 14	Eficiência de amplificação dos <i>primers</i> desenhados	82
Tabela 15	Dados de expressão relativa dos genes candidatos. ND: Valor não determinado de expressão relativa devido a desvio-padrão maior do que 0,4	84
Tabela 16	Número de genes específicos da cada componente tumoral (DCIS e IDC) de cada biblioteca, classificados pelas categorias <i>RefSeq</i> , <i>Unigene</i> e <i>EST</i>	86
Tabela 17	Número de genes específicos da cada componente tumoral (DCIS e IDC) de cada biblioteca, classificados pelas categorias <i>RefSeq</i> , <i>Unigene</i> e <i>EST</i> , após inspeção visual para exclusão de genes ribossomais e mitocondriais	87
Tabela 18	Quantificação de aRNA após 2 <i>rounds</i> de amplificação por transcrição <i>in vitro</i>	88
Tabela 19	Valores da correlação de incorporação de fluoróforos entre a lâmina controle e a lâmina <i>swap</i> , de acordo com cada amostra	92
Tabela 20	Genes diferencialmente expressos identificados através da análise conjunta dos 6 casos por <i>microarray</i>	93
Tabela 21	Genes com mesmo comportamento em pelo menos dois casos	95
Tabela 22	Percentual de genes confirmados por <i>microarray</i> como diferencialmente expressos entre IDC e DCIS no grupo independente de 6 casos pareados	95

LISTA DE ABREVIATURAS

° C	<i>Graus Celcius</i>
μFD	<i>Microfaraday</i>
μg	<i>Micrograma</i>
μL	<i>Microlitro</i>
μM	<i>Micromolar</i>
ACTB	<i>Actin-Beta</i>
BCR	<i>Break Cluster Region</i>
BLAT	<i>BLAST-like Alignment Tool</i>
BLAST	<i>Basic Local Alignment Tool</i>
BSA	<i>Bovine Serum Albumin</i>
cDNA	<i>Complementar DNA</i>
Cy3	<i>Cyanine 3 dye</i>
Cy5	<i>Cyanine 5 dye</i>
Ct	<i>Cycle threshold</i>
DEPC	<i>Dietilpirocarbonato</i>
DNA	<i>Desoxyribonucleic Acid</i>
dNTP	<i>Desoxirribonucleosídeos-Trifosfato</i>
ds-cDNA	<i>Double Strand cDNA</i>
DTT	<i>Ditiotreitol</i>
EDTA	<i>Ethylene Diamine TetrAcetic Acid</i>
EST	<i>Expressed Sequence Tag</i>
GAPDH	<i>Glyceraldehyde-3-phosphate dehydrogenase</i>
HPRT	<i>Hipoxantina Fosforribosil Transferase</i>
HCl	<i>Hydrogen Chloride</i>
INCA	<i>Instituto Nacional do Câncer</i>
Kb	<i>Kilobase</i>
Kv	<i>Kilovolts</i>
LOH	<i>Loss of Heterozigosis</i>
MgCl₂	<i>Magnesium Chloride</i>
mL	<i>Mililitro</i>

mM	Milimolar
mRNA	<i>Messenger RNA</i>
NaAc	<i>Sodium Acetate</i>
NaCl	<i>Sodium Chloride</i>
NaOH	<i>Sodium Hydroxide</i>
NCBI	<i>National Center for Biotechnology Information</i>
ng	Nanograma
nM	Nanomolar
nm	Nanômetro
OHMS	Unidade de resistência de um condutor
pb	Pares de base
PCR	<i>Polymerase Chain Reaction</i>
pH	Potencial Hidrogeniônico
PHRED	<i>Software</i> que analisa qualidade de sequências nucleotídicas
RaSH	<i>Rapid Subtraction Hybridization</i>
RNA	<i>Ribonucleic acid</i>
RNAse	Ribonuclease
rNTP	Ribonucleosídeos-Trifosfato
rpm	Rotações por minuto
RT-PCR	<i>Reverse Transcription PCR</i>
SAGE	<i>Serial Analysis of Gene Expression</i>
SDS	Sulfato Dodecil de Sódio
SSC	Solução Salina de Citrato de Sódio
TE	Tris-EDTA
TNM	<i>Tumor Node Metastasis</i>
Tris	<i>Trishydroxymethylaminomethane</i>
TS	<i>Template Switch</i>
U	Unidade

ÍNDICE

1	INTRODUÇÃO	1
1.1	Bases genéticas do câncer de mama	3
1.2	A Progressão Tumoral em Câncer de Mama	8
1.3	O Processo de Metástase Tumoral	11
1.3.1	O microambiente tumoral	16
1.4	Metodologias para Avaliação de Expressão Gênica	17
1.4.1	Análises em Grande Escala	17
1.4.2	RT_PCR em tempo real	20
1.5	Estudos Clínicos Orientados	22
1.6	Justificativa	24
2	OBJETIVOS	26
2.1	Objetivo Geral	26
2.2	Objetivos Específicos	26
3	MATERIAL E MÉTODOS	27
3.1	Seleção dos Casos	27
3.2	Coloração das Lâminas Histológicas para Microdissecção por LCM	27
3.3	Microdissecção por LCM (<i>Laser Capture Microdissection</i>)	28
3.4	Extração de RNA Total	30
3.5	Adaptação do Protocolo de Construção de Bibliotecas Subtrativas RASH a Partir de RNA Amplificado	30
3.5.1	Amplificação de MRNA	30
3.5.2	Clivagem dos cDNAs com a enzima de restrição <i>Dpn II</i>	38
3.5.3	Ligação das amostras DCIS e IDC aos adaptadores XDPN12 e XDPN14	39
3.5.4	PCR-Teste da ligação dos adaptadores	42
3.5.5	PCR em larga escala	42
3.5.6	Clivagem dos cDNAs com a enzima de restrição <i>Xho I</i>	43

3.5.7	Hibridização subtrativa	43
3.5.8	Clonagem	44
3.5.9	Transformação de bactérias <i>Escherichia coli</i> e sequenciamento dos clones	46
3.6	Submissão das seqüências ao <i>Pipeline Geap-Bdc</i>	46
3.7	RT-PCR para verificação da extensão do fragmento	48
3.7.1	Genes candidatos selecionados	48
3.7.2	Genes normalizadores	49
3.8	RT_PCR em tempo real	51
3.8.1	Eficiência dos <i>primers</i>	52
3.8.2	Quantificação Relativa	53
3.9	cDNA <i>MICROARRAY</i>	53
3.9.1	Construção da plataforma	53
3.9.2	Marcação e hibridização	55
3.9.3	Captura das imagens e quantificação do sinal	58
3.9.4	Pré-análise das lâminas hibridizadas	58
3.9.5	Análise estatística	62
4	RESULTADOS	64
4.1	Seleção dos casos	64
4.2	Adaptação do protocolo RASH para RNA amplificado	66
4.2.1	Amplificação de mRNA	67
4.2.2	Bibliotecas RaSH construídas	70
4.3	Classificação das seqüências geradas	71
4.4	Análise comparativa das bibliotecas RASH	75
4.4.1	Identificação de Transcritos Específicos de cada Biblioteca	75
4.5	Seleção de genes para validação por RT_PCR em tempo real	78
4.6	Validação em grande escala através de <i>microarray</i>	86
4.6.1	Pré-análise das lâminas	89
4.6.2	Identificação de genes diferencialmente expressos	92
5	DISCUSSÃO	97
6	CONCLUSÕES	113

ANEXOS

- Anexo 1** Genes identificados como específicos de cada biblioteca RaSH.
- Anexo 2** Dados de expressão relativa dos genes normalizadores.
- Anexo 3** Valores individuais dos *Cts* das duplicatas de cada gene em cada amostra e os valores de eficiência de amplificação obtidos.
- Anexo 4** Genes confirmados por microarray como diferencialmente expressos entre IDC e DCIS por caso pareado.

1 INTRODUÇÃO

O câncer de mama é o tipo de câncer feminino mais freqüente. E, além disso, é também a causa de maior número de mortes entre as mulheres no mundo inteiro por câncer (HORTOBAGYI et al. 2005), correspondendo à cerca de 400 mil mortes por ano (PARKIN et al. 2005). Aproximadamente 90% de todas as mortes por câncer se devem à formação de metástase (CAVALLARO e CHRISTOFORI 2004). Em 2003 esse tipo de câncer foi responsável por cerca de 41.600 novos casos e 9.300 óbitos e há estimativas de que em 2006 houve cerca de 48.930 novos casos para cada 100 mil mulheres, em todo o país (Ministério da Saúde 2006).

A classificação histológica do câncer de mama baseia-se no tipo de tumor, estadio e grau (CALLAGY et al. 2003). Quanto ao tipo de tumor, pode ser dividido em carcinoma Ductal e Lobular. O carcinoma Ductal compreende a grande maioria das neoplasias malignas de mama e abrange os componentes invasivo (*invasive ductal carcinoma - IDC*) e o *in situ* (*ductal carcinoma in situ - DCIS*). Embora IDC corresponda à maioria dos casos de tumores mamários, no qual pode ou não haver a presença do componente *in situ*, o número de DCIS tem aumentado significativamente nas últimas décadas, correspondendo à cerca de 20% de todos os cânceres de mama (LEONARD e SWAIN 2004), devido aos avanços desenvolvidos nos exames diagnósticos. O carcinoma *in situ* é o resultado de uma proliferação de células epiteliais tumorais dentro de ductos mamários, não apresentando evidência de invasão (LEONARD e

SWAIN 2004); enquanto que o invasivo apresenta invasão de células tumorais em tecidos adjacentes, sangue e linfa (LESTER 2005).

A classificação e o estadiamento do câncer geralmente seguem o sistema **TNM** (Tumor/Nódulo/Metástase). Esse sistema foi criado pela UICC (*Union Internationale contre le Cancer*) e pelo AJCC (*The American Joint Committee on Cancer*). Tal sistema agrupa diversos tipos tumorais de acordo com as seguintes características: (1) T - Tamanho do tumor, (2) N - Envolvimento de nódulo linfático e (3) M - Ocorrência de metástase à distância do tumor primário.

O estadiamento pode ser clínico (cTNM), realizado normalmente a partir de exame físico e diagnóstico por imagem (evidências anteriores ao processo cirúrgico), ou histológico (patológico) (pTNM) quando é realizado por exame histopatológico da peça cirúrgica e dos linfonodos, para detectar a presença ou ausência de metástases em linfonodos regionais.

Em relação ao estadio, os tipos tumorais podem se agrupar em cinco classes. Existe o estadio 0, o qual abrange o carcinoma *in situ*, sem a presença do carcinoma invasivo; o estadio I, câncer invasivo mas localizado; estadio II, câncer invasivo localmente limitado ou espalhado regionalmente; estadio III, carcinoma invasivo localmente extensivo ou espalhado regionalmente; e, o estadio IV, câncer invasivo extensivo com presença de metástase à distância; sendo que do estadio I a IV pode haver presença do componente *in situ*.

Quanto ao grau, tanto DCIS quanto IDC são ainda subclassificados, cada um deles, em baixo, intermediário e alto grau (DCIS-1, 2 e 3; IDC-1, 2 e 3 -bem, médio e pouco diferenciado, respectivamente); conferindo um gradiente

de pouca para alta agressividade tumoral, em direção ao menos diferenciado (MA et al. 2003; SONTAG e AXELROD 2005).

1.1 BASES GENÉTICAS DO CÂNCER DE MAMA

Com o avanço tecnológico e, conseqüentemente, com o aprimoramento de metodologias de manipulação de moléculas de DNA, foi possível descrever muitos tipos de alterações gênicas associadas a diversos tipos tumorais.

Os mecanismos moleculares envolvidos nos tumores de mama compreendem basicamente alterações em oncogenes, como por exemplo, ciclinas, MYC, ERBB2, BCL2; e em genes supressores de tumor, como por exemplo, BRCA1, BRCA2, CDK e TP53. Enquanto que os primeiros (oncogenes) regulam positivamente o ciclo celular, os últimos (genes supressores) regulam negativamente. Desta forma, alterações nestes genes podem provocar, respectivamente, ganho e perda de funções relacionadas ao ciclo celular. Estas alterações podem ocorrer em nível do genoma, transcriptoma e proteoma.

Alterações moleculares têm sido relatadas, por estudos de perda de heterozigose (LOH), para alguns genes supressores de tumor, como por exemplo, BRCA1 e BRCA2. Estes genes expressam proteínas envolvidas no processo de reparo de moléculas de DNA (NAROD e FOULKES 2004). Tal inativação pode ser na linhagem germinativa (câncer hereditário) ou somática (câncer esporádico). O câncer hereditário de mama representa cerca de 5 a 10% de todos os casos de câncer de mama diagnosticados, e os riscos de câncer de mama conferido por mutações em BRCA1 e BRCA2 são,

respectivamente, de 8.39 e 2.66% para idade de 40 anos, e, de 47.45 e 31.85% para idade de 75 anos (FU et al. 2007). No entanto o risco de desenvolver câncer não é o mesmo para todos os portadores de mutações nesses genes. Este risco é influenciado por outros fatores, como por exemplo, cofatores ambientais e hormonais (NAROD e FOULKES 2004).

Outro gene supressor de tumor bem estudado é o TP53, o qual codifica a proteína p53 presente nas células normais na forma não-mutante. A proteína p53 tem um papel chave em mediar respostas celulares a vários tipos de estresse, que em geral causam danos ao DNA. Sua função é prevenir que células com estes danos continuem a proliferar, bloqueando o ciclo celular e estimulando a apoptose das mesmas (LACROIX et al. 2006). Quando ocorrem danos no DNA, algumas quinases são ativadas e fosforilam p53, permitindo que ele se desligue da proteína Mdm2. Como resultado, p53 se acumula na célula e ativa a transcrição do gene p21, o qual inativa quinases dependentes de ciclinas (*Cdk*) envolvidas na regulação da transição da fase G1 para S, bloqueando a progressão do ciclo celular; e além disso, p53 ativado estimula o processo de apoptose (ALBERTS et al. 2002b). Alterações neste gene resultam num p53 inativo, incapaz de bloquear o ciclo celular e a estimular o processo de apoptose, levando à proliferação de células cujo DNA pode apresentar grande número de mutações, resultando numa maior agressividade do tumor (SIMÃO et al. 2002). As alterações nesse gene ocorrem em cerca de 30% dos carcinomas mamários (BORRESEN-DALE 2003) e estão associados a tumores negativos para a expressão dos receptores de estrógeno e positivos para amplificação do gene ERBB2 (BORRESEN-DALE 2003).

O ERBB2 (*erythroblastic leukemia viral oncogene homolog 2*), um oncogene muito estudado, também conhecido como NEU (*neuroblastoma derived oncogene homolog*) e HER2 (*herstatin*) é um membro da família de receptores de fatores de crescimento epidermal (*EGF*) e é amplificado ou superexpresso em cerca de 25 a 30% dos carcinomas invasivos de mama (TIMMS et al. 2002). Sua superexpressão está associada a formas mais agressivas de tumor, sendo considerada como um marcador de pior prognóstico (GUO et al. 2006). Esta superexpressão, a qual pode estar presente tanto em DCIS quanto em IDC, pode gerar uma alteração na especificidade de ligação desses receptores com outros membros de sua família. Isto pode resultar na formação aberrante de homo e heterodímeros, tornando sua sinalização intracelular alterada (ZAHND et al. 2006). Diversos estudos têm encontrado alterações no *status* de ERBB2, como consequência de amplificação de seu *locus* gênico (PUPUTTI et al. 2006) e mudanças nos níveis de mRNA e proteína (POTEMSKI et al. 2006). Um anticorpo monoclonal recombinante, o *trastuzumab* (herceptina) (CARTER et al. 1992), foi desenvolvido para inibir a atividade do receptor ERBB2, impedindo uma sinalização intracelular que promova a proliferação celular aberrante. Isso tem possibilitado seu uso atual em terapias adjuvantes.

Existem diversas alterações moleculares que envolvem ganho e perda de material genético, resultando num desequilíbrio alélico em câncer de mama. Diversos estudos têm abordado este aspecto utilizando a técnica *CGH* (*Comparative Genomic Hybridization* – Hibridação Genômica Comparativa) (KALLIONIEMI et al. 1992). Estes autores encontraram uma amplificação, com aumento de 8 vezes do gene ERBB2 na linhagem celular de câncer de mama

SK-BR-3. Outra metodologia bastante utilizada é *Array-CGH* (*Array-based Comparative Genomic Hybridization*) (PINKEL et al. 1998), de maior resolução na detecção de alterações genômicas, em relação à técnica convencional de CGH.

Estas técnicas têm proporcionado análises globais do genoma de células tumorais de mama em função da detecção de amplificações e deleções em nível genômico, os quais compreendem ganhos de porções gênicas nos cromossomos 1q e 8q, assim como perda nos cromossomos 8p, 16q e 17p (DEVILLEE e CORNELISSE 1994). A amplificação da região 20q.13.2 tem sido relatada por diversos pesquisadores como associada a vários tipos de tumores. WERNER et al. (1999) encontraram amplificação desta região em amostras microdissecadas a laser de hiperplasia ductal, hiperplasia ductal atípica e carcinomas ductal *in situ* e invasivo de mama, sugerindo que este fenômeno deve desempenhar um papel na tumorigênese mamária. Mais recentemente, NAMBA et al. (2006), a partir de experimentos de microarray e de array-CGH, encontraram maiores diferenças genômicas e em nível de expressão gênica para estágios mais iniciais do câncer de mama.

Além dos fatores genéticos, outros fatores podem desempenhar um papel no desenvolvimento do câncer de mama. Um deles, de pronunciada importância, é o estado hormonal, como já citado anteriormente. Os hormônios esteróides sexuais estrógeno e progesterona têm um crítico papel no desenvolvimento do câncer de mama, e, além disso, ambos os hormônios e seus receptores têm sido implicados como fatores que influenciam a progressão de tumores mamários (CORDERA e JORDAN 2006).

Os estrogênios regulam diversas atividades nas células tumorais mamárias, incluindo a proliferação celular (RAI et al. 2005). Esta função é promovida quando o estrógeno interage de maneira direta com os receptores nucleares de estrógeno (*estrogen receptor* – ER), e esses por sua vez, interagem com seqüências regulatórias do DNA, o que promove a transcrição de genes que codificam proteínas participantes do metabolismo de proliferação celular (RAI et al. 2005); uma situação que influencia na tumorigênese mamária (BRODIE et al. 2003). Vários estudos correlacionam a expressão positiva do ER como um indicativo de bom prognóstico. Esta relação é válida, baseando-se na idéia de que estes tumores são estrógeno-dependentes, ou seja, como existem receptores para o estrógeno, estes se ligam ao estrógeno de maneira pronunciada, desencadeando a informação molecular para proliferação celular. Sendo assim, o bloqueio da ação do estrógeno impediria esse processo. E isto é feito através do tratamento com tamoxifeno, droga que inibide a enzima aromatase, a qual é responsável pela fase final da biossíntese do estrógeno (especificamente do estradiol) (HARKONEN et al. 2003).

Outros inibidores de aromatase, além de tamoxifeno, também têm sido utilizados, tais como, anastrozol, exemestano e letrozol (DIXON 2006). Tais inibidores de aromatases são considerados como bastante eficazes no tratamento adjuvante de câncer de mama (DIXON 2006).

A progesterona apresenta efeitos contraditórios na glândula mamária. SUMIDA et al. (2004), por exemplo, verificaram a partir de ensaios zimográficos que o tratamento com o hormônio progesterona em linhagens celulares transfectadas com receptores de progesterona tanto do tipo A quanto do tipo B inibe de maneira acentuada a atividade da metaloproteinase-9 (MMP9), a qual

age na degradação da matriz extracelular (*ECM - extracellular matrix*) ao realizar a proteólise de colágeno-IV. Isso resulta numa diminuição do potencial metastático destas linhagens celulares. Por outro lado, outros estudos mostraram que terapias de reposição hormonal com estrógeno e progesterona resultam em maior risco de câncer de mama, em relação a uma terapia com apenas o hormônio estrógeno (CAMPAGNOLI et al. 2005).

1.2 A PROGRESSÃO TUMORAL EM CÂNCER DE MAMA

A progressão tumoral em câncer de mama tem sido descrita como um processo que ocorre em várias etapas, manifestando-se através de uma sequência de estágios patologicamente definidos. Embora seja amplamente aceito que o carcinoma mamário inicia-se como uma forma pré-maligna conhecida como hiperplasia ductal atípica (*atypical ductal hyperplasia – ADH*), progredindo para a forma pré-invasiva conhecida como DCIS e finalmente progredindo para a forma invasora conhecida como IDC, ainda não está totalmente esclarecida a base molecular desta progressão (MA et al. 2003), ou seja, não está totalmente estabelecido que DCIS seja um precursor obrigatório na progressão para IDC. Da mesma forma, ainda permanece incerto se as lesões hiperplásicas em mama, especialmente aquelas com atipia, referem-se a precursores ou apenas a marcadores de alto risco de câncer de mama, já que existem ambos os casos: uma lesão hiperplásica que progride para um DCIS bastante agressivo e uma lesão hiperplásica que nunca progride para um DCIS (ALLRED et al. 2001).

Diferentes estudos mostram evidências que suportam ambas as hipóteses. As análises moleculares globais que visam comparar os componentes pré-malignos e malignos de carcinoma mamário suportam a idéia de que DCIS é precursor de IDC. Por exemplo, pesquisadores como ZHUANG et al. (1995) e AMARI et al. (2003), a partir de estudos de LOH nos dois componentes de carcinoma ductal (*in situ* e invasivo) obtiveram dados que mostram perdas alélicas idênticas e semelhantes entre os dois componentes, respectivamente. Ambos os trabalhos sugerem que DCIS pode ser um precursor de IDC. Mais recentemente, LARSON et al. (2006) mostraram dados que corroboram com esta idéia; a partir de análises de LOH -também chamado de desequilíbrio alélico- encontraram suporte de que a hiperplasia ductal atípica é uma precursora direta das formas malignas de câncer de mama. Outros como VAN'T VEER et al. (2002), MA et al. (2003), WEIGELT et al. (2003) e FOLGUEIRA et al. (2006), a partir de experimentos de microarray, verificaram uma grande semelhança no perfil global de expressão gênica dos vários tipos de lesão que representam a progressão do câncer de mama, corroborando com os dados anteriores.

Apesar do padrão geral de expressão gênica ser muito similar em lesões *in situ* e invasivo de carcinoma de mama, importantes diferenças entre estes componentes também foram encontradas em nível de expressão gênica e de algumas proteínas específicas. Por exemplo, MAXHIMER et al. (2005), através de estudos histoquímicos, descreveram uma maior expressão em componente invasivo em relação ao *in situ* de Heparanase-1 (HPR1) (uma endoglicosidase que degrada especificamente o proteoglicano sulfato de heparan na matriz extracelular de mama), sugerindo seu papel na progressão de DCIS para IDC.

Outra expressão diferencial, também maior no componente invasivo, foi relatada para o fator de transcrição BP1, a partir de análises imunohistoquímicas (MAN et al. 2005). Estes autores relataram que em tecidos com os componentes DCIS e IDC associados, a expressão de BP1 foi de 21, 46, e 81% em tecido hiperplásico, tumor *in situ* e invasivo, respectivamente; sugerindo que a expressão de BP1 pode refletir ou influenciar diretamente os processos de progressão e/ou invasão tumoral. O BP1 é um membro novo da família gênica *homeobox* DLX, e tem sido implicado na regulação de diversas vias metabólicas (FU et al. 2003).

Recentemente, alguns estudos -incluindo nosso grupo- caracterizaram molecularmente, através de *microarray*, a transição de DCIS para IDC, analisando células dos 2 componentes extraídos de uma mesma amostra tumoral (SCHUETZ et al. 2006; Castro et al. em preparação). A utilização de células tumorais do ducto mamário e de lesões invasoras de uma mesma amostra, isto é, com o mesmo *background* genético, aumenta a chance de encontrar genes realmente envolvidos no mecanismo de invasão em câncer de mama. Esses trabalhos relatam genes que devem desempenhar papéis importantes nesse processo, tais como PLEKHC1, MEF2C, BGN, POSTN (SCHUETZ et al. 2006); e LUM e CRABP2 (Castro et al. em preparação) e que podem ser testados como marcadores de invasão ou de prognóstico em carcinoma ductal de mama.

Assim, o entendimento molecular da biologia de DCIS e de sua transição para IDC pode facilitar a compreensão dos primeiros eventos da progressão do carcinoma ductal de mama.

1.3 O PROCESSO DE METÁSTASE TUMORAL

Recentes estudos têm abordado os eventos moleculares envolvidos com o processo metastático relatando diversos fatores envolvidos, tanto direta quanto indiretamente. Basicamente, estes estudos têm considerado quatro etapas necessárias para ocorrer o processo metastático de diversas neoplasias. Tais etapas incluem, respectivamente, (1) o processo de invasão celular (invasão tumoral), primeira etapa que é representada pela penetração de células epiteliais no tecido mesenquimal adjacente, a partir da passagem pela membrana basal; (2) intravasão celular, que é a passagem das células neoplásicas para os vasos sanguíneos e/ou linfáticos; (3) extravasão, que é a passagem destas células da corrente sanguínea e/ou de vasos linfáticos para tecidos adjacentes; e (4) micrometástase, que é a formação de um tumor secundário em órgão adjacente ou distante do tumor primário (YANG et al. 2004). Esses estudos têm explorado esse tema com a finalidade de compreender melhor cada uma dessas etapas.

A invasão tumoral, de maneira geral, ocorre em consequência de 3 processos conjuntos: a perda de adesão entre as células epiteliais; o ganho de um fenótipo mais invasivo; e a secreção de proteases que degradam a membrana basal do epitélio mamário (MAN e SANG 2004), favorecendo assim, a migração de células neoplásicas para tecidos adjacentes.

A adesão entre células epiteliais é exercida por um processo conhecido como mecanismo juncional, o qual é capaz de promover contatos célula-célula e célula-matriz extracelular. Os principais grupos de moléculas de adesão celular são as caderinas; integrinas; moléculas de adesão celular da

superfamília Ig (*Ig-superfamily CAMs*), as quais apresentam vários domínios estruturalmente similares às imunoglobulinas; e, aquelas que são ligantes de carboidratos, como por exemplo, as selectinas (ALBERTS et al. 2002a).

As caderinas correspondem a uma superfamília de proteínas transmembranas responsáveis pela manutenção da integridade da estrutura epitelial e endotelial, à medida que realizam contatos célula-célula através de interações homofílicas entre seus homodímeros protéicos extracelulares. É uma proteína que interage indiretamente com o citoesqueleto, através de ligações entre seu domínio citoplasmático e as proteínas ancoradoras α , β , δ -catenina (placoglobina), α -actinina e vinculina (ALBERTS et al. 2002a).

A manutenção da adesão celular, assim como a invasão de células epiteliais em tecidos mesenquimais adjacentes, é regulada por diversos mecanismos moleculares. Alguns destes são os mesmos presentes num processo conhecido como *EMT* (*epithelial-mesenchymal transition* – transição epitélio-mesênquima) (THIERY e SLEEMAN 2006). Tal processo corresponde a uma série de eventos, como por exemplo, mudanças no comportamento e na arquitetura celular. As células epiteliais perdem suas características e começam a apresentar características de células mesenquimais, adotando um fenótipo fibroblasto-*like* e atividade invasiva (CONACCI-SORRELL et al. 2002; NELSON e NUSSE 2004). Um exemplo destas mudanças é o processo conhecido como *Cadherin-Switch*, onde ocorre perda de expressão de E-caderina (marcador epitelial) e expressão *de novo* de N-caderina e caderina-11 (marcadores mesenquimais) (CAVALLARO e CHRISTOFORI 2004, THIERY e SLEEMAN 2006). Esta série de eventos que leva a diminuição de E-caderina, inicia-se com a atividade de membros de famílias de fatores de crescimento,

como por exemplo, *TGFB*, *FGF*, *EGF* e *SF/HGF*. A ligação com esses fatores resulta na ativação de vários efetores intracelulares, tais como membros da família de pequenas *GTPases* (*Ras*, *Rho*, *Rac* e *Cdc42*) e membros da família tirosina-quinase *Src*. Tais efetores regulam a desorganização de complexos juncionais e mudanças na organização do citoesqueleto (THIERY e SLEEMAN 2006). Além disso, também ativam reguladores transcricionais como SNAIL (atualmente conhecido como SNAI1) e SLUG (atualmente conhecido como SNAI2), os quais reprimem a expressão gênica de E-caderina (THIERY e SLEEMAN 2006). A ativação de SNAI é dependente da via de sinalização WNT, que quando ativada inibe sua fosforilação, o que resulta em aumento do seu nível celular e de sua atividade, reprimindo a expressão de E-caderina (YOOK et al. 2005). Isto resulta na perda dos complexos juncionais do epitélio dependentes de E-caderina, o que também pode resultar na interrupção do seqüestro, por parte de E-caderina, de β -catenina citoplasmática. Esta por sua vez, pode migrar para o núcleo, o que torna possível sua ligação a LEF-1/TCF4 (*lymphoid-enhancer-binding factor/T-cell factor 4*) e o deslocamento da proteína Groucho. Isso permite a ativação do mecanismo de transcrição de vários genes alvos da via de sinalização WNT (THIERY e SLEEMAN 2006), os quais em uma célula diferenciada deveriam permanecer inativados.

É crescente o número de estudos que visam um melhor entendimento do processo EMT. BOYER et al. (2000) relatam o papel do processo EMT nas primeiras etapas da invasão e metástase de células tumorais. Este evento também tem sido associado com as mudanças moleculares observadas nas células neoplásicas metastáticas mais agressivas (THIERY 2003).

YANG et al. (2004) encontraram uma elevada expressão do fator de transcrição TWIST em células tumorais primárias, as quais eram capazes de penetrar os vasos sanguíneos. Relatam também que a expressão ectópica desse gene promove uma perda de adesão celular, a qual é mediada pela atividade da proteína E-caderina, ativação de marcadores mesenquimais e indução da motilidade celular, sugerindo que TWIST contribui com a metástase através do processo EMT.

Além disso, durante a invasão tumoral, à medida que as células epiteliais perdem adesão e alteram sua arquitetura, iniciam uma degradação da membrana basal, concomitantemente com um aumento na expressão de algumas integrinas.

As integrinas correspondem a uma família de proteínas transmembranas responsáveis pela ligação célula-matriz extracelular através de interações heterofílicas entre seus heterodímeros α e β extracelulares. Sendo assim, sua expressão em epitélio normal está confinada às células que promovem contato com a membrana basal. Seus principais ligantes de ECM são: colágeno-IV, fibronectina, vitronectina, laminina e heparan sulfato (ALBERTS et al. 2002a). Assim como as caderinas, também são proteínas que interagem indiretamente com o citoesqueleto de actina, através de ligações entre seu domínio citoplasmático e as proteínas ancoradoras α -actinina, vinculina, tensina e paxilina (HOOD e CHERESH 2002).

À medida que se ligam ao citoesqueleto, as integrinas formam *clusters* e desenvolvem um pequeno complexo focal, ao qual podem se agregar mais integrinas, resultando no desenvolvimento de uma estrutura conhecida como contato focal (FRIEDL e WOLF 2003). Ao formar este contato focal com a

ECM, a célula neoplásica ganha um fenótipo mais invasivo, pois estes contatos focais são dinâmicos e podem se formar rapidamente em novas localizações da célula à medida que esta necessite se mover. E dependendo do tipo celular e da matriz extracelular esta estrutura pode ser regulada por diferentes tipos de integrinas.

As integrinas também estão envolvidas com a regulação da atividade de proteases da ECM. As células produzem, ativam e liberam diversos tipos diferentes de proteases que especificamente degradam moléculas da ECM. E este processo deve ser finamente controlado, para que a degradação da ECM seja suficiente para a passagem das células, mas não tão elevada que impeça a tração celular. Dois exemplos de membros da família de metaloproteases de matriz (*matrix metalloproteinase* - MMP) são MMP2 e MMP9, os quais possuem a maior atividade enzimática contra colágeno do tipo IV, este que é o principal constituinte da membrana basal (HOOD e CHERESH 2002). Um maior nível de atividade proteolítica de MMP2 foi encontrado na superfície de células invasoras de carcinoma mamário do que em células não neoplásicas, devido à ativação de MMP2 pela ligação à integrina $\alpha\beta3$ (BROOKS et al. 1996).

Portanto, durante a invasão, a integrina $\alpha\beta3$ não funciona somente como um receptor de adesão/migração, mas também como ativador de proteases que são necessárias para a degradação da ECM (HOOD e CHERESH 2002), e consequentemente, para uma maior invasão da célula neoplásica em tecidos adjacentes.

1.3.1 O microambiente tumoral

Alguns estudos têm abordado a influência que algumas células adjacentes podem apresentar no desenvolvimento e progressão de um tumor. Tais células representam o microambiente tumoral. Estes trabalhos demonstram que o crescimento e a diferenciação celular, a polaridade de células epiteliais normais e de carcinomas mamários, e principalmente, o comportamento invasivo, são influenciados por células estromais adjacentes, incluindo fibroblastos, miofibroblastos, leucócitos e células mioepiteliais (RADISKY et al. 2001). Os miofibroblastos têm sido implicados no remodelamento da matriz extracelular, cicatrização de feridas e inflamações crônicas, assim como, na invasão de câncer de mama. Além disso, infiltração linfocítica, fibrose, angiogênese e linfogênese têm mostrado grande importância, como características prognósticas, no processo de invasão tumoral em neoplasias mamárias (ALLINEN et al. 2004). Estes mesmos autores relataram a relação de algumas quimioquinas com o processo de invasão em câncer de mama. Tais quimioquinas são proteínas que atuam no transporte de leucócitos. Neste estudo, os pesquisadores encontraram uma superexpressão das quimioquinas CXCL14 e CXCL12 em células mioepiteliais e em miofibroblastos tumorais, respectivamente. Tais proteínas encontraram-se ligadas a seus receptores nas células epiteliais, aumentando a proliferação, migração e invasão tumoral; atuando na tumorigênese, possivelmente, como fatores parácrinos. A quimioquina CXCL12 tem sido previamente relacionada à metástase (KANG et al. 2003), mas sua alta expressão em DCIS sugere também um papel nos estágios iniciais da tumorigênese (ALLINEN et al. 2004). Outro estudo sugere que ao invés de secreção de proteases para degradação

das membranas mio-epitelial e basal, o processo de invasão tumoral ocorre com morte localizada de células da camada mioepitelial (localizada acima da membrana basal), que quando somada a lesões do microambiente externo, e também a alterações genéticas nesta população celular, resultará no recrutamento de fatores que causarão a degradação da membrana basal (MAN e SANG 2004).

Esses relatos mostram a complexidade do processo de invasão e metástase tumoral, onde vários processos devem ser considerados como críticos para a regulação deste mecanismo.

1.4 METODOLOGIAS PARA AVALIAÇÃO DE EXPRESSÃO GÊNICA

1.4.1 Análises em grande escala

Diversas metodologias têm sido desenvolvidas com o intuito de fornecer informações em nível transcricional em larga escala, como por exemplo duas metodologias de grande destaque como a técnica de “Análise Serial de Expressão Gênica” (*Serial Analysis of Gene Expression – SAGE*) (VELCULESCU et al. 1995) e a de “Microarray” (SCHENA et al. 1995). Ambas as abordagens são métodos de identificação de genes diferencialmente expressos em diferentes tecidos, mas que divergem em importantes aspectos.

A metodologia de SAGE baseia-se na identificação de curtas seqüências nucleotídicas que representam fragmentos de transcritos. Estas curtas seqüências de 10, ou até mesmo de 21 nucleotídeos (LongSAGE) (SAHA et al. 2002), são chamadas de “tags” (etiquetas). Embora sejam seqüências curtas, acredita-se que cada tag tenha uma alta probabilidade de representar

especificamente um gene. Esta técnica representa uma medida direta do nível de expressão gênica, através da capacidade de identificação do número de *tags* existentes na amostra estudada. É uma técnica bastante sensível por ser capaz de identificar um número muito grande de transcritos, mesmo aqueles com baixo nível de expressão. Mas, por outro lado é uma técnica laboriosa, de alto custo e que necessita de grande quantidade de mRNA (2,5 a 5,0 µg), não sendo apropriada para a análise de um grande número de amostras (MADDEN et al. 2000; YE et al. 2002). Já a metodologia de *Microarray* difere marcadamente da metodologia de SAGE em dois aspectos. Só é possível analisar a expressão dos genes cujas moléculas de DNA estão imobilizadas numa plataforma e, além disso, é uma medida indireta de quantificação de transcritos, pois depende da incorporação de fluoróforos, como por exemplo, Cy3 e Cy5 (MADDEN et al. 2000; YE et al. 2002), moléculas equivalentes que emitem fluorescência com comprimentos de onda distintos, ou ainda moléculas radioativas. A grande vantagem dessa metodologia é a possibilidade de avaliar um grande número de amostras que representam uma classe em estudo, permitindo estimar a variabilidade genética na população estudada, diminuindo a possibilidade de identificar genes diferencialmente expressos associados à característica genética distinta de uma das amostras que não tenha relação com o aspecto estudado.

Outra abordagem utilizada que permite a identificação de genes diferencialmente expressos em diferentes tecidos é a técnica de Hibridação Subtrativa (JIANG et al. 2000). Esta é uma metodologia que se baseia na construção de uma biblioteca de cDNA, da qual serão subtraídos transcritos que são comuns aos dois tecidos. Isto acaba revelando a presença daqueles

que são mais ou unicamente expressos no tecido que teve os transcritos comuns subtraídos. Além disso, essa metodologia é capaz de identificar transcritos ainda não descritos na literatura. Recentemente foi desenvolvido um aprimoramento desta técnica, que torna esta abordagem rápida, conhecida como “Rápida Hibridação Subtrativa” (RaSH) (JIANG et al. 2000). Ela possui a vantagem de necessitar de pouca quantidade de mRNA (1,0 μ g), e além disso, é uma metodologia tecnicamente pouco laboriosa que simplifica significativamente a subtração de bibliotecas de cDNAs em relação a outros protocolos, tornando-a um método rápido e que barateia custos (JIANG et al. 2000).

Este método de hibridação subtrativa facilita a remoção das seqüências comumente expressas entre duas populações de transcritos, posteriormente a uma etapa de desnaturação e renaturação. Ela é baseada numa subtração de massa, alterando-se a razão da quantidade de cDNA da biblioteca que se quer subtrair (*driver*), ou seja, da amostra que se pretende retirar da amostragem final, em relação a quantidade de cDNA da biblioteca que se quer identificar os genes diferencialmente expressos (*tester*). Desta forma, a subtração ocorre com a ligação das extremidades dos cDNAs *tester* num vetor, ambos previamente clivados com uma enzima de restrição. Este simples passo da subtração, juntamente com a alteração da proporção de massa molecular, torna esta técnica diferente de qualquer outra usada em outros protocolos de subtração de biblioteca de cDNAs (JIANG et al. 2000). Em seguida, os fragmentos resultantes da subtração são clonados e sequenciados, identificando assim, os prováveis genes que estão mais expressos na população de interesse (*tester*).

1.4.2 RT_PCR em tempo real

A metodologia conhecida como *Real-time* – reverse transcription - *polymerase chain reaction* (*real-time RT-PCR*) (RT-PCR em tempo real) (GIBSON et al. 1996) têm sido atualmente muito utilizada por sua alta sensibilidade, boa reprodutibilidade e amplo espectro de quantificação do nível de expressão gênica (PFAFFL 2002). É baseada em um sistema de detecção fluorescente, capaz de fornecer a medida simultânea de expressão gênica em diferentes amostras; mesmo quando a amostra apresenta uma pequena quantidade de mRNA, como por exemplo, a partir de apenas 1 célula ou a partir de apenas 1 molécula de DNA (LOCKEY et al. 1998). Uma desvantagem desta metodologia é na investigação exploratória, pois não é apropriada para analisar o nível de expressão gênica de um número muito grande de transcritos. Atualmente alguns produtos disponíveis no mercado permitem analisar genes que representam vias bioquímicas completas, fornecendo um retrato bioquímico do grupo de amostra em questão (*RT² Profiler™ PCR Array* – *SuperArray Bioscience Corporation*). Mas de qualquer forma é uma metodologia menos exploratória que as demais já citadas.

Em geral, pode-se utilizar duas estratégias de quantificação na análise dos resultados produzidos por essa metodologia: a quantificação absoluta (1) e a relativa (2). A quantificação absoluta de mRNA é determinada a partir de uma curva padrão estabelecida a partir de quantidades precisamente conhecidas de material. Essa curva é utilizada como referência na quantificação das amostras testadas. Já a quantificação relativa é baseada na razão de expressão entre um gene alvo e um gene referência, sendo adequada para experimentos que investigam mudanças fisiológicas em níveis de expressão gênica (PFAFFL

2002). A segunda forma de quantificação (quantificação relativa) é a mais comumente utilizada. Os genes referência ou calibrador normalmente são genes de manutenção celular, conhecidos como genes endógenos ou frequentemente chamados de genes constitutivos (*housekeeping genes*). Em muitos estudos a expressão destes genes é considerada como invariável entre diferentes tecidos e são utilizados como normalizadores sem uma validação apropriada. Estudos recentes têm mostrado que suas expressões individuais podem variar como resultado do crescimento tumoral, hipóxia, ou por tratamentos experimentais (SCHMITTGEN e ZAKRAJSEK 2000; JANSSENS et al. 2004), podendo alterar a correta interpretação dos resultados. Recentemente VANDESOMPELE et al. (2002) desenvolveu um *software* conhecido como *geNorm* para calcular o melhor normalizador para um determinado grupo de amostras.

O *software geNorm*, de livre acesso e visualizado a partir do *software Microsoft Excel*, calcula a estabilidade individual de cada gene em relação ao seu grupo de genes candidatos a normalizadores em uma determinada amostra. Esta estabilidade é calculada de acordo com a similaridade em seu perfil de expressão através de uma comparação par-a-par. O gene com o maior valor M (medida de variação de expressão gênica) é excluído do grupo. Posteriormente, esta mesma análise é repetida até a seleção do melhor normalizador e/ou conjunto de genes que apresentam o menor valor M para este determinado conjunto de amostras é utilizado para normalizar os dados (VANDESOMPELE et al. 2002). Dessa forma os dados são normalizados utilizando a média geométrica dos genes mais estáveis no grupo de amostras estudado, como um fator de normalização. Esse método de normalização

trouxe inúmeros benefícios para superar a dificuldade na escolha de genes normalizadores e está sendo amplamente utilizado (SZABO et al. 2004; LARIONOV et al. 2005).

Devido à alta sensibilidade e reprodutibilidade da metodologia de RT_PCR em tempo real, ela vem sendo bastante utilizada na validação técnica de genes candidatos a estarem diferencialmente expressos, identificados por técnicas que permitem uma análise em larga escala. É crescente a utilização do RT_PCR quantitativo na avaliação de genes candidatos em um grande número de amostras para correlacionar com características clínicas e assim identificar marcadores de diagnóstico tumoral ou prognóstico para ser utilizado na rotina clínica (DE KOK et al. 2002). STATHOPOULOU et al. (2006) propuseram aplicar a metodologia para monitorar e quantificar células epiteliais circulantes em sangue periférico de pacientes operados de câncer de mama. Outros como HUGHES et al. (2006) demonstraram que o RT_PCR em tempo real foi capaz de caracterizar, através da amplificação de marcadores moleculares previamente selecionados, linfonodos sentinelas mamários acometidos por metástase, com uma acurácia superior às análises intra-operatórias de linfonodos sentinelas, podendo auxiliar os diagnósticos patológicos.

1.5 ESTUDOS CLÍNICOS ORIENTADOS

A elucidação dos eventos genéticos que explicam o início e a progressão do câncer de mama é prejudicada por aspectos inerentes a delineamentos experimentais que utilizam linhagens celulares. Estudos

utilizando linhagens celulares representam um material genético que não sofreu os mesmos eventos moleculares sofridos pelo tecido tumoral que as originou, não permitindo uma correlação segura entre o perfil de expressão gênica e a morfologia tumoral. Já os estudos que utilizam tecido tumoral, são mais interessantes para estudos clínicos orientados, pois permitem uma correlação mais acurada. Tais estudos eram limitados no passado pela dificuldade em capturar populações puras de células a partir de um tecido tumoral complexo (BONNER et al. 1997). Atualmente, com o desenvolvimento da técnica de LCM (*laser capture microdissection*) (EMMERT-BUCK et al. 1996) essa dificuldade pôde ser superada. Essa técnica permite a coleta rápida, precisa e seletiva de um mesmo tipo celular a partir de tecidos heterogêneos (morfologicamente complexos), os quais sofreram a influência de seu microambiente tecidual nativo (EMMERT-BUCK et al. 1996); aumentando a acurácia e a sensibilidade de estudos moleculares. Vários grupos de pesquisa (SGROI et al. 1999; MA et al. 2003; SCHUETZ et al. 2006), incluindo o nosso Castro et al. (em preparação), têm utilizado esta metodologia em lesões tumorais de mama, que mimetizam a instalação e progressão da doença, com a finalidade de obter uma análise mais fidedigna do perfil de expressão da célula tumoral dentro do contexto de sua morfologia, do que do perfil de expressão do tecido como um todo.

Uma dificuldade inerente de células microdissecadas a laser é a ínfima quantidade de mRNA adquirida. Com isso, surge a necessidade de amplificação deste mRNA. Este procedimento é comumente realizado através de 2 *rounds* de transcrição *in vitro*, sendo que um dos métodos mais utilizados é baseado na metodologia proposta por WANG et al. (2000), a qual resulta numa amplificação

linear que aumenta a quantidade de mRNA na ordem de $5,0 \times 10^2$ a $1,0 \times 10^3$ vezes (ZHAO et al. 2002), não apresentando alteração na representatividade dos transcritos em estudos de microarray (GOMES et al. 2003). No entanto, essa metodologia não tem sido explorada em outras análises de expressão gênica.

1.6 JUSTIFICATIVA

Considerando a alta incidência do câncer de mama no Brasil e no mundo, e que a maioria dos tipos de câncer de mama é do grupo ductal, ressalta-se a importância de melhor caracterizar os genes candidatos a participarem do processo de invasão de células tumorais para o tecido adjacente nesse tipo de câncer, melhor caracterizando a transição para o processo invasivo. Para essa finalidade, nossa proposta neste estudo é a identificação de genes diferencialmente expressos entre DCIS e IDC, identificando putativos genes chaves envolvidos nos primeiros eventos de progressão da doença, que poderão ser testados como marcador potencial de invasão e/ou prognóstico em carcinoma ductal de mama. Como há vários relatos na literatura, inclusive de nosso grupo, que o padrão geral de expressão desses 2 componentes, oriundos de uma mesma amostra, é extremamente similar quando analisado através de microarray, nosso objetivo foi aplicar outra metodologia mais sensível, que pudesse selecionar genes diferencialmente expressos e identificar genes novos e sem o prévio conhecimento de estarem envolvidos nesse processo. Assim aplicamos a metodologia de bibliotecas subtrativas RaSH, utilizando células microdissecadas a *laser* dos dois componentes celulares oriundas de uma única amostra tumoral. Esta abordagem aumenta a

chance de identificação de genes envolvidos com o processo de invasão celular, uma vez que o *background* genético das duas populações de células é o mesmo. Dessa forma, nesse trabalho a metodologia RaSH foi adaptada para RNA amplificado oriundo de quantidades ínfimas de RNA total, as quais foram provenientes de células tumorais microdissecadas a laser.

2 OBJETIVOS

2.1 OBJETIVO GERAL

- Identificar genes-chaves envolvidos no processo de invasão, que possam ser testados como potenciais marcadores moleculares de invasão e/ou prognóstico em carcinoma ductal de mama.

2.2 OBJETIVOS ESPECÍFICOS

- Adaptar a metodologia RASH para sua aplicação em amostras de RNA amplificado.
- Construir bibliotecas de cDNA subtrativas (RaSH) dos componentes tumorais DCIS (*tester*)/IDC (*driver*) e IDC (*tester*)/DCIS (*driver*) de uma mesma amostra tumoral.
- Identificar os genes específicos de cada biblioteca RASH.
- Confirmar a diferença de expressão de genes candidatos através da técnica de RT_PCR em tempo real, nas mesmas amostras utilizadas para confecção das bibliotecas
- Avaliar a expressão dos genes candidatos em um grupo independente de amostras, através da imobilização de todos os candidatos em lâmina de vidro (*microarray*).

3 MATERIAL E MÉTODOS

3.1 SELEÇÃO DOS CASOS

Cinco casos de carcinoma ductal invasivo de mama foram pré-selecionados a partir do Banco de Tumores do Centro de Tratamento e Pesquisa - Hospital A. C. Camargo que preenchiam os seguintes critérios: 1) informações sobre a presença do componente DCIS, 2) pacientes com menos de 50 anos de idade, 3) amostra retirada cirurgicamente antes de tratamento hormonal ou quimioterápico e 4) presença de linfonodos acometidos por células neoplásicas.

3.2 COLORAÇÃO DAS LÂMINAS HISTOLÓGICAS PARA MICRODISSECÇÃO POR LCM

Os cortes histológicos foram realizados em condições de baixa temperatura (cerca de 30° C negativos) para fins de preservação da integridade do RNA total do tecido tumoral, sendo que cada uma das lâminas recebeu 0.4 µm de espessura de tecido em sua superfície. Foi realizada, antes da microdissecção, a coloração de uma das lâminas histológicas com hematoxilina e eosina (H.E.) possibilitando a confirmação da presença dos dois componentes tumorais (DCIS e IDC). As demais lâminas não receberam nenhum tipo de coloração e foram mantidas a 80° C negativos até o momento da microdissecção. Minutos antes da microdissecção a laser, estas lâminas foram submetidas à coloração com azul de toluidina, segundo as seguintes

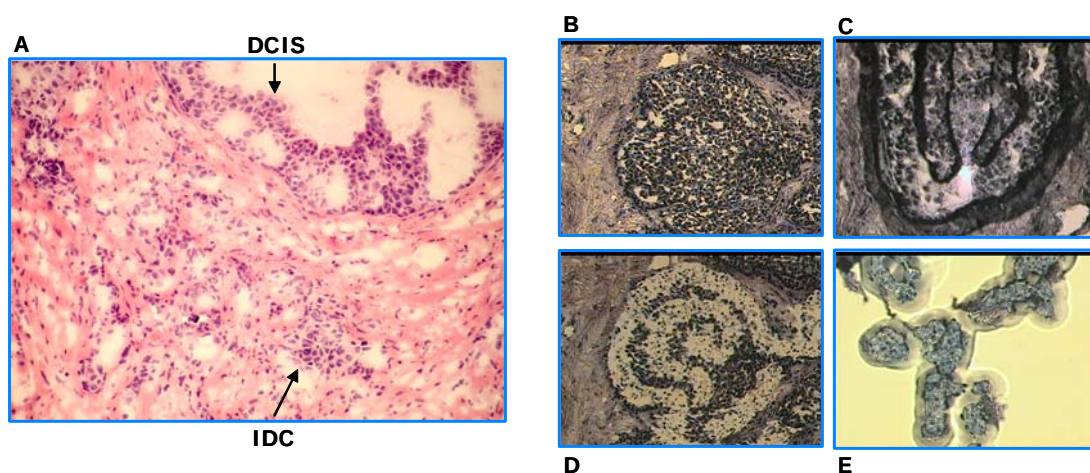
etapas: 1) Imersão da lâmina em etanol 75% (álcool diluído com H₂O DEPC) por 30 segundos para fixação do tecido; 2) Imersão da lâmina em H₂O DEPC por 30 segundos para lavagem; 3) Coloração da lâmina com a adição de 100 µL de corante azul de toluidina por 20 segundos; 4) Imersão da lâmina em H₂O DEPC para lavagem do corante por 30 segundos; 5) Imersão da lâmina em etanol 75, 95 e 100% por 30 segundos em cada um deles, para desidratação do tecido; e 6) Imersão da lâmina em xilol por 5 minutos para diafanização.

3.3 MICRODISSECÇÃO POR LCM (LASER CAPTURE MICRODISSECTION)

Tanto a confirmação da presença dos 2 componentes de lesão nas amostras utilizadas quanto a microdissecção a laser das células de interesse foram realizadas pela patologista Cynthia A. B. T. Osório, do grupo do Dr. Fernando Soares. Para tanto, utilizou-se o aparelho *Pixcell Laser Capture Microdissection System (Arcturus Systems for Microgenomics)*, o qual pertence ao Departamento de Pesquisa do Hospital A.C. Camargo. Este sistema consiste basicamente de um microscópio invertido, ao qual é acoplada uma fonte de raio laser infravermelho de baixa potência e uma câmera ligada a um monitor e um computador. A lâmina com o corte histológico é fixada ao *chariot* por um mecanismo de sucção a vácuo. O feixe de raio laser incide sobre uma pequena peça plástica, chamada *cap*, a qual é revestida por um filme termoplástico. Este *cap* fica posicionado entre o feixe de raio laser e o corte histológico. Com o disparo do feixe de *laser*, o filme termoplástico recebe esta energia luminosa e esquenta. Com isso, ele expande e se adere à célula escolhida. Quando o

pulso de raio laser cessa, o filme resfria sofrendo retração, e assim, leva consigo a célula aderida à superfície do *cap*. A figura 1 ilustra o procedimento de microdissecção a laser.

Diversas lâminas histológicas foram reservadas para a captura de células dos dois componentes tumorais DCIS e IDC. Foram efetuados, em média, cerca de 15 mil disparos de raios laser para a microdissecção de células de cada um dos 2 componentes, para cada amostra tumoral selecionada. O diâmetro do feixe de raios laser foi ajustado para 15 μm . Desta forma, a cada disparo era possível coletar de uma a duas células de cada componente DCIS e IDC; de acordo com o pulso utilizado, o qual foi ajustado para 0,2 segundos -o disparo de raios laser com um pulso de 0,2 segundos torna mais fácil a captura de um número maior de células por *cap* (*CapSureTM HS LCM Caps*) num menor tempo. O tempo máximo de captura estabelecido por prévios experimentos de nosso laboratório, para obtenção de RNA sem prejuízo na sua integridade, foi de 30 minutos.



Legenda: A: Coloração por Hematoxilina-Eosina para confirmação da presença de ambos os componentes tumorais. B: Ducto antes da microdissecção. C: Ducto sendo microdissecado. D: Ducto após microdissecção. E: Células de DCIS aderidas ao *cap*.

Figura 1 - Microdissecção a laser de DCIS e IDC de uma mesma amostra tumoral.

3.4 EXTRAÇÃO DE RNA TOTAL

Ao final da operação, os *caps* contendo apenas células provenientes de DCIS e os *caps* contendo apenas células provenientes de IDC foram retirados com as células aderidas a sua superfície.

Para obtenção do extrato celular, estes *caps* receberam um tampão de extração celular (*PicoPureTM RNA Isolation Kit*), sendo mantidos a 42° C por 30 minutos. Após os *caps* terem sido acoplados a tubos plásticos e terem sido invertidos, procedeu-se com uma centrifugação dos mesmos a 800 rpm por 2 minutos. Desta forma foi possível coletar o lisado celular e, posteriormente a isso, seguiu-se com a extração do RNA total dos dois componentes tumorais. A extração de RNA total seguiu o protocolo descrito no kit de extração *PicoPureTM RNA Isolation Kit*), seguindo recomendações do fabricante.

3.5 ADAPTAÇÃO DO PROTOCOLO DE CONSTRUÇÃO DE BIBLIOTECAS SUBTRATIVAS RASH A PARTIR DE RNA AMPLIFICADO

3.5.1 Amplificação de mRNA

A metodologia de amplificação de mRNA utilizada está descrita em GOMES et al. (2003), com algumas modificações para amostras microdissecadas a *laser* SARAIVA (2005), das quais se obtém ínfimas quantidades de RNA total (da ordem de poucos nanogramas). Essas modificações consistiram em adição de carreadores como glicogênio para aRNA (20 mg/mL) e acrilamida linear para cDNA (5 mg/mL). Para utilização

[illegible]

Logo após o término da reação de síntese da primeira fita de cDNA, procedeu-se com a reação de síntese da segunda fita, baseado na metodologia *template switch* (TS) (MATZ et al. 1999). Para isso, usou-se 200 ng do oligonucleotídeo TS_DpnII, o qual também possuía uma sequência alvo para a enzima de restrição *Dpn* II e uma trinca de guaninas na extremidade 3' que alinham na sequência de citosinas (as quais são depositadas pela enzima transcriptase reversa SuperScript II) da extremidade 3' da primeira fita de cDNA recém sintetizada. O oligonucleotídeo TS_DpnII está apresentado a seguir, onde o sítio de reconhecimento da enzima DpnII está representado em azul - 5' AAG CAG TGG TAA CAA CGC AGA GAT CGC GGG 3' [30 pb]. Esta segunda reação contou com todo o volume da reação de síntese de 1ª fita de cDNA, 200 ng do oligonucleotídeo TS_DpnII, 1X *Advantage PCR Buffer*, 0,2 mMol/L dNTP, 1,4 U RNase H (*Promega*), 1X *Advantage Polimerase cDNA mix* (*Clontech*). Esta reação foi submetida as seguintes condições: 37° C por 10 minutos para digestão do mRNA, 94° C por 3 minutos para desnaturação, 65° C por 5 minutos para alinhamento do *primer* e 75° C por 30 minutos para extensão.

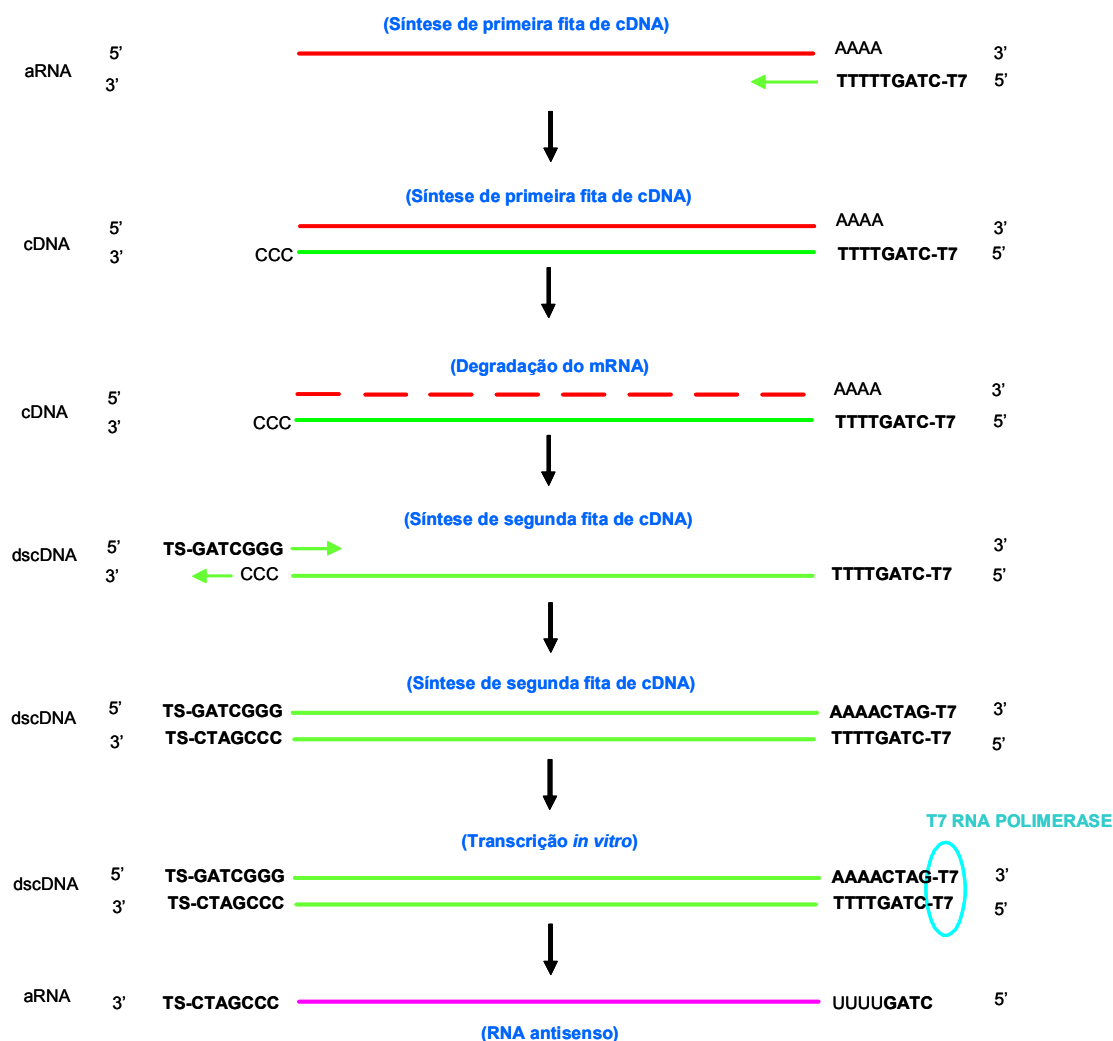
Após a inativação da reação com a adição de 2,5 mMol/L NaOH e 0,5 mMol/L EDTA e incubação a 65° C por 10 minutos, foi realizada sua purificação através da adição de volume 1:1 de fenol: clorofórmio: álcool isoamílico (25: 24: 1) com valor de pH igual a 8,0, com vortex por 2 minutos; manteve-se em temperatura ambiente por 10 minutos com posterior centrifugação a 14.000 rpm por 15 minutos. O sobrenadante foi aspirado e dispensado em novo tubo com adição de um volume 1:1 de clorofórmio com vortex por 2 minutos; manteve-se a temperatura ambiente por 10 minutos com posterior centrifugação a 14000 rpm por 15 minutos. O sobrenadante foi aspirado e

dispensado em novo tubo para sofrer precipitação, onde recebeu 3 volumes de etanol absoluto, 1/10 do volume final de acetato de sódio (NaAc) 3M com valor de pH igual a 5.2, e 5 ng de acrilamida linear,,usado aqui como carreador da precipitação. Manteve-se a 80° C negativos por 1 hora com posterior centrifugação a 14000 rpm por 30 minutos a 4° C. Por fim, foram realizadas cinco lavagens do *pellet* de cDNA com etanol 70%.

O cDNA foi ressuspendido em 12,0 µL de H₂O DEPC. Deste volume foram reservados 1,0 µL para quantificação em espectrofotômetro *GeneQuant* (Amersham Biosciences) e 1,0 µL como molde para uma reação de RT-PCR, assumindo-se que a extensão do fragmento relativo ao gene GAPDH evidencia indiretamente a integridade do RNA total. É importante ressaltar que este tipo de monitoramento foi necessário por que a quantidade de RNA total obtida foi insuficiente para ser avaliada por eletroforese em gel de agarose. Assim, os produtos amplificados por RT-PCR das amostras DCIS e IDC foram submetidos a uma eletroforese em gel de acrilamida, o qual foi corado com nitrato de prata (SANGUINETTI et al. 1994). Essa reação foi realizada com 1X PCR *buffer*, 1,5 mMol/L MgCl₂, 0,2 mMol/L dNTP, 0,4 µMol/L *primer forward*, 0,4 µMol/L *primer reverse*, 1 U Taq DNA polimerase (*Invitrogen*).

Para as reações que apresentaram extensão dos fragmentos específicos (RNA de qualidade satisfatória), foram realizadas reações de transcrição *in vitro* utilizando os cDNAs dupla fita. Essa reação constituiu-se de todo o volume de cDNA sintetizado, 75 mMol/L rNTP mix (A, C, G e UTP), 1X *reaction buffer*, 2,5 µL mix da enzima (polimerase do fago T7 e Inibidor de RNase).

Esta reação foi mantida a 37° C por 6 horas gerando RNA amplificado antisenso (aRNA). Todas as etapas desse procedimento estão representadas na figura 2. Após o término da reação de transcrição *in vitro*, esta foi submetida a uma purificação com a adição de 1,0 mL do reagente TRIzol® Reagent (*Sigma–Aldrich Corporation*). Depois de centrifugar a reação a 12000 g por 15 minutos a 4° C, foi adicionado ao sobrenadante 1/10 de NaAc e 10 ng do carreador glicogênio e este foi precipitado com adição de isopropanol num volume de 1:1. Após lavagem, o aRNA foi ressuspendido em 7,0 µL de H₂O DEPC, sendo que 1,0 µL deste volume foi reservado para quantificação em espectrofotômetro *GeneQuant* (*Amersham Biosciences*). Depois da reação de amplificação o aRNA foi submetido novamente a uma reação de síntese de cDNA, por duas metodologias diferentes.



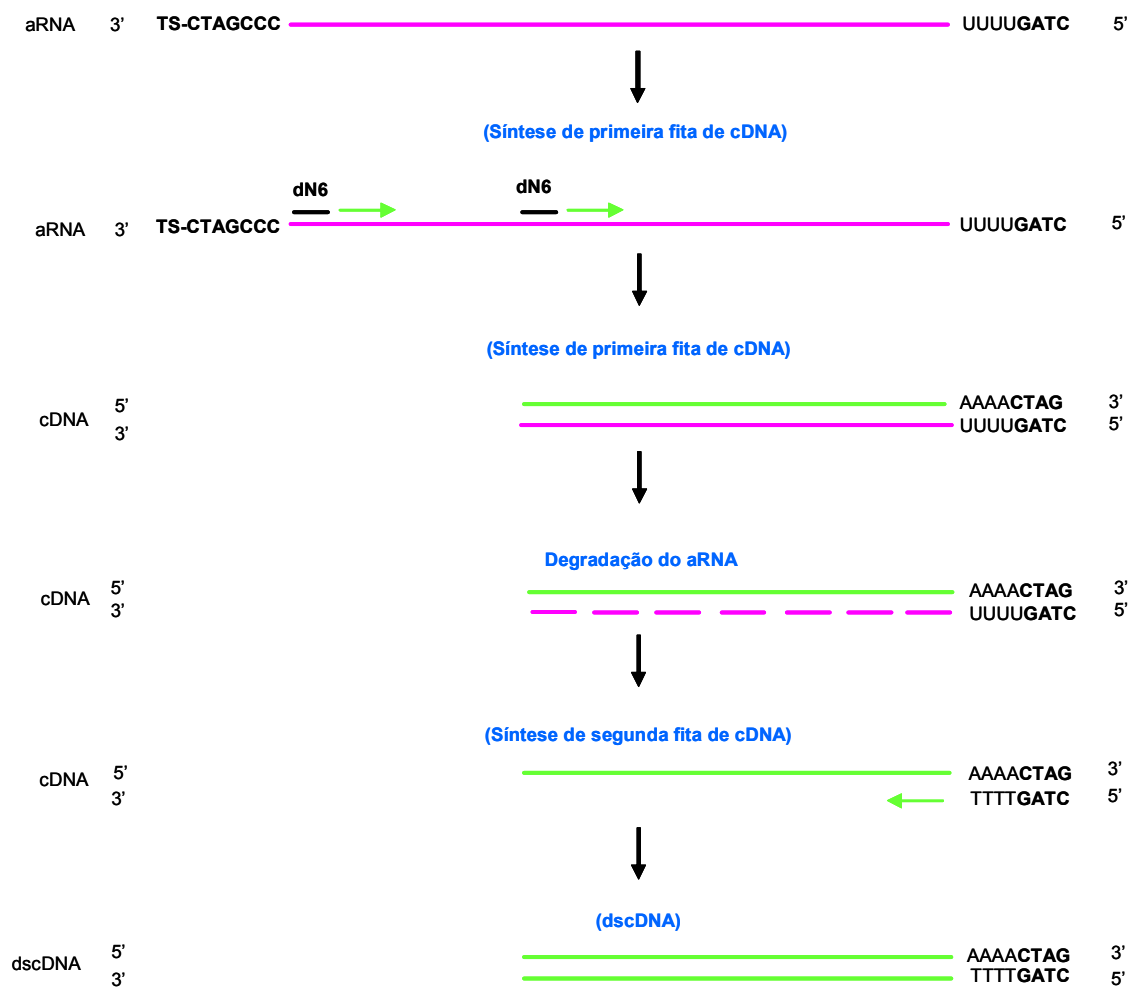
Legenda: Ilustração esquemática da metodologia de síntese de cDNA com os oligonucleotídeos dT24-T7-DpnII e TS-DpnII e 1 *round* de amplificação de mRNA por transcrição *in vitro*, adaptada para a construção da biblioteca de RASH.

Figura 2 – Síntese de cDNA e amplificação por transcrição *in vitro* de mRNA.

Na primeira, para a reação de síntese de primeira fita de cDNA, foram adicionados 9 µg do oligonucleotídeo *random primer* (dN6) aos 6,0 µL de aRNA (figura 3). Esta reação seguiu as mesmas condições da primeira reação (item 3.5.1). Na segunda metodologia, para a reação de síntese de primeira fita de cDNA foram adicionados 300 ng do oligonucleotídeo TS_DpnII aos 6,0 µL de aRNA, para alinhamento na região 3' dos transcritos que apresentam o sítio complementar ao *primer* TS (resultado da transcrição *in vitro*) (figura 4). As

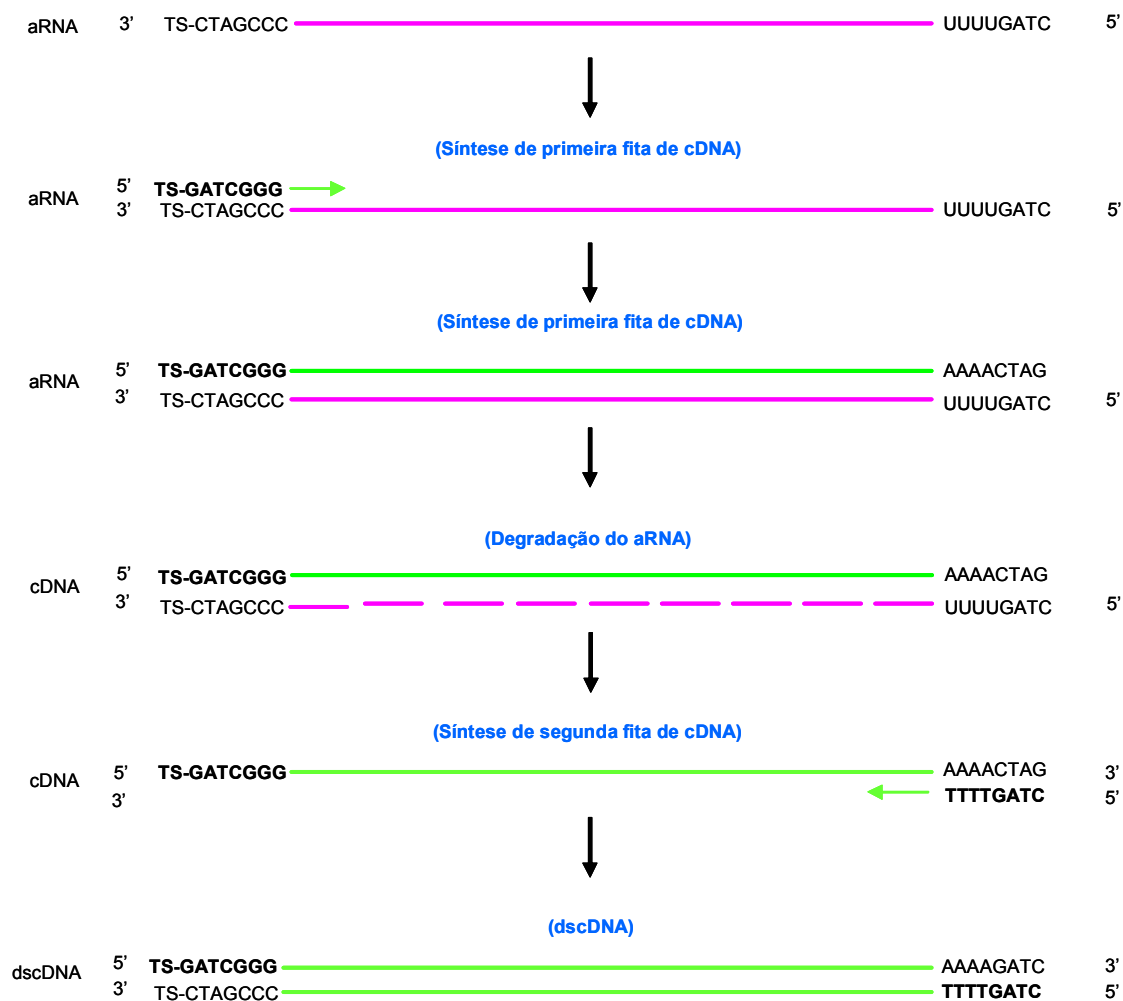
condições desta reação também seguiram as condições da primeira reação (item 3.5.1).

Terminada esta reação, procedeu-se com a reação de síntese da segunda fita de cDNA. Nesta nova reação foram adicionados 250 ng do oligonucleotídeo dT24-T7_DpnII. Este procedimento foi o mesmo para as duas metodologias. As condições desta reação seguiram como descritas no item 3.5.1. Ao término, seguiu-se com uma purificação e precipitação (Item 3.5.1). O cDNA foi ressuspendido em 12,0 µL de H₂O DEPC. Todas as etapas da síntese de cDNA utilizando o oligonucleotídeo *random primer* (dN6) e o TS_DpnII a partir do RNA antisense amplificado estão representadas nas figuras 3 e 4, respectivamente. Deste volume foram reservados 1,0 µL como molde para uma reação de PCR para monitorar a qualidade do ds-cDNA, a partir do aRNA, através de nova amplificação do gene GAPDH. Além disso, metade do volume restante (5,5 µL) de ds-cDNA de cada população de transcritos foi reservado para posteriormente ser submetido à experimentos de confirmação de diferença de expressão por RT_PCR em tempo real.



Legenda: Ilustração esquemática da metodologia de amplificação de mRNA adaptada para a construção da biblioteca de RASH utilizando *random primer* na síntese da primeira fita após o ciclo de transcrição *in vitro*.

Figura 3 - Metodologia de amplificação de mRNA por *random primer*.



Legenda: Ilustração esquemática da metodologia de amplificação de mRNA adaptada para a construção da biblioteca de RASH utilizando TS_DpnII na síntese da primeira fita após o ciclo de transcrição *in vitro*.

Figura 4 - Metodologia de amplificação de mRNA por *template switch*.

3.5.2 Clivagem dos cDNAs com a enzima de restrição *Dpn II*

A metodologia RaSH inicia-se com a clivagem das duas populações de cDNA dupla fita com a enzima de restrição *Dpn II*. Embora a enzima *Dpn II* apresente um corte freqüente (a sequência nucleotídica alvo corresponde à 4 nucleotídeos, estando presente teoricamente a cada 256 pb num transcrito), foi inserido 1 sítio de reconhecimento dessa enzima nos oligonucleotídeos utilizados na amplificação, para aumentar a chance do transcrito apresentar os

2 sítios de restrição permitindo sua representação nas bibliotecas RaSH. Esse procedimento deve favorecer principalmente a metodologia que utiliza o *random primer* na síntese de primeira fita a partir de RNA amplificado, a qual produz fragmentos com tamanhos menores.

Desta forma, as duas populações de transcritos DCIS e IDC foram clivadas com a enzima de restrição *Dpn* II (*New England Biolabs*). O controle desta reação de restrição foi realizado através de uma reação de clivagem de um fragmento de DNA de 459 pb, o qual possui um único sítio alvo para *Dpn* II, resultando em 2 fragmentos, um de 127 e outro de 332 pb. Assim, nas reações de clivagem de DCIS e IDC foram utilizadas 180 U de enzima para clivar 6 µg de DNA, num volume de 180,0 µL de reação; enquanto que para a reação controle foram utilizadas 15 U da enzima para clivar 0,5 µg de DNA, num volume de 15,0 µL de reação, mantendo assim a mesma proporção de unidade de enzima por quantidade de DNA e por volume de reação. As reações foram mantidas por 3 horas a 37° C.

Após a inativação da reação de clivagem, mantendo a reação a 65° C por 20 minutos, a reação foi submetida a uma purificação e a uma precipitação pelo método de fenol-clorofórmio como citado no item 3.5.1.

3.5.3 Ligação das amostras DCIS e IDC aos adaptadores XDPN12 e XDPN14

Os adaptadores XDPN12 e XDPN14 foram adicionados juntos à reação de ligação numa concentração final de 10 µM. Todo o volume de DNA digerido com a enzima *Dpn* II foi submetido à reação de ligação a esses adaptadores, e o volume final da reação foi ajustado para 40,0 µL, mas sem a adição da

enzima T4 DNA Ligase. A Reação de ligação dos adaptadores foi realizada com 10 $\mu\text{Mol/L}$ *primer* XDPN12, 10 $\mu\text{Mol/L}$ *primer* XDPN14, 1X *buffer* e 20,0 μL DNA.

Antes de adicionar a enzima de ligação, fez-se a diminuição gradual da temperatura para que os adaptadores se ligassem entre si de maneira correta, antes da ligação destes às moléculas de DNA. Sendo assim, a reação foi mantida a 55° C por 1 minuto e, posteriormente, a temperatura foi decaindo 2° C a cada 2 minutos até a reação atingir 30° C. Depois disso, a temperatura decaiu em 2° C a cada 4 minutos. Um esquema deste processo pode ser visto a seguir:

54° C por 2'
 52° C por 2'
 50° C por 2'
 48° C por 2'
 46° C por 2'
 44° C por 2'
 42° C por 2'
 40° C por 2'
 38° C por 2'
 36° C por 2'
 34° C por 2'
 32° C por 2'
 30° C por 2'
 28° C por 4'
 26° C por 4'
 24° C por 4'
 22° C por 4'
 20° C por 4'
 18° C por 4'
 16° C por 4'
 14° C por 12 horas

Ao atingir a temperatura de 14° C, reservou-se 1,0 μL de cada amostra para controle posterior (sem-ligase). Então, foram adicionadas 2000 U da enzima T4 DNA Ligase (*New England Biolabs*) em cada amostra. Com isso, o volume final da reação foi de 40,0 μL ; sendo mantida por 12 horas a 14° C.

5' CTGATCACTCGAGAGATC..... GATCTCTCGAGT 3' Xho I
 3' ← CTAGAGAGCTCACTAGTC 5'

5' CTGATCACTCGAGAGATC → Xho I
 3' TGAGCTCTCTAG.....CTAGAGAGCTCACTAGTC

3.5.4 PCR-Teste da ligação dos adaptadores

Para verificar se a reação de ligação dos adaptadores havia funcionado, foi realizada uma reação de PCR. Para isso, foram adicionados 60,0 µL de LoTE (Low TE) (Invitrogen) aos 40,0 µL de cada uma das duas reações de ligação. Nesta reação de PCR foram utilizados 2,0 µL dos 100,0 µL de cada reação de ligação, com a finalidade de servir como DNA molde. Da mesma forma, foi utilizado 1,0 µL da amostra sem-ligase para controle. A reação de PCR foi realizada com o uso de 10 µMol/L *primer* XDPN18, 1X PCR *buffer*, 0,1 µMol/L dNTP, 1,5 mMol/L MgCl₂, 2,5 U Taq DNA polimerase (*Invitrogen*). Esta reação foi submetida às seguintes condições: 72° C por 5 minutos; 25 ciclos de 94° C por 1 minuto, 55° C por 1 minuto e 72° C por 1 minuto; acrescido de 1 minuto a 72° C. O *primer* XDPN18 apresentava uma sequência alvo para a enzima de restrição *Xho* I.

3.5.5 PCR em larga escala

Após a confirmação da ligação dos adaptadores através da reação de PCR-Teste, foram realizadas 20 reações réplicas de PCR para DCIS e IDC de cada amostra, utilizando o mesmo procedimento descrito acima. Depois de somar as 20 reações de cada amostra, essas foram submetidas a uma purificação e uma precipitação (Item 3.5.1). Ressuspendeu-se os DNAs de cada amostra em 100,0 µL de H₂O Mili-Q, com posterior quantificação em espectrofotômetro *UV/VIS Spectrometer Lambda Bio10 (Perkin Elmer)*.

3.5.6 Clivagem dos cDNAs com a enzima de restrição *Xho* I

Nesta reação a enzima de restrição *Xho* I foi utilizada para clivar 10 µg de DNA de cada um dos dois componentes tumorais das duas amostras. Estas populações que foram clivadas receberam o nome de *tester*. Esta reação distinguiu a população *tester* daquela que não foi clivada com a enzima de restrição *Xho* I, as quais foram chamadas de *driver*. Entretanto, houve a digestão de apenas uma porção das moléculas de DCIS e IDC das duas amostras. A reação de clivagem foi mantida a 37° C por 6 horas e contou com 10 µg de DNA (DCIS ou IDC), 1X *Xho* I *buffer* e 20 U *Xho* I (*New England Biolabs*).

Mais uma vez as reações foram purificadas e precipitadas (Item 3.5.1). Então, os *pellets* de DNA foram ressuspensos em 30,0 µL de H₂O.

3.5.7 Hibridização subtrativa

A hibridização subtrativa baseou-se na diferença de massa entre as duas populações (cinquenta vezes mais massa molecular de uma população (*driver*) em relação à outra (*tester*)). A reação seguiu com a adição de 5 µg do componente DCIS (*driver*) à 100 ng de IDC (*tester*) (SUB1). O inverso também foi realizado (SUB2). Estas populações de DNA foram mantidas em um tampão composto por 40% formamida, 0,2% SDS, 50 mM Tris-HCl pH 7,5 e 0,5 M NaCl, o qual minimiza a excisão de ácidos nucleicos em hibridizações incubadas por longos períodos, e, também minimiza alinhamentos inespecíficos. A reação de hibridização subtrativa das diferentes populações de transcritos (DCIS e IDC) foi realizada como descrita abaixo:

SUB1	SUB2
Tampão de Hibridização – 16,6 µL	Tampão de Hibridização – 16,6 µL
Amostra Driver = DCIS – 5,0 µL (5 µg)	Amostra Driver = IDC – 5,0 µL (5 µg)
Amostra Tester = IDC – 0,4 µL (100 ng)	Amostra Tester = DCIS – 0,4 µL (100 ng)
22 µL	22 µL

Esta reação foi mantida a 100° C por 5 minutos para desnaturação das moléculas de DNA dupla fita e depois a 42° C por 48 horas para renaturação das moléculas dupla fita e conseqüente formação de moléculas híbridas (uma fita da molécula de DNA da população *driver* com uma fita da molécula de DNA da população *tester*). A diferença de massa entre as populações *driver* e *tester* favorece a formação de moléculas duplas *driver/tester* e *driver/driver*. Desta forma, desfavorece a formação de moléculas duplas *tester/tester* pelo fato de estarem em menor quantidade, sendo que sua formação deve corresponder a moléculas específicas ou super-representadas (genes super-expressos) na população de transcritos *tester*, esta última tendo sido clivada com a enzima de restrição *Xho* I. Posteriormente procedeu-se mais uma vez com uma purificação e uma precipitação (Item 3.5.1) do *pellet* de DNA, o qual foi ressuspensionado em 20,0 µL de LoTE.

3.5.8 Clonagem

Após o procedimento de desnaturação e renaturação (hibridização) (Item 3.5.7), a clonagem pôde ser feita com a ligação das moléculas de DNA clivadas com a enzima de restrição *Xho* I, em plasmídeo p-ZERO 2 (*Invitrogen*) previamente clivado com a enzima de restrição *Xho* I. A reação de clonagem, que foi mantida a 16° C por 3 horas, ocorreu com 3,0 µL de DNA, 0,5 µg do vetor p-ZERO 2, 1X *Ligase buffer*, 4 U T4 DNA ligase (*Invitrogen*). Somente é

possível ligar no plasmídeo moléculas de DNA fita dupla que foram clivadas com *Xho* I, oriundas portanto da população *tester*.

Um esquema de todo o procedimento de construção das bibliotecas pode ser visto na figura 4. Após o término da reação, novamente seguiu-se com purificação e precipitação (Item 3.5.1) da reação de ligação, que foi ressuspensa em um volume de 6,0 μ L.

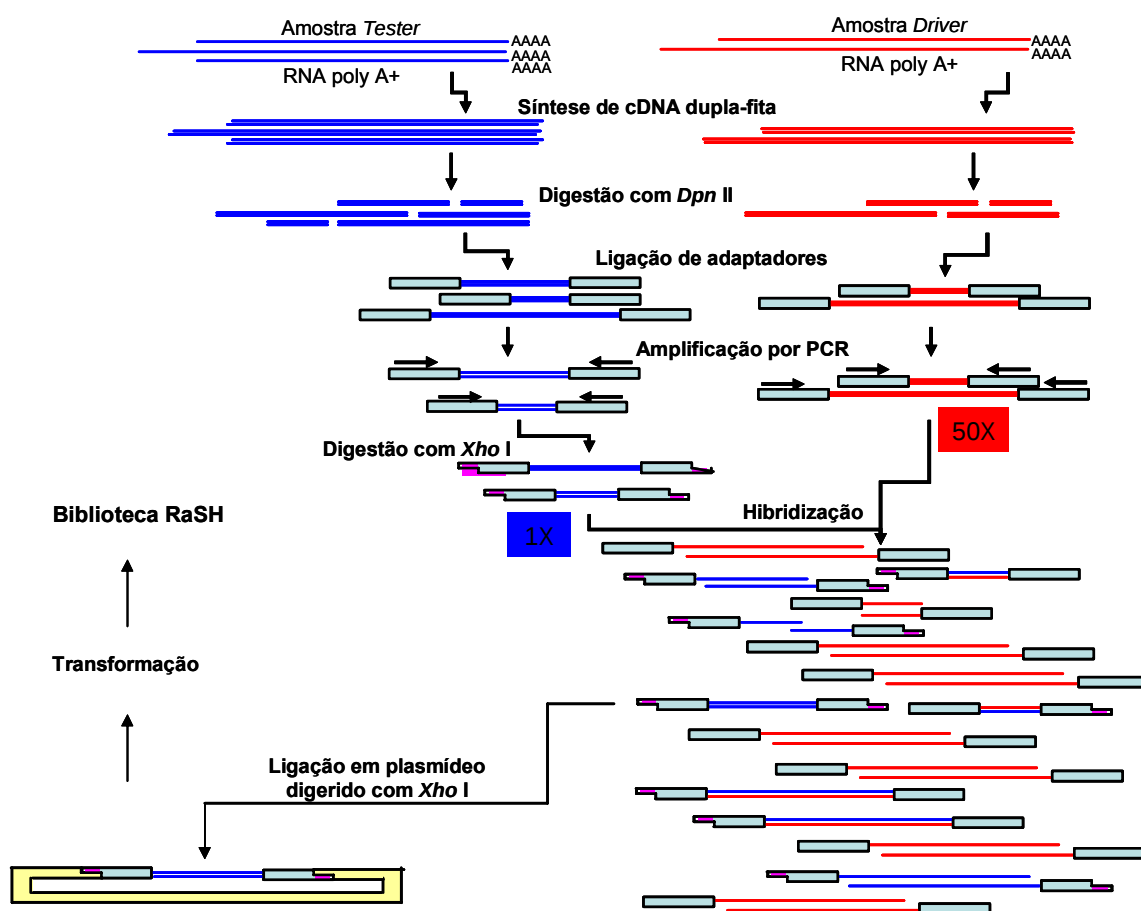


Figura 5 - Ilustração esquemática da metodologia RaSH de construção de biblioteca subtrativa de cDNA.

3.5.9 Transformação de bactérias *Escherichia coli* e sequenciamento dos clones

Após a transformação por eletroporação (2,5 KV, 25 μ FD, 200 OHMS) de bactérias *E. coli* TOP10 (*Invitrogen*) com 1 μ L da reação de ligação de cada uma das bibliotecas, os transformantes foram cultivados a 37° C sob agitação de 200 rpm por 1 hora, e logo em seguida, foram plaqueadas em LB-LowSalt-Ágar contendo o antibiótico zeocina (50 μ g/ μ L) e mantidas por 12 horas a 37° C para crescimento de colônias individuais.

A seleção dos clones recombinantes baseou-se no sistema p-ZERO 2 (*Invitrogen*). O sítio de múltipla clonagem do vetor p-ZERO 2 ao receber um inserto de DNA acaba por interromper a expressão do gene CDB; o qual expressa uma proteína letal para a bactéria. Sendo assim, teoricamente, todas as colônias que crescerem é porque possuem um vetor recombinante. Os clones recombinantes foram confirmados através de reações de PCR de colônia. Os produtos de PCR de colônia foram submetidos a uma eletroforese em gel de agarose 1% corado com brometo de etídeo. Os clones que apresentaram tamanhos maiores do que o esperado para o vetor sem inserto (cerca de 100 pb) foram posteriormente submetidos à reação de sequenciamento em seqüenciador ABI 3100, utilizando-se para isso o kit *BigDye Terminator v3.1 cycle sequencing* (*Applied Biosystems*).

3.6 SUBMISSÃO DAS SEQÜÊNCIAS AO PIPELINE GEAP-BDC

As sequências obtidas foram submetidas à análise *in silico*, utilizando para isso um *pipeline* disponível na página online GEAP (*Generic EST*

Annotation Pipeline) (<http://guarani.fmrp.usp.br/cgi-bin/geap/index.cgi>) criada pelo Dr. Wilson A. Silva Junior, do Departamento de Genética da Universidade de São Paulo - *campus* Ribeirão Preto.

Através do *pipeline* as seqüências foram analisadas em relação à qualidade do sequenciamento cujo critério foi PHRED ≥ 20 em mais de 100 pb. Seqüências cuja qualidade não alcançou o critério estabelecido foram reunidas na categoria *Low-Quality*. As seqüências de boa qualidade foram submetidas à análise para exclusão daquelas que apresentaram seqüências repetitivas e do vetor utilizado na clonagem. Em seguida essas seqüências foram submetidas à outra análise de classificação, através de alinhamento com seqüências depositadas em um banco de dados local. Essa análise categorizou as seqüências quanto à seqüências mitocondriais, bacterianas, leveduras e ribossomais, excluindo-as das classificações posteriores.

Em seguida o restante das seqüências, a partir do alinhamento destas em um banco local de seqüências, foi agrupado de forma hierárquica nas categorias *Reference Sequences* (*RefSeq*) (<http://www.ncbi.nlm.nih.gov/RefSeq>), *Unigene*, *EST*, *Non-redundant* (NR) e por fim *No-match*, quando não foi possível classificá-las em nenhuma das categorias anteriores. Posteriormente estas seqüências foram submetidas a uma triagem manual para identificar falsos positivos, os quais foram excluídos de análises posteriores.

3.7 RT-PCR PARA VERIFICAÇÃO DA EXTENSÃO DO FRAGMENTO

Para verificação da amplificação de fragmentos específicos utilizando os pares de oligonucleotídeos desenhados para cada gene, foram conduzidas reações de RT-PCR. Essas foram realizadas com 1X PCR *buffer*, 1,5 mMol/L MgCl₂; 1 Mol/L Betaína, 0,4 mMol/L *primer forward*, 0,4 mMol/L *primer reverse*, 1 U *Platinum Taq DNA Polymerase (Invitrogen)* e cDNA da linhagem HB4a (O'HARE et al. 1991). O programa apresentou as seguintes etapas: 4 minutos a 94° C, seguidos de 35 ciclos de 45 segundos a 94° C, 45 segundos a 60° C e 1 minuto a 72° C. Por fim, 6 minutos a 72° C. Em seguida, alíquotas de 5,0 µL foram submetidas a uma eletroforese em gel de acrilamida 8%. A amplificação específica foi confirmada através da avaliação do tamanho esperado de cada fragmento.

3.7.1 Genes candidatos selecionados

Os *primers* para os genes candidatos foram desenhados com o uso do software *Primer Express 2.0*, do aparelho *ABI Prism® 7300 Sequence Detection System (Applied Biosystems)* de maneira a alinhar em éxons diferentes, evitando resultados artefatuais resultantes de contaminação genômica. Além disso, foram seguidas as recomendações sugeridas pelo fabricante do aparelho (*Applied Biosystems*), como por exemplo, temperatura de alinhamento dos *primers* próxima a 60° C; cerca de 60% de sua estrutura composta por bases nitrogenadas pirimidinas citosina e guanina; oligonucleotídeo com cerca de 20 bases e o tamanho do fragmento a ser

amplificado (*Amplicon*), sempre que possível, entre 50 a 150 pb. Os *primers* desenhados para os oito genes selecionados e os *amplicons* que estes *primers* amplificam, estão representados na tabela 1.

Tabela 1 - Genes selecionados, as seqüências dos *primers* que amplificam tais genes e o tamanho do fragmento amplificado (*Amplicon*).

Gene	Primer Forward	Primer Reverse	Amplicon
ANKR30A	5' CCTGAAGCCTACAGACATAAAATAACA 3'	5' GATGGGAATCCTGTGCAGCT 3'	80 pb
COPB2	5' GCTATTCCCTGCTGGTTTCAGT 3'	5' CTCAAAACGATGCTCAGGATCTG 3'	188 pb
EIF1	5' CCGCAGGCCGTTTCC 3'	5' CATCACCCCTTACTTGCATCAGC 3'	125 pb
FN1	5' CTCAGAGATACCATCAGAGAAC 3'	5' CTTCTCTCTCAGCCTGTAC 3'	86 pb
KRT19	5' GGTTCATGGCCGAGCAGAAC 3'	5' CCTCCCGGTTCAATTCTTCAC 3'	74 pb
METTL3	5' GGCATGATTGAAAGACTATCTC 3'	5' GAACCGTGCAACCACATC 3'	146 pb
RDBP	5' GATGGAGTCAGCAGATCAGGC 3'	5' CCAGTAGCGGCATCCAGC 3'	111 pb
TMTC3	5' GGTTATTCAATCCCACTCCAGTA 3'	5' CTTGTAGGAGTTGGGCTTACAGC 3'	72 pb

3.7.2 Genes normalizadores

Foram testados 4 genes endógenos que são normalmente utilizados como controles internos, com base na literatura (VANDESOMPELE et al. 2002) como normalizadores das reações de amplificação, sendo eles: ACTB (*beta-actina*), BCR (*breakpoint cluster region*), GAPDH (*glyceraldehyde-3-phosphate dehydrogenase*) e HPRT1 (*hypoxanthine phosphoribosyltransferase-1*). Foram desenhados oligonucleotídeos para amplificação desses genes, em éxons diferentes utilizando o *software Primer Express 2.0*, do aparelho *ABI Prism® 7300 Sequence Detection System (Applied Biosystems)*. As seqüências dos oligonucleotídeos desenhados e o tamanho do fragmento amplificado podem ser observados na Tabela 2.

Tabela 2 - Genes normalizadores com suas seqüências e tamanho do fragmento amplificado (*Amplicon*).

Gene	Primer Forward	Primer Reverse	Amplicon
ACTB	5' GGCACCCAGCACAATGAAG 3'	5' CCGATCCACACGGAGTACTTG 3'	66 pb
BCR	5' CCTTCGACGTCAATAACAAGGA 3'	5' CCTGCGATGGCGTTCAC 3'	67 pb
GAPDH	5' GAAGGTGAAGGTCGGA 3'	5' GGGTCATTGATGGCAAC 3'	102 pb
HPRT1	5' GAACGTCTTGCTCGAGATGTGA 3'	5' TCCAGCAGGTCAGCAAAGAAT 3'	101 pb

Para a escolha do(s) melhor(es) normalizador(es) foi utilizada uma ferramenta disponível *on line* (<http://medgen.ugent.be/~jvdesomp/genorm/>), desenvolvida para o software *Microsoft Excel* e que é conhecida como *geNorm* (VANDESOMPELE et al. 2002). O *geNorm* determina quais são os genes normalizadores mais estáveis em um conjunto de amostras de cDNA e calcula, quando possível, um fator de normalização da expressão gênica para cada amostra com base na média geométrica dos valores obtidos para os diferentes normalizadores selecionados.

A estabilidade dos genes normalizadores é determinada partindo do princípio de que dois genes normalizadores ideais possuem razões de expressão idênticas em todas as amostras de cDNA do estudo. O *geNorm* avalia a razão de expressão de cada gene normalizador em relação aos demais normalizadores sempre aos pares, permitindo o cálculo do desvio-padrão desses valores transformados em escala logarítmica. E a partir desses valores, o *geNorm* calcula um valor M de variação, que é a média do desvio-padrão de cada gene em particular em relação aos demais genes normalizadores. Assim, quanto menor o valor M -o qual indica variação- maior é a estabilidade daquele gene normalizador naquela amostra. A cada normalizador excluído o programa automaticamente atribui um novo valor referente à estabilidade dos demais normalizadores nas amostras de cDNA.

Esse processo ocorre diversas vezes até que permaneçam apenas os genes cujos níveis de expressão sejam os mais estáveis determinando, assim, a combinação adequada de normalizadores a serem utilizados nas análises de RT_PCR em tempo real.

3.8 RT_PCR EM TEMPO REAL

As reações foram realizadas no aparelho *ABI Prism® 7300 Sequence Detection System (Applied Biosystems)* e a metodologia de análise foi a partir de SYBR® Green (*Applied Biosystems*).

Esta metodologia se baseia na incorporação de fluoróforos por parte das duplas fitas de DNA. À medida que ocorre a amplificação, maior é a quantidade de DNA sintetizada na amostra; sendo assim, maior será a incorporação e consequentemente maior será o nível de fluorescência emitida, a qual é captada pelo aparelho.

A padronização da reação foi realizada seguindo recomendações do fabricante (*Applied Biosystems*), as quais incluem a utilização de 3 diferentes concentrações finais de *primer* (200, 400 e 800 nM) combinadas com duas diluições de cDNA (0,1 ng e 0,3 ng – valores referentes à quantidade de RNA total utilizada na reação de transcrição reversa). As reações foram realizadas em duplicata.

Foi incluída no estudo uma amostra calibradora, que correspondeu a um RNA submetido a 1 *round* de amplificação por transcrição *in vitro*, proveniente da linhagem celular HB4a (O'HARE et al. 1991). Esta amostra foi utilizada como uma referência nos cálculos de expressão relativa para os valores de *Cts*

dos genes normalizadores e alvos, permitindo que estes dados fossem submetidos à análise do *software geNorm* para a seleção do melhor gene normalizador.

Somente foram aceitos valores de desvio-padrão menores do que 0,4 entre as duplicatas. As concentrações dos iniciadores foram padronizadas considerando dois fatores: a utilização das menores concentrações de *primers* a fim de evitar a formação de dímeros, e a obtenção dos menores valores de C_T (*Threshold Cycle* – ciclo no qual é detectada a fluorescência) possíveis. As condições de RT_PCR seguiram as condições de ciclagem sugeridas pelo fornecedor. Ao final da amplificação, adicionou-se uma etapa de 20 minutos de duração, na qual a temperatura aumentou gradualmente de 60° C para 95° C a 0,2° C por segundo, com a contínua aquisição da fluorescência (MORRISON et al. 1998). Esta etapa possibilita uma análise da curva de dissociação, utilizada para confirmação da especificidade da reação de amplificação. As reações foram realizadas em placas 96 wells - *MicroAmp® Optical* (Applied Biosystems).

3.8.1 Eficiência dos *primers*

Para o cálculo da eficiência de amplificação dos fragmentos, foi realizado, para cada par de *primers*, uma reação de amplificação utilizando cinco diluições seriadas de cDNA, em duplicata (100; 10; 1; 0,1 e 0,01 ng), utilizando como molde um RNA amplificado por 1 *round* de transcrição *in vitro* proveniente da linhagem celular HB4a (O'HARE et al. 1991). A partir dos resultados da reação foi construída uma curva-padrão, a qual relaciona os C_t s com o valor das quantidades de cDNA, numa escala logarítmica de base 10.

O valor correspondente ao coeficiente angular (a) da equação da reta da curva padrão ($y = ax + b$), foi utilizado para identificar a eficiência de cada par de *primers*, a partir da seguinte fórmula matemática: **Eficiência:** $10^{(-1/a)} - 1$.

3.8.2 Quantificação Relativa

O cálculo da quantificação relativa de expressão de cada gene nas amostras tumorais DCIS e IDC foi realizado com o uso do modelo matemático proposto por PFAFFL (2001). Este modelo considera a eficiência de amplificação dos oligonucleotídeos no cálculo da expressão relativa, permitindo a quantificação relativa da expressão quando tais eficiências são diferentes entre si. Assim, os dados foram normalizados pela média dos *Cts* do gene normalizador, que foi identificado pelo *software geNorm*, como sendo o gene mais estável entre as 4 amostras testadas. Dessa forma, a partir da fórmula abaixo, calculou-se a diferença de expressão relativa entre IDC e DCIS.

$$\text{Razão} = \frac{(1 + \text{Eficiência gene alvo})^{(Ct \text{ amostra calibradora} - Ct \text{ amostra IDC})}}{(1 + \text{Eficiência gene normalizador})^{(Ct \text{ amostra calibradora} - Ct \text{ amostra IDC})}}$$

3.9 cDNA MICROARRAY

3.9.1 Construção da plataforma

Foi construída uma plataforma de cDNA com as seqüências específicas de cada biblioteca RASH, as quais foram selecionadas de acordo com a presença específica em cada um dos componentes tumorais DCIS e IDC. Desta forma, seqüências que estavam presentes em ambos os componentes

de uma mesma amostra, ou mesmo em amostras diferentes, não fizeram parte desta seleção.

Os elementos que compõem a lâmina passaram inicialmente por um *pipeline*, desenvolvido pela Dra. Helena Brentani, que visa controlar hibridização cruzada. A extensão dos fragmentos ocorreu através de uma amplificação por PCR de cultura, contendo 1X PCR *Buffer*, 1,5 mMol/L MgCl₂, 0,2 mMol/L dNTP, 0,2 μMol/L de cada *primer* M13 *forward* – 5' GTA AAA CGA CGG CCA G 3' e *reverse* – 5' CAG GAA ACA GCT ATG AC 3' e 1 U Taq DNA polimerase (*Phoneutria*). O programa da reação seguiu as seguintes etapas: 4 minutos a 95° C, seguidos de 35 ciclos de 45 segundos a 95° C, 1 minuto a 60° C e 4 minutos a 72° C. A extensão final foi de 7 minutos a 72° C. Os fragmentos foram submetidos a uma eletroforese em gel de agarose 1%, corado com brometo de etídeo, para verificação de fragmento único. Os fragmentos de banda única foram, então, purificados utilizando uma placa de filtro *Sephadex G50 Fine* com 96 *weels* (*Millipore*), por centrifugação a 900 rpm a 4° C por 5 minutos. Posteriormente, esses fragmentos foram precipitados com dois volumes de etanol absoluto adicionado de 0,3 M de acetato de sódio, valor de pH 6,0 e armazenados a temperatura de – 20° C por 24 horas. Em seguida, foram submetidos à centrifugação a 4° C por 35 minutos, a 4000 rpm. Os precipitados foram lavados com etanol 70% e centrifugados novamente a 4000 rpm por 20 minutos. Após a secagem, foram ressuspensos em 60,0 μL de TE pH 7,5.

Aproximadamente 100 picolitros de cada fragmento (em média 15×10^{-3} ng) foram arranjados em lâmina de vidro *Corning Gaps II* (*Corning*) sem replicata, com o auxílio do robô *Flexys* (*Genome Solutions, UK*). As etapas de

imobilização dos *amplicons* na lâmina de vidro foram coordenadas pelo Dr. Alex Fiorini de Carvalho do grupo do Dr. Luiz Fernando Lima Reis (LICR). Os fragmentos foram imobilizados em grupos de 48 *subarrays*, dispostos em 4 colunas com 12 fileiras cada um, e cada *subarray* contendo 49 *spots*, dispostos em 7 colunas com 7 fileiras cada um. Em cada *subarray* foram imobilizados 3 controles negativos (ausência de material), 5 *blank spots* (DMSO) e 1 controle positivo de marcação (cDNA correspondente a um fragmento do *lambda fago Q gene*).

3.9.2 Marcação e hibridização

As amostras de RNA amplificado referentes à população de células de carcinoma ductal *in situ* (DCIS) e invasivo (IDC) de amostras tumorais que foram utilizadas nas reações de marcação e hibridização são provenientes do projeto de doutorado da aluna Nádia Castro, do nosso grupo de pesquisa. Este grupo de amostras constou de 6 casos pareados de carcinoma ductal *in situ* e invasivo de mama (tabela 3).

Tabela 3 - Casos pareados de carcinoma ductal *in situ* e invasivo de mama utilizadas na hibridização com os transcritos específicos de DCIS e IDC identificados por RaSH.

Características histológicas e imunohistoquímicas	Caso 2	Caso 3	Caso 13	Caso 25	Caso 75	Caso 81
Estadio clínico	IIA	IIA	IIA	IIA	IIB	IIB
Estadiamento	T0N0M0	T1N0M0	T2N0M0	T2N0M0	T2N0M0	T2N1M0
Receptor de estrógeno	+	+	+	+	-	-
Receptor de progesterona	+	-	+	+	-	-
Grau nuclear	3	3	2	3	2	3
Grau SBR	II	II	II	II	II	II
p53	+	-	+	+	+	+
c-erb-B2	+	+	+	+	+	+

Todas as amostras de DCIS e IDC foram hibridizadas com um RNA referência, que consistiu de um aRNA proveniente da linhagem celular HB4a (O'HARE et al. 1991). O uso da referência produz uma medida relativa de expressão para cada gene, possibilitando comparações de experimentos independentes. O sistema usado foi o *dye-swap*, o qual consiste na hibridização de uma amostra duas vezes, usando a inversão de corantes. Essa inversão dos fluoróforos é realizada para minimizar erros decorrentes das propriedades específicas de cada fluoróforo cianina (Cy3 e Cy5) que apresentam eficiências de incorporação distintas na síntese de primeira fita do cDNA a partir do RNA. As sondas de cDNA fita simples foram obtidas pela reação de transcrição reversa, utilizando aRNA (submetido a 2 *rounds* de amplificação) como molde.

Inicialmente 4 µg de aRNA das amostras DCIS e IDC, e 3,5 µg de aRNA referência foram misturados com 9 µg de oligonucleotídeo *random primer* dN6 e 1,0 µL de RNA mix controle (mistura de DNA de diferentes cepas do vírus *B. subtilis*); o volume da reação foi completado com água DEPC para 15,9 µL. Após uma incubação a 70° C por 10 minutos, foram adicionados 1X *SuperScript* II *buffer*, 10 mMol/L DTT, 10 mM dCTP, 25 mM demais desoxirribonucleotídeos (dATP, dTTP e dGTP), 20 U RNase IN (*Promega*), 400 U *SuperScript* II (*Invitrogen*) e 2,0 µL de cianina. A reação foi mantida por 2 horas a 42° C.

Posteriormente, a reação foi inativada com a adição de 25 mMol/L EDTA e 50 mN NaOH, mantendo-se a 70° C por 20 minutos. Em seguida, procedeu-se com a adição de 50 mN HCl para neutralização.

Em seguida, a reação foi purificada em coluna Autoseq G50 (*Amersham Biosciences*), e então submetida à reação de hibridização.

A reação de hibridização contou com a mistura das duas reações de cDNAs Cy3 e Cy5 purificadas. Em seguida adicionou-se 4 µg poly-A DNA (*Amersham Biosciences*), 4 µg Cot1 DNA (*Invitrogen*), 20 µg BSA. Então, o volume foi diminuído em *speed-vac* e houve adição de 50,0 µL de tampão de hibridização previamente aquecido a 70° C, 25,0 µL de formamida deionizada, 9,5 µL 50X solução *Denhardt's* e 10 µg de esperma de salmão. Esta reação foi mantida sob agitação a 45° C enquanto se procedia com a lavagem das lâminas com água deionizada, as quais tinham sido pré-hibridizadas por mais de 6 horas a 42° C em tampão contendo 5X SSC; 0,2% SDS; 1% BSA e solução 5X *Denhardt's*.

As amostras marcadas com fluoróforos distintos, amostra-teste e referência, são misturadas e hibridizadas numa mesma lâmina. Após a preparação das lâminas, o cDNA marcado foi aquecido a 95° C por 5 minutos e, em seguida, adicionado à câmara de hibridização; a qual foi mantida sob agitação a 42° C de 16 a 24 horas.

Posteriormente à hibridização, as lâminas foram lavadas com tampões pré-aquecidos a 42° C, sendo uma vez com tampão 2X SSC por 15 minutos, duas vezes 0,1X SSC; 0,1% SDS por 15 minutos, e por fim, duas vezes 0,1 X SSC por 15 minutos. Em seguida, as lâminas foram submetidas à digitalização de imagens.

3.9.3 Captura das imagens e quantificação do sinal

Os sinais gerados pela hibridização das amostras foram capturados pelo *ScanArray Express* (Perkin Elmer Life Sciences, Inc., EUA), um *scanner a laser* confocal com canal duplo. Esse sistema tem dois *lasers* com comprimento de ondas diferentes, que estimulam especificamente o Cy3 ou o Cy5. A leitura dos fluoróforos é feita separadamente e a composição da imagem é realizada pelo programa. Quando as moléculas são excitadas, emitem um sinal proporcional ao número de moléculas presentes, que é captado pelo *scanner*, sendo essa medida proporcional à quantidade de RNA mensageiro nas amostras (Tabela 4). A quantificação desse sinal ocorre devido à conversão do comprimento de onda emitido pelos fluoróforos excitados em energia elétrica, produzindo um sinal mensurável. Como a hibridização é competitiva, os valores de fluorescência obtidos revelam níveis relativos de expressão de cada transcrito na amostra-teste e na amostra-referência.

Tabela 4 - Características específicas de cada fluoróforo utilizado.

Fluoróforo	Excitação (nm)	Emissão (nm)	Coloração
Cy3	550	570	verde
Cy5	649	670	vermelho

3.9.4 Pré-análise das lâminas hibridizadas

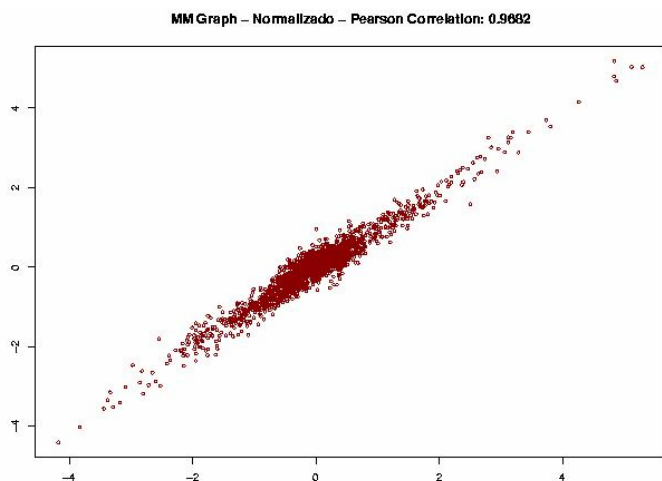
Após a captura de sinal pelo *scanner*, os resultados foram submetidos a análise de qualidade, denominadas de pré-análise, as quais estão apresentadas a seguir.

A Dendrograma

Refere-se a uma clusterização hierárquica com distância euclidiana não supervisionada. Essa clusterização leva em consideração os dados de expressão de todos os genes da lâmina. Essa análise funciona como um controle da parte experimental, permitindo a visualização das similaridades das amostras. Assim, as lâminas controle e *swap*, uma vez que são réplicas com inversão de corantes, normalmente devem apresentar maior semelhança no dendrograma quando comparada à outra amostra do projeto.

B MM Plot

O gráfico mostra a correlação entre a lâmina controle e a lâmina *swap* através de uma correlação de *Pearson*, levando em consideração a eficiência de incorporação dos fluoróforos (Cy3 e Cy5). A correlação de *Pearson* determina uma medida do grau de correlação linear entre duas variáveis aleatórias. Para uma lâmina ser considerada de boa qualidade essa correlação tem que ser superior a 80% após a normalização. Isso é estabelecido para amostras cujas referências não apresentam grande similaridade no padrão geral de expressão com as amostras testes, que é justamente o caso desse estudo em que a referência utilizada é uma amostra da linhagem celular HB4a. Na figura 6 está apresentado o *MM plot* normalizado de uma das amostras utilizadas nesse estudo.

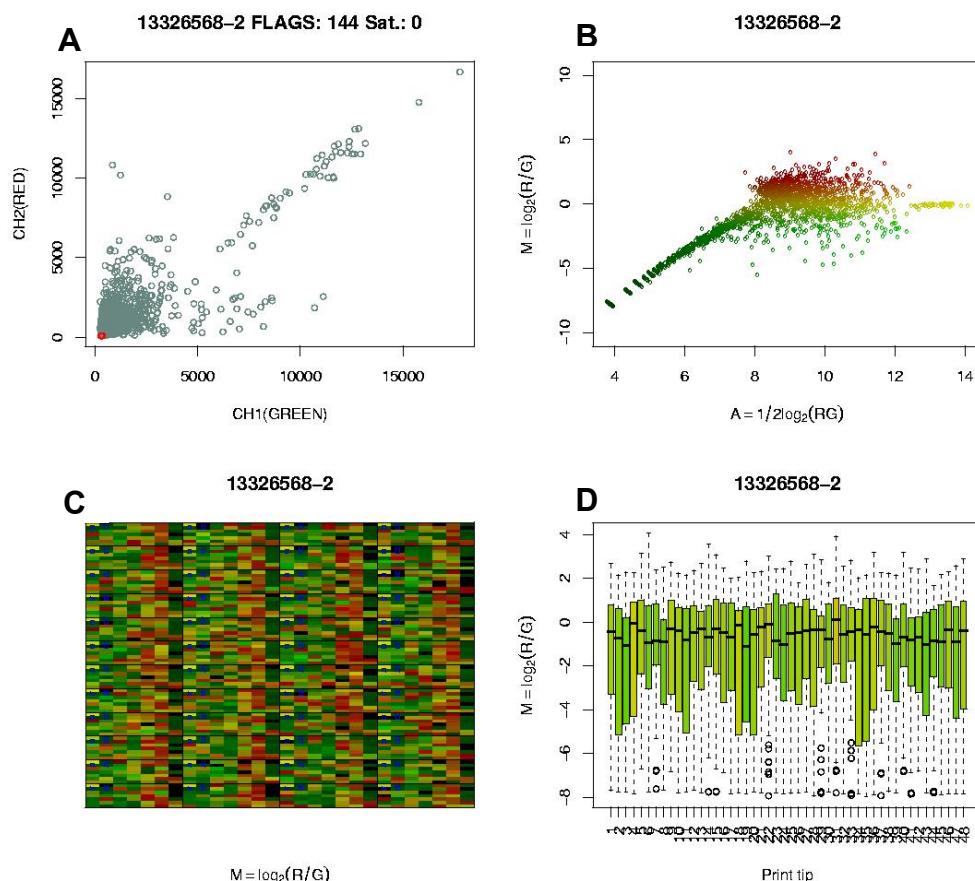


Legenda: Eixo x - lâmina controle: amostra teste-Cy3/amostra referência-Cy5; eixo y - lâmina *swap*: amostra referência-Cy3/amostra teste-Cy5. O gráfico mostra a dispersão de pontos representantes da relação entre os valores de intensidade para cada gene das lâminas controle e *swap*.

Figura 6 - *MM plot* da amostra 13-DCIS.

C Ma, Log X Log e Boxplot

Para esse controle de qualidade são medidos os valores de intensidade de fluorescência para cada gene, após a subtração de sinal do *background* usando escala *log* base 2. Um *plot* cartesiano é construído com *M* na coordenada e *A* na abcissa, onde $M = \log_2 (R/G)$ e $A = \log_2 [(\sqrt{R \times G})]$, ou seja, pode-se observar a variação das razões de fluoróforos dependente da intensidade dos pontos. Os *Box plots* consistem de caixas com uma linha central (mediana dos dados) e duas caudas, a superior representando o percentil 75 e a inferior representando o percentil 25 (ALLISON et al. 2006). Os gráficos mostram as intensidades de Cy3 e Cy5 da mesma lâmina (figura 7), sem nenhum tipo de normalização.



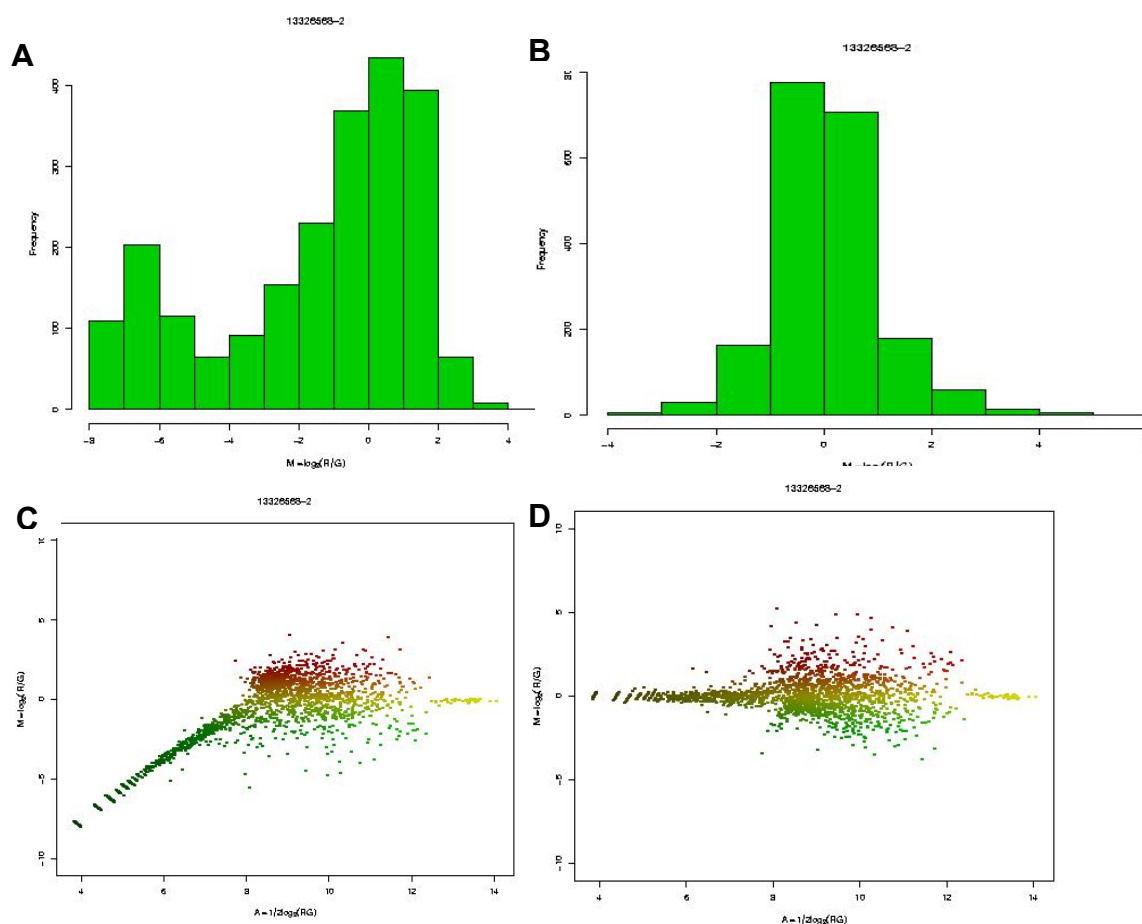
Legenda - A: Relação da intensidade bruta da lâmina (Cy3/Cy5). *Flags*: Número de pontos retirados por apresentarem sinal de intensidade menor do que o sinal de *background*. *Sat*: Número de pontos retirados por apresentarem sinal maior do que o aceito (63000 BTUs). B: *MA plot*: gráfico representando a razão (Cy3/Cy5) no eixo y e a média das intensidades dos corantes em escala logarítmica no eixo x. C: Gráfico representativo da lâmina. Pontos em azul representam genes que foram excluídos pelo programa *Scannarray* ou manualmente. D: *BoxPlot*: Representa a dispersão de pontos de cada *subarray* (1 *subarray* contém 49 *spots*). O bloco verde representa a concentração de 75% dos pontos, segundo a razão de intensidade.

Figura 7 - Representação da fluorescência de todos os genes de uma lâmina.

D Normalização das Lâminas

A proposta de normalização é minimizar variações sistemáticas introduzidas experimentalmente nas medidas da expressão gênica, para que as diferenças biológicas quantificadas possam ser distinguidas com menos artefatos. A normalização por *Lowess* (*local weight regression*) foi utilizada neste projeto. Para cálculo do *Lowess* definem-se janelas com intervalos pré-definidos, que variam de 0 a 100%. Nesse projeto, foi utilizado intervalo de

20%, ou seja, a análise de regressão foi realizada com 20% dos pontos ao redor do ponto em questão (Figura 8).



Legenda: A e C: Representação dos dados antes de serem normalizados. A: histograma que mostra a distribuição da medida relativa dos dados de uma amostra-teste e sua referência. C: dispersão dos pontos referentes a medida relativa dos dados amostra-teste e sua referência. B e D: Representação dos dados normalizados.

Figura 8 - Gráficos representativos da normalização por Lowess.

3.9.5 Análise estatística

O valor de intensidade do *background* local foi subtraído do valor de intensidade de expressão do *foreground* obtido para cada *spot*, sendo considerados apenas os pontos que apresentaram valor positivo. Após essa etapa, todos os valores foram trabalhados na escala logarítmica de base 2. A normalização dos dados foi realizada através do método de Lowess.

Para identificação dos genes diferencialmente expressos nos componentes IDC e DCIS, as médias dos valores de intensidade entre as réplicas de cada experimento de marcação foram submetidas ao teste *t de Student* pareado. O critério de seleção dos genes diferencialmente expressos foi p-valor menor do que 0,05 para esta análise. Para identificação dos genes diferencialmente expressos em cada par de amostras (IDC e DCIS), foi estabelecido como *cut-off* uma diferença de expressão de 1,5 para mais ou menos expresso, utilizando a medida de intensidade média dos experimentos controle e *swap*. As análises de qualidade das lâminas e análises matemáticas para identificação dos genes diferencialmente expressos foram realizadas pela Dra. Helena P. Brentani do Laboratório de Bioinformática do Hospital do Câncer A. C. Camargo.

4 RESULTADOS

4.1 SELEÇÃO DOS CASOS

Cinco casos de carcinoma ductal de mama foram pré-selecionados a partir do Banco de Tumores do Centro de Tratamento e Pesquisa - Hospital A. C. Camargo, como candidatos a material de estudo. A pré-seleção foi baseada nos seguintes critérios: 1) presença de ambos os componentes DCIS e IDC, 2) pacientes com menos de 50 anos de idade, 3) amostra retirada cirurgicamente antes de tratamento hormonal ou quimioterápico e 4) presença de linfonodos acometidos por células neoplásicas. A inclusão do quarto critério foi baseada na pressuposição de que a presença de metástase regional garante que esteja ativada a programação genética para as etapas do processo metastático (invasão tumoral, intravasão e extravasão), tornando mais provável a identificação de genes envolvidos com o processo metastático completo. Os cinco casos previamente selecionados estão apresentados na Tabela 5.

Tabela 5 - Casos previamente selecionados como material de estudo.

Casos	Estadio clínico	Estadiamento (TNM)	Nº linfonodos acometidos/ Nº linfonodos examinados
1	IIIB	pT _{4b} N ₁ M ₀	10/34
2	I	pT _{1c} N ₁ M ₀	1/14
3	IIB	pT ₂ N ₁ M ₀	2/31
4	IIIB	pT ₄ N ₂ M ₀	20/23
5	IIB	pT ₂ N ₁ M ₀	2/35

A partir dos casos da Tabela 5, foram gerados cortes histológicos para possibilitar a coleta mínima de 2 *caps* com células capturadas de cada componente tumoral.

Todos os casos foram submetidos à microdissecção a laser com a finalidade de se obter uma população homogênea de células tanto de DCIS quanto de IDC de um mesmo caso. Foram capturadas cerca de 15 mil células de cada componente tumoral DCIS e IDC, sendo posteriormente submetidas à extração de RNA total. Todas elas foram submetidas ao procedimento de 1 *round* de amplificação de mRNA, seguindo o protocolo baseado em GOMES et al. (2003), com adaptações para quantidades da ordem de poucos nanogramas de RNA total (menos do que 50 ng) (SARAIVA 2005). No entanto, para os casos 3, 4 e 5 houve perda de material em diferentes etapas do procedimento de amplificação de mRNA. Os casos 3 e 4 apresentaram uma quantidade insuficiente de RNA amplificado para a continuação do processo; e o caso 5 apresentou problemas na etapa de checagem de clivagem do ds-cDNA com a enzima de restrição *Dpn* II, resultando em quantidades insuficientes. Dessa forma somente os casos 1 e 2 apresentaram quantidade suficiente de material para serem utilizadas em etapas posteriores, além de apresentarem integridade do mRNA. Sendo assim, estas foram utilizadas na adaptação do protocolo de construção de bibliotecas subtrativas RaSH para aRNA.

As características clínicas das pacientes cujas amostras tumorais foram utilizadas para o estudo, e as características anatomopatológicas e imunohistoquímicas destas amostras estão apresentadas na tabela 6.

Tabela 6 - Características clínicas das pacientes, e, anatomopatológicas e imunohistoquímicas das amostras selecionadas para estudo.

Características	Caso 1	Caso 2
Idade	45 anos	44 anos
Mama acometida	Esquerda	Direita
Menarca	14 anos	13 anos
Pré-Menopausa	Sim	Sim
Gestações	Ausente	1ª gestação com 24 anos
Câncer Familiar	Histórico negativo	Histórico negativo
Invasão Sanguínea e Linfática	Presentes	Presentes
Receptor de Estrógeno	Positivo (40% das células)	Positivo (80% das células)
Receptor de Progesterona	Positivo (30% das células)	Positivo (40% das células)
P53	Positivo	Negativo
C-erb-B2	Negativo (+1)	Negativo (+1)
Índice Mitótico	15 para 10 campos	5 para 10 campos
Grau Nuclear	3	3
Grau SBR	II	II

4.2 ADAPTAÇÃO DO PROTOCOLO RASH PARA RNA AMPLIFICADO

A técnica de LCM resulta em pouca quantidade de RNA total. Estima-se uma obtenção de cerca de 30 ng de mRNA para cada 100 mil células microdissecadas (FULLER et al. 2003). Isso gera a necessidade de realizar uma amplificação do RNA mensageiro (mRNA) a fim de aumentar a quantidade de cDNA a ser utilizada posteriormente na etapa de construção das bibliotecas subtrativas RaSH. Para isso, foi necessário implementar pequenas alterações na metodologia de amplificação de mRNA. Para tanto, foram inseridos sítios de reconhecimento da enzima de restrição *Dpn II* nos oligonucleotídeos dT(24)-T7 e *template switch* (TS), os quais são utilizados na metodologia de amplificação do mRNA.

Este procedimento foi adotado com o intuito de aumentar a possibilidade de representar um maior número de transcritos amplificados, uma vez que só poderão ser identificados na biblioteca os ds-cDNAs que possuírem 2 sítios de reconhecimento da enzima de restrição *Dpn II*.

As amostras foram submetidas a 1 *round* de amplificação por transcrição *in vitro* e o ds-cDNA proveniente do RNA amplificado foi sintetizado.

4.2.1 Amplificação de mRNA

Assumindo que a estimativa de FULLER et al. (2003) esteja correta, de que a cada 100 mil células microdissecadas a laser é possível obter cerca de 30 ng de mRNA, e considerando que foram cerca de 22 mil células capturadas (15 mil disparos – 1,5 células/disparo), foi possível obter cerca de 7 ng de mRNA de cada amostra. E a partir de 1 *round* de amplificação por transcrição *in vitro* obtivemos no mínimo 470 ng de aRNA de cada amostra, representando um aumento no número de moléculas de 67 vezes (tabela 7).

Tabela 7 - Quantificação de aRNA após 1 *round* de amplificação por transcrição *in vitro*.

Caso	Diluição (1:100)	230	260	280	R (260/ 280)	Concentração (ng/ µL)	Quantidade (µg)
Caso número 1	DCIS	0,281	0,018	0,012	1,5	72	0,50
	IDC	0,027	0,029	0,021	1,38	116	0,82
Caso número 2	DCIS	0,017	0,017	0,007	2,4	68	0,47
	IDC	0,379	0,024	0,009	2,6	96	0,67

A integridade dos cDNAs a partir do RNA total foi analisada por RT-PCR a partir de amplificação de um fragmento correspondente ao gene GAPDH, o qual possui um tamanho de 296 pb (GAPDHF 5' CTGCACCACCAAGTGCTTA 3' e 5' CATGACGGCAGGTCAGGTC 3'). A síntese de primeira fita de cDNA a

partir de aRNA foi realizada seguindo duas metodologias diferentes. Ambas as metodologias de amplificação que foram utilizadas apresentam vantagens e desvantagens, uma em relação à outra.

A metodologia que utiliza o oligonucleotídeo TS se baseia na atividade da enzima *SuperscriptTM II Reverse Transcriptase (Invitrogen)*, uma transcriptase reversa do tipo *Moloney Murine Leukemia Virus reverse transcriptase (M-MLV RT)*, que é capaz de adicionar alguns nucleotídeos (a maior parte deles são nucleotídeos citosina) à extremidade 3' das moléculas de cDNA recém sintetizadas, na medida em que alcança a extremidade 5' da molécula molde de mRNA. Desta forma, o oligonucleotídeo TS, o qual possui repetições de nucleotídeos guanina em sua extremidade 3', é capaz de se alinhar às repetições de citosinas e permanecer anexado à extremidade 3' do cDNA. Com isso, essa estratégia promove uma troca de fita molde (*template switching effect*) sendo que o oligonucleotídeo possibilita a síntese da segunda fita de cDNA. Assim, este efeito serve como base para a obtenção de amostras de cDNAs amplificados enriquecidos com sequências inteiras de transcritos (*full-length sequences*) (Matz et al. 1999). Por ser capaz de gerar transcritos inteiros, e portanto mais longos, essa estratégia aumenta a probabilidade dos transcritos apresentarem 2 sítios *Dpn II*, permitindo sua clonagem e identificação nas bibliotecas RASH. Embora uma desvantagem seja o fato de que não são todas as moléculas de cDNA que recebem uma sequência de citosinas na extremidade 3' pelo fato da transcriptase reversa não ser capaz de estender o início de todos os transcritos, principalmente os mais longos. Se tais transcritos não apresentarem essa cauda de citosinas, não podem ser representados, pois desta forma não há uma sequência complementar no RNA

antisense em que o *primer* TS possa se alinhar para sintetizar a primeira fita de cDNA.

Já a metodologia que utiliza o oligonucleotídeo dN6 possibilita a síntese de uma maior quantidade de moléculas de cDNA. Como é um *primer* composto por 6 nucleotídeos, e cuja sequência é randômica, pode facilmente se alinhar em vários trechos dos transcritos amplificados. Sendo assim esta metodologia é capaz de, teoricamente, representar todos os transcritos amplificados de uma amostra. No entanto, tais transcritos apresentam um tamanho relativamente pequeno, quando comparado com transcritos que são amplificados pela metodologia que utiliza o oligonucleotídeo TS. Isso ocorre por que o trecho de cDNA representado pelo *random primer* é correspondente ao trecho mais próximo à extremidade 3'. E por serem pequenos, os transcritos selecionados por dN6 possuem uma menor probabilidade de apresentarem 2 sítios *Dpn* II, o que impossibilitaria sua clonagem, e conseqüentemente, sua identificação nas bibliotecas RaSH.

Perante as vantagens e desvantagens das duas metodologias e pelo fato de que não há relatos na literatura a respeito da construção de bibliotecas subtrativas baseadas em amostras amplificadas que utilizam oligonucleotídeos dN6 e/ou TS, as duas metodologias foram desenvolvidas com o intuito de comparar qual delas seria capaz de resultar numa melhor adaptação do protocolo de construção de bibliotecas subtrativas – RaSH.

Para as amostras DCIS e IDC do caso 1 (Tabela 5) a síntese de cDNA primeira fita a partir do RNA amplificado por 1 *round* de transcrição *in vitro*, foi realizada com o uso do oligonucleotídeo *random primer* (dN6) (Figura 3)

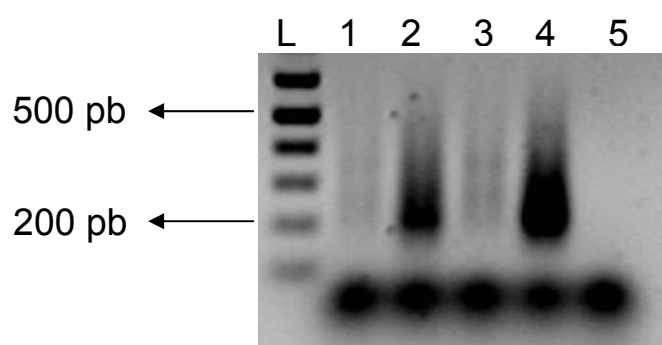
enquanto que para o caso 2 (Tabela 5) foi utilizado o oligonucleotídeo TS-DpnII (Figura 4).

4.2.2 Bibliotecas RaSH construídas

Os fragmentos gerados com a clivagem dos cDNAs pela enzima de restrição *Dpn* II foram ligados aos adaptadores XDPN12 e XDPN14. Estes adaptadores serviram de ancoragem para o oligonucleotídeo XDPN18 amplificar estes fragmentos por PCR. A ligação dos adaptadores às moléculas de cDNA de ambos os componentes DCIS e IDC foi obtida com sucesso para as amostras de ambos os casos tumorais. Este resultado pode ser observado na figura 9, onde as amostras DCIS e IDC que não receberam ligase (controles da reação) não tiveram a ligação dos adaptadores, não havendo amplificação dos cDNAs; enquanto que as amostras que receberam ligase, tiveram amplificação dos cDNAs. Tal amplificação possibilitou a obtenção de quantidades suficientes de fragmentos (Tabela 8), para uma reação de hibridização subtrativa adequada, com posterior clonagem. Aproximadamente 500 clones foram submetidos à reação de PCR de colônia para posterior sequenciamento dos fragmentos obtidos. A partir de cada amostra tumoral foram geradas duas bibliotecas subtrativas (*IDC-tester/DCIS-driver* e *DCIS-tester/IDC-driver*), totalizando 4 bibliotecas. Estas foram nomeadas da seguinte forma: RHID01, RHSD01, RHID02 e RHSD02; em que “RH”, “ID ou SD” e “01 ou 02” significam “RaSH”, “IDC ou DCIS” e “caso 1 ou 2”, respectivamente.

Tabela 8 - Quantificação dos fragmentos de DNA amplificados por PCR.

Caso	Diluição (1:250)	260	280	R	C (ng/ μ L)	Quantidade aproximada (μ g)
1	DCIS	0,087	0,060	1,46	1098	100
	IDC	0,084	0,055	1,50	1053	100
2	DCIS	0,099	0,057	1,74	495	500
	IDC	0,119	0,070	1,70	595	600



Legenda: Gel de agarose 1% corado com brometo de etídeo. L-Ladder 1 kb plus (Invitrogen);, 1-DCIS sem ligase; 2-DCIS com ligase; 3-IDC sem ligase; 4-IDC com ligase; 5-controle negativo (NO).

Figura 9 - Eficiência da ligação dos adaptadores XDPN12 e XDPN14 às moléculas de cDNA dos componentes DCIS e IDC do caso 1.

4.3 CLASSIFICAÇÃO DAS SEQUÊNCIAS GERADAS

Após o sequenciamento de aproximadamente 500 clones de cada biblioteca RaSH, obteve-se 2062 sequências, as quais foram submetidas à análise pelo *pipeline* GEAP-BDC e visualizada por uma página disponibilizada pelo grupo do Dr. Wilson A. Silva Junior (Figura 10).

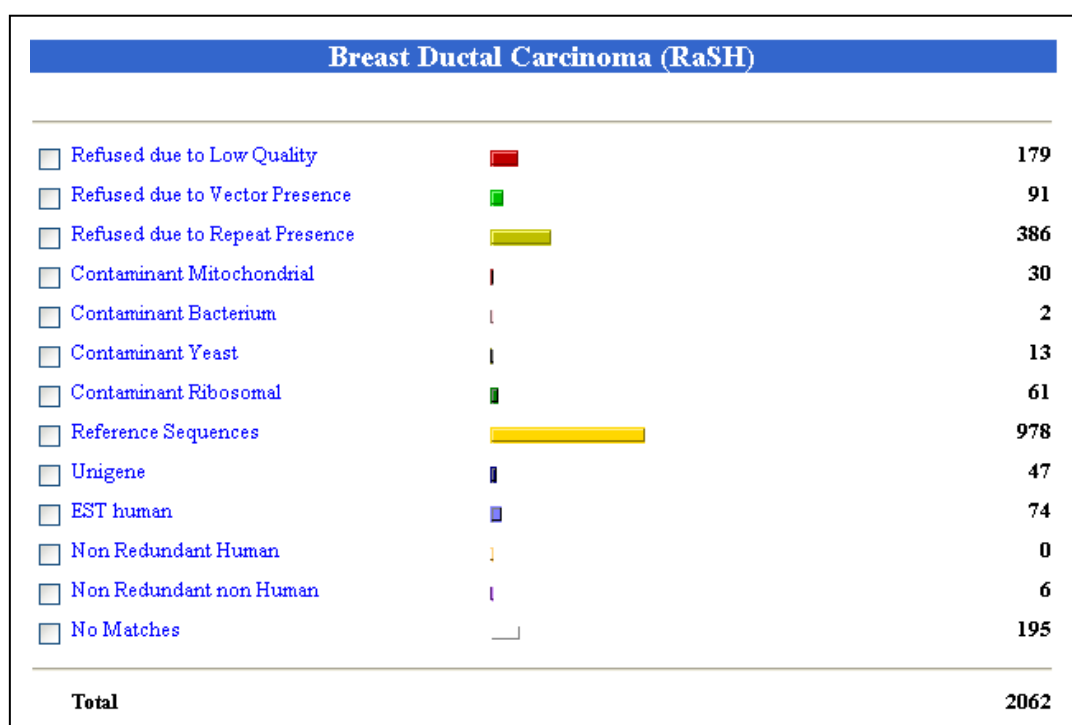


Figura 10 - Resultado da análise do *pipeline* GEAP-BDC das sequências submetidas das 4 bibliotecas.

As 2062 seqüências foram analisadas através do *pipeline* em relação à qualidade do sequenciamento, cujo critério foi PHRED ≥ 20 em mais de 100 pb. O *pipeline* classificou hierarquicamente as sequências da seguinte maneira: seqüências cuja qualidade não alcançou o critério estabelecido foram reunidas na categoria *Low-Quality* e retiradas das análises subsequentes. Esta categoria correspondeu a 179 seqüências (cerca de 8,7% de todas as seqüências submetidas). Do total de seqüências de boa qualidade foram excluídas de análises posteriores aquelas correspondentes a seqüências do vetor utilizado na clonagem e seqüências repetitivas, o que correspondeu a 91 (4,4%) e 386 (18,7%) seqüências, respectivamente. Em seguida, essas 1406 seqüências restantes de boa qualidade foram submetidas a uma análise de classificação, através do alinhamento com seqüências depositadas em um banco de dados

local. Essa análise categorizou as seqüências quanto a presença de seqüências mitocondriais (30 seqüências - 1,4%), bacterianas (2 seqüências - 0,096%), leveduras (13 seqüências - 0,63%) e ribossomais (61 seqüências - 2,9%), excluindo-as das classificações posteriores, conforme está apresentado na Figura 9.

Em seguida, as 1300 seqüências que não se agruparam nas categorias anteriores foram alinhadas com um banco local de seqüências, sendo agrupadas nas categorias *RefSeq* (<http://www.ncbi.nlm.nih.gov/RefSeq>), o que correspondeu a 978 seqüências (47,4%); *Unigene*, correspondendo a 47 seqüências (2,2%); *EST*, com 74 seqüências (3,5%); *Non-redundant Human* (*NR*), com 6 seqüências (0,2%) e, por fim, *No-match*, com 195 seqüências (9,4%), quando não foi possível classificá-las em nenhuma das categorias anteriores.

As 195 seqüências correspondentes à categoria *No-match* foram submetidas à inspeção visual através da ferramenta *BLASTn* (<http://www.ncbi.nlm.nih.gov/BLAST>), para alinhamento com bancos de dados *non-redundant* (*nr*) e *expressed sequence tag* (*EST*) disponíveis. O resultado destas análises mostrou que 23 delas apresentaram alinhamento com seqüências conhecidas. Destas, 11 alinharam com mRNAs e 12 com *ESTs*, embora este alinhamento tenha sido em pequenos trechos dos transcritos, em relação ao tamanho dos fragmentos. O restante dos *No-match* (172 seqüências) apresentou trechos com repetições de nucleotídeos, principalmente repetições de adeninas e timinas. Sendo assim, como nenhum gene novo foi encontrado nessa categoria de seqüências, todas elas foram excluídas de análises posteriores.

A Tabela 9 mostra o número total de seqüências de cada biblioteca, além das seqüências relativas às categorias *RefSeq*, *Unigene* e *EST* identificadas.

Tabela 9 - Número de seqüências relativas as categorias *RefSeq*, *Unigene* e *ESTs*, de acordo com cada biblioteca RaSH, e o percentual de seqüências destas categorias por biblioteca.

Biblioteca RaSH	Reference sequence	Unigene	EST	Percentual em relação ao total de seqüências obtidas em cada biblioteca
RHID01	182	11	15	48,3% (208/430)
RHSD01	208	11	15	48,7% (234/480)
RHID02	440	10	12	74% (462/624)
RHSD02	148	15	32	37% (195/528)
Total	978	47	74	53,2% (1099/2062)

As seqüências dessas categorias (*RefSeq*, *Unigene* e *EST*) foram clusterizadas a fim de se identificar o número de genes que representavam. A tabela 10 apresenta o número de genes representados por cada categoria, de acordo com cada biblioteca.

Tabela 10 – Número de genes classificados pelas categorias *RefSeq*, *Unigene* e *EST*, de acordo com cada biblioteca.

Biblioteca RaSH	Reference sequence	Unigene	EST	Total
RHID01	96	11	14	121
RHSD01	116	7	13	136
RHID02	180	8	12	200
RHSD02	96	15	32	143
Total	448	41	71	560

4.4 ANÁLISE COMPARATIVA DAS BIBLIOTECAS RASH

4.4.1 Identificação de Transcritos Específicos de cada Biblioteca

A página *GEAP-BDC* disponibiliza uma ferramenta que permite comparar as sequências identificadas em cada biblioteca individualmente.

É possível observar nas figuras 11 e 12 parte do resultado da análise comparativa das bibliotecas DCIS e IDC provenientes dos dois casos (pacientes 1 e 2).

RefSeq		RHID01 (IDC-1)	RHSD01 (DCIS-1)
NM_020532	Homo sapiens reticulon 4 (RTN4), transcript variant 1, mRNA		1
NM_052997	Homo sapiens ankyrin repeat domain 30A (ANKRD30A), mRNA		1
NM_001723	Homo sapiens dystonin (DST), transcript variant 1e, mRNA		1
NM_004104	Homo sapiens fatty acid synthase (FASN), mRNA	1	
NM_203463	Homo sapiens LAG1 longevity assurance homolog 6 (S. cerevisiae) (LASS6), mRNA	2	
NM_022036	Homo sapiens G protein-coupled receptor, family C, group 5, member C (GPCR5C), transcript variant 1, mRNA	2	
NM_181783	Homo sapiens transmembrane and tetrapeptide repeat containing 3 (TMTC3), mRNA	1	
NM_031208	Homo sapiens fumarylacetoacetate hydrolase domain containing 1 (FAHD1), transcript variant 2, mRNA	1	
NM_000618	Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1), mRNA		1

Legenda: (Caso 1), segundo o *pipeline GEAP BDC*. À esquerda: *Acession numbers*; ao centro: *Official names*; à direita: nomes das bibliotecas com o número de identificações de cada transcrito; RHID01 - IDC/*tester* e DCIS/*driver*; RHSD01 – DCIS/*tester* e IDC/*driver*.

Figura 11 - Representação de parte da comparação das duas bibliotecas construídas com as amostras do primeiro paciente.

RefSeq		RHID02 (IDC-2)	RHSD02 (DCIS-2)
NM_052886	Homo sapiens mal, T-cell differentiation protein 2 (MAL2), mRNA	1	
NM_003038	Homo sapiens solute carrier family 1 (glutamate/neutral amino acid transporter), member 4 (SLC1A4), mRNA	1	
NM_052997	Homo sapiens ankyrin repeat domain 30A (ANKRD30A), mRNA		7
NM_003017	Homo sapiens splicing factor, arginine/serine-rich 3 (SFRS3), mRNA	1	1
NM_001642	Homo sapiens amyloid beta (A4) precursor-like protein 2 (APLP2), mRNA		1
NM_181783	Homo sapiens transmembrane and tetrapeptide repeat containing 3 (TMTC3), mRNA	1	
NM_003299	Homo sapiens heat shock protein 90kDa beta (Gp94), member 1 (HSP90B1), mRNA	2	
NM_022068	Homo sapiens family with sequence similarity 38, member B (FAM38B), mRNA		1
NM_001293	Homo sapiens chloride channel, nucleotide-sensitive, 1A (CLNS1A), mRNA	2	
NM_002489	Homo sapiens NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa (NDUFA4), nuclear gene encoding mitochondrial protein, mRNA		3

Legenda: (Caso 2), segundo o *pipeline* GEAP BDC. À esquerda: *Acession numbers*; ao centro: *Official names*; à direita: nomes das bibliotecas com o número de identificações de cada transcrito; RHID02 - IDC/tester e DCIS/driver; RHSD02 – DCIS/tester e IDC/driver.

Figura 12 - Representação de parte da comparação das duas bibliotecas construídas com as amostras do segundo paciente.

Desta forma, puderam ser feitas várias comparações entre as bibliotecas, como por exemplo, entre as bibliotecas dos componentes IDC e DCIS das mesmas amostras na comparação duas a duas (como apresentado nas Figuras 11 e 12), com o intuito de identificar os genes específicos de cada uma, portanto aqueles potencialmente mais expressos em cada um dos componentes tumorais IDC e DCIS. Os resultados dessas comparações, com o número de genes específicos classificados pelas categorias *RefSeq*, *Unigene* e *EST*, estão apresentados nas Tabelas 11A e 11B para as amostras 1 e 2, respectivamente.

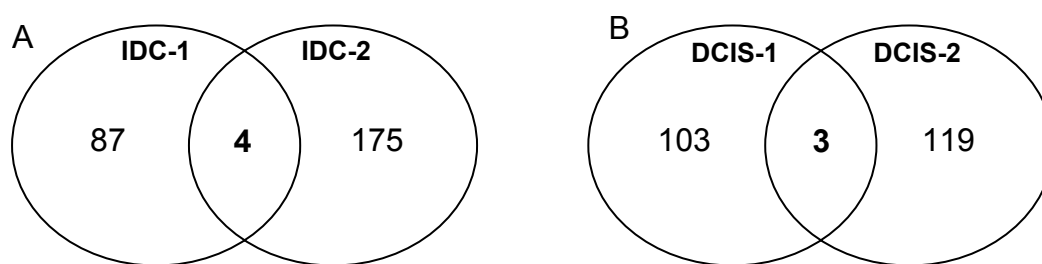
Tabela 11A - Número de genes específicos da cada componente tumoral, DCIS e IDC do caso 1, classificados pelas categorias *RefSeq*, *Unigene* e *EST*.

Categoria dos genes específicos identificados	RHID01	RHSD01
<i>RefSeq</i>	67	87
<i>Unigene</i>	11	7
<i>EST</i>	13	12
Total	91	106

Tabela 11B - Número de genes específicos da cada componente tumoral, DCIS e IDC do caso 2, classificados pelas categorias *RefSeq*, *Unigene* e *EST*.

Categoria dos genes específicos identificados	RHID02	RHSD02
<i>RefSeq</i>	160	76
<i>Unigene</i>	8	15
<i>EST</i>	11	31
Total	179	122

Outra análise realizada envolveu a comparação dos genes específicos de cada componente tumoral (IDC e DCIS) entre as duas amostras para identificar genes comuns entre as duas bibliotecas de mesmo componente tumoral, isto é, com comportamento concordante. Assim, conforme apresentado na figura 13, a comparação entre os genes de cada componente entre as amostras dos dois casos revelou somente 4 genes comuns às bibliotecas RHID01 e RHID02, ou seja, foram específicos do componente IDC nos dois casos; e, 3 genes foram comuns às bibliotecas RHSD01 e RHSD02, ou seja, foram específicos do componente DCIS nos dois casos, totalizando 7 genes com mesmo comportamento.



Legenda: Número de genes representados pelas categorias *RefSeq*, *Unigene* e *EST* específicos de cada biblioteca e comuns entre seus pares. A: Bibliotecas da amostra 1; B: Bibliotecas da amostra 2.

Figura 13 - Genes comuns entre os mesmos componentes tumorais dos dois casos.

4.5 SELEÇÃO DE GENES PARA VALIDAÇÃO POR RT_PCR EM TEMPO REAL

A seleção dos genes candidatos à validação por RT_PCR em tempo real baseou-se nos 7 genes específicos que foram identificados nos mesmos componentes dos dois casos. Desses, 4 foram comuns em RHID01 e RHID02: COPB2 (*Homo sapiens coatomer protein complex, subunit beta 2*), EIF1 (*Homo sapiens eukaryotic translation initiation factor 1*), RDBP (*Homo sapiens RNA binding protein*) e TMTC3 (*Homo sapiens transmembrane and tetratricopeptide repeat containing 3*); e 3 foram comuns em RHSD01 e RHSD02: ANKRD30A (*Homo sapiens ankyrin repeat domain 30A*), KRT19 (*Homo sapiens keratin 19*) e METTL3 (*Homo sapiens methyltransferase like 3*) (Tabela 12).

Tabela 12 - Genes selecionados para validação por RT_PCR em tempo real. Genes específicos de cada componente tumoral e comuns entre os casos 1 e 2.

<i>Official name</i>	<i>Acession number</i>	Componente tumoral identificado
COPB2	004766.1	IDC
EIF1	005801.3	IDC
RDBP	002904.5	IDC
TMTC3	181783.1	IDC
ANKRD30A	052997.2	DCIS
KRT19	002276.3	DCIS
METTL3	019852.3	DCIS

Além disso, também foi selecionado o gene fibronectina - variante transcricional 1 (*Acession Number*: NM_212482; *Official Name*: FN1). Embora clones referentes ao gene fibronectina tenham sido encontrados nas bibliotecas *tester-IDC* e *tester-DCIS* do caso 2 (identificado uma vez em RHID02 e uma vez em RHSD02), foi identificado como específico da biblioteca *tester-IDC* do caso 1 (identificado 6 vezes). Sobretudo, devido à função já descrita desse gene, de participar de mecanismos de adesão e migração celular, além de estar associado ao processo de metástase tumoral (IOACHIM et al. 2002), ele foi selecionado para ser confirmado por RT_PCR em tempo real.

Dessa forma, foram selecionados 8 genes para terem a diferença de expressão confirmada através de RT_PCR em tempo real nos 2 componentes IDC e DCIS. As funções dos genes específicos nos dois casos, incluindo a função de FN1 (específico somente na amostra 1) estão descritas no Quadro 1.

Quadro 1 - Função dos 8 genes selecionados para confirmação por RT_PCR em tempo real.

Official name	Acession number	Função do gene
COPB2	004766	Atividade de transporte protéico
EIF1	005801	Atividade de fator de iniciação de transcrição
RDBP	002904	Atividade de ligação à RNA, nucleotídeos e proteínas
TMTC3	181783	Atividade de ligação
ANKRD30A	052997	Atividade como molécula de adesão celular e de fator de transcrição
KRT19	002276	Constituição estrutural de citoesqueleto
METTL3	019852	Atividade de ligação a RNA e atividade metiltransferase
FN1	212482	Constituinte de matriz extracelular

Para confirmação dos dados encontrados nas bibliotecas RaSH por RT_PCR em tempo real, os 8 genes candidatos selecionados foram analisados nas mesmas amostras utilizadas na confecção das bibliotecas, utilizando os cDNAs fita dupla reservados antes da reação de transcrição *in vitro* (item 3.5.1), sendo portanto uma validação técnica.

Como não foi possível padronizar o gene ANKR30A, uma vez que os oligonucleotídeos desenhados não amplificaram nenhum fragmento específico deste gene, ele não foi submetido a experimentos posteriores de RT_PCR quantitativo. A quantidade de cDNA e as concentrações finais dos oligonucleotídeos utilizados para a amplificação dos 4 genes normalizadores e 7 genes candidatos foi obtida após a padronização (item 3.7.1) e os valores estão apresentados na Tabela 13.

Tabela 13 - Padronização da quantidade de cDNA e das concentrações de *primers* para reação de RT_PCR em tempo real.

Gene	0,1 ng de cDNA amplificado (HB4a)		
	200 nM	400 nM	800 nM
	<i>Primer</i>	<i>Primer</i>	<i>Primer</i>
ACTB		x	
BCR			x
GAPDH			x
HPRT1		x	
COPB2		x	
EIF1		x	
FN1			x
KRT19	x		
METTL3		x	
RDBP		x	
TMTC3		x	

Sendo assim, seguiu-se para os experimentos de cálculo da eficiência com 7 genes candidatos e 4 genes normalizadores, utilizando cDNA molde proveniente da linhagem celular HB4a.

Para o gene ACTB o valor de eficiência não pôde ser calculado. A curva de dissociação do gene normalizador ACTB mostrou 2 picos em todas as diluições, indicando que houve a amplificação de um fragmento inespecífico.

Os valores da eficiência de amplificação dos *primers* desenhados estão descritos na Tabela 14.

Tabela 14 - Eficiência de amplificação dos *primers* desenhados.

Gene*	Slope (a)*	Correlação		Pontos na
		(R2)*	Eficiência*	Curva*
BCR	-3.23	0,96	0.96	4
GAPDH	-3.25	0,98	0.97	4
HPRT1	-3.45	0,99	0.94	4
COPB2	-3.86	0,98	0.81	4
EIF1	-3.44	0,99	0.95	5
KRT19	-2,90	0,96	0,80	5
FN1	-3.46	0,99	0.94	5
MTTL3	-3.72	0,99	0.85	3
RDBP	-3.27	0,99	0.98	5
TMTC3	-3.41	0,91	0.96	5

Gene*: nome do gene; *Slope**: coeficiente angular da reta (a); Correlação R2*: correlação de Pearson dos valores obtidos em relação ao *Ct* e as diferentes diluições de cada amostra; Eficiência*: valor de eficiência de cada *primer*, obtido através da fórmula que considera o coeficiente angular da reta; Pontos da curva*: número de pontos referentes a diluições utilizados para obtenção da reta.

Dessa forma seguiu-se com os experimentos de RT_PCR quantitativo utilizando os cDNAs oriundos dos componentes tumorais IDC e DCIS dos dois casos para os 10 genes (3 genes normalizadores e 7 genes candidatos).

No entanto, durante os experimentos de validação, o gene TMTC3 não apresentou amplificação em nenhuma amostra do estudo provavelmente por apresentar baixa expressão, e desta forma, não podendo ser analisado, foi excluído de análises posteriores.

Para identificação dos genes normalizadores mais estáveis entre as amostras analisadas nesse estudo foi utilizado o programa *geNorm* (VANDESOMPELE et al. 2002). Desta forma, os genes BCR, GAPDH e HPRT1 foram submetidos à análise do *geNorm*. Apesar de todos os genes testados apresentarem valor de medida de estabilidade de expressão gênica dentro do índice recomendado pelo programa ($M < 1,5$), decidiu-se utilizar o gene GAPDH

como normalizador para este conjunto de amostras, por ter se mostrado o gene normalizador mais estável, pois apresentou o menor valor de variação M (M=0.794) (Figura 14).

A média obtida dos valores de *Cts* de cada gene normalizador está apresentada em Anexo 2.

	BCR	GAPDH	HPRT1
DCIS-1	8,21E -03	8,17E -03	8,37E -02
IDC-1	8,39E -01	3,09E -01	1,67E +00
DCIS-2	3,46E -02	8,95E -03	7,35E -02
IDC-2	2,58E -01	8,85E -02	1,64E +00
M < 1.5	1,010	0,794	0,957

Legenda: M-Valor de variação do gene no conjunto de amostras.

Figura 14 - Análise do *software geNorm* sobre o papel dos genes BCR, GAPDH e HPRT1 como normalizadores neste conjunto de amostras.

Assim seguiu-se com os cálculos de quantificação de expressão relativa dos 6 genes selecionados (COPB2, EIF1, FN1, KRT19, METTL3 e RDBP) nas amostras tumorais DCIS e IDC, utilizando para isso a média das duplicatas dos *cycles threshold* (*Cts*) dos genes de interesse e do gene GAPDH. Este cálculo foi feito segundo a fórmula proposta por PFAFFL (2000), que leva em consideração a eficiência individual de cada gene. Só foram analisados genes que apresentaram valores de *Cts* (duplicatas) com desvio-padrão menor do que 0,4.

A Tabela 15 mostra os valores de expressão relativa IDC/DCIS obtidos nas reações de RT_PCR em tempo real, dos diferentes genes candidatos nas

amostras dos casos 1 e 2. Os valores individuais das duplicatas de cada gene em cada amostra e os valores de eficiência obtidos estão apresentados no Anexo 3.

Tabela 15 - Dados de expressão relativa dos genes candidatos. ND: Valor não determinado de expressão relativa devido a desvio-padrão maior do que 0,4.

Gene	Amostra	Média Cts amostras	Desvio-Padrão dos Cts	Expressão relativa IDC/DCIS	Identificado na biblioteca
METTL3	IDC-1	25.18	0.07	-4,27	DCIS
	DCIS-1	31.08	0.06		
	IDC-2	27.20	0.09	ND	
	DCIS-2	29.86	0.51		
KRT19	IDC-1	27.23	0.26	-1,27	DCIS
	DCIS-1	33.00	0.17		
	IDC-2	29.99	1.71	ND	
	DCIS-2	36.69	3.53		
EIF1	IDC-1	25.35	0.01	-1.33	IDC
	DCIS-1	32.54	0.35		
	IDC-2	28.60	0.18	1,53	
	DCIS-2	32.68	0.07		
FN1	IDC-1	24.37	0.04	0,92	IDC
	DCIS-1	29.71	0.22		
	IDC-2	23.3	0.15	5,24	
	DCIS-2	29.24	0.11		
COPB2	IDC-1	29.70	0.07	ND	IDC
	DCIS-1	33.65	0.95		
	IDC-2	31.33	0.00	ND	
	DCIS-2	35.02	0.52		
RDBP	IDC-1	21.99	0.09	1,19	IDC
	DCIS-1	27.53	0.09		
	IDC-2	22.38	0.04	0,88	
	DCIS-2	25.56	0.13		

Considerando que os padrões gerais de expressão dos componentes DCIS e IDC de uma mesma amostra tumoral são extremamente similares,

observados através de experimentos de microarray pelo nosso grupo (Castro et al. em preparação) e pela literatura (SCHUETZ et al. 2006), estabelecemos um valor mínimo de diferença de expressão de 1,5 para considerar um gene como diferencialmente expresso na análise de RT_PCR em tempo real. Dessa forma, aqueles que apresentaram um desvio-padrão maior do que 0,4 entre as duplicatas foram desconsiderados, pois valores maiores do que 0,4 representam diferenças de expressão de 1,3, se o gene apresentar eficiência de amplificação de 100% ($2^{0,4} = 1,3$). Assim, dos 6 genes avaliados por RT_PCR quantitativo, para 5 deles foi possível obter a expressão relativa em pelo menos uma amostra. E destes, 3 foram confirmados como diferencialmente expressos entre IDC e DCIS, sendo eles os genes EIF1, FN1 e METTL3.

Os genes EIF1 e FN1 foram confirmados por RT_PCR em tempo real como mais expressos em IDC em relação a DCIS. O gene EIF1 apresentou-se 1,53 vezes mais expresso em IDC em relação à DCIS quando analisado no caso 2. Assim como o gene FN1 foi confirmado como 5,24 vezes mais expresso em IDC em relação a DCIS também no caso 2. Enquanto que o gene METTL3 foi confirmado como 4,27 vezes mais expresso em DCIS em relação à IDC quando analisado no caso 1.

Como houve um baixo índice de confirmação dos resultados pelo RT_PCR em tempo real, provavelmente por que as diferenças são pequenas entre os 2 componentes, decidimos realizar a confirmação de todos os genes candidatos que foram identificados como específicos de DCIS e IDC nas amostras avaliadas por RASH. Para isso utilizamos a metodologia de

microarray, que possibilita a análise de todos os genes em um único experimento.

4.6 VALIDAÇÃO EM GRANDE ESCALA ATRAVÉS DE MICROARRAY

Com o intuito de avaliar a diferença de expressão entre IDC e DCIS de um número maior de genes, decidiu-se realizar uma validação em grande escala por *microarray*. Para isso foram selecionados os genes específicos de cada componente tumoral (DCIS e IDC) de cada caso. Assim, do total de 560 genes identificados individualmente nas 4 bibliotecas (Tabela 9 - item 4.3), representados pelas categorias *RefSeq*, *Unigene* e *EST*, foram identificados 443 genes específicos de cada uma das 4 bibliotecas (Tabela 16).

Tabela 16 - Número de genes específicos da cada componente tumoral (DCIS e IDC) de cada biblioteca, classificados pelas categorias *RefSeq*, *Unigene* e *EST*.

Biblioteca RaSH	Reference Sequence	Unigene	EST	Total de sequências específicas
RHID01	54	10	12	76
RHSD01	77	6	12	95
RHID02	146	7	11	164
RHSD02	64	14	30	108
Total	341	37	65	443

Esses genes foram submetidos a uma inspeção visual mais detalhada para seleção daqueles que seriam imobilizados na plataforma. Durante essa etapa foram identificados 19 genes ribossomais e mitocondriais, os quais foram excluídos de análises posteriores. Assim, identificou-se 424 genes específicos

de uma só ou comuns em duas bibliotecas do mesmo componente tumoral. A tabela 17 apresenta os números dos genes classificados pelas categorias *RefSeq*, *Unigene* e *EST*, específicos de cada biblioteca, em que os genes comuns estão representados em apenas uma das duas bibliotecas do mesmo componente tumoral, sendo assim, o número total apresentado na Tabela 17 é de 417 genes.

Tabela 17 - Número de genes específicos da cada componente tumoral (DCIS e IDC) de cada biblioteca, classificados pelas categorias *RefSeq*, *Unigene* e *EST*, após inspeção visual para exclusão de genes ribossomais e mitocondriais.

Biblioteca RaSH	Reference Sequence	Unigene	EST	Total de seqüências específicas
RHID01	52	9	10	71
RHSD01	71	6	11	88
RHID02	139	7	10	156
RHSD02	60	13	29	102
Total	319	35	60	417

Os 417 transcritos específicos de cada uma das 4 bibliotecas estão apresentados no Anexo 1.

Dos 417 transcritos identificados como diferencialmente expressos, específicos de cada biblioteca, 414 foram imobilizados numa lâmina de vidro. Não foi possível a imobilização de 3 deles, devido a impossibilidade de sua amplificação por PCR a partir de colônias. Nessa plataforma foram também imobilizados outros clones pertencentes a outros projetos em desenvolvimento no nosso grupo totalizando 1911 seqüências que correspondem a genes humanos. No presente estudo as demais seqüências só foram analisadas para

avaliação da qualidade experimental do microarray e nos cálculos de normalização dos dados, com a finalidade de aumentar o poder dessas análises.

Inicialmente se pretendia analisar a expressão desses 414 genes nas amostras dos mesmos casos usados para confeccionar as bibliotecas, para confirmação técnica através de microarray dos dados de RASH. Para isso, metade da quantidade de ds-cDNA reservada para os experimentos de validação foi submetida a um segundo *round* de amplificação por transcrição *in vitro* -a outra metade deste material foi submetida aos experimentos de RT_PCR em tempo real. No entanto, as quantidades obtidas de RNA amplificado em 2 *rounds* de transcrição *in vitro* foram insuficientes para prosseguir com a metodologia de microarray (tabela 18); provavelmente por ineficiência da reação de transcrição *in vitro* ou por perda de material na purificação do aRNA.

Tabela 18 - Quantificação de aRNA após 2 *rounds* de amplificação por transcrição *in vitro*.

Caso	Diluição (1:100)	230	260	280	R (260/ 280)	Concentração (ng/μL)	Quantidade (μg)
Caso número 1	DCIS	0,741	0,051	0,034	1,5	204	2,0
	IDC	0,752	0,068	0,041	1,65	272	2,7
Caso número 2	DCIS	0,153	0,021	0,012	1,75	84	0,8
	IDC	0,242	0,034	0,021	1,62	136	1,3

Assim partimos para a utilização de um grupo independente de 6 casos, os quais constituíam amostras pareadas DCIS e IDC, as quais haviam sido processadas da mesma maneira que as amostras utilizadas na construção das

bibliotecas RaSH. Essas amostras tiveram as células dos componentes DCIS e IDC microdissecadas a *laser* e submetidas a 2 *rounds* de transcrição *in vitro* para possibilitar os experimentos de *microarray*, totalizando 12 amostras utilizadas nesse estudo. As amostras foram hibridizadas em duplicata utilizando o sistema inverso de incorporação dos corantes Cy3 e Cy5 (*dye-swap*). Dessa forma, as 24 amostras foram co-hibridizadas com uma referência comum (RNA oriundo da linhagem celular HB4a) (O'HARE et al. 1991).

4.6.1 Pré-análise das lâminas

Para avaliação da qualidade das lâminas confeccionadas e do experimento de *microarray*, incluindo todas as etapas, tais como pré-hibridização, hibridização, lavagem e as condições da varredura do sinal através do *scanner*, as lâminas foram submetidas à pré-análises de qualidade utilizadas pela nossa infra-estrutura de *microarray*, conforme detalhado em material e métodos (item 3.9.4). Como todas as lâminas apresentaram qualidade satisfatória em todos os itens da pré-análise, todas elas foram utilizadas para as análises matemáticas para identificação dos genes diferencialmente expressos.

A Figura 15 mostra uma clusterização hierárquica baseada no padrão geral de expressão encontrado, e a partir desta, é possível notar que houve uma maior semelhança no padrão geral entre as lâminas controle e *swap* para 10 das 12 amostras; o que já era esperado, pois estas lâminas são apenas réplicas experimentais com inversão de corantes. No entanto, para duas amostras esse pareamento não foi perfeito. A amostra 25, do componente tumoral IDC, teve sua lâmina *swap* mais semelhante com o par de lâminas

controle e *swap* do componente tumoral DCIS, ao invés de apresentar uma maior semelhança com a lâmina controle de seu mesmo componente. E a amostra 75 teve as lâminas controle dos componentes DCIS e IDC mais semelhantes entre si, assim como as lâminas *swap* dos 2 componentes mais semelhantes entre si, ao invés de apresentar uma maior semelhança entre as lâminas controle e *swap* do mesmo componente tumoral. Ou seja, para as amostras 25 e 75, a diferença na incorporação dos fluoróforos (Cy3 e Cy5) foi maior do que a própria diferença no perfil de expressão gênica, indicando que este resultado é consequência de uma alta semelhança no padrão geral de expressão gênica desses componentes tumorais IDC e DCIS, os quais tiveram suas células microdissecadas de uma mesma amostra tumoral.

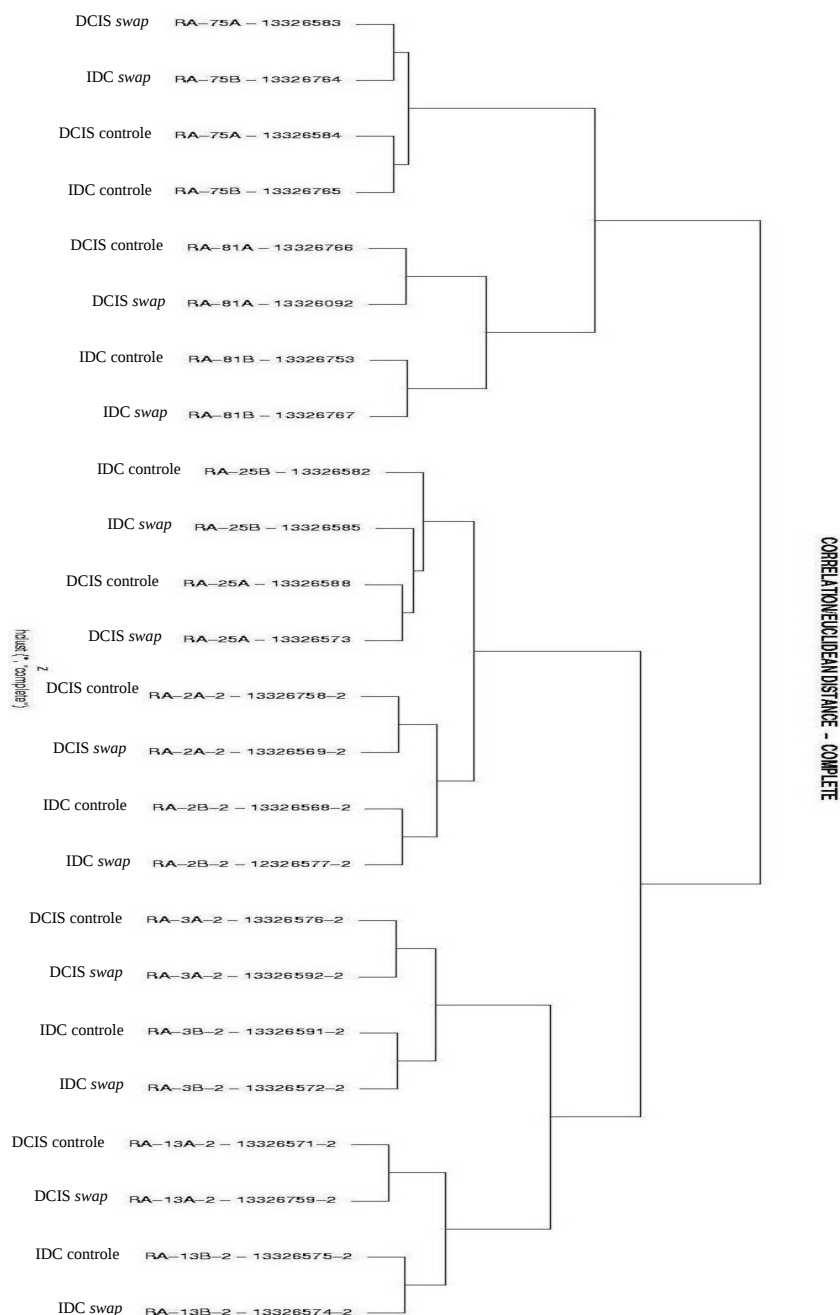


Figura 15 - Clusterização hierárquica. Dendrograma das lâminas controle e swap das 12 amostras tumorais (6 amostras DCIS e 6 amostras IDC).

Os valores obtidos da correlação de *Pearson* (R^2) entre as lâminas controle e swap de cada amostra estão apresentadas na Tabela 19. Esses valores determinam uma medida de relacionamento entre duas variáveis aleatórias, sendo que todas elas apresentaram valores maiores do que 0,8

(valor mínimo aceito no estudo), o que reflete que a qualidade do experimento foi satisfatória.

Tabela 19 - Valores da correlação de incorporação de fluoróforos entre a lâmina controle e a lâmina *swap*, de acordo com cada amostra.

Amostras	Correlação Pós-Normalização
2A	0,94
2B	0,95
3A	0,95
3B	0,95
13A	0,98
13B	0,93
25A	0,92
25B	0,89
75A	0,94
75B	0,93
81A	0,92
81B	0,92

4.6.2 Identificação de genes diferencialmente expressos

As análises matemáticas para identificação dos genes diferencialmente expressos foram realizadas com os seis casos pareadas de carcinoma ductal *in situ* e invasivo de mama. Nesse estudo foram avaliados os 414 transcritos identificados como específicos de cada uma das 4 bibliotecas RaSH construídas. O valor de intensidade do *background* local foi subtraído do valor de intensidade do *foreground* obtido para cada *spot*, sendo considerados apenas os pontos que apresentaram valor positivo. Após esta etapa todos os valores foram trabalhados em escala logarítmica de base 2. A normalização dos dados foi realizada através do método de *Lowess*. Os valores relativos

normalizados de cada gene foram submetidos ao teste T de *Student* pareado. Foram considerados genes diferencialmente expressos aqueles que apresentaram um p-valor pareado menor do que 0,05. Desta forma, a análise conjunta dos 6 casos, possibilitou a identificação de 8 genes diferencialmente expressos entre DCIS e IDC, com *fold change* variando entre 1,13 e 1,33, sendo que 5 deles apresentaram comportamento concordante com os dados de RaSH, mostrando uma baixa concordância com os dados de RaSH (1,20% - 5/414). A Tabela 20 apresenta os 5 genes com seus respectivos *folds*. No entanto, é interessante lembrar que neste caso não se trata de uma validação técnica, mas sim de uma validação biológica, na qual houve a utilização de um grupo independente de amostras, já sendo esperado que o nível de concordância diminuísse de forma expressiva.

Tabela 20 - Genes diferencialmente expressos identificados através da análise conjunta dos 6 casos por *microarray*.

Genes diferencialmente expressos	<i>Fold</i>	p-valor pareado	Componente - Biblioteca	<i>Acession Number</i>
EST Human	1,13	0,01	DCIS-1	BE122677
Unigene	1,13	0,01	DCIS-2	DB324154
EST Human	1,33	0,04	DCIS-2	CN414308
ARS2	-1,15	0,03	IDC-2	NM_182800
TRIP6	-1,22	0,04	IDC-2	NM_003302

O câncer de mama é uma doença muito heterogênea, e pacientes com as mesmas características anatomopatológicas podem apresentar um comportamento clínico distinto quanto à evolução da doença e resposta a tratamentos (PORTER et al. 2003). Essa diferença pode estar refletida nas diferenças no nível de expressão gênica entre amostras de IDC e DCIS de

diferentes pacientes. Isso pode explicar o baixo índice de confirmação com os dados de RaSH, no conjunto de 6 amostras pareadas distintas.

Assim, nós avaliamos individualmente os casos pareados, com o objetivo de identificar genes que apresentam diferença de expressão nos componentes *in situ* e invasivo, não de forma sistemática entre os 2 componentes, mas sim diferenças no nível de expressão que ocorrem somente em determinados grupos de amostras, que podem estar associadas a determinada característica clínica da mesma. Para isso foi realizada uma comparação individual em cada par considerando o valor médio das replicatas, controle e *swap*, entre os dois componentes IDC/DCIS. Nesse caso, foram considerados diferencialmente expressos aqueles genes que apresentaram uma diferença de expressão (*fold change*) maior ou igual a 1,50. Esta análise está apresentada em Anexo 4 e permitiu a identificação de 114 transcritos diferencialmente expressos, sendo 67 down-regulados e 47 up-regulados. Destes, 15 transcritos apresentaram o mesmo comportamento em pelo menos dois diferentes pares de casos, sendo 12 *down*-regulados e 3 *up*-regulados em IDC. Um resultado que chamou a atenção foi que todos esses 15 transcritos foram oriundos das bibliotecas das amostras do caso 2, as quais foram adaptadas para RNA amplificado a partir da utilização do oligonucleotídeo TS_DpnII. Dentre esses, a EST CN414308 e o gene TFF1 foram identificados em 3 diferentes casos. Os genes com mesmo comportamento em mais de um caso podem ser observados na Tabela 21.

Tabela 21 - Genes com mesmo comportamento em pelo menos dois casos.

Up-regulados em IDC	Casos	Gene
LGEA-RHID02-070406-014-F06	Casos 2 e 75	Homo sapiens glutathione S-transferase M3 (brain) (GSTM3), mRNA
LGEA-RHID02-260406-016-B11	Casos 2 e 13	Homo sapiens CDK5 regulatory subunit associated protein 3 (CDK5RAP3), transcript variant 1, mRNA
LGEA-RHID02-090506-017-G09	Casos 13 e 75	Homo sapiens connective tissue growth factor (CTGF), mRNA
Down-regulados em IDC	Casos	Gene
LGEA-RHSD02-020506-025-G08	Casos 2 e 81	Homo sapiens amyloid beta (A4) precursor-like protein 2 (APLP2), mRNA
LGEA-RHSD02-130306-021-H07	Casos 2 e 13	EST T09065
LGEA-RHSD02-230306-022-B06	Casos 2 e 3	PREDICTED: Homo sapiens hypothetical protein LOC647524, transcript variant 2 (LOC647524), Mrna
LGEA-RHSD02-280406-024-D11	Casos 2 e 13	Homo sapiens nephronectin (NPNT), mRNA
LGEA-RHSD02-020506-025-C10	Casos 3 e 81	Homo sapiens period homolog 2 (Drosophila) (PER2), transcript variant 1, Mrna
LGEA-RHSD02-270606-026-E06	Casos 3 e 81	Homo sapiens phospholipase C, beta 4 (PLCB4), transcript variant 2, mRNA
LGEA-RHSD02-120406-023-C01	Casos 13 e 25	Homo sapiens F-box and WD-40 domain protein 4 (FBXW4), Mrna
LGEA-RHSD02-120406-023-C06	Casos 3, 13 e 25	EST CN414308
LGEA-RHSD01-230306-023-D12	Casos 3 e 81	EST CA415848
LGEA-RHSD02-020506-025-E03	Casos 3 e 81	EST BU684023
LGEA-RHSD02-130306-021-H11	Casos 3 e 81	EST AI546885
LGEA-RHSD02-270606-026-C02	Casos 3, 13 e 81	Homo sapiens trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in) (TFF1), mRNA

Com o intuito de verificar se houve diferença entre as duas metodologias utilizadas na confecção das bibliotecas RaSH que utilizou amostras de dois pacientes quanto ao índice de confirmação de genes diferencialmente expressos entre IDC e DCIS, foi analisada a proporção de transcritos confirmados em cada um dos 6 casos tumorais (Tabela 22).

Tabela 22 - Percentual de genes confirmados por microarray como diferencialmente expressos entre IDC e DCIS no grupo independente de 6 casos pareados.

Biblioteca	Caso 2	Caso 3	Caso 13	Caso 25	Caso 75	Caso 81
RaSH 1	2,54%	8,28%	3,82%	6,33%	0,63%	3,18%
RaSH 2	2,33%	9,33%	8,94%	5,83%	0,77%	6,22%

A Tabela 22 mostra que não houve diferença significativa entre os índices de validação a partir de bibliotecas RaSH construídas através de duas estratégias diferentes de amplificação de mRNA.

5 DISCUSSÃO

O câncer de mama apresenta uma alta frequência na população brasileira e mundial, resultando num grande número de mortes, sendo que o processo de metástase é sua principal causa. Assim, o entendimento molecular das etapas que participam deste processo, como por exemplo, a invasão do tumor em tecidos adjacentes, que é a primeira delas, é muito importante para auxiliar na identificação de potenciais marcadores moleculares de invasão tumoral e de prognóstico. Nesse contexto, este trabalho teve o intuito de identificar genes diferencialmente expressos da transição do carcinoma ductal de mama *in situ* para o invasivo, que podem ser candidatos a participarem do processo de invasão neste tipo de tumor.

A escolha da metodologia utilizada, bibliotecas subtrativas de cDNA (RaSH), foi baseada no fato de ser uma metodologia alternativa em relação a outras comumente utilizadas na análise de diferença de expressão gênica em larga escala, como por exemplo, microarray. A metodologia RaSH é sensível e sua aplicação não requer um conhecimento prévio dos genes estudados, podendo identificar genes novos e/ou transcritos raros, isto é, com um baixo nível de expressão (JIANG et al. 2000; BOUKERCHE et al. 2004), que podem estar envolvidos com determinado processo biológico. Neste contexto, sua aplicação é bastante conveniente na identificação de genes envolvidos no processo de invasão tumoral em carcinoma ductal de mama, podendo não só identificar genes diferencialmente expressos entre os dois componentes, ductal *in situ* e invasivo, mas também identificar genes novos e outros ainda não

relatados na literatura como envolvidos com esse processo. Diversos trabalhos que utilizaram a metodologia de biblioteca subtrativa RaSH relataram que ela é eficaz não somente na identificação de genes diferencialmente expressos, mas também daqueles cuja expressão é exclusiva de uma população de transcritos (JIANG et al. 2000).

Outra característica importante no delineamento desse estudo foi a utilização da metodologia de microdissecção a laser, que permitiu a captura de populações puras de células dos dois componentes, ductal *in situ* e invasivo, a partir de uma mesma amostra tumoral. Este delineamento traz vantagens em relação à utilização de culturas celulares ou linhagens imortalizadas que representam um material genético que sofreu outros eventos moleculares distintos daqueles do tecido tumoral que as originou. Além disso, possibilita a obtenção de uma correlação mais acurada entre o perfil de expressão gênica da célula tumoral dentro do contexto de sua morfologia original (CARRARO et al. *in press*). Por estes motivos, este delineamento experimental é mais apropriado na busca por marcadores moleculares de invasão tumoral e de prognóstico. Assim, com sua utilização pudemos capturar apenas as células de interesse de cada componente tumoral (DCIS e IDC), gerando dados representativos da célula tumoral de interesse e não do tumor como um todo, evitando a coleta de células estromais não neoplásicas, o que interfere no padrão de expressão gênica.

Embora essa metodologia tenha vantagens relacionadas à captura específica de células, a desvantagem é que a quantidade de RNA obtida é muito pequena, insuficiente para estudos de análise de expressão gênica, como por exemplo, na construção de bibliotecas subtrativas RaSH. Desta

forma, neste trabalho a metodologia de construção das bibliotecas necessitou ser adaptada para RNA amplificado para gerar quantidades de RNA suficientes para sua aplicação.

Como este trabalho foi o primeiro a estabelecer uma metodologia de análise de expressão gênica por biblioteca subtrativa de cDNA a partir de quantidades de RNA total da ordem de poucos nanogramas, houve carência de informações na literatura a respeito de qual metodologia de amplificação de mRNA seria mais eficiente para sua adaptação.

Recentemente foi desenvolvida uma metodologia de biblioteca subtrativa de cDNA que utiliza amostras de RNA amplificado (LIU et al. 2006). Este método, chamado de *Subtractive Transcription-based Amplification of mRNA (STAR)* foi empregado na análise de genes diferencialmente expressos em tecido cerebral afetado pela doença de Alzheimer. No entanto, tal metodologia utilizou transcritos previamente clonados em bibliotecas de cDNA. Dessa forma, a utilização da metodologia de amplificação de mRNA por transcrição *in vitro* foi aplicada como um procedimento que visou eliminar a etapa de amplificação por PCR, sendo capaz de aumentar a população de transcritos e de maneira linear, resultando em quantidades suficientes desses transcritos para confecção da biblioteca com menor probabilidade de introdução de vieses.

Em nosso estudo, como o objetivo era a aplicação da metodologia de amplificação diretamente na população de mRNA de uma amostra, isto é, transcritos não previamente clonados, foram testados 2 protocolos de amplificação que diferiram na síntese de primeira fita de cDNA a partir de RNA amplificado por 1 *round* de transcrição *in vitro*: (1) com a utilização de oligonucleotídeo *random primer* (dN6) e (2) com a utilização de

oligonucleotídeo *template switch* (TS). Durante o processo de confecção das bibliotecas, nenhum protocolo apresentou alguma vantagem significativa em relação ao outro. Esta etapa foi igualmente adaptada seguindo ambas as metodologias, e não houve diferença na quantidade de massa de aRNA sintetizado entre os dois protocolos de amplificação.

Dessa forma foram obtidas 2.062 seqüências a partir de 4 bibliotecas que utilizaram os 2 protocolos distintos de amplificação, correspondendo a 443 transcritos específicos de cada biblioteca, sendo 7 deles com comportamento concordante nas amostras dos dois casos. Embora a diferença molecular entre DCIS e IDC de uma mesma amostra tumoral seja realmente bastante sutil (SCHUETZ et al. 2006; Castro et al. em preparação), a utilização da metodologia RASH pôde revelar um número relativamente grande de genes candidatos a serem diferentemente expressos. No entanto, o pequeno número de genes comuns entre as amostras dos dois casos (7 genes), suporta a idéia de que a utilização das duas metodologias de amplificação representou diferencialmente as populações de transcritos nas amostras dos dois casos.

Por exemplo, o uso de TS gera transcritos maiores em relação ao uso de dN6, sendo que nem todos os cDNAs são igualmente representados. Essa característica inerente da estratégia pode levar a uma baixa representatividade dos genes mais longos devido a baixa eficiência da transcriptase reversa em sintetizar fragmentos longos de cDNA a partir de mRNA (MATZ et al. 1999). Isso reflete na representatividade dessas moléculas, uma vez que elas não recebem a trinca de citosinas, necessária para o alinhamento do *primer*. Por outro lado, na estratégia que utilizou o oligonucleotídeo dN6, alguns transcritos podem ter tido sua representatividade prejudicada durante a confecção da

biblioteca. Isso se deve ao fato de que esses transcritos, por serem curtos, não possuem o sítio de reconhecimento da enzima *Dpn* II. Esses vieses na representatividade dos transcritos devem ter influenciado na identificação de genes comuns às duas bibliotecas do mesmo componente tumoral.

A seleção dos candidatos à validação por RT_PCR em tempo real foi baseada na escolha dos transcritos específicos de cada par de biblioteca (DCIS-*tester* e IDC-*driver* e vice-versa) e comuns entre os mesmos tipos de bibliotecas (DCIS-*tester* do caso 1 e DCIS-*tester* do caso 2, e, IDC-*tester* do caso 1 e IDC-*tester* do caso 2) das amostras das duas pacientes, resultando na seleção dos genes ANKRD30A, COPB2, EIF1, KRT19, METTL3, RDBP, TMTC3; e na função, no caso do gene FN1. Para confirmar a diferença de expressão desses genes através de RT_PCR em tempo real, considerando a grande semelhança no perfil de expressão gênica dos componentes DCIS e IDC de uma mesma amostra tumoral, decidimos ser estridentes no desvio-padrão dos valores de *Cts* das duplicatas de cada gene analisado, estabelecendo um limite de 0,4, que é menor do que o valor de 0,5 normalmente recomendado (NOLAN et al. 2006). Ao assumir este valor, estamos aceitando um ruído experimental de 1,3 vezes (no caso de eficiência de amplificação de 100%). Sendo assim, as diferenças de expressão de 1,5 - valor do *cut-off* adotado nesse estudo- mesmo que pequenas, tornam-se mais confiáveis.

Considerando esse critério, foi possível concluir os experimentos de validação somente para os genes EIF1, FN1 e METTL3; sendo possível confirmar a diferença de expressão em ensaios quantitativos para os três. O gene EIF1 apresentou-se 1,53 vezes mais expresso em IDC em relação à

DCIS. Já o gene FN1 foi confirmado como 5,24 mais expresso em IDC em relação a DCIS. Enquanto que o gene METTL3 foi confirmado como 4,27 vezes mais expresso em DCIS em relação a IDC.

Os genes EIF1 e METTL3 não apresentam nenhum relato na literatura sobre seus papéis em câncer de mama. E sendo assim, a confirmação como genes diferencialmente *up*-regulados em IDC e DCIS, respectivamente, sugere que eles podem ser novos genes envolvidos no processo de invasão em carcinoma ductal de mama. Enquanto isso, a identificação da *up*-regulação do gene FN1 em IDC, vem a corroborar com a literatura, pois este gene tem a função de promover adesão célula-matriz extracelular, sendo assim, é considerado como um marcador mesenquimal. E a célula neoplásica epitelial que apresenta uma alta expressão deste gene, apresenta um alto potencial invasivo (IOACHIM et al. 2002).

Os valores de diferença de expressão relativamente baixos encontrados neste estudo corroboram com achados da literatura, os quais estudaram os 2 componentes de uma mesma amostra tumoral, onde inexistia uma diferença causada por diferentes *backgrounds* genéticos. Por exemplo, SCHUETZ et al. (2006) avaliando 18 genes por RT_PCR em tempo real em 4 casos pareados DCIS-IDC, encontraram que 64% dos valores de *fold change* não ultrapassaram o valor de 4 vezes, sendo que o menor valor obtido de diferença de expressão entre os 2 componentes foi de 1,12 para o gene LATS2 e o maior valor de 15,44 para o gene GREM1. No trabalho realizado pelo nosso grupo (CASTRO et al. em preparação), avaliando 2 genes, LUM e CRABP2, em 6 casos pareados DCIS-IDC através de RT_PCR em tempo real, foi identificado que 45% dos valores de *fold change* nos 6 casos testados não ultrapassaram o

valor de 4, sendo que o menor valor obtido de diferença de expressão entre os dois componentes foi de 1,4.

Como o número de genes confirmados por RT_PCR quantitativo foi pequeno e devido a pouca quantidade de cDNA disponível para validação de um número maior de genes candidatos, decidimos realizar uma validação em larga escala por *microarray*, avaliando todo o grupo de genes identificados pelo RaSH como específicos de cada um dos componentes DCIS e IDC nos dois casos. Assim, após a inspeção visual mais detalhada e exclusão de transcritos ribossomais e mitocondriais, que não tinham sido filtrados pelo *pipeline GEAP BDC*, foram selecionados 417 transcritos específicos das bibliotecas RaSH, dos quais 414 foram imobilizados em plataforma de *microarray* e avaliados em 6 casos pareados de carcinoma ductal de mama.

Os dados dos experimentos de *microarray* foram submetidos a duas análises: uma análise pareada conjunta e outra individual destes 6 casos. A análise pareada conjunta dos 6 casos, através de teste t-pareado, confirmou a diferença de expressão de 5 transcritos.

Comparando os nossos dados de bibliotecas subtrativas com dados produzidos por outros grupos, nosso estudo mostrou um número inferior de genes confirmados por outras metodologias quantitativas. No estudo realizado por LIU et al. (2006) foram identificados 800 genes específicos de um total de 1000 analisados. Dos 800 específicos, foram selecionados randomicamente 40 genes para serem avaliados por RT_PCR em tempo real, havendo a confirmação de diferença de expressão para 80% deles, enquanto que nossos dados de RT_PCR quantitativo mostraram a confirmação da diferença de expressão para 50% dos 4 genes avaliados. Essa diferença no índice de

confirmação pode ser resultado da diferente metodologia de amplificação utilizada e também da maior diferença no perfil transcricional entre as populações de RNA subtraídas. Talvez uma opção para diminuir o número de falsos positivos no nosso estudo seria realizar outros *rounds* de subtração e abolir a etapa de amplificação por PCR.

Já no estudo realizado por BOUKERCHE et al. (2004), os autores obtiveram um índice de confirmação de diferença de expressão por *Northern blotting* para 11% dos 72 genes previamente identificados como diferencialmente expressos por *Northern blotting*-reverso. Nos dados de confirmação gerados por análise de microarray nosso índice de validação foi de 1,2%, no entanto nesse caso se tratou de uma validação biológica, uma vez que foi utilizado um grupo independente de amostras.

Os dados de BOUKERCHE et al. (2004) também mostraram valores superiores de diferença de expressão em relação ao nosso estudo (-1,22 a 1,33). Mas há uma diferença importante entre esses trabalhos: BOUKERCHE et al. (2004) utilizaram linhagens celulares de melanoma M4Beu e TIP26R, de baixo e alto potencial metastático, respectivamente; o que pode resultar na identificação de uma alteração mais acentuada do padrão de expressão gênica entre as duas populações de transcritos.

Para avaliar se a biblioteca subtrativa teria enriquecido o conjunto de genes candidatos a estarem diferencialmente expressos, comparamos nossos dados de percentual de diferença de expressão em relação ao número total de genes analisados com outro estudo do nosso grupo, em que os genes avaliados não haviam sido submetidos a nenhuma técnica de enriquecimento, como foi o caso da análise de diferença de expressão por microarray a partir de

transcritos previamente identificados como diferencialmente expressos pela metodologia de RaSH. Para isso, comparamos os dados deste estudo com a análise de expressão por microarray realizada por Castro et al. (em preparação) utilizando uma plataforma customizada com 4.608 genes humanos (BRENTANI et al. 2005). Nesse estudo, foram identificados 33 genes diferencialmente expressos, correspondendo a um percentual de 0,71%. Esses resultados mostram que aparentemente as bibliotecas RaSH não enriqueceram significativamente nosso grupo de genes, pois nosso percentual foi de 1,2%. A etapa de amplificação por PCR, aliada ao fato dos dois padrões de expressão serem extremamente semelhantes, pôde ter contribuído para a geração de alguns artefatos.

Além disso, outro fator relevante que pode ter contribuído para um baixo índice de confirmação de diferença de expressão é o fato dos dois casos utilizados na confecção das bibliotecas de cDNA serem positivos para metástase linfonodal, enquanto que dos 6 casos utilizados para hibridização nos experimentos de microarray, 5 deles eram negativos para metástase linfonodal.

Em relação aos valores relativamente baixos de *fold change* encontrados na análise conjunta dos 6 casos através de *microarray* (máximo de 1,33), os dados de nosso estudo estão em concordância com os dados encontrados por Castro et al. (em preparação), cujos valores não ultrapassaram 2,85. Esses dados reforçam que as diferenças de expressão gênica são muito tênues nesses dois grupos de amostras (DCIS e IDC) e somente para pequenos grupos de genes. Além disso, esses baixos valores de diferença de expressão são normalmente ocasionados pelas sucessivas

normalizações durante as análises de dados de expressão, sendo que nos experimentos de RT_PCR quantitativo esses valores são geralmente maiores.

Entre os 5 genes confirmados pela análise de microarray nesse estudo, estão duas *ESTs* (BE122677 e CN414308), 1 *Unigene* (DB324154) e 2 genes cujos mRNAs já estão caracterizados (ARS2 e TRIP6); sendo que apenas um destes apresenta relatos na literatura sobre seu envolvimento com processo de invasão tumoral. Este gene é o TRIP6 (*Acession number: NM_003302*), sendo encontrado em nossos dados como mais expresso em IDC em relação à DCIS. Sua função está associada à formação de adesões focais, estruturas envolvidas no processo de invasão (FRIEDL e WOLF 2003); mas ainda não foi relatado como envolvido com câncer de mama.

O gene TRIP6 (*thyroid hormone receptor interactor 6*) codifica uma proteína que interage com receptores de hormônios tireóideos, os quais são fatores de transcrição dependentes de hormônio. YI e BECKERLE (1998) mapearam TRIP6 na região 7q22 e relataram sua associação com o processo de tumorigênese pelo fato de que a deleção desta região é encontrada em alguns casos de leiomioma uterino. Esses mesmos autores relatam que este gene codifica duas cópias de um domínio LIM. Este domínio é uma estrutura dupla de *zinc-fingers* responsável por mediar interações protéicas. De acordo com sua sequência nucleotídica, TRIP6 foi agrupado na família zixina, as quais estão envolvidas em vias de sinalização celular e organização do citoesqueleto.

Recentes estudos mostraram que a proteína codificada pelo gene TRIP6 se liga ao receptor do ácido lisofosfatídico (*Lysophosphatidic acid - LPA*); sendo que esta interação é específica entre o receptor LPA2 e TRIP6, não ocorrendo interação com os receptores LPA1 e LPA3 (XU et al. 2004). LPA tem

a capacidade de induzir o rearranjo do citoesqueleto de actina e auxiliar no agrupamento de complexos focais e na formação de adesão focal, a qual é dependente da via de sinalização Rho e mediada pela via de sinalização de integrinas (PANETTI et al. 2001). O recrutamento de TRIP6 para a membrana citoplasmática, o qual é dependente de LPA, promove seu endereçamento para adesões focais e co-localização com fibras de actina do citoesqueleto. Neste processo, TRIP6 se associa com alguns componentes da adesão focal como paxilina, quinases de adesão focal, Src e p130^{cas}. Estas proteínas formam complexos com moléculas de sinalização, Grb2 e Crk, determinando respostas celulares como sinalização mitogênica, locomoção e sobrevivência celular (HANKS e POLTE 1997). XU et al. (2004), através de estudos de duplo-híbrido, ensaios de migração haptotática e RNA-i, com a linhagem celular SKOV-3, de câncer de ovário, mostraram que TRIP6 aumenta a migração celular mediada por LPA e que a inibição de sua expressão reduz esta migração.

Até o presente momento, nenhum estudo foi publicado a respeito do papel deste gene em carcinoma ductal mamário. Mas com estes achados, sugerimos TRIP6 como um novo gene supostamente envolvido com a invasão tumoral em câncer de mama e que deve ser testado como marcador molecular de invasão ou prognóstico. Os demais genes também são candidatos a participarem do processo de invasão, mesmo ainda não tendo sido relatado na literatura nenhum envolvimento com esse processo.

Enquanto isso, a análise pareada individual dos casos confirmou a diferença de expressão de um número maior de genes e com valores de *fold change* também maiores do que o encontrado na análise pareada conjunta dos 6 casos. Foram identificados 114 transcritos com diferença de expressão entre

IDC e DCIS. Sendo que a maior diferença encontrada foi de uma *EST* que se apresentou mais expressa em DCIS em relação à IDC com um *fold* de 6,40 vezes. Dentre os 114 transcritos, focamos em 15 genes que foram identificados como diferencialmente expressos em pelo menos dois casos, reforçando o envolvimento desses no processo de invasão tumoral em carcinoma ductal de mama, já que tais casos são diferentes daqueles utilizados na confecção das bibliotecas RaSH. Desses 15 genes, 9 deles apresentam caracterização do mRNA, 3 (CDK5RAP3, CTGF e GSTM3) são *up*-regulados em IDC e 6 (APLP2, FBXW4, NPNT, PER2, PLCB4 e TFF1) são *down*-regulados em IDC. Dentre os 6 transcritos que não apresentam caracterização de seus mRNAs, 5 *ESTs* (AI546885, BU684023, CA415848, CN414308 e T09065) e uma proteína hipotética (LOC647524) são *down*-regulados em IDC, sendo novos genes candidatos, sem função conhecida, que podem estar envolvidos no processo de invasão em câncer de mama.

O gene TFF1, *down*-regulado em IDC pelos dados de RaSH, também conhecido como pS2, foi encontrado como *down*-regulado em IDC em 3 dos 6 casos por microarray. Tem sido relatado que o nível de transcrição desse gene é induzido por estrógeno, sendo o mais abundante na linhagem celular MCF-7, a qual apresenta uma alta expressão do receptor de estrógeno (BROOKS et al. 1973). CORTE et al. (2006), através de ensaios imunoradiométricos em 1031 pacientes com câncer de mama, analisaram os níveis intratumorais do mRNA de TFF1, relatando que eles são bastante variáveis entre as pacientes. Encontraram que os altos níveis de mRNA de TFF1 relacionaram-se com uma maior sobrevida global para o grupo de pacientes que receberam tratamento

adjuvante com tamoxifeno, indicando que seu nível de mRNA tem um significativo valor preditivo nestes casos.

São crescentes as evidências que indicam um papel das fosfolipases C (PLC) na invasão e metástase do câncer de mama (BERTAGNOLO et al. 2006). Diversos trabalhos encontraram uma alta expressão gênica para diferentes isoformas do gene PLC. Em contraste, nossos dados de RaSH identificaram a *down*-regulação da isoforma beta-4 (PLCB4) no componente IDC, o que foi confirmado por microarray em dois casos independentes. Esta proteína cataliza a conversão de fosfatidil-inositol-4,5-bifosfato em inositol-1,4,5-trifosfato e diacilglicerol, e a família protéica PLC está implicada em vias de sinalização como *WNT* e *gap junction*. BERTAGNOLO et al. (2006), através de ensaios de imunohistoquímica em 77 tumores de mama, encontraram uma alta expressão da isoforma beta-2 na grande maioria dos tumores, em relação ao tecido normal. Após analisar os dados clínicos e histopatológicos destas amostras, sugeriram que sua alta expressão prediz um prognóstico desfavorável. Em nosso trabalho, a identificação da *down*-regulação em IDC da isoforma beta-4 sugere um papel distinto desta em relação a outros membros da família PLC ou apenas um dado artefactual. Assim, esse resultado indica a necessidade de uma validação por RT_PCR em tempo real a fim de avaliar seu real *status* de expressão gênica em carcinoma ductal de mama.

A identificação da *up*-regulação em IDC do gene CTGF pelas bibliotecas RaSH foi confirmada pelos experimentos de microarray em dois casos independentes. O fator de crescimento do tecido conectivo (CTGF) é capaz de estimular a expressão de várias moléculas de matriz extracelular, e é conhecido como um potente agente angiogênico (KONDO et al. 2002). Sua *up*-

regulação tem sido encontrada em certos tumores devido ao seu papel chave no controle da proliferação e migração de células endoteliais (SHIMO et al. 2001). Mas apresenta um papel controverso na tumorigênese, segundo a literatura. XIE et al. (2001) encontraram um elevado nível de mRNA de CTGF em 52% dos 44 tumores primários de mama analisados por RT_PCR em tempo real, sugerindo seu envolvimento na progressão do câncer de mama, já que o alto nível de expressão de CTGF foi associado com características mais avançadas, como por exemplo, estadió e *status* linfonodal. Enquanto que JIANG et al. (2004) encontraram baixos níveis protéicos de CTGF, por imunohistoquímica, em tumores primários de mama.

Dentre os genes sem caracterização do mRNA que foram identificados em pelo menos dois casos, a *EST* CN414308 foi confirmada como *down*-regulada em IDC em 3 dos 6 casos analisados. Por isso, e pelo fato de ter sido confirmada como menos expressa em IDC através da análise pareada conjunta dos 6 casos (*fold change* de 1,33), sugerimos este transcrito como um novo gene supostamente envolvido com o processo de invasão tumoral em carcinoma ductal de mama.

Entre os genes que confirmaram os dados de RASH em apenas um dos 6 casos analisados, alguns têm sido relatados como relacionados com o câncer. Exemplos destes genes, confirmados como *down*-regulados em IDC, são ANKRD30A e PLA2G12A, em desacordo com a literatura; e, ANXA1, SFRP1, ITGB3BP, corroborando com a literatura. Exemplos desses genes, confirmados como *up*-regulados em IDC são THBS2 e RAB22A. O gene RDBP, que foi identificado nas bibliotecas RASH como específico de IDC, foi confirmado por RT_PCR quantitativo nos dois casos e também por microarray

como mais expresso no componente IDC no caso 81 (o único dentre os 6 casos cujos linfonodos apresentavam-se acometidos por metástase), faz dele um candidato promissor a desempenhar um importante papel no processo de invasão tumoral em carcinoma ductal mamário.

As análises de expressão de diferentes estágios da progressão do carcinoma de mama têm encontrado grandes semelhanças no nível do transcriptoma entre DCIS e IDC, dando fortes evidências que alterações na expressão gênica que conferem um potencial invasivo já estão presentes em estágios pré-invasivos da doença (MA et al. 2003). No entanto, apesar do padrão geral ser extremamente semelhante, inegavelmente as diferenças moleculares mais pronunciadas devem estar presentes em apenas alguns genes da célula tumoral, que juntamente com a influência do microambiente, são responsáveis pela alteração da morfologia e progressão da doença. Apesar dessa grande semelhança em termos gerais de expressão, o pequeno número de genes responsáveis por esta alteração morfológica deve desempenhar um papel chave no processo, e a identificação desses genes pode contribuir para sua melhor compreensão e auxiliar na identificação de marcadores moleculares de invasão e/ou de prognóstico para esse tipo de tumor.

Neste sentido, com o intuito de contribuir para essa finalidade, a utilização da metodologia RaSH possibilitou a identificação de genes candidatos, alguns com função já relatada como envolvida com câncer de mama e outros de função ainda desconhecida, sugerindo que eles podem desempenhar um papel chave no processo de invasão tumoral em carcinoma

ductal de mama, sendo candidatos a serem testados como marcadores moleculares de invasão ou prognóstico.

6 CONCLUSÕES

1. A adaptação da metodologia RaSH para RNA amplificado (aRNA) foi obtida com êxito utilizando duas estratégias diferentes de síntese de primeira fita de cDNA a partir de RNA amplificado, sem vantagem aparente de nenhuma delas durante o processo de confecção.
2. Os resultados obtidos pela análise das 4 bibliotecas subtrativas de cDNA RaSH, confeccionadas a partir de uma população homogênea de células dos componentes tumorais (DCIS e IDC) dos 2 casos de carcinoma ductal de mama mostraram que os 2 tipos de lesão são muito semelhantes em nível molecular apesar da morfologia distinta.
3. Foram identificados 417 genes específicos das 4 bibliotecas confeccionadas, sendo genes candidatos a estarem diferencialmente expressos na transição do carcinoma ductal *in situ* para o invasivo, sendo que somente um baixo percentual desses foi confirmado por outras metodologias de análise de expressão gênica, tais como RT_PCR em tempo real e microarray.
4. A *up*-regulação no componente invasor dos genes EIF1 e FN1, e a *down*-regulação no componente *in situ* do gene METTL3 foram confirmados por experimentos de RT_PCR quantitativo nas amostras utilizadas para confecção das bibliotecas de RASH.

5. Cinco genes candidatos (TRIP6, ARS2, CN414308, BE122677 e DB324154) tiveram sua diferença de expressão confirmada através de experimentos de microarray utilizando 6 casos independentes, sugerindo que são potenciais candidatos a estarem envolvidos no processo de invasão tumoral em carcinoma ductal mamário.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. **Molecular biology of the cell**. 4th ed. New York: Garland Science; 2002a. Cell junction, cell adhesion, and the extracellular matrix; p.1065-125.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. **Molecular biology of the cell**. 4th ed. New York: Garland Science; 2002b. The cell cycle and Programmed Cell Death; p.983-1026.

Allinen M, Beroukhi R, Cai L, et al. Molecular characterization of the tumor microenvironment in breast cancer. **Cancer Cell** 2004; 6:17-32.

Allred DC, Mohsin SK, Fuqua SA. Histological and biological evolution of human premalignant breast disease. **Endocr Relat Cancer** 2001; 8:47-61.

Amari M, Moriya T, Ishida T, et al. Loss of heterozygosity of asynchronous lesions of ductal carcinoma in situ and invasive ductal carcinoma of the human breast. **Jpn J Clin Oncol** 2003; 33:556-62.

Bertagnolo V, Benedusi M, Querzoli P, et al. PLC-beta2 is highly expressed in breast cancer and is associated with a poor outcome: a study on tissue microarrays. **Int J Oncol** 2006; 28:863-72.

Bonner RF, Emmert-Buck M, Cole K, et al. Laser capture microdissection: molecular analysis of tissue. **Science** 1997; 278:1481-83.

Borresen-Dale AL. TP53 and breast cancer. **Hum Mutat**. 2003; 21:292-300.

Boukerche H, Su ZZ, Kang DC, Fisher PB. Identification and cloning of genes displaying elevated expression as a consequence of metastatic progression in human melanoma cells by rapid subtraction hybridization. **Gene** 2004; 343:191-201.

Boyer B, Valles AM, Edme N. Induction and regulation of epithelial-mesenchymal transitions. **Biochem Pharmacol** 2000; 60:1091-9.

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

Brodie AH, Jelovac D, Long B. The intratumoral aromatase model: studies with aromatase inhibitors and antiestrogens. **J Ster Biochem Mol Biol** 2003; 86:283-8.

Callagy G, Cattaneo E, Daigo Y, et al. Molecular classification of breast carcinomas using tissue microarrays. **Diagn Mol Pathol** 2003; 12:27-34.

Campagnoli C, Abba C, Ambroggio S, Peris C. Pregnancy, progesterone and progestins in relation to breast cancer risk. **J Steroid Biochem Mol Biol** 2005; 97(5):441-50.

Carraro DM, Brentani HP, Soares FA, Reis LF, Brentani RR. From tissue samples to tumor markers. In: Appasani K, organizador. **Bioarrays: from basics to diagnostics**. New Jersey: Humana Press; 2006. *in press*

Carter P, Presta L, Gorman CM, et al. Humanization of an anti-p185HER2 antibody for human cancer therapy. **PNAS** 1992; 89:4285-9.

Cavallaro U, Christofori G. Multitasking in tumor progression: signaling functions of cell adhesion molecules. **Ann N Y Acad Sci** 2004; 1014:58-66.

Conacci-Sorrell M, Zhurinsky J, Ben-ze'ev A. The cadherin-catenin adhesion system in signaling and câncer. **J Clin Invest** 2002; 109:987-91.

Cordera F, Jordan VC. Steroid receptors and their role in the biology and control of breast cancer growth. **Semin Oncol** 2006; 33:631-41.

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

Devilee P, Cornelisse CJ. Somatic genetic changes in human breast cancer. **Biochim Biophys Acta** 1994; 1198:113-30.

Dixon JM. Aromatase inhibitors in early breast cancer therapy: a variety of treatment strategies. **Expert Opin Pharmacother** 2006; 7:2465-79.

Emmert-Buck MR, Bonner RF, Smith PD, et al. Laser capture microdissection. **Science** 1996; 274:998-1001.

Folgueira MA, Brentani H, Katayama ML. Gene expression profiling of clinical stages II and III breast cancer. **Braz J Med Biol Res** 2006; 39:1101-13.

Friedl P, Wolf K. Tumour-cell invasion and migration: diversity and escape mechanisms. **Nat Rev Cancer** 2003; 3:362-74.

Fuller AP, Palmer-Toy D, Erlander MG, Sgroi DC. Laser capture microdissection and advanced molecular analysis of human breast cancer. **J Mammary Gland Biol Neoplasia** 2003; 8:335-45.

Fu R, Harris EL, Helfand M, Nelson HD. Estimating risk of breast cancer in carriers of BRCA1 and BRCA2 mutations: a meta-analytic approach. **Stat Med** 2007; 26:1775-87.

Fu SW, Schwartz A, Stevenson H, et al. Correlation of expression of BP1, a homeobox gene, with estrogen receptor status in breast cancer. **Breast Cancer Res** 2003; 5:R82-7.

Gibson UE, Heid CA, Williams PM. A novel method for real time quantitative RT-PCR. **Genome Res** 1996; 6:995-1001.

Gomes LI, Silva RL, Stolf BS, et al. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. **Anal Biochem** 2003; 321:244-51.

Guo W, Pylayeva Y, Pepe A, et al. Beta 4 integrin amplifies ErbB2 signaling to promote mammary tumorigenesis. **Cell** 2006; 126:489-502.

Hanks SK, Polte TR. Signaling through focal adhesion kinase. **Bioessays** 1997; 19:137-45.

Harkonen P. Paracrine prolactin may cause prostatic problems. **Endocrinology** 2003; 144:2266-8.

Hood JD, Cheresch DA. Role of integrins in cell invasion and migration. **Nat Rev Cancer** 2002; 2:91-100.

Hortobagyi GN, de la Garza Salazar J, Pritchard K, et al. The global breast cancer burden: variations in epidemiology and survival. **Clin Breast Cancer** 2005; 6:391-401.

Hughes SJ, Xi L, Raja S, et al. A rapid, fully automated, molecular-based assay accurately analyzes sentinel lymph nodes for the presence of metastatic breast cancer. **Ann Surg** 2006; 243:389-98.

Ioachim E, Charchanti A, Briasoulis E, et al. Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and

laminin in breast cancer: their prognostic value and role in tumour invasion and progression. **Eur J Cancer** 2002; 38:2362-70.

Janssens N, Janicot M, Perera T, Bakker A. Housekeeping genes as internal standards in cancer research. **Mol Diagn** 2004; 8:107-13.

Jiang H, Kang D, Alexandre D, Fisher PB. RaSH, a rapid subtraction approach for identifying and cloning differentially expressed genes. **PNAS** 2000; 97:12684-9.

Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. **Science** 1992; 258:818-21.

Kang Y, Siegel PM, Shu W, et al. A multigenic program mediating breast cancer metastasis to bone. **Cancer Cell** 2003; 3:537-49.

Kondo S, Kubota S, Shimo T, et al. Connective tissue growth factor increased by hypoxia may initiate angiogenesis in collaboration with matrix metalloproteinases. **Carcinogenesis** 2002; 23:769-76.

Lacroix M, Toillon RA, Leclercq G. p53 and breast cancer, an update. **Endocr Relat Cancer** 2006; 13:293-325.

Larionov A, Krause A, Miller W. A standard curve based method for relative real time PCR data processing. **BMC Bioinformatics** 2005; 21; 6:62.

Larson PS, de las Morenas A, Cerda SR, Bennett SR, Cupples LA, Rosenberg CL. Quantitative analysis of allele imbalance supports atypical ductal hyperplasia lesions as direct breast cancer precursors. **J Pathol** 2006; 209:307-16.

Leonard GD, Swain SM. Ductal carcinoma *in situ*, complexities and challenges. **J Natl Cancer Inst** 2004; 96:906-20.

Lester SC. The breast. In: KumarV, Abbas AK, Fausto N, editors. **Robbins and Cotran pathologic basis of disease**. 7th ed. Philadelphia: Elsevier Saunders; 2005. p.1119-54.

Liu QY, Sooknanan RR, Malek LT, et al. Novel subtractive transcription-based amplification of mRNA (STAR) method and its application in search of rare and differentially expressed genes in AD brains. **BMC Genomics** 2006; 7:286.

Lockey C, Otto E, Long Z. Real-time fluorescence detection of a single DNA molecule. **Biotechniques** 1998; 24:744-6.

Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. **PNAS** 2003; 100:5974-79.

Madden SL, Wang CJ, Landes G. Serial analysis of gene expression: from gene discovery to target identification. **DDT Rev** 2000; 5:415-25.

Man YG, Sang QXA. The significance of focal myoepithelial cell layer disruptions in human breast tumor invasion: a paradigm shift from the “protease-centered” hypothesis. **Exp Cell Res** 2004; 301:103-18.

Man YG, Fu SW, Schwartz A, Pinzone JJ, Simmens SJ, Berg PE. Expression of BP1, a novel homeobox gene, correlates with breast cancer progression and invasion. **Breast Cancer Res Treat** 2005; 90:241-7.

Matz M, Shagin D, Bogdanova E et al. Amplification of cDNA ends based on template-switching effect and step-out PCR. **Nucleic Acids Res** 1999; 27:1558-60.

Maxhimer JB, Pesce CE, Stewart RA, Gattuso P, Prinz RA, Xu X. Ductal carcinoma in situ of the breast and heparanase-1 expression: a molecular explanation for more aggressive subtypes. **J Am Coll Surg** 2005; 200:328-35.

Ministério da Saúde. Secretaria de Assistência à Saúde. Instituto Nacional do Câncer. **Estimativa/2006: incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2005.

Morrison TB, Wei JJ, Wittwer CT. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. **Biotechniques** 1998; 24:954-8, 960, 962.

Namba R, Maglione JE, Davis RR, et al. Heterogeneity of mammary lesions represent molecular differences. **BMC Cancer** 2006; 6:275.

Narod SA, Foulkes WD. *BRCA1* and *BRCA2*: 1994 and beyond. **Nature Rev Cancer** 2004; 4:665-76.

Nelson WJ, Nusse R. Convergence of Wnt, β -catenin, and cadherin pathways. **Science**. 2004; 303:1483-7.

O'Hare MJ, Ormerod MG, Monaghan P, Lane EB, Gusterson BA. Characterization in vitro of luminal and myoepithelial cells isolated from the human mammary gland by cell sorting. **Differentiation** 1991; 46:209-21.

Panetti TS, Magnusson MK, Peyruchaud O, et al. Modulation of cell interactions with extracellular matrix by lysophosphatidic acid and sphingosine 1-phosphate. **Prostaglandins** 2001; 64:93-106.

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. **CA Cancer J Clin** 2005; 55:74-108.

Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. **Nucleic Acids Res** 2001; 29:e45

Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. **Nat Genet** 1998; 20:207-11.

Porter D, Lahti-Domenici J, Keshaviah A, et al. Molecular markers in ductal carcinoma *in situ* of the breast. **Mol Cancer Res** 2003; 1:362-75.

Potemski P, Pluciennik E, Bednarek AK, et al. A comparative assessment of HER2 status in operable breast cancer by real-time RT-PCR and by immunohistochemistry. **Med Sci Monit** 2006;12:MT57-61.

Puputti M, Sihto H, Isola J, Butzow R, Joensuu H, Nupponen NN. Allelic imbalance of HER2 variant in sporadic breast and ovarian cancer. **Cancer Genet Cytogenet** 2006; 167:32-8.

Radisky D, Hagios C, Bissel MJ. Tumors are unique organs defined by abnormal signaling and context. **Semin Cancer Biol** 2001; 11:87-95.

Rai D, Frolova A, Frasor J, Carpenter AE, Katzenellenbogen BS. Distinctive Actions of Membrane Targeted versus Nuclear Localized Estrogen Receptors in Breast Cancer Cells. **Mol Endocrinol** 2005; 19:1606-17.

Saraiva TF. **Estabelecimento de protocolo de amplificação de RNA a partir de amostras tumorais microdissecadas a laser**. Botucatu; 2005. [Monografia-Universidade Estadual Paulista Júlio de Mesquita Filho].

Sanguinetti CJ, Dias N, Simpson AJG. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. **Biotechniques** 1994 17:3-6.

Saha S, Sparks AB, Rago C, et al. Using the transcriptome to annotate the genome. **Nat Biotechnol** 2002; 20:508-12.

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

Schmittgen TD, Zakrajsek BA. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. **J Biochem Biophys Methods** 2000; 46:69-81.

Schuetz CS, Bonin M, Clare SE. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. **Cancer Res** 2006; 66:5278-86.

Shimo T, Kubota S, Kondo S, et al. Connective tissue growth factor as a major angiogenic agent that is induced by hypoxia in a human breast cancer cell line. **Cancer Lett** 2001; 174:57-64.

Simão TA, Ribeiro FS, Amorim LM, et al. TP53 mutations in breast cancer tumors of patients from Rio de Janeiro, Brazil: association with risk factors and tumor characteristics. **Int J Cancer** 2002; 101:69-73.

Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR Jr, Elkahloun AG. In vivo gene expression profile analysis of human breast cancer progression. **Cancer Res** 1999; 59:5656-61.

Sontag L, Axelrod DE. Evaluation of pathways for progression of heterogeneous breast tumors. **J Theor Biol** 2005; 232:179-89.

Stathopoulou A, Ntoulia M, Perraki M, et al. A highly specific real-time RT-PCR method for the quantitative determination of CK-19 mRNA positive cells in peripheral blood of patients with operable breast cancer. **Int J Cancer** 2006; 119:1654-9.

Sumida T, Itahana Y, Hamakawa H, Desprez PY. Reduction of human metastatic breast cancer cell aggressiveness on introduction of either form a or b of the progesterone receptor and then treatment. **Cancer Res** 2004; 64:7886-92.

Szabo A, Perou CM, Karaca M, Perreard L, Quackenbush JF, Bernard PS. Statistical modeling for selecting housekeeper genes. **Genome Biol** 2004; 5:R59.

Thiery JP. Epithelial-mesenchymal transitions in development and pathologies. **Curr Opin Cell Biol** 2003; 15:740-6.

Thiery JP, Sleeman JP. Complex networks orchestrate epithelial-mesenchymal transitions. **Nat Rev Mol Cell Biol** 2006; 7:131-42.

Timms JF, White SL, O'Hare MJ, Waterfield MD. Effects of ErbB-2 overexpression on mitogenic signalling and cell cycle progression in human breast luminal epithelial cells. **Oncogene** 2002; 21:6573-86.

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

Van't Veer LJ, Dai H, Van de Viver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. **Nature** 2002; 415: 484-5.

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

Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM. High-fidelity mRNA amplification for gene profiling. **Nat Biotechnol** 2000; 18:457-9.

Weigelt B, Glass AM, Wessels LFA, Witteveen AT, Peterse JL, van't Veer L. Gene expression profiles of primary breast tumors maintained in distant metastases. **PNAS** 2003; 26:15901-5.

Werner M, Mattis A, Aubele M, et al. 20q13.2 amplification in intraductal hyperplasia adjacent to in situ and invasive ductal carcinoma of the breast. **Virchows Arch** 1999; 435:469-72.

Xie D, Nakachi K, Wang H, Elashoff R, Koeffler HP. Elevated levels of connective tissue growth factor, WISP-1, and CYR61 in primary breast cancers associated with more advanced features. **Cancer Res** 2001; 61:8917-23.

Xu J, Lai YL, Lin WC, Lin FT. TRIP6 enhances lysophosphatidic acid-induced cell migration by interacting with lysophosphatidic acid 2 receptor. **J Biol Chem** 2004; 279:10459-68.

Yang J, Mani SA, Donaher JL, et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. **Cell** 2004; 117:927-39.

Ye SQ, Lavoie T, Usher DC, Zhang LQ. Microarray, SAGE and their applications to cardiovascular diseases. **Cell Res** 2002; 12:105-15.

Yi J, Beckerle MC. The human TRIP6 gene encodes a LIM domain protein and maps to chromosome 7q22, a region associated with tumorigenesis. **Genomics** 1998; 49:314-6.

Yook JI, Li XY, Ota I, Fearon ER, Weiss SJ. Wnt-dependent regulation of the E-cadherin repressor snail. **J Biol Chem** 2005; 280:11740-8.

Zahnd C, Pecorari F, Straumann N, Wyler E, Pluckthun A. Selection and characterization of Her2 binding-designed ankyrin repeat proteins. **J Biol Chem** 2006; 281:35167-75.

Zhao H, Hastie T, Whitfield ML, Borresen-Dale AL, Jeffrey SS. Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis. **BMC Genomics** 2002; 3:31.

Zhuang Z, Merino MJ, Chuaqui R, Liotta LA, Emmert-Buck MR. Identical allelic loss on chromosome 11q13 in microdissected in situ and invasive human breast cancer. **Cancer Res** 1995; 55:467-71.

ANEXOS

Anexo 1 - Genes identificados como específicos de cada biblioteca RaSH.

CLONE NAME	GENE SYMBOL	FUNCTION	GENE DESCRIPTION
LGEA-RHID01-080306-013-A02	GNB2	signal transduction	Homo sapiens guanine nucleotide binding protein (G protein), beta polypeptide 2 (GNB2), mRNA
LGEA-RHID01-080306-013-A06	MYEOV2	unknown	tgnl UG Hs#S16885895 Homo sapiens cDNA FLJ44859 fis, clone BRALZ2003330 /gb=AK126809 /gi=34533443 /ug=Hs.293884 /len=1790
LGEA-RHID01-080306-013-B04	DDX21	unknown	Homo sapiens DEAD (Asp-Glu-Ala-Asp) box polypeptide 21 (DDX21), mRNA
LGEA-RHID01-080306-013-B12	RAB22A	protein transport, small GTPase mediated signal transduction	Homo sapiens RAB22A, member RAS oncogene family (RAB22A), mRNA
LGEA-RHID01-080306-013-C01	GNAS	signal transduction	Homo sapiens GNAS complex locus (GNAS), transcript variant 2, mRNA
LGEA-RHID01-080306-013-C03	BG681172	unknown	tgnl UG Hs#S3554536 602629310F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4753874 5', mRNA sequence /clone=IMAGE:4753874 /clone_end=5' /gb=BG681172 /gi=13912569 /ti=44253428 /ug=Hs.581355 /len=847
LGEA-RHID01-080306-013-C11	FUS	biological process unknown	Homo sapiens fusion (involved in t(12;16) in malignant liposarcoma) (FUS), transcript variant 1, mRNA
LGEA-RHID01-080306-013-C12	PSMD12	unknown	Homo sapiens proteasome (prosome, macropain) 26S subunit, non-ATPase, 12 (PSMD12), transcript variant 1, mRNA
LGEA-RHID01-080306-013-D01	CACHD1	unknown	BX098142 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGp99814743 ; IMAGE:327637, mRNA sequence
LGEA-RHID01-080306-013-D03	ZNF148	transcription	Homo sapiens zinc finger protein 148 (pH2-52) (ZNF148), mRNA
LGEA-RHID01-080306-013-D04	AK026101	unknown	tgnl UG Hs#S2653948 Homo sapiens cDNA: FLJ22448 fis, clone HRC09541 /gb=AK026101 /gi=10438842 /ug=Hs.582106 /len=1460
LGEA-RHID01-080306-013-E03	RPS6KA2	protein amino acid phosphorylation	tgnl UG Hs#S11132537 BX109074 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGp998G024024, mRNA sequence /clone=IMAGp998G024024 ; IMAGE:1587481 /gb=BX109074 /gi=27835605 /ug=Hs.135686 /len=621
LGEA-RHID01-080306-013-F12	APC	Wnt receptor signaling pathway, cell adhesion	Homo sapiens adenomatosis polyposis coli (APC), mRNA
LGEA-RHID01-080306-013-G10	VPS24	protein transport	Homo sapiens vacuolar protein sorting 24 (yeast) (VPS24), transcript variant 2, mRNA
LGEA-RHID01-080306-013-H01	PTDSS1	phospholipid biosynthesis	UI-CF-EC1-abn-f-19-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone UI-CF-EC1-abn-f-19-0-UI 3', mRNA sequence
LGEA-RHID01-080306-013-H04	JTB	unknown	Homo sapiens jumping translocation breakpoint (JTB), mRNA
LGEA-RHID01-080306-013-H09	IFI6	immune response	Homo sapiens interferon, alpha-inducible protein (clone IFI-6-16) (G1P3), transcript variant 3, mRNA
LGEA-RHID01-201205-012-A02	HSMPP8	cell cycle	tgnl UG Hs#S18152396 Homo sapiens M-phase phosphoprotein, mpp8 (HSMPP8), mRNA /cds=p(10,2592) /gb=NM_017520 /gi=41055988 /ug=Hs.269654 /len=3234
LGEA-RHID01-201205-012-C01	NDUFB2	mitochondrial electron transport	Homo sapiens NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa (NDUFB2), nuclear gene encoding mitochondrial protein, mRNA
LGEA-RHID01-201205-012-F01	RPS16	protein biosynthesis	Homo sapiens ribosomal protein S16 (RPS16), mRNA
LGEA-RHID01-230306-014-A02	CD656502	unknown	AGENCOURT_14537079 NIA Human H1 Embryonic Stem Cell cDNA Library (Long) Homo sapiens cDNA clone IMAGE:30421083 5', mRNA sequence
LGEA-RHID01-230306-014-A06	FASN	fatty acid biosynthesis	Homo sapiens fatty acid synthase (FASN), mRNA
LGEA-RHID01-230306-014-B03	SMARCE1	regulation of transcription from RNA polymerase II promoter	Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1 (SMARCE1), mRNA
LGEA-RHID01-230306-014-B06	NM_020213	unknown	Homo sapiens poly (ADP-ribose) polymerase family, member 6 (PARP6), transcript variant 1, mRNA

LGEA-RHID01-230306-014-B10	LASS6	unknown	Homo sapiens LAG1 longevity assurance homolog 6 (S. cerevisiae) (LASS6), mRNA
LGEA-RHID01-230306-014-C03	SAPS3	unknown	BX103072 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGp998N234379 ; IMAGE:1723990, mRNA sequence
LGEA-RHID01-230306-014-C05	CANX	protein secretion, protein folding	Homo sapiens calnexin (CANX), transcript variant 1, mRNA
LGEA-RHID01-230306-014-C06	GTF2F2	transcription	Homo sapiens general transcription factor IIF, polypeptide 2 (30kD subunit) (GTF2F2), mRNA
LGEA-RHID01-230306-014-D03	ACOT2	lipid metabolism	Homo sapiens acyl-CoA thioesterase 2 (ACOT2), mRNA
LGEA-RHID01-230306-014-D07	KIAA1598	unknown	MR3-NN0220-281000-004-a02 NN0220 Homo sapiens cDNA, mRNA sequence
LGEA-RHID01-230306-014-D09	ARID4B	unknown	Homo sapiens AT rich interactive domain 4B (RBP1- like) (ARID4B), transcript variant 2, mRNA
LGEA-RHID01-230306-014-D10	C1orf27	unknown	Homo sapiens chromosome 1 open reading frame 27 (C1orf27), mRNA
LGEA-RHID01-230306-014-E04	IMMP2L	proteolysis	to77d02.x1 NCL_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2184291 3', mRNA sequence
LGEA-RHID01-230306-014-E07	PRKD2	intracellular signaling cascade	Homo sapiens protein kinase D2 (PRKD2), mRNA
LGEA-RHID01-230306-014-F01	NM_001018022	unknown	Homo sapiens similar to FKSG62 (LOC389286), mRNA
LGEA-RHID01-230306-014-F03	SLC12A2	ion transport, amino acid transport	Homo sapiens solute carrier family 12 (sodium/potassium/chloride transporters), member 2 (SLC12A2), mRNA
LGEA-RHID01-230306-014-F12	XM_931731	unknown	PREDICTED: Homo sapiens similar to zinc finger and BTB domain containing 8, transcript variant 3 (LOC653121), mRNA
LGEA-RHID01-230306-014-G07	PCBP1	mRNA metabolism	Homo sapiens poly(rC) binding protein 1 (PCBP1), mRNA
LGEA-RHID01-230306-014-G11	CSNK1E	DNA repair	Homo sapiens casein kinase 1, epsilon (CSNK1E), transcript variant 2, mRNA
LGEA-RHID01-230306-014-H01	AI925925	unknown	wh12b12.x1 NCL_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2380511 3', mRNA sequence
LGEA-RHID01-230306-014-H02	PPGB	proteolysis regulation of transcription	Homo sapiens protective protein for beta-galactosidase (galactosialidosis) (PPGB), mRNA
LGEA-RHID01-230306-014-H03	PEG3		Homo sapiens paternally expressed 3 (PEG3), mRNA
LGEA-RHID01-230306-014-H11	USP4	ubiquitin-dependent protein catabolism	Homo sapiens ubiquitin specific peptidase 4 (proto-oncogene) (USP4), transcript variant 2, mRNA
LGEA-RHID01-270606-016-A01	HAGH	unknown	Homo sapiens fumarylacetoacetate hydrolase domain containing 1 (FAHD1), transcript variant 2, mRNA
LGEA-RHID01-270606-016-A04	WDR76	unknown	Homo sapiens WD repeat domain 76 (WDR76), mRNA
LGEA-RHID01-270606-016-B02	SF3B14	nuclear mRNA splicing, via spliceosome	Homo sapiens splicing factor 3B, 14 kDa subunit (SF3B14), mRNA
LGEA-RHID01-270606-016-C04	IPO7	protein import into nucleus	tgnl UG Hs#S4806911 Homo sapiens cDNA FLJ90427 fis, clone NT2RP3000481, highly similar to Homo sapiens RanBP7/importin 7 mRNA /gb=AK074908 /gi=22760659 /ug=Hs.523470 /len=3781
LGEA-RHID01-270606-016-C05	NFAT5	regulation of transcription	Homo sapiens nuclear factor of activated T-cells 5, tonicity-responsive (NFAT5), transcript variant 3, mRNA
LGEA-RHID01-270606-016-C11	PRKCBP1	regulation of transcription, DNA-dependent	Homo sapiens protein kinase C binding protein 1 (PRKCBP1), transcript variant 1, mRNA
LGEA-RHID01-270606-016-D02	CSTF1	mRNA polyadenylation, mRNA cleavage	602441968F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4557737 5', mRNA sequence
LGEA-RHID01-270606-016-E03	BQ690425	unknown	tgnl UG Hs#S4594897 AGENCOURT_8176614 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6252524 5', mRNA sequence /clone=IMAGE:6252524 /clone_end=5' /gb=BQ690425 /gi=21815741 /ti=142921322 /ug=Hs.103548 /len=1108
LGEA-RHID01-270606-016-E08	CAPN1	positive regulation of cell proliferation	Homo sapiens chromosome 8 open reading frame 76 (C8orf76), mRNA

LGEA-RHID01-270606-016-E10	F2	platelet activation, regulation of progression through cell cycle	Homo sapiens KIAA0652 (KIAA0652), mRNA
LGEA-RHID01-270606-016-F04	XR_001013	unknown	PREDICTED: Homo sapiens hypothetical LOC541471 (LOC541471), misc RNA
LGEA-RHID01-270606-016-F11	WDR61	unknown	Homo sapiens WD repeat domain 61 (WDR61), mRNA
LGEA-RHID01-270606-016-F12	SRPK1	regulation of mRNA processing	Homo sapiens SFRS protein kinase 1 (SRPK1), mRNA
LGEA-RHID01-270606-016-G10	YARS	apoptosis, cell motility	Homo sapiens tyrosyl-tRNA synthetase (YARS), mRNA
LGEA-RHID01-270606-016-H07	EIF3S5	protein biosynthesis	Homo sapiens eukaryotic translation initiation factor 3, subunit 5 epsilon, 47kDa (EIF3S5), mRNA
LGEA-RHID01-300306-015-A03	ERBB4	cell proliferation	MCF7RNAL19110TR Human MCF7 breast cancer cell line near full length normalized library (MCF7_EST) Homo sapiens cDNA clone MCF7_RNA_L_19_110, mRNA sequence
LGEA-RHID01-300306-015-A04	SAR1A	intracellular protein transport	Homo sapiens SAR1 gene homolog A (S. cerevisiae) (SAR1A), mRNA
LGEA-RHID01-300306-015-A06	PHB2	regulation of transcription	Homo sapiens prohibitin 2 (PHB2), mRNA
LGEA-RHID01-300306-015-A07	CA775602	unknown	io87h03.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6133492 5', mRNA sequence
LGEA-RHID01-300306-015-B01	CTSH	proteolysis	Homo sapiens cathepsin H (CTSH), transcript variant 2, mRNA
LGEA-RHID01-300306-015-B06	GPRC5C	signal transduction	Homo sapiens G protein-coupled receptor, family C, group 5, member C (GPRC5C), transcript variant 1, mRNA
LGEA-RHID01-300306-015-B08	TBN	unknown	tgnl UG Hs#S4621990 Homo sapiens mRNA; cDNA DKFZp313E0720 (from clone DKFZp313E0720) /gb=AL833401 /gi=21734038 /ug=Hs.556116 /len=2989
LGEA-RHID01-300306-015-C10	MTL5	response to oxidative stress, anti-apoptosis	Homo sapiens metallothionein-like 5, testis-specific (tesmin) (MTL5), transcript variant 1, mRNA
LGEA-RHID01-300306-015-E05	DA077127	unknown	tgnl UG Hs#S28317635 DA077127 BRACE2 Homo sapiens cDNA clone BRACE2029397 5', mRNA sequence /clone=BRACE2029397 /clone_end=5' /gb=DA077127 /gi=80510716 /ug=Hs.545288 /len=559
LGEA-RHID01-300306-015-F09	PRDM4	cell proliferation, regulation of transcription, signal transduction	Homo sapiens PR domain containing 4 (PRDM4), mRNA
LGEA-RHID01-300306-015-H07	NKTR	protein folding	Homo sapiens natural killer-tumor recognition sequence (NKTR), transcript variant 2, mRNA
LGEA-RHID01-300306-015-H10	HIP1	endocytosis	Homo sapiens huntingtin interacting protein 1 (HIP1), mRNA
LGEA-RHSD01-060306-022-A11	RTN4	regulation of apoptosis	Homo sapiens reticulon 4 (RTN4), transcript variant 1, mRNA
LGEA-RHSD01-060306-022-A08	HNRPAB	mRNA editing	Homo sapiens heterogeneous nuclear ribonucleoprotein A/B (HNRPAB), transcript variant 2, mRNA
LGEA-RHSD01-060306-022-A09	CB242937	unknown	UI-CF-FN0-agb-f-13-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone UI-CF-FN0-agb-f-13-0-UI 3', mRNA sequence
LGEA-RHSD01-060306-022-B05	BRD7	unknown	Homo sapiens bromodomain containing 7 (BRD7), mRNA
LGEA-RHSD01-060306-022-B06	AF448858	unknown	tgnl UG Hs#S16507502 Homo sapiens hypothetical short protein mRNA, complete cds /cds=p(229,432) /gb=AF448858 /gi=33187757 /ug=Hs.411490 /len=2862
LGEA-RHSD01-060306-022-B08	CN484301	unknown	tgnl UG Hs#S19724297 hw45c10.y2 Human primary human ocular pericytes. Unamplified (hw) Homo sapiens cDNA clone hw45c10 5', mRNA sequence /clone=hw45c10 /clone_end=5' /gb=CN484301 /gi=46565805 /ug=Hs.268803 /len=618
LGEA-RHSD01-060306-022-B11	SFRP1	cell differentiation, Wnt receptor signaling pathway	Homo sapiens secreted frizzled-related protein 1 (SFRP1), mRNA
LGEA-RHSD01-060306-022-C06	AA207202	unknown	tgnl UG Hs#S574033 zq84g02.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone IMAGE:648338 3', mRNA sequence /clone=IMAGE:648338 /clone_end=3' /gb=AA207202 /gi=1802712 /ug=Hs.571775 /len=626
LGEA-RHSD01-060306-022-C10	PLP2	chemotaxis, cytokine and chemokine mediated signaling pathway	Homo sapiens proteolipid protein 2 (colonic epithelium- enriched) (PLP2), mRNA

LGEA-RHSD01-060306-022-D01	CAT	response to oxidative stress	Homo sapiens catalase (CAT), mRNA
LGEA-RHSD01-060306-022-D02	BC033025	unknown	tgn UG Hs#S4554472 Homo sapiens phosphatidic acid phosphatase type 2 domain containing 1B, mRNA (cDNA clone MGC:32924 IMAGE:5267610), complete cds /cds=p(205,876) /gb=BC033025 /gi=21542540 /ug=Hs.567619 /len=2204
LGEA-RHSD01-060306-022-E01	SPEN	regulation of transcription, DNA-dependent	Homo sapiens spen homolog, transcriptional regulator (Drosophila) (SPEN), mRNA
LGEA-RHSD01-060306-022-E02	PIGG	GPI anchor biosynthesis	Homo sapiens phosphatidylinositol glycan, class G (PIGG), mRNA
LGEA-RHSD01-060306-022-E03	RPS15A	protein biosynthesis	Homo sapiens ribosomal protein S15a (RPS15A), transcript variant 2, mRNA
LGEA-RHSD01-060306-022-E04	RPL34	protein biosynthesis	Homo sapiens ribosomal protein L34 (RPL34), transcript variant 2, mRNA
LGEA-RHSD01-060306-022-E05	PCID1	unknown	Homo sapiens dendritic cell protein (hfl-B5), mRNA
LGEA-RHSD01-060306-022-E08	XM_936529	unknown	PREDICTED: Homo sapiens similar to phosphodiesterase 4D interacting protein isoform 1 (LOC647466), mRNA
LGEA-RHSD01-060306-022-E09	METTL3	nucleotide and nucleic acid metabolism, RNA methylation	Homo sapiens methyltransferase like 3 (METTL3), mRNA
LGEA-RHSD01-060306-022-F05	TATDN1	unknown	Homo sapiens TatD DNase domain containing 1 (TATDN1), mRNA
LGEA-RHSD01-060306-022-F07	BI224335	unknown	602940661F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5103831 5', mRNA sequence
LGEA-RHSD01-060306-022-F09	AW088654	unknown	xd10g01.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2593392 3', mRNA sequence
LGEA-RHSD01-060306-022-G09	PCSK1	metabolism, cell-cell signaling	Homo sapiens proprotein convertase subtilisin/kexin type 1 (PCSK1), mRNA
LGEA-RHSD01-060306-022-G10	C2orf25	unknown	Homo sapiens chromosome 2 open reading frame 25 (C2orf25), mRNA
LGEA-RHSD01-060306-022-H05	HLA-B	antigen processing	Homo sapiens major histocompatibility complex, class I, B (HLA-B), mRNA
LGEA-RHSD01-060306-022-H11	WDR33	spermatogenesis	Homo sapiens WD repeat domain 33 (WDR33), transcript variant 2, mRNA
LGEA-RHSD01-180406-026-A05	BBS4	unknown	Homo sapiens Bardet-Biedl syndrome 4 (BBS4), mRNA
LGEA-RHSD01-180406-026-A08	MATN3	extracellular matrix organization and biogenesis	Homo sapiens matrilin 3 (MATN3), mRNA
LGEA-RHSD01-180406-026-A10	BRWD1	biological process unknown	Homo sapiens bromodomain and WD repeat domain containing 1 (BRWD1), transcript variant 2, mRNA
LGEA-RHSD01-180406-026-B06	TCP1	tubulin folding, protein folding	Homo sapiens t-complex 1 (TCP1), transcript variant 2, mRNA
LGEA-RHSD01-180406-026-C01	DA646975	unknown	DA646975 MAMMA1 Homo sapiens cDNA clone MAMMA1002224 5', mRNA sequence
LGEA-RHSD01-180406-026-C06	BE122677	unknown	01fh94 Human Epidermal Keratinocyte Subtraction Library-Upregulated Transcripts Homo sapiens cDNA clone 01fh94 5' similar to Mus musculus L1 repeat, Tf subfamily, member 9 (L1Md-Tf9), mRNA sequence
LGEA-RHSD01-180406-026-C09	IGF1	positive regulation of cell proliferation, cell motility	Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1), mRNA
LGEA-RHSD01-180406-026-C10	KIAA0776	unknown	Homo sapiens KIAA0776 (KIAA0776), mRNA
LGEA-RHSD01-180406-026-D01	UQCR	electron transport	Homo sapiens ubiquinol-cytochrome c reductase, 6.4kDa subunit (UQCR), mRNA
LGEA-RHSD01-180406-026-D03	BU839374	unknown	tgn UG Hs#S5935804 AGENCOURT_8184241 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6263776 5', mRNA sequence /clone=IMAGE:6263776 /clone_end=5' /gb=BU839374 /gi=24023769 /ti=147858777 /ug=Hs.572133 /len=1255
LGEA-RHSD01-180406-026-D04	TANK	signal transduction	Homo sapiens TRAF family member-associated NFKB activator (TANK), transcript variant 1, mRNA

LGEA-RHSD01-180406-026-D10	SMOC2	unknown	Homo sapiens SPARC related modular calcium binding 2 (SMOC2), mRNA
LGEA-RHSD01-180406-026-D12	PPP4R1	signal transduction	Homo sapiens protein phosphatase 4, regulatory subunit 1 (PPP4R1), mRNA
LGEA-RHSD01-180406-026-E05	FEN1	DNA replication	Homo sapiens flap structure-specific endonuclease 1 (FEN1), mRNA
LGEA-RHSD01-180406-026-E08	ITGB3BP	cell adhesion, signal transduction	Homo sapiens integrin beta 3 binding protein (beta3-endonexin) (ITGB3BP), mRNA
LGEA-RHSD01-180406-026-F01	SLC39A8	unknown	Homo sapiens solute carrier family 39 (zinc transporter), member 8 (SLC39A8), mRNA
LGEA-RHSD01-180406-026-F07	DCAKD	unknown	Homo sapiens dephospho-CoA kinase domain containing (DCAKD), mRNA
LGEA-RHSD01-180406-026-F10	C1S	innate immune response	Homo sapiens complement component 1, s subcomponent (C1S), transcript variant 2, mRNA
LGEA-RHSD01-180406-026-F11	XM_936381	unknown	PREDICTED: Homo sapiens hypothetical protein LOC642228 (LOC642228), mRNA
LGEA-RHSD01-180406-026-G09	SEPHS2	selenocysteine biosynthesis	Homo sapiens selenophosphate synthetase 2 (SEPHS2), mRNA
LGEA-RHSD01-180406-026-G10	CR736439	unknown	CR736439 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE998A093975 ; IMAGE:1568528 3', mRNA sequence
LGEA-RHSD01-180406-026-H10	BF968136	unknown	602269123F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357445 5', mRNA sequence
LGEA-RHSD01-180406-026-H11	AA358397	unknown	EST67265 Fetal lung III Homo sapiens cDNA 3' end, mRNA sequence
LGEA-RHSD01-180406-026-H12	APOBEC3B	unknown	Homo sapiens apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B (APOBEC3B), mRNA
LGEA-RHSD01-230306-023-A05	XM_944191	unknown	PREDICTED: Homo sapiens similar to ribosomal protein L13a, transcript variant 5 (LOC648326), mRNA
LGEA-RHSD01-230306-023-B02	RNF103	central nervous system development	Homo sapiens ring finger protein 103 (RNF103), mRNA
LGEA-RHSD01-230306-023-B07	FAM83D	tRNA aminoacylation for protein translation	Homo sapiens family with sequence similarity 83, member D (FAM83D), mRNA
LGEA-RHSD01-230306-023-D07	ACTB	unknown	Homo sapiens actin, beta (ACTB), mRNA
LGEA-RHSD01-230306-023-D08	BTAF1	negative regulation of transcription	Homo sapiens BTAF1 RNA polymerase II, B-TFIID transcription factor-associated, 170kDa (Mot1 homolog, S. cerevisiae) (BTAF1), mRNA
LGEA-RHSD01-230306-023-D09	LOC645737	unknown	PREDICTED: Homo sapiens similar to 60S ribosomal protein L7, transcript variant 1 (LOC645737), mRNA
LGEA-RHSD01-230306-023-D12	CA415848	unknown	UI-H-FE0-bbo-k-04-0-UI.s1 NCI_CGAP_FE0 Homo sapiens cDNA clone UI-H-FE0-bbo-k-04-0-UI 3', mRNA sequence
LGEA-RHSD01-230306-023-E08	TRIP12	protein ubiquitination	Homo sapiens thyroid hormone receptor interactor 12 (TRIP12), mRNA
LGEA-RHSD01-230306-023-G12	SAA1	regulation of protein secretion, positive regulation of cell adhesion	Homo sapiens serum amyloid A1 (SAA1), transcript variant 1, mRNA
LGEA-RHSD01-230306-023-H03	DNAJA1	protein folding	Homo sapiens DnaJ (Hsp40) homolog, subfamily A, member 1 (DNAJA1), mRNA
LGEA-RHSD01-230306-023-H04	SEC63	protein targeting to membrane	Homo sapiens SEC63-like (S. cerevisiae) (SEC63), mRNA
LGEA-RHSD01-230306-023-H10	XR_000921	unknown	PREDICTED: Homo sapiens similar to ribosomal protein L13a; 60S ribosomal protein L13a; 23 kD highly basic protein, transcript variant 1 (LOC387930), misc RNA
LGEA-RHSD01-230306-023-H12	OGFRL1	unknown	Homo sapiens opioid growth factor receptor-like 1 (OGFRL1), mRNA
LGEA-RHSD01-310306-024-B03	ANXA1	anti-apoptosis, cell motility	Homo sapiens annexin A1 (ANXA1), mRNA
LGEA-RHSD01-310306-024-B06	ARFGEF1	exocytosis	Homo sapiens ADP-ribosylation factor guanine nucleotide-exchange factor 1(brefeldin A-inhibited) (ARFGEF1), mRNA
LGEA-RHSD01-310306-024-C07	PRDX4	I-kappaB phosphorylation	Homo sapiens peroxiredoxin 4 (PRDX4), mRNA
LGEA-RHSD01-310306-024-C08	OSBPL8	steroid metabolism, lipid transport	Homo sapiens oxysterol binding protein-like 8 (OSBPL8), transcript variant 2, mRNA

LGEA-RHSD01-310306-024-D05	PPIA	protein folding	Homo sapiens peptidylprolyl isomerase A (cyclophilin A) (PPIA), transcript variant 2, mRNA
LGEA-RHSD01-310306-024-D12	BMPR2	transmembrane receptor protein serine/threonine kinase signaling pathway	tgnt UG Hs#S1728677 Homo sapiens bone morphogenetic protein receptor, type II (serine/threonine kinase) (BMPR2), mRNA /cds=p(528,3644) /gb=NM_001204 /gi=72376969 /ug=Hs.471119 /len=11449
LGEA-RHSD01-310306-024-E12	BU676595	unknown	UI-CF-DU1-aak-b-08-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone UI-CF-DU1-aak-b-08-0-UI 3', mRNA sequence
LGEA-RHSD01-310306-024-F05	SYF2	biological process unknown	Homo sapiens SYF2 homolog, RNA splicing factor (S. cerevisiae) (SYF2), transcript variant 2, mRNA
LGEA-RHSD01-310306-024-F09	RPL23	protein biosynthesis	Homo sapiens ribosomal protein L23 (RPL23), mRNA
LGEA-RHSD01-310306-024-F10	XM_934508	unknown	PREDICTED: Homo sapiens family with sequence similarity 91, member A2, transcript variant 6 (FAM91A2), mRNA
LGEA-RHSD01-310306-024-G02	CCND1	regulation of progression through cell cycle	Homo sapiens cyclin D1 (CCND1), mRNA
LGEA-RHSD01-310306-024-G11	PPP1CA	cell cycle	Homo sapiens protein phosphatase 1, catalytic subunit, alpha isoform (PPP1CA), transcript variant 3, mRNA
LGEA-RHSD01-310306-024-H02	RPL32	protein biosynthesis	Homo sapiens ribosomal protein L32 (RPL32), transcript variant 3, mRNA
LGEA-RHSD01-310306-025-A09	EPRS	protein biosynthesis	Homo sapiens glutamyl-prolyl-tRNA synthetase (EPRS), mRNA
LGEA-RHSD01-310306-025-A10	SDHB	electron transport	Homo sapiens succinate dehydrogenase complex, subunit B, iron sulfur (lp) (SDHB), mRNA
LGEA-RHSD01-310306-025-A12	C20orf108	unknown	Homo sapiens chromosome 20 open reading frame 108 (C20orf108), mRNA
LGEA-RHSD01-310306-025-B01	HSP90AA1	protein folding	Homo sapiens heat shock protein 90kDa alpha (cytosolic), class A member 1 (HSP90AA1), transcript variant 2, mRNA
LGEA-RHSD01-310306-025-C02	DST	integrin-mediated signaling pathway, cell adhesion, actin cytoskeleton organization and biogenesis	Homo sapiens dystonin (DST), transcript variant 1e, mRNA
LGEA-RHSD01-310306-025-C04	XRN2	microtubule-based process, DNA repair	Homo sapiens 5'-3' exoribonuclease 2 (XRN2), mRNA
LGEA-RHSD01-310306-025-C12	GNB2L1	signal transduction	Homo sapiens guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 (GNB2L1), mRNA
LGEA-RHSD01-310306-025-D01	DA405060	unknown	DA405060 BRTHA3 Homo sapiens cDNA clone BRTHA3009813 5', mRNA sequence
LGEA-RHSD01-310306-025-E10	NQO1	electron transport	Homo sapiens NAD(P)H dehydrogenase, quinone 1 (NQO1), transcript variant 3, mRNA
LGEA-RHSD01-310306-025-G05	CHI3L1	carbohydrate metabolism	Homo sapiens chitinase 3-like 1 (cartilage glycoprotein-39) (CHI3L1), mRNA
LGEA-RHSD01-310306-025-H02	AOF2	regulation of transcription, DNA-dependent	Homo sapiens amine oxidase (flavin containing) domain 2 (AOF2), transcript variant 2, mRNA
LGEA-RHSD01-310306-025-H06	RAB6A	protein transport	Homo sapiens RAB6A, member RAS oncogene family (RAB6A), transcript variant 2, mRNA
LGEA-RHID02-070406-014-A05	SFRS11	RNA splicing	Homo sapiens splicing factor, arginine/serine-rich 11 (SFRS11), mRNA
LGEA-RHID02-070406-014-A07	PITPNA	transport	Homo sapiens phosphatidylinositol transfer protein, alpha (PITPNA), mRNA
LGEA-RHID02-070406-014-A12	POMP	unknown	Homo sapiens chromosome 13 open reading frame 12 (C13orf12), mRNA
LGEA-RHID02-070406-014-B01	KIAA1324	unknown	Homo sapiens KIAA1324 (KIAA1324), mRNA
LGEA-RHID02-070406-014-B05	EIF1	protein biosynthesis	Homo sapiens eukaryotic translation initiation factor 1 (EIF1), mRNA
LGEA-RHID02-070406-014-B06	AK023481	unknown	tgnt UG Hs#S2651528 Homo sapiens cDNA FLJ13419 fis, clone PLACE1002115 /gb=AK023481 /gi=10435427 /ug=Hs.421621 /len=2290
LGEA-RHID02-070406-014-B09	BC053563	unknown	tgnt UG Hs#S15970791 Homo sapiens microRNA 21, mRNA (cDNA clone IMAGE:6190064) /gb=BC053563 /gi=31565528 /ug=Hs.444569 /len=3404

LGEA-RHID02-070406-014-C01	SAT	unknown	Homo sapiens spermidine/spermine N1-acetyltransferase (SAT), mRNA
LGEA-RHID02-070406-014-C02	HLA-DRA	immune response	Homo sapiens major histocompatibility complex, class II, DR alpha (HLA-DRA), mRNA
LGEA-RHID02-070406-014-C04	SPCS2	signal peptide processing	Homo sapiens signal peptidase complex subunit 2 homolog (S. cerevisiae) (SPCS2), mRNA
LGEA-RHID02-070406-014-C11	VEGF	cell proliferation, angiogenesis, cell migration, signal transduction	Homo sapiens vascular endothelial growth factor (VEGF), transcript variant 4, mRNA
LGEA-RHID02-070406-014-D07	AU136967	unknown	AU136967 PLACE1 Homo sapiens cDNA clone PLACE1005453 5', mRNA sequence
LGEA-RHID02-070406-014-D12	HLA-DOA	immune response	Homo sapiens major histocompatibility complex, class II, DO alpha (HLA-DOA), mRNA
LGEA-RHID02-070406-014-E04	ZNF317	regulation of transcription, DNA-dependent	Homo sapiens zinc finger protein 317 (ZNF317), mRNA
LGEA-RHID02-070406-014-F04	GRSF1	mRNA polyadenylation	Homo sapiens G-rich RNA sequence binding factor 1 (GRSF1), mRNA
LGEA-RHID02-070406-014-F06	GSTM3	metabolism, establishment of blood-nerve barrier	Homo sapiens glutathione S-transferase M3 (brain) (GSTM3), mRNA
LGEA-RHID02-070406-014-F10	FTH1	immune response, cell proliferation	Homo sapiens ferritin, heavy polypeptide 1 (FTH1), mRNA
LGEA-RHID02-070406-014-G06	RPS2	protein biosynthesis	Homo sapiens ribosomal protein S2 (RPS2), mRNA
LGEA-RHID02-070406-014-G07	RBBP8	unknown	Homo sapiens retinoblastoma binding protein 8 (RBBP8), transcript variant 2, mRNA
LGEA-RHID02-070406-014-G11	FST	negative regulation of follicle-stimulating hormone secretion	Homo sapiens follistatin (FST), transcript variant FST344, mRNA
LGEA-RHID02-070406-014-H05	CD59	immune response, blood coagulation	Homo sapiens CD59 molecule, complement regulatory protein (CD59), transcript variant 1, mRNA
LGEA-RHID02-070406-014-H08	NAP1L1	positive regulation of cell proliferation	Homo sapiens nucleosome assembly protein 1-like 1 (NAP1L1), transcript variant 2, mRNA
LGEA-RHID02-070406-014-H12	POSTN	skeletal development, cell adhesion	Homo sapiens periostin, osteoblast specific factor (POSTN), mRNA
LGEA-RHID02-090506-017-A09	VPS13A	protein localization, Golgi to endosome transport	Homo sapiens vacuolar protein sorting 13A (yeast) (VPS13A), transcript variant C, mRNA
LGEA-RHID02-090506-017-B07	SNF1LK	regulation of progression through mitotic cell cycle	Homo sapiens SNF1-like kinase (SNF1LK), mRNA
LGEA-RHID02-090506-017-C01	AB046792	unknown	tgnl UG Hs#S2649414 Homo sapiens mRNA for KIAA1572 protein, partial cds /cds=p(1091,2569) /gb=AB046792 /gi=10047208 /ug=Hs.5638 /len=5609
LGEA-RHID02-090506-017-C03	TNFAIP3	apoptosis	Homo sapiens tumor necrosis factor, alpha-induced protein 3 (TNFAIP3), mRNA
LGEA-RHID02-090506-017-C04	KIAA0090	unknown	Homo sapiens KIAA0090 (KIAA0090), mRNA
LGEA-RHID02-090506-017-C06	COPB2	ER to Golgi vesicle-mediated transport, Golgi to ER intra-Golgi vesicle-mediated transport	Homo sapiens coatamer protein complex, subunit beta 2 (beta prime) (COPB2), mRNA
LGEA-RHID02-090506-017-D06	FCGR2B	immune response, signal transduction	Homo sapiens heat shock 70kDa protein 6 (HSP70B') (HSPA6), mRNA
LGEA-RHID02-090506-017-D07	HSP90AB1	protein folding	Homo sapiens heat shock protein 90kDa alpha (cytosolic), class B member 1 (HSP90AB1), mRNA
LGEA-RHID02-090506-017-D09	EGR2	regulation of transcription	Homo sapiens early growth response 2 (Krox-20 homolog, Drosophila) (EGR2), mRNA
LGEA-RHID02-090506-017-D10	ITPR1	signal transduction	Homo sapiens inositol 1,4,5-triphosphate receptor, type 1 (ITPR1), mRNA
LGEA-RHID02-090506-017-D11	TLOC1	protein transport	Homo sapiens translocation protein 1 (TLOC1), mRNA
LGEA-RHID02-090506-017-E03	PRPF18	RNA splicing	Homo sapiens PRP18 pre-mRNA processing factor 18 homolog (yeast) (PRPF18), mRNA

LGEA-RHID02-090506-017-E06	TMEM98	unknown	Homo sapiens transmembrane protein 98 (TMEM98), transcript variant 2, mRNA
LGEA-RHID02-090506-017-E08	TSN	DNA recombination	Homo sapiens translin (TSN), mRNA
LGEA-RHID02-090506-017-E11	UBE1	DNA replication, ubiquitin cycle	Homo sapiens ubiquitin-activating enzyme E1 (A1S9T and BN75 temperature sensitivity complementing) (UBE1), transcript variant 2, mRNA
LGEA-RHID02-090506-017-F01	PCMTD1	protein modification	Homo sapiens protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 1 (PCMTD1), mRNA
LGEA-RHID02-090506-017-F03	ACAD11	electron transport	Homo sapiens acyl-Coenzyme A dehydrogenase family, member 11 (ACAD11), mRNA
LGEA-RHID02-090506-017-F06	ORC5L	DNA replication	Homo sapiens origin recognition complex, subunit 5-like (yeast) (ORC5L), transcript variant 1, mRNA
LGEA-RHID02-090506-017-F12	CCT8	protein folding	Homo sapiens chaperonin containing TCP1, subunit 8 (theta) (CCT8), mRNA
LGEA-RHID02-090506-017-G01	PCBP2	mRNA metabolism	Homo sapiens poly(rC) binding protein 2 (PCBP2), transcript variant 1, mRNA
LGEA-RHID02-090506-017-G06	EXT2	signal transduction	Homo sapiens exostoses (multiple) 2 (EXT2), transcript variant 2, mRNA
LGEA-RHID02-090506-017-G09	CTGF	cell motility	Homo sapiens connective tissue growth factor (CTGF), mRNA
LGEA-RHID02-090506-017-H08	RAB13	cell adhesion, small GTPase mediated signal transduction	Homo sapiens RAB13, member RAS oncogene family (RAB13), mRNA
LGEA-RHID02-090506-017-H10	DA393184	unknown	DA393184 BRTHA2 Homo sapiens cDNA clone BRTHA2031546 5', mRNA sequence
LGEA-RHID02-130306-011-A01	BTF3	transcription	Homo sapiens basic transcription factor 3 (BTF3), transcript variant 1, mRNA
LGEA-RHID02-130306-011-A06	EIF4A2	protein biosynthesis	Homo sapiens eukaryotic translation initiation factor 4A, isoform 2 (EIF4A2), mRNA
LGEA-RHID02-130306-011-B02	PAIP2	regulation of translation	Homo sapiens poly(A) binding protein interacting protein 2 (PAIP2), transcript variant 2, mRNA
LGEA-RHID02-130306-011-B07	ZNF160	transcription, hemopoiesis	Homo sapiens zinc finger protein 160 (ZNF160), transcript variant 2, mRNA
LGEA-RHID02-130306-011-B09	BM990717	unknown	UI-H-DI0-atr-m-17-0-UI.s1 NCI_CGAP_DI0 Homo sapiens cDNA clone IMAGE:5863480 3', mRNA sequence
LGEA-RHID02-130306-011-C10	CYFIP1	unknown	Homo sapiens cytoplasmic FMR1 interacting protein 1 (CYFIP1), transcript variant 1, mRNA
LGEA-RHID02-130306-011-C12	NM_005463	RNA processing	tgnl UG Hs#S1729670 Homo sapiens heterogeneous nuclear ribonucleoprotein D-like (HNRPDL), transcript variant 1, mRNA /cds=p(581,1843) /gb=NM_005463 /gi=14110410 /ug=Hs.527105 /len=3514
LGEA-RHID02-130306-011-D06	KIAA0446	transport	Homo sapiens KIAA0446 gene product (KIAA0446), mRNA
LGEA-RHID02-130306-011-D07	DDX58	innate immune response	Homo sapiens DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (DDX58), mRNA
LGEA-RHID02-130306-011-D10	RDBP	transcription, biological process unknown	Homo sapiens RD RNA binding protein (RDBP), mRNA
LGEA-RHID02-210306-012-A08	ZNF281	transcription	Homo sapiens zinc finger protein 281 (ZNF281), mRNA
LGEA-RHID02-210306-012-B01	PEX7	protein transport	Homo sapiens peroxisomal biogenesis factor 7 (PEX7), mRNA
LGEA-RHID02-210306-012-B02	CNOT6	transcription	Homo sapiens CCR4-NOT transcription complex, subunit 6 (CNOT6), mRNA
LGEA-RHID02-210306-012-B03	PPIC	signal transduction	Homo sapiens peptidylprolyl isomerase C (cyclophilin C) (PPIC), mRNA
LGEA-RHID02-210306-012-B07	TMEM106B	unknown	Homo sapiens transmembrane protein 106B (TMEM106B), mRNA
LGEA-RHID02-210306-012-C02	AW956036	unknown	EST368106 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence
LGEA-RHID02-210306-012-C09	ATP5E	ATP synthesis coupled proton transport	Homo sapiens ATP synthase, H ⁺ transporting, mitochondrial F1 complex, epsilon subunit (ATP5E), nuclear gene encoding mitochondrial protein, mRNA
LGEA-RHID02-210306-012-D04	BTN3A3	unknown	Homo sapiens butyrophilin, subfamily 3, member A3 (BTN3A3), transcript variant 2, mRNA
LGEA-RHID02-210306-012-D08	RAI17	transcription	Homo sapiens retinoic acid induced 17 (RAI17), mRNA

LGEA-RHID02-210306-012-D10	AV719766	unknown	AV719766 GLC Homo sapiens cDNA clone GLCEAA11 5', mRNA sequence
LGEA-RHID02-210306-012-E02	ZC3H14	unknown	Homo sapiens nuclear protein UKp68 (FLJ11806), transcript variant 4, mRNA
LGEA-RHID02-210306-012-E09	RPS4X	protein biosynthesis	Homo sapiens ribosomal protein S4, X-linked (RPS4X), mRNA
LGEA-RHID02-210306-012-E10	YWHAZ	unknown	Homo sapiens tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (YWHAZ), transcript variant 2, mRNA
LGEA-RHID02-210306-012-E11	CN293284	unknown	17000600111590 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence
LGEA-RHID02-210306-012-E12	PYGL	carbohydrate metabolism	Homo sapiens phosphorylase, glycogen; liver (Hers disease, glycogen storage disease type VI) (PYGL), mRNA
LGEA-RHID02-210306-012-F07	MBNL1	embryonic development (sensu Mammalia)	Homo sapiens muscleblind-like (Drosophila) (MBNL1), transcript variant 6, mRNA
LGEA-RHID02-210306-012-F10	COPS6	unknown	Homo sapiens COP9 constitutive photomorphogenic homolog subunit 6 (Arabidopsis) (COPS6), mRNA
LGEA-RHID02-210306-012-F11	CD46	innate immune response	Homo sapiens CD46 molecule, complement regulatory protein (CD46), transcript variant j, mRNA
LGEA-RHID02-210306-012-F12	NCOR1	regulation of transcription	Homo sapiens nuclear receptor co-repressor 1 (NCOR1), mRNA
LGEA-RHID02-210306-012-G01	GEMIN6	spliceosome assembly	Homo sapiens gem (nuclear organelle) associated protein 6 (GEMIN6), mRNA
LGEA-RHID02-210306-012-G03	LOC201895	unknown	Homo sapiens hypothetical protein LOC201895 (LOC201895), mRNA
LGEA-RHID02-210306-012-G10	ENTPD5	unknown	Homo sapiens ectonucleoside triphosphate diphosphohydrolase 5 (ENTPD5), mRNA
LGEA-RHID02-210306-012-H04	ACLY	metabolism	Homo sapiens ATP citrate lyase (ACLY), transcript variant 1, mRNA
LGEA-RHID02-210306-012-H07	IGFBP7	regulation of cell growth	Homo sapiens insulin-like growth factor binding protein 7 (IGFBP7), mRNA
LGEA-RHID02-210306-012-H10	B3GNT1	poly-N-acetyllactosamine biosynthesis	Homo sapiens UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 6 (B3GNT6), mRNA
LGEA-RHID02-250406-015-A02	NOL10	unknown	Homo sapiens nucleolar protein 10 (NOL10), mRNA
LGEA-RHID02-250406-015-A04	SLC44A1	transport	Homo sapiens solute carrier family 44, member 1 (SLC44A1), transcript variant 2, mRNA
LGEA-RHID02-250406-015-B01	MAL2	unknown	Homo sapiens mal, T-cell differentiation protein 2 (MAL2), mRNA
LGEA-RHID02-250406-015-B02	CILP	nucleotide and nucleic acid metabolism	Homo sapiens cartilage intermediate layer protein, nucleotide pyrophosphohydrolase (CILP), mRNA
LGEA-RHID02-250406-015-B09	TPM1	regulation of muscle contraction, cell motility	Homo sapiens tropomyosin 1 (alpha) (TPM1), transcript variant 7, mRNA
LGEA-RHID02-250406-015-C05	CFB	proteolysis	Homo sapiens complement factor B (CFB), mRNA
LGEA-RHID02-250406-015-C06	LOC653559	unknown	PREDICTED: Homo sapiens similar to Rho-associated protein kinase 1 (Rho-associated, coiled-coil containing protein kinase 1) (p160 ROCK-1) (p160ROCK) (LOC653559), mRNA
LGEA-RHID02-250406-015-D09	BQ303893	unknown	RC1-BT0254-270700-114-g08_1 BT0254 Homo sapiens cDNA, mRNA sequence
LGEA-RHID02-250406-015-D11	MORF4L1	regulation of cell growth	Homo sapiens mortality factor 4 like 1 (MORF4L1), transcript variant 2, mRNA
LGEA-RHID02-250406-015-D12	KIAA0152		Homo sapiens KIAA0152 (KIAA0152), mRNA
LGEA-RHID02-250406-015-E02	FUT8	embryonic development (sensu Mammalia)	Homo sapiens fucosyltransferase 8 (alpha (1,6) fucosyltransferase) (FUT8), transcript variant 4, mRNA
LGEA-RHID02-250406-015-E04	SRP9	protein targeting	Homo sapiens signal recognition particle 9kDa (SRP9), mRNA
LGEA-RHID02-250406-015-E05	SLC1A4	transport	Homo sapiens solute carrier family 1 (glutamate/neutral amino acid transporter), member 4 (SLC1A4), mRNA
LGEA-RHID02-250406-015-E07	AW835509	unknown	QV0-LT0015-180200-127-c09 LT0015 Homo sapiens cDNA, mRNA sequence
LGEA-RHID02-250406-015-E08	RBM6	RNA processing	Homo sapiens RNA binding motif protein 6 (RBM6), mRNA
LGEA-RHID02-250406-015-E09	KBTBD2	unknown	Homo sapiens kelch repeat and BTB (POZ) domain containing 2 (KBTBD2), mRNA
LGEA-RHID02-250406-015-F03	IWS1	IWS1	Homo sapiens IWS1 homolog (S. cerevisiae) (IWS1), mRNA

LGEA-RHID02-250406-015-F06	EIF4G2	cell cycle arrest	Homo sapiens eukaryotic translation initiation factor 4 gamma, 2 (EIF4G2), mRNA
LGEA-RHID02-250406-015-F08	BC094756	unknown	tgn UG Hs#S24639348 Homo sapiens myotubularin related protein 11, mRNA (cDNA clone IMAGE:30349334), complete cds /cds=p(293,1039) /gb=BC094756 /gi=63100308 /ug=Hs.425144 /len=4896
LGEA-RHID02-250406-015-F09	NMI	inflammatory response	Homo sapiens N-myc (and STAT) interactor (NMI), mRNA
LGEA-RHID02-250406-015-G01	TMED10	intracellular protein transport	Homo sapiens transmembrane emp24-like trafficking protein 10 (yeast) (TMED10), mRNA
LGEA-RHID02-250406-015-G06	MLL5	regulation of transcription, DNA-dependent	Homo sapiens myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila) (MLL5), mRNA
LGEA-RHID02-250406-015-G08	SERPINA3	inflammatory response	Homo sapiens serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3 (SERPINA3), mRNA
LGEA-RHID02-250406-015-G10	CLNS1A	transport	Homo sapiens chloride channel, nucleotide-sensitive, 1A (CLNS1A), mRNA
LGEA-RHID02-250406-015-H03	SP1	transcription	Homo sapiens Sp1 transcription factor (SP1), mRNA
LGEA-RHID02-250406-015-H06	ARS2	response to arsenic	Homo sapiens ARS2 protein (ARS2), transcript variant 2, mRNA
LGEA-RHID02-250406-015-H11	TMTC3	Protein binding	Homo sapiens transmembrane and tetratricopeptide repeat containing 3 (TMTC3), mRNA
LGEA-RHID02-260406-016-A06	BAG3	anti-apoptosis	Homo sapiens BCL2-associated athanogene 3 (BAG3), mRNA
LGEA-RHID02-260406-016-A07	C2orf4	unknown	Homo sapiens chromosome 2 open reading frame 4 (C2orf4), mRNA
LGEA-RHID02-260406-016-A09	SH3BGRL	unknown	Homo sapiens SH3 domain binding glutamic acid-rich protein like (SH3BGRL), mRNA
LGEA-RHID02-260406-016-A11	CA12	one-carbon compound metabolism	Homo sapiens carbonic anhydrase XII (CA12), transcript variant 2, mRNA
LGEA-RHID02-260406-016-B01	C3orf62	unknown	Homo sapiens chromosome 3 open reading frame 62 (C3orf62), mRNA
LGEA-RHID02-260406-016-B04	IRXL1	unknown	Homo sapiens iroquois homeobox protein-like 1 (IRXL1), mRNA
LGEA-RHID02-260406-016-B06	ATXN10	unknown	Homo sapiens ataxin 10 (ATXN10), mRNA
LGEA-RHID02-260406-016-B07	COX4I1	electron transport	Homo sapiens cytochrome c oxidase subunit IV isoform 1 (COX4I1), mRNA
LGEA-RHID02-260406-016-B08	TRIP6	focal adhesion formation	Homo sapiens thyroid hormone receptor interactor 6 (TRIP6), mRNA
LGEA-RHID02-260406-016-B11	CDK5RAP3	cell proliferation	Homo sapiens CDK5 regulatory subunit associated protein 3 (CDK5RAP3), transcript variant 1, mRNA
LGEA-RHID02-260406-016-C01	POLS	cell cycle	Homo sapiens polymerase (DNA directed) sigma (POLS), mRNA
LGEA-RHID02-260406-016-C02	PEBP1	unknown	Homo sapiens phosphatidylethanolamine binding protein 1 (PEBP1), mRNA
LGEA-RHID02-260406-016-C04	USP9X	ubiquitin cycle	Homo sapiens ubiquitin specific peptidase 9, X-linked (USP9X), transcript variant 4, mRNA
LGEA-RHID02-260406-016-C07	CD74	signal transduction, cell proliferation	Homo sapiens CD74 molecule, major histocompatibility complex, class II invariant chain (CD74), transcript variant 2, mRNA
LGEA-RHID02-260406-016-D02	STOM	unknown	Homo sapiens stomatin (STOM), transcript variant 2, mRNA
LGEA-RHID02-260406-016-D06	MGC4562	unknown	Homo sapiens hypothetical protein MGC4562 (MGC4562), mRNA
LGEA-RHID02-260406-016-D08	KIAA1961	unknown	Homo sapiens KIAA1961 gene (KIAA1961), transcript variant 2, mRNA
LGEA-RHID02-260406-016-D11	STEAP2	Golgi to plasma membrane transport	Homo sapiens six transmembrane epithelial antigen of the prostate 2 (STEAP2), mRNA
LGEA-RHID02-260406-016-E05	ATF4	transcription	Homo sapiens activating transcription factor 4 (tax-responsive enhancer element B67) (ATF4), transcript variant 2, mRNA
LGEA-RHID02-260406-016-E06	A1569340	unknown	tr79h09.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2224577 3' similar to TR:Q63358 Q63358 MYOSIN HEAVY CHAIN. ;contains Alu repetitive element;contains element MER22 repetitive element ;, mRNA sequence
LGEA-RHID02-260406-016-E07	EEF1A1	protein biosynthesis	Homo sapiens eukaryotic translation elongation factor 1 alpha 1 (EEF1A1), mRNA

LGEA-RHID02-260406-016-E08	U66589	unknown	tgnl UG Hs#S553543 Human ribosomal protein L5 pseudogene mRNA, complete cds /gb=U66589 /gi=1575566 /ug=Hs.180946 /len=3582
LGEA-RHID02-260406-016-E10	PPP1CB	cell cycle	Homo sapiens protein phosphatase 1, catalytic subunit, beta isoform (PPP1CB), transcript variant 3, mRNA
LGEA-RHID02-260406-016-F01	MAP3K1	protein amino acid phosphorylation	PREDICTED: Homo sapiens mitogen-activated protein kinase kinase kinase 1 (MAP3K1), mRNA
LGEA-RHID02-260406-016-F02	HSP90B1	anti-apoptosis	Homo sapiens heat shock protein 90kDa beta (Grp94), member 1 (HSP90B1), mRNA
LGEA-RHID02-260406-016-F04	TAPBP	retrograde vesicle-mediated transport, Golgi to ER	Homo sapiens TAP binding protein (tapasin) (TAPBP), transcript variant 1, mRNA
LGEA-RHID02-260406-016-F07	KIAA1598		Homo sapiens KIAA1598 (KIAA1598), mRNA
LGEA-RHID02-260406-016-F12	SLC40A1	morphogenesis	Homo sapiens solute carrier family 40 (iron-regulated transporter), member 1 (SLC40A1), mRNA
LGEA-RHID02-260406-016-G12	AB209172	unknown	tgnl UG Hs#S24303232 Homo sapiens mRNA for solute carrier family 6 (neurotransmitter transporter, taurine), member 6 variant protein /cds=p(228,1331) /gb=AB209172 /gi=62087923 /ug=Hs.529488 /len=6328
LGEA-RHID02-260406-016-H03	THBS2	cell adhesion	Homo sapiens thrombospondin 2 (THBS2), mRNA
LGEA-RHID02-310306-013-A11	PTTG1IP	protein import into nucleus	Homo sapiens pituitary tumor-transforming 1 interacting protein (PTTG1IP), mRNA
LGEA-RHID02-310306-013-B02	CD518345	unknown	AGENCOURT_14371227 NIH_MGC_181 Homo sapiens cDNA clone IMAGE:30407573 5', mRNA sequence
LGEA-RHID02-310306-013-B04	DIP2C	metabolism	Homo sapiens DIP2 disco-interacting protein 2 homolog C (Drosophila) (DIP2C), mRNA
LGEA-RHID02-310306-013-B06	RAP1A	cell cycle	Homo sapiens RAP1A, member of RAS oncogene family (RAP1A), transcript variant 1, mRNA
LGEA-RHID02-310306-013-B08	SKP1A	biological process unknown	Homo sapiens S-phase kinase-associated protein 1A (p19A) (SKP1A), transcript variant 2, mRNA
LGEA-RHID02-310306-013-C12	SPAG6	spermatid development	Homo sapiens sperm associated antigen 6 (SPAG6), transcript variant 2, mRNA
LGEA-RHID02-310306-013-D03	MLPH	intracellular protein transport	Homo sapiens melanophilin (MLPH), mRNA
LGEA-RHID02-310306-013-D05	FCER1G	cell surface receptor linked signal transduction	Homo sapiens Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide (FCER1G), mRNA
LGEA-RHID02-310306-013-E01	GALNAC4S-6ST	hexose biosynthesis	Homo sapiens B cell RAG associated protein (GALNAC4S-6ST), mRNA
LGEA-RHID02-310306-013-E08	C19ORF10	unknown	Homo sapiens chromosome 19 open reading frame 10 (C19orf10), mRNA
LGEA-RHID02-310306-013-E12	PIGY	unknown	Homo sapiens phosphatidylinositol glycan, class Y (PIGY), mRNA
LGEA-RHID02-310306-013-F11	ARHGEF16	unknown	Homo sapiens Rho guanine exchange factor (GEF) 16 (ARHGEF16), mRNA
LGEA-RHID02-310306-013-G05	RPL3	protein biosynthesis	Homo sapiens ribosomal protein L3 (RPL3), transcript variant 1, mRNA
LGEA-RHID02-310306-013-G08	ST6GAL1	development	Homo sapiens ST6 beta-galactosamide alpha-2,6-sialyltransferase 1 (ST6GAL1), transcript variant 1, mRNA
LGEA-RHID02-310306-013-G12	GLUL	regulation of neurotransmitter levels	Homo sapiens glutamate-ammonia ligase (glutamine synthetase) (GLUL), transcript variant 3, mRNA
LGEA-RHID02-310306-013-H10	CHMP4A	protein transport	Homo sapiens chromatin modifying protein 4A (CHMP4A), mRNA
LGEA-RHSD02-020506-025-A11	LOC148137	unknown	Homo sapiens hypothetical protein BC017947 (LOC148137), mRNA
LGEA-RHSD02-020506-025-B07	BX648535	unknown	tgnl UG Hs#S16818116 Homo sapiens mRNA; cDNA DKFZp779N0931 (from clone DKFZp779N0931) /gb=BX648535 /gi=34367697 /ug=Hs.188879 /len=4349
LGEA-RHSD02-020506-025-B08	KLHDC2	unknown	Homo sapiens kelch domain containing 2 (KLHDC2), mRNA
LGEA-RHSD02-020506-025-C09	BG167469	unknown	tgnl UG Hs#S3201835 602345777F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4455666 5', mRNA sequence /clone=IMAGE:4455666 /clone_end=5' /gb=BG167469 /gi=12674251 /ti=45223968 /ug=Hs.455755 /len=1165

LGEA-RHSD02-020506-025-C10	PER2	transcription, signal transduction	Homo sapiens period homolog 2 (Drosophila) (PER2), transcript variant 1, mRNA
LGEA-RHSD02-020506-025-D02	LOC653188	unknown	PREDICTED: Homo sapiens similar to Beta-glucuronidase precursor, transcript variant 15 (LOC653188), mRNA
LGEA-RHSD02-020506-025-D04	H13414	unknown	tgnl UG Hs#S206876 yj15a04.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:148782 3', mRNA sequence /clone=IMAGE:148782 /clone_end=3' /gb=H13414 /gi=878234 /ug=Hs.571960 /len=418
LGEA-RHSD02-020506-025-E03	BU684023	unknown	UI-CF-EN0-acn-d-01-0-UI.s1 UI-CF-EN0 Homo sapiens cDNA clone UI-CF-EN0-acn-d-01-0-UI 3', mRNA sequence
LGEA-RHSD02-020506-025-E04	DA489876	unknown	DA489876 FCBBF3 Homo sapiens cDNA clone FCBBF3000980 5', mRNA sequence
LGEA-RHSD02-020506-025-E05	CV369880	unknown	PM2-MT0075-270700-003-h10 MT0075 Homo sapiens cDNA, mRNA sequence
LGEA-RHSD02-020506-025-E07	PLA2G12A	biological process unknown	Homo sapiens phospholipase A2, group X1A (PLA2G12A), mRNA
LGEA-RHSD02-020506-025-F12	AL833615	unknown	tgnl UG Hs#S4621776 Homo sapiens mRNA; cDNA DKFZp686D0673 (from clone DKFZp686D0673) /gb=AL833615 /gi=21734262 /ug=Hs.588618 /len=3169
LGEA-RHSD02-020506-025-G03	FLJ44796	unknown	Homo sapiens FLJ44796 protein (FLJ44796), mRNA
LGEA-RHSD02-020506-025-G07	FAM38B	unknown	Homo sapiens family with sequence similarity 38, member B (FAM38B), mRNA
LGEA-RHSD02-020506-025-G08	APLP2	G-protein coupled receptor protein signaling pathway	Homo sapiens amyloid beta (A4) precursor-like protein 2 (APLP2), mRNA
LGEA-RHSD02-020506-025-G10	BX538341	unknown	tgnl UG Hs#S16056254 Homo sapiens mRNA; cDNA DKFZp686C13222 (from clone DKFZp686C13222) /gb=BX538341 /gi=31874840 /ug=Hs.388565 /len=5116
LGEA-RHSD02-020506-025-H01	TRAPPC2	transcription	Homo sapiens trafficking protein particle complex 2 (TRAPPC2), transcript variant 1, mRNA
LGEA-RHSD02-020506-025-H06	BF751901	unknown	RC3-BN0034-111100-116-e09 BN0034 Homo sapiens cDNA, mRNA sequence
LGEA-RHSD02-120406-023-A04	PGM5P2	unknown	Homo sapiens phosphoglucomutase 5 pseudogene 2 (PGM5P2) on chromosome 9
LGEA-RHSD02-120406-023-A07	DB347221	unknown	tgnl UG Hs#S29664339 DB347221 TKIDN2 Homo sapiens cDNA clone TKIDN2002329 3', mRNA sequence /clone=TKIDN2002329 /clone_end=3' /gb=DB347221 /gi=83214192 /ug=Hs.589156 /len=485
LGEA-RHSD02-120406-023-B02	EIF4G3	regulation of translational initiation	Homo sapiens eukaryotic translation initiation factor 4 gamma, 3 (EIF4G3), mRNA
LGEA-RHSD02-120406-023-B08	ASNSD1	asparagine biosynthesis	Homo sapiens asparagine synthetase domain containing 1 (ASNSD1), mRNA
LGEA-RHSD02-120406-023-B09	AW151986	unknown	xf71d01.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2623489 3', mRNA sequence
LGEA-RHSD02-120406-023-C01	FBXW4	Wnt receptor signaling pathway	Homo sapiens F-box and WD-40 domain protein 4 (FBXW4), mRNA
LGEA-RHSD02-120406-023-C05	BM932001	unknown	UI-E-EJ1-ajl-g-02-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone UI-E-EJ1-ajl-g-02-0-UI 5', mRNA sequence
LGEA-RHSD02-120406-023-C06	CN414308	unknown	17000583090059 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence
LGEA-RHSD02-120406-023-C09	DB324154	unknown	tgnl UG Hs#S29788477 DB324154 NT2RI2 Homo sapiens cDNA clone NT2RI2022704 3', mRNA sequence /clone=NT2RI2022704 /clone_end=3' /gb=DB324154 /gi=83461437 /ug=Hs.590521 /len=547
LGEA-RHSD02-120406-023-C12	CD511652	unknown	tgnl UG Hs#S15910490 AGENCOURT_14353941 NIH_MGC_187 Homo sapiens cDNA clone IMAGE:30405306 5', mRNA sequence /clone=IMAGE:30405306 /clone_end=5' /gb=CD511652 /gi=31443370 /ti=342055218 /ug=Hs.364299 /len=795

LGEA-RHSD02-120406-023-D02	DN913290	unknown	MCF7RNAL16M23TR Human MCF7 breast cancer cell line near full length normalized library (MCF7_EST) Homo sapiens cDNA clone MCF7_RNA_L_16_M23, mRNA sequence
LGEA-RHSD02-120406-023-D04	KCTD3	unknown	Homo sapiens potassium channel tetramerisation domain containing 3 (KCTD3), mRNA
LGEA-RHSD02-120406-023-D07	BX649033	unknown	tgnl UG Hs#S16817617 Homo sapiens mRNA; cDNA DKFZp686K1098 (from clone DKFZp686K1098) /gb=BX649033 /gi=34368205 /ug=Hs.586476 /len=2977
LGEA-RHSD02-120406-023-D11	BX089752	unknown	BX089752 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE998G195403 ; IMAGE:2186538, mRNA sequence
LGEA-RHSD02-120406-023-E03	GREB1	biological process unknown	Homo sapiens GREB1 protein (GREB1), transcript variant a, mRNA
LGEA-RHSD02-120406-023-E05	PILRB	transmembrane receptor protein tyrosine kinase activation (dimerization)	Homo sapiens paired immunoglobulin-like type 2 receptor beta (PILRB), transcript variant 3, mRNA
LGEA-RHSD02-120406-023-E06	INADL	unknown	Homo sapiens InaD-like (Drosophila) (INADL), transcript variant 4, mRNA
LGEA-RHSD02-120406-023-F02	CN414853	unknown	17000600846810 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence
LGEA-RHSD02-120406-023-F05	BP418539	unknown	BP418539 Homo sapiens small intestine Homo sapiens cDNA clone HIE01697r 3', mRNA sequence
LGEA-RHSD02-120406-023-F11	LOC648013	unknown	PREDICTED: Homo sapiens similar to zinc finger protein 700 (LOC648013), mRNA
LGEA-RHSD02-120406-023-F12	DA645859	unknown	DA645859 MAMMA1 Homo sapiens cDNA clone MAMMA1000120 5', mRNA sequence
LGEA-RHSD02-130306-021-E05	CASC4	unknown	Homo sapiens cancer susceptibility candidate 4 (CASC4), transcript variant 1, mRNA
LGEA-RHSD02-130306-021-F04	ANKHD1	unknown	Homo sapiens ankyrin repeat and KH domain containing 1 (ANKHD1), transcript variant 1, mRNA
LGEA-RHSD02-130306-021-F07	AV651843	unknown	AV651843 GLC Homo sapiens cDNA clone GLCCUB04 3', mRNA sequence
LGEA-RHSD02-130306-021-G01	AZGP1	cell adhesion, negative regulation of cell proliferation	Homo sapiens alpha-2-glycoprotein 1, zinc (AZGP1), mRNA
LGEA-RHSD02-130306-021-G06	CCDC22	unknown	Homo sapiens coiled-coil domain containing 22 (CCDC22), mRNA
LGEA-RHSD02-130306-021-G10	H91436	unknown	yu96h03.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:241109 5', mRNA sequence
LGEA-RHSD02-130306-021-G11	PICALM	protein complex assembly	Homo sapiens phosphatidylinositol binding clathrin assembly protein (PICALM), transcript variant 2, mRNA
LGEA-RHSD02-130306-021-H07	T09065	unknown	EST06958 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBO62 5' end, mRNA sequence
LGEA-RHSD02-130306-021-H10	KIAA0492	unknown	PREDICTED: Homo sapiens KIAA0492 protein (KIAA0492), mRNA
LGEA-RHSD02-130306-021-H11	AI546885	unknown	PN2.1_09_E05.r mynorm Homo sapiens cDNA 5', mRNA sequence
LGEA-RHSD02-230306-022-A01	NISCH	unknown	Homo sapiens nischarin (NISCH), mRNA
LGEA-RHSD02-230306-022-A04	BX648473	unknown	tgnl UG Hs#S16818178 Homo sapiens mRNA; cDNA DKFZp686E0455 (from clone DKFZp686E0455) /gb=BX648473 /gi=34367635 /ug=Hs.293653 /len=4340
LGEA-RHSD02-230306-022-A08	WDR42A	unknown	Homo sapiens WD repeat domain 42A (WDR42A), mRNA
LGEA-RHSD02-230306-022-A10	KRT18	morphogenesis	Homo sapiens keratin 18 (KRT18), transcript variant 2, mRNA
LGEA-RHSD02-230306-022-B04	NM_001950	unknown	tgnl UG Hs#S1726668 Homo sapiens E2F transcription factor 4, p107/p130-binding (E2F4), mRNA /cds=p(64,1305) /gb=Nm_001950 /gi=44829053 /ug=Hs.108371 /len=2100
LGEA-RHSD02-230306-022-B06	LOC647524	unknown	PREDICTED: Homo sapiens hypothetical protein LOC647524, transcript variant 2 (LOC647524), mRNA
LGEA-RHSD02-230306-022-C02	BX497700	unknown	DKFZp779D1835_r1 779 (synonym: hncc1) Homo sapiens cDNA clone DKFZp779D1835 5', mRNA sequence

LGEA-RHSD02-230306-022-C12	CAPN2	proteolysis	Homo sapiens calpain 2, (mII) large subunit (CAPN2), mRNA
LGEA-RHSD02-230306-022-D05	KRT19	sarcomere organization	Homo sapiens keratin 19 (KRT19), mRNA
LGEA-RHSD02-230306-022-D07	AW137203	unknown	UI-H-BI1-acq-f-07-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone IMAGE:2715396 3', mRNA sequence
LGEA-RHSD02-230306-022-E09	AI908823	unknown	QV-BT191-010499-024 BT191 Homo sapiens cDNA, mRNA sequence
LGEA-RHSD02-230306-022-E11	ACTA2	unknown	Homo sapiens actin, alpha 2, smooth muscle, aorta (ACTA2), mRNA
LGEA-RHSD02-230306-022-F06	CXORF45	unknown	Homo sapiens chromosome X open reading frame 45 (CXorf45), mRNA
LGEA-RHSD02-230306-022-F07	NCOA7	cell wall catabolism	Homo sapiens nuclear receptor coactivator 7 (NCOA7), mRNA
LGEA-RHSD02-230306-022-G02	AK023601	unknown	tgnl UG Hs#S2651647 Homo sapiens cDNA FLJ13539 fis, clone PLACE1006640 /gb=AK023601 /gi=10435580 /ug=Hs.587547 /len=1822
LGEA-RHSD02-230306-022-G03	LOC654192	unknown	PREDICTED: Homo sapiens similar to ankyrin repeat domain 30A (LOC654192), mRNA
LGEA-RHSD02-230306-022-H02	BQ435770	unknown	AGENCOURT_7908273 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6102522 5', mRNA sequence
LGEA-RHSD02-230306-022-H06	CD687728	unknown	EST4249 human nasopharynx Homo sapiens cDNA, mRNA sequence
LGEA-RHSD02-270606-026-B09	VEZT	unknown	Homo sapiens vezatin, adherens junctions transmembrane protein (VEZT), mRNA
LGEA-RHSD02-270606-026-C02	TFF1	carbohydrate metabolism	Homo sapiens trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in) (TFF1), mRNA
LGEA-RHSD02-270606-026-C09	DA318042	unknown	DA318042 BRHIP3 Homo sapiens cDNA clone BRHIP3009569 5', mRNA sequence
LGEA-RHSD02-270606-026-D04	AU148184	unknown	AU148184 MAMMA1 Homo sapiens cDNA clone MAMMA1002820 3', mRNA sequence
LGEA-RHSD02-270606-026-D06	NUMBL	nervous system development	Homo sapiens numb homolog (Drosophila)-like (NUMBL), mRNA
LGEA-RHSD02-270606-026-D07	PLCG1	intracellular signaling cascade	Homo sapiens phospholipase C, gamma 1 (PLCG1), transcript variant 2, mRNA
LGEA-RHSD02-270606-026-D08	LOC441763		PREDICTED: Homo sapiens hypothetical LOC441763 (LOC441763), mRNA
LGEA-RHSD02-270606-026-D10	HLXB9	regulation of transcription from RNA polymerase II promoter	Homo sapiens homeobox HB9 (HLXB9), mRNA
LGEA-RHSD02-270606-026-E05	ALDH1A1	metabolism	Homo sapiens aldehyde dehydrogenase 1 family, member A1 (ALDH1A1), mRNA
LGEA-RHSD02-270606-026-E06	PLCB4	intracellular signaling cascade	Homo sapiens phospholipase C, beta 4 (PLCB4), transcript variant 2, mRNA
LGEA-RHSD02-270606-026-F01	RGC32	unknown	Homo sapiens response gene to complement 32 (RGC32), mRNA
LGEA-RHSD02-270606-026-F08	SNAP91	unknown	Homo sapiens synaptosomal-associated protein, 91kDa homolog (mouse) (SNAP91), mRNA
LGEA-RHSD02-270606-026-F11	SSR3	cotranslational protein targeting to membrane	Homo sapiens signal sequence receptor, gamma (translocon-associated protein gamma) (SSR3), mRNA
LGEA-RHSD02-270606-026-G03	GRN	signal transduction	Homo sapiens granulin (GRN), transcript variant 1, mRNA
LGEA-RHSD02-270606-026-G06	UBAP2L	unknown	Homo sapiens ubiquitin associated protein 2-like (UBAP2L), mRNA
LGEA-RHSD02-270606-026-G11	AW293174	unknown	UI-H-BW0-aII-c-12-0-UI.s1 NCI_CGAP_Sub6 Homo sapiens cDNA clone IMAGE:2729470 3', mRNA sequence
LGEA-RHSD02-280406-024-A04	ASXL1	transcription	Homo sapiens additional sex combs like 1 (Drosophila) (ASXL1), mRNA
LGEA-RHSD02-280406-024-A12	EPHX2	inflammatory response	Homo sapiens epoxide hydrolase 2, cytoplasmic (EPHX2), mRNA
LGEA-RHSD02-280406-024-B02	AI418471	unknown	tg48e06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2112034 3' similar to gb:X57346 14-3-3 PROTEIN BETA (HUMAN);, mRNA sequence
LGEA-RHSD02-280406-024-B05	MALAT1	unknown	Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA) (MALAT1) on chromosome 11

LGEA-RHSD02-280406-024-B07	CDC16	cell proliferation	Homo sapiens CDC16 cell division cycle 16 homolog (S. cerevisiae) (CDC16), mRNA
LGEA-RHSD02-280406-024-B11	AL122062	unknown	tgnl UG Hs#S1972344 Homo sapiens mRNA; cDNA DKFZp727M031 (from clone DKFZp727M031) /gb=AL122062 /gi=6102854 /ug=Hs.196437 /len=2637
LGEA-RHSD02-280406-024-C08	FAF1	apoptosis	Homo sapiens Fas (TNFRSF6) associated factor 1 (FAF1), transcript variant 1, mRNA
LGEA-RHSD02-280406-024-D01	BF678333	unknown	602084956F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4249212 5', mRNA sequence
LGEA-RHSD02-280406-024-D05	STATIP1	unknown	Homo sapiens signal transducer and activator of transcription 3 interacting protein 1 (STATIP1), mRNA
LGEA-RHSD02-280406-024-D06	FLJ11806	unknown	Homo sapiens nuclear protein UKp68 (FLJ11806), transcript variant 2, mRNA
LGEA-RHSD02-280406-024-D08	STIP1	response to stress	Homo sapiens stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organizing protein) (STIP1), mRNA
LGEA-RHSD02-280406-024-D11	NPNT		Homo sapiens nephronectin (NPNT), mRNA
LGEA-RHSD02-280406-024-E04	NR2C2	regulation of transcription, DNA-dependent	Homo sapiens nuclear receptor subfamily 2, group C, member 2 (NR2C2), mRNA
LGEA-RHSD02-280406-024-E12	R22320	unknown	yh26d10.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:130867 3', mRNA sequence
LGEA-RHSD02-280406-024-F05	CK821278	unknown	ig51b01.y5 HR85 islet Homo sapiens cDNA clone IMAGE:5594808 5', mRNA sequence
LGEA-RHSD02-280406-024-F11	SNTB2	actin binding	Homo sapiens syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2) (SNTB2), transcript variant 2, mRNA
LGEA-RHSD02-280406-024-G04	DA812605	unknown	DA812605 OVARC1 Homo sapiens cDNA clone OVARC1001812 5', mRNA sequence
LGEA-RHSD02-280406-024-G05	ANKRD30A	regulation of transcription, DNA-dependent	Homo sapiens ankyrin repeat domain 30A (ANKRD30A), mRNA
LGEA-RHSD02-280406-024-G08	BRD8	regulation of cell growth	Homo sapiens bromodomain containing 8 (BRD8), transcript variant 1, mRNA

Anexo 2 - Dados de expressão relativa dos genes normalizadores.

Gene	Amostra	Ct médio entre as duplicatas
GAPDH	IDC-1	27.16
	DCIS-1	32.51
	IDC-2	29.00
	DCIS-2	32.38
BCR	IDC-1	28.67
	DCIS-1	35.52
	IDC-2	30.42
	DCIS-2	33.39
HPRT	IDC-1	29.71
	DCIS-1	34.20
	IDC-2	29.74
	DCIS-2	34.39

Anexo 3 - Valores individuais dos Cts das duplicatas de cada gene em cada amostra e os valores de eficiência de amplificação obtidos.

Genes	Eficiência	Amostra	Valores de Ct	Desvio-padrão
METTL3	0.85	IDC-1	25.13	0.07
		IDC-1	25.24	
		DCIS-1	31.04	0.06
		DCIS-1	31.12	
		IDC-2	27.13	0.09
		IDC-2	27.27	
		DCIS-2	30.23	0.51
		DCIS-2	29.5	
COPB2	0.81	IDC-1	29.65	0.07
		IDC-1	29.75	
		DCIS-1	34.33	0.95
		DCIS-1	32.98	
		IDC-2	31.33	0.01
		IDC-2	31.34	
		DCIS-2	34.65	0.52
		DCIS-2	35.39	
EIF1	0.95	IDC-1	25.35	0.01
		IDC-1	25.36	
		DCIS-1	32.29	0.35
		DCIS-1	32.79	
		IDC-2	28.74	0.18
		IDC-2	28.47	
		DCIS-2	32.63	0.07
		DCIS-2	32.73	
FN1	0.94	IDC-1	24.41	0.04
		IDC-1	24.34	
		DCIS-1	29.87	0.22
		DCIS-1	29.56	
		IDC-2	23.41	0.15
		IDC-2	23.19	
		DCIS-2	29.16	0.11
		DCIS-2	29.32	
RDBP	0.98	IDC-1	21.89	0.09
		IDC-1	22.02	
		DCIS-1	27.60	0.09
		DCIS-1	27.46	
		IDC-2	22.36	0.04
		IDC-2	22.41	
		DCIS-2	25.66	0.13
		DCIS-2	25.46	

Anexo 4 - Genes confirmados por microarray como diferencialmente expressos entre IDC e DCIS por caso pareado.

CASO 2	GENES
LGEA-RHSD01-230306-023-B02	Homo sapiens ring finger protein 103 (RNF103), Mrna
LGEA-RHSD01-230306-023-G12	Homo sapiens serum amyloid A1 (SAA1), transcript variant 1, mRNA
LGEA-RHSD02-020506-025-G08	Homo sapiens amyloid beta (A4) precursor-like protein 2 (APLP2), mRNA
LGEA-RHSD02-130306-021-H07	EST T09065
LGEA-RHSD02-230306-022-B06	PREDICTED: Homo sapiens hypothetical protein LOC647524, transcript variant 2 (LOC647524), Mrna
LGEA-RHSD02-280406-024-D11	Homo sapiens nephronectin (NPNT), mRNA
LGEA-RHID01-080306-013-E03	UNIGENE BX109074
LGEA-RHID01-201205-012-F01	Homo sapiens ribosomal protein S16 (RPS16), Mrna
LGEA-RHID02-070406-014-F06	Homo sapiens glutathione S-transferase M3 (brain) (GSTM3), mRNA
LGEA-RHID02-070406-014-H08	Homo sapiens nucleosome assembly protein 1-like 1 (NAP1L1), transcript variant 2, Mrna
LGEA-RHID02-250406-015-E08	Homo sapiens RNA binding motif protein 6 (RBM6), mRNA
LGEA-RHID02-260406-016-B11	Homo sapiens CDK5 regulatory subunit associated protein 3 (CDK5RAP3), transcript variant 1, mRNA
LGEA-RHID02-310306-013-E01	Homo sapiens B cell RAG associated protein (GALNAC4S-6ST), mRNA
CASO 3	GENES
LGEA-RHSD01-060306-022-D01	Homo sapiens catalase (CAT), Mrna
LGEA-RHSD01-060306-022-G10	Homo sapiens chromosome 2 open reading frame 25 (C2orf25), mRNA
LGEA-RHSD01-060306-022-H05	Homo sapiens major histocompatibility complex, class I, B (HLA-B), mRNA
LGEA-RHSD01-060306-022-H11	Homo sapiens WD repeat domain 33 (WDR33), transcript variant 2, mRNA
LGEA-RHSD01-180406-026-A10	Homo sapiens bromodomain and WD repeat domain containing 1 (BRWD1), transcript variant 2, mRNA
LGEA-RHSD01-180406-026-C01	EST DA646975
LGEA-RHSD01-230306-023-D12	EST CA415848
LGEA-RHSD01-310306-024-E12	EST BU676595
LGEA-RHSD02-020506-025-A11	Homo sapiens hypothetical protein BC017947 (LOC148137), Mrna
LGEA-RHSD02-020506-025-B07	UNIGENE BX648535
LGEA-RHSD02-020506-025-C10	Homo sapiens period homolog 2 (Drosophila) (PER2), transcript variant 1, Mrna
LGEA-RHSD02-020506-025-D04	UNIGENE H13414
LGEA-RHSD02-020506-025-E03	EST BU684023
LGEA-RHSD02-120406-023-C06	EST CN414308
LGEA-RHSD02-120406-023-E03	Homo sapiens GREB1 protein (GREB1), transcript variant a, Mrna
LGEA-RHSD02-120406-023-F12	EST DA645859
LGEA-RHSD02-130306-021-H11	EST AI546885
LGEA-RHSD02-230306-022-B06	PREDICTED: Homo sapiens hypothetical protein LOC647524, transcript variant 2 (LOC647524), Mrna
LGEA-RHSD02-270606-026-C02	Homo sapiens trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in) (TFF1), mRNA
LGEA-RHSD02-270606-026-E06	Homo sapiens phospholipase C, beta 4 (PLCB4), transcript variant 2, mRNA
LGEA-RHSD02-270606-026-G06	Homo sapiens ubiquitin associated protein 2-like (UBAP2L), mRNA
LGEA-RHSD02-280406-024-D01	EST BF678333
LGEA-RHSD02-280406-024-D06	Homo sapiens nuclear protein UKp68 (FLJ11806), transcript variant 2, mRNA
CASO 13	GENES
LGEA-RHSD01-180406-026-E08	Homo sapiens integrin beta 3 binding protein (beta3-endonexin) (ITGB3BP), mRNA
LGEA-RHSD02-020506-025-D02	PREDICTED: Homo sapiens similar to Beta-glucuronidase precursor, transcript variant 15 (LOC653188), mRNA
LGEA-RHSD02-020506-025-E05	EST CV369880
LGEA-RHSD02-020506-025-H06	EST BF751901
LGEA-RHSD02-120406-023-C01	Homo sapiens F-box and WD-40 domain protein 4 (FBXW4), Mrna
LGEA-RHSD02-120406-023-C05	EST BM932001
LGEA-RHSD02-120406-023-C06	EST CN414308
LGEA-RHSD02-120406-023-E06	Homo sapiens InaD-like (Drosophila) (INADL), transcript variant 4, Mrna
LGEA-RHSD02-130306-021-H07	EST T09065
LGEA-RHSD02-230306-022-B04	UNIGENE Homo sapiens E2F transcription factor 4
LGEA-RHSD02-230306-022-E09	EST AI908823
LGEA-RHSD02-270606-026-C02	Homo sapiens trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in) (TFF1), mRNA
LGEA-RHSD02-280406-024-B05	Homo sapiens metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA) (MALAT1) on chromosome 11
LGEA-RHSD02-280406-024-D11	Homo sapiens nephronectin (NPNT), mRNA
LGEA-RHSD02-280406-024-G05	Homo sapiens ankyrin repeat domain 30A (ANKRD30A), mRNA
LGEA-RHID01-080306-013-A02	Homo sapiens guanine nucleotide binding protein (G protein), beta polypeptide 2 (GNB2), Mrna

LGEA-RHID01-080306-013-B12	Homo sapiens RAB22A, member RAS oncogene family (RAB22A), mRNA
LGEA-RHID01-201205-012-C01	Homo sapiens NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa (NDUFB2), nuclear gene encoding mitochondrial protein, mRNA
LGEA-RHID01-230306-014-D09	Homo sapiens AT rich interactive domain 4B (RBP1- like) (ARID4B), transcript variant 2, mRNA
LGEA-RHID01-230306-014-D10	Homo sapiens chromosome 1 open reading frame 27 (C1orf27), mRNA
LGEA-RHID02-090506-017-D10	Homo sapiens inositol 1,4,5-triphosphate receptor, type 1 (ITPR1), mRNA
LGEA-RHID02-090506-017-G09	Homo sapiens connective tissue growth factor (CTGF), mRNA
LGEA-RHID02-250406-015-A04	Homo sapiens solute carrier family 44, member 1 (SLC44A1), transcript variant 2, mRNA
LGEA-RHID02-250406-015-E04	Homo sapiens signal recognition particle 9kDa (SRP9), mRNA
LGEA-RHID02-260406-016-A11	Homo sapiens carbonic anhydrase XII (CA12), transcript variant 2, mRNA
LGEA-RHID02-260406-016-B08	Homo sapiens thyroid hormone receptor interactor 6 (TRIP6), mRNA
LGEA-RHID02-260406-016-B11	Homo sapiens CDK5 regulatory subunit associated protein 3 (CDK5RAP3), transcript variant 1, mRNA
LGEA-RHID02-260406-016-D06	Homo sapiens hypothetical protein MGC4562 (MGC4562), mRNA
LGEA-RHID02-260406-016-F12	Homo sapiens solute carrier family 40 (iron-regulated transporter), member 1 (SLC40A1), mRNA

CASO 25

GENES

LGEA-RHSD01-060306-022-E02	Homo sapiens phosphatidylinositol glycan, class G (PIGG), Mrna
LGEA-RHSD01-180406-026-G10	EST CR736439
LGEA-RHSD01-180406-026-H10	EST BF968136
LGEA-RHSD01-230306-023-H04	Homo sapiens SEC63-like (S. cerevisiae) (SEC63), Mrna
LGEA-RHSD02-120406-023-C01	Homo sapiens F-box and WD-40 domain protein 4 (FBXW4), Mrna
LGEA-RHSD02-120406-023-C06	EST CN414308
LGEA-RHSD02-280406-024-B11	UNIGENE AL122062
LGEA-RHSD02-280406-024-F05	EST CK821278
LGEA-RHID01-080306-013-C01	Homo sapiens GNAS complex locus (GNAS), transcript variant 2, Mrna
LGEA-RHID01-080306-013-G10	Homo sapiens vacuolar protein sorting 24 (yeast) (VPS24), transcript variant 2, mRNA
LGEA-RHID01-080306-013-H04	Homo sapiens jumping translocation breakpoint (JTB), mRNA
LGEA-RHID01-270606-016-B02	Homo sapiens splicing factor 3B, 14 kDa subunit (SF3B14), mRNA
LGEA-RHID01-270606-016-D02	EST BG399854
LGEA-RHID01-270606-016-F04	PREDICTED: Homo sapiens hypothetical LOC541471 (LOC541471), misc RNA
LGEA-RHID02-070406-014-A07	Homo sapiens phosphatidylinositol transfer protein, alpha (PITPNA), Mrna
LGEA-RHID02-070406-014-B06	UNIGENE AK023481
LGEA-RHID02-070406-014-B09	UNIGENE BC053563
LGEA-RHID02-070406-014-C02	Homo sapiens major histocompatibility complex, class II, DR alpha (HLA-DRA), Mrna
LGEA-RHID02-130306-011-D06	Homo sapiens KIAA0446 gene product (KIAA0446), mRNA
LGEA-RHID02-210306-012-H07	Homo sapiens insulin-like growth factor binding protein 7 (IGFBP7), mRNA
LGEA-RHID02-250406-015-B02	Homo sapiens cartilage intermediate layer protein, nucleotide pyrophosphohydrolase (CILP), mRNA
LGEA-RHID02-260406-016-A07	Homo sapiens chromosome 2 open reading frame 4 (C2orf4), mRNA
LGEA-RHID02-260406-016-C07	Homo sapiens CD74 molecule, major histocompatibility complex, class II invariant chain (CD74), transcript variant 2, mRNA
LGEA-RHID02-260406-016-F07	Homo sapiens KIAA1598 (KIAA1598), mRNA
LGEA-RHID02-260406-016-H03	Homo sapiens thrombospondin 2 (THBS2), mRNA

CASO 75

GENES

LGEA-RHSD01-180406-026-F01	Homo sapiens solute carrier family 39 (zinc transporter), member 8 (SLC39A8), Mrna
LGEA-RHID02-070406-014-F06	Homo sapiens glutathione S-transferase M3 (brain) (GSTM3), mRNA
LGEA-RHID02-090506-017-G09	Homo sapiens connective tissue growth factor (CTGF), mRNA

CASO 81

GENES

LGEA-RHSD01-060306-022-B11	Homo sapiens secreted frizzled-related protein 1 (SFRP1), mRNA
LGEA-RHSD01-180406-026-C09	Homo sapiens insulin-like growth factor 1 (somatomedin C) (IGF1), mRNA
LGEA-RHSD01-230306-023-D12	EST CA415848
LGEA-RHSD01-310306-024-B03	Homo sapiens annexin A1 (ANXA1), mRNA
LGEA-RHSD01-310306-024-C08	Homo sapiens oxysterol binding protein-like 8 (OSBPL8), transcript variant 2, mRNA
LGEA-RHSD02-020506-025-C10	Homo sapiens period homolog 2 (Drosophila) (PER2), transcript variant 1, mRNA
LGEA-RHSD02-020506-025-E03	EST BU684023
LGEA-RHSD02-020506-025-E07	Homo sapiens phospholipase A2, group XIA (PLA2G12A), mRNA
LGEA-RHSD02-020506-025-G08	Homo sapiens amyloid beta (A4) precursor-like protein 2 (APLP2), mRNA
LGEA-RHSD02-120406-023-D02	EST DN913290
LGEA-RHSD02-120406-023-D07	UNIGENE BX649033
LGEA-RHSD02-130306-021-H11	EST AI546885
LGEA-RHSD02-230306-022-E11	Homo sapiens actin, alpha 2, smooth muscle, aorta (ACTA2), mRNA
LGEA-RHSD02-270606-026-C02	Homo sapiens trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in) (TFF1), mRNA
LGEA-RHSD02-270606-026-E06	Homo sapiens phospholipase C, beta 4 (PLCB4), transcript variant 2, mRNA
LGEA-RHID02-070406-014-F10	Homo sapiens ferritin, heavy polypeptide 1 (FTH1), Mrna
LGEA-RHID02-130306-011-A06	Homo sapiens eukaryotic translation initiation factor 4A, isoform 2 (EIF4A2),

	mRNA
LGEA-RHID02-130306-011-D07	Homo sapiens DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (DDX58), mRNA
LGEA-RHID02-130306-011-D10	Homo sapiens RD RNA binding protein (RDBP), mRNA
LGEA-RHID02-250406-015-G01	Homo sapiens transmembrane emp24-like trafficking protein 10 (yeast) (TMED10), mRNA
LGEA-RHID02-250406-015-G10	Homo sapiens chloride channel, nucleotide-sensitive, 1A (CLNS1A), mRNA
