

**IDENTIFICAÇÃO DE MARCADORES MOLECULARES
PARA O CÂNCER DE MAMA POR MEIO DA ANÁLISE
DA EXPRESSÃO GÊNICA**

MARCIA DELLAMANO

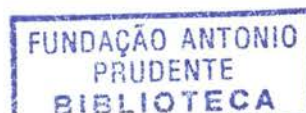
**Dissertação apresentada à Fundação Antônio Prudente
para a obtenção do título de Mestre em Ciências**

Área de concentração: Oncologia

**Orientadora:
Dra. Otávia Luisa Silva Damas de Caballero**

**Co-Orientador:
Dr. André Luiz Vettore**

**São Paulo
2004**



FICHA CATALOGRÁFICA

Preparada pela Biblioteca do Centro de Tratamento e Pesquisa

Hospital do Câncer A. C. Camargo

Dellamano, Marcia

Identificação de marcadores moleculares para o câncer de mama por meio da análise da expressão gênica / Marcia Dellamano. – São Paulo, 2004.

111p.

Dissertação(mestrado)–Fundação Antônio Prudente.

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

Orientadora: Otávia Luisa Silva Damas de Caballero.

Descritores: 1. NEOPLASIAS MAMÁRIAS/genética. 2. EXPRESSÃO GÊNICA/genética. 3. PERFIL DA EXPRESSÃO GÊNICA. 4. MARCADORES BIOLÓGICOS.

DEDICATÓRIA

Dedico esta dissertação em primeiro lugar aos meus pais, por tudo que me ensinaram, pelo exemplo de força e dedicação, pela paciência, pelo carinho, por todo amor e por acreditarem nos meus sonhos e me ajudarem a alcançá-los, dedico com todo meu amor a vocês!!

Dedico também às duas melhores irmãs do mundo, pela paciência, pela amizade, principalmente pela força no decorrer deste trabalho e por todo carinho.

Dedico ao meu sobrinho que faz a vida de toda família ainda mais feliz e unida.

Dedico à minha orientadora e ao meu co-orientador por acreditarem em mim, pela paciência, pelo aprendizado, pela amizade, pela confiança e por todo carinho.

AGRADECIMENTOS

Agradeço à minha querida orientadora Dra Otávia Caballero pela confiança em mim e em meu trabalho, por tudo que me ensinou em seu trabalho tão competente, pela paciência, pelo carinho, pela dedicação, pela amizade e pelo agradável convívio durante esse período.

Agradeço ao meu querido co-orientador Dr. André Luiz Vettore pela imensa contribuição neste trabalho e em todo meu aprendizado, por ter acreditado em mim, pelas valiosas explicações, por sua dedicação, pelo carinho e pela amizade .

Agradeço a Mariana Brait pela grande amizade, pelas idas de final de semana ao laboratório, pela divertida companhia, por todos seus ensinamentos, pelas opiniões compartilhadas e ainda que muito longe, por sempre estar pronta para me ajudar e ter se tornado uma verdadeira irmã, fazendo minha vida em São Paulo muito mais fácil. Agradeço também a seu pai e suas duas mães por me tratarem com tanto carinho e pelas dicas culturais. Agradeço ainda a tia Rê, por ter me adotado como sua sobrinha, pelo carinho e pela força positiva de sempre!

Agradeço a Barbara Pereira de Mello pela grande e competente ajuda, pelo agradável convívio, pela amizade, por ter compartilhado comigo este trabalho e principalmente por toda paciência nas explicações corridas entre os experimentos.

Agradeço a Andréa Seixas pela grande amizade, pelo aprendizado, principalmente no começo do meu trabalho (dupla "*Forward e Reverse*"), pelo agradável convívio, pela paciência, pelas conversas sinceras, pelo carinho e principalmente pelo tabule, pela sopa e pelas barrinhas integrais!

Agradeço ao grande amigo e companheiro de laboratório Daniel Onofre Vidal, pelo divertido convívio, pela amizade, pela paciência na leitura deste trabalho e principalmente pelo auxílio salvador na formatação das tabelas.

A Luciane Luciane Tsukamoto Kagohara pela grande amizade e por toda ajuda.

Agradeço a todo pessoal do laboratório pela agradável convivência e grande amizade: Adriana Bulgarelli pelas opiniões compartilhadas "em tempo real" e pelo

divertido convívio; Danielle Renzoni por não se importar com meus papéis em cima de sua bancada, pelas conversas, pela companhia na copa durante o almoço; Fabio Piccoli pela amizade, pelos cds, pelas dicas de cinema e pelas experiências nos “bancos de dados” compartilhadas; Fabrício Carvalho por sempre estar disposto a ajudar; Letícia Olops e Roberta Félix pela amizade e pelo carinho; Maria Isabel Achatz pelo exemplo de uma ótima profissional e “super-mãe”; Mariana Maschietto pela amizade e pelo convívio sempre tão tranquilo; Simone Sredni por todo carinho, pela amizade, pelo interesse e pela cuidadosa microdissecção das amostras congeladas; Fernanda Góes, Dinamar Gaspar, Marcelo Cardoso e Luiz Fernando Lopes pelo divertido convívio.

Agradeço ao Dr. André Carvalho pelas explicações estatísticas, pelas tabelas do excel e pela disposição em ajudar.

Agradeço a Valéria Paixão pela paciência, pelo convívio no laboratório, pelo carinho, pela competente separação das amostras.

A Fabiana Bettoni, pela enorme amizade, pelo divertido convívio diário em casa e no trabalho, pelas conversas, pelas pipocas salvadoras nos finais de semana, por estar sempre disposta a me ajudar e principalmente pela compreensão durante a etapa final deste trabalho.

A Raphael Bessa, por sua amizade e principalmente por toda ajuda quando mais precisei, obrigada pela atenção e pela paciência!

A Lilian Pires pela descoberta de uma nova amizade, pelos momentos divertidos e pelo exemplo de uma mãe dedicada.

A Ana Paula Felipe pelo divertido convívio em casa, pela amizade e pela compreensão durante a etapa final deste trabalho.

Agradeço a Dra. Helena Brentani do Laboratório de Biologia Computacional pela valiosa e fundamental colaboração neste trabalho, pelas análises no banco de *SAGE*, pela paciência, por estar sempre disposta a ajudar. Agradeço também ao seu grupo de alunos pelo auxílio e pelas análises nos bancos de dados.

Agradeço ao Dr. Fernando Augusto Soares do Departamento de Anatomia Patológica pela importante colaboração, pelo competente trabalho e pela disposição em me ajudar.

Agradeço ao Dr. Hugo Campos do departamento de Anatomia Patológica pela revisão das lâminas e por sempre estar disposto a ajudar. Agradeço também a Dra. Isabela Werneck da Cunha, deste mesmo departamento, pela importante contribuição.

Agradeço a Carlos Nascimento, Miyuki Silva e Severino Silva por cortarem as amostras com organização e rapidez

Agradeço a Suely Nonogaki, por toda preocupação com esse trabalho e por toda sua colaboração.

Agradeço a Dra. Maria Mitzi Brentani da Faculdade de Medicina da Universidade de São Paulo pelas linhagens celulares HB4A e HB4aC5.2 cedidas ao instituto.

Agradeço a Jane Kaiano pela amizade, pelas dicas e ensinamentos e por seu jeito prático e rápido.

Agradeço a Inês Nishimoto pela descoberta de uma grande amizade e por estar sempre disposta em ajudar.

Agradeço ao Dr. Andrew Simpson por ter me permitido fazer parte do Laboratório de Genética do Câncer e pela colaboração neste trabalho.

Agradeço ao Dr. Ricardo Renzo Brentani, diretor do Instituto Ludwig e Presidente do Hospital do Câncer A.C. Camargo pelas excelentes condições de trabalho.

Agradeço às meninas da sala de cultura de células, Regina Nomizo pela importante ajuda, pela grande amizade e pelas conversas sempre tão sinceras e Roseli Mendonça pelos meios de cultura.

Agradeço ao diretor da pós-graduação Dr. Luiz Fernando Lima Reis e às funcionárias Ana Maria Kuninari, Marcia Hiratani e Elaine Viana, por todo carinho, pela atenção e por sempre estarem dispostas a ajudar.

Agradeço à toda equipe da biblioteca (Suely Francisco, Maria Adrina M. Bassols, Francyne Pólen G. Lima e Rosinéia Aguiar Carneiro) pelo competente trabalho, pelo carinho e atenção, pela enorme ajuda, pela formatação da dissertação e pelas correções feitas na bibliografia.

Agradeço ao Departamento de Mastologia do Hospital do Câncer A.C. Camargo pela importante colaboração.

Agradeço aos funcionários do Serviço de Arquivo Médico e Estatístico do Hospital do Câncer A.C. Camargo em especial a Sra Hirde Contesini, pela ajuda e disponibilização dos prontuários.

Agradeço aos membros da banca de qualificação desta dissertação: Dr. Fernando Augusto Soares, Dra. Dirce Maria Carraro, Dra Eloiza Helena Tajara da Silva e Dra. Gisele Wally Braga Colleoni pelo tempo atribuído à leitura dos relatórios e pelas sempre proveitosas sugestões.

Agradeço a CAPES pela bolsa de estudos concedida, a FAPESP pelo apoio financeiro através do CEPID e ao Instituto Ludwig de Pesquisa Sobre o Câncer pelo apoio financeiro e pela oportunidade de desenvolver nosso projeto.

Agradeço a *Applied Biosystems* por todo suporte técnico, pelos valiosos treinamentos, pelas dicas e principalmente pelo uso do ABI7000 para a realização das reações de RT-PCR em tempo real.

Agradeço a todos do Instituto Ludwig, pela ajuda, pelo convívio no dia-a-dia e pelas amizades conquistadas.

Agradeço a todos que de alguma maneira contribuíram para a realização deste trabalho.

Agradeço a Banca Julgadora da Dissertação pelo tempo atribuído à leitura desta dissertação.

AGRADECIMENTOS ESPECIAIS

Agradeço ao meu pai e minha mãe pelo infinito amor e dedicação e por tudo o que eles fazem por mim.

Agradeço às minhas queridas irmãs, por todo amor e amizade e ao sobrinho “mais lindo do mundo”!!!

Agradeço a todos da minha família que me acolheram de maneira tão carinhosa em São Paulo, pelos tempos de hospedagem, pelas conversas sinceras e pelos divertidos domingos.

Agradeço ao meu cunhado e sua família pela amizade e por sempre estarem dispostos a ajudar.

Agradeço às minhas eternas amigas: Luciana Peters, Janaína Radispiel e Débora Bruno, Beatriz Paixão e ao amigo Regis Ribeiro pela imensa amizade, por todo apoio, pelos bons tempos de faculdade e por tornarem minha vida melhor.

Agradeço a Mariana Brait por tudo o que fez por mim, pelo meu aprendizado e principalmente pela grande amizade.

Agradeço às amigas: Andréa Seixas e Luciane Tsukamoto Kagohara e ao amigo Daninel Onofre Vidal, pela gigante ajuda e paciência na impressão deste trabalho.

Agradeço a todos que deram sua pequena ou grande contribuição para que eu chegasse até aqui.

RESUMO

Dellamano, M. **Identificação de marcadores moleculares para o câncer de mama por meio da análise da expressão gênica**. São Paulo; 2004. [Dissertação de Mestrado-Fundação Antonio Prudente].

O câncer de mama é a neoplasia mais freqüente que acomete as mulheres do mundo todo. Embora exista um considerável progresso no diagnóstico precoce e no tratamento desse tipo de tumor, a mortalidade por essa doença ainda permanece relativamente alta. Tumores de mama com aparentemente o mesmo tipo histológico e estadiamento variam amplamente em sua resposta às terapias disponíveis. A identificação de genes diferencialmente expressos no câncer de mama em relação ao tecido normal é de extrema importância para o entendimento da biologia do processo de tumorigênese e na identificação de novos potenciais marcadores de diagnóstico, prognóstico e no desenvolvimento de alvos terapêuticos. Neste trabalho, a análise do perfil da expressão gênica por meio de utilização dos dados disponibilizados pelo banco de *SAGE* resultou na identificação de 175 genes, os quais apresentaram um aumento da expressão de pelo menos quatro vezes em relação ao tecido normal. Levando-se em consideração a função celular, o processo biológico em que o gene está envolvido, a disponibilidade de anticorpos comerciais, e a localização celular, dezenove genes (*ALPPL2*, *BIRC5*, *CAD*, *CELSR2*, *CTSF*, *DGCR6*, *EFNA4*, *EI24*, *FN3KR3*, *GATA3*, *GATA4*, *HMG20B*, *PTPRF*, *QSCN6*, *SCARB2*, *STARD10*, *ZC3HDC1*, *KIAA0141* e *KIF22*) foram selecionados para terem sua expressão diferencial avaliada por meio de RT-PCR em tempo real em vinte e duas amostras de tecidos tumorais de mama, em estágio inicial de desenvolvimento da doença (menores do que dois centímetros) e em uma mistura de amostras não tumorais. A quantificação da expressão desses genes revelou que onze deles (*ALPPL2*, *CELSR2*, *FN3KR3*, *GATA3*, *GATA4*, *HMG20B*, *PTPRF*, *QSCN6*, *SCARB2*, *STARD10* e *ZC3HDC1*) apresentaram-se com aumento da expressão de pelo menos duas vezes em mais de 25% das amostras testadas, em relação à mistura de tecidos normais. Os genes *ALPPL2*, *CELSR2*, *GATA 3* e *HMG20B*, apresentaram aumento de expressão em mais de 50% das amostras analisadas. A correlação entre o aumento de expressão

dos genes *CELSR2*, *GATA3*, *GATA4* e *KIAA0141* e a positividade para a imunodeteção de receptor de estrógeno se mostrou estatisticamente significativa (com valores de p iguais a: 0,0270; 0,0223; 0,0146 e 0,0325, respectivamente). A associação entre a expressão do gene *CELSR2* e a positividade para a imunodeteção do receptor de progesterona se mostrou estatisticamente significativa ($p=0,0278$). Os resultados apresentados no presente estudo demonstraram que a estratégia de escolha de genes a partir da análise de banco de dados públicos disponíveis *on line* foi adequada, por terem sido validados como apresentando aumento de expressão por meio de RT-PCR em tempo real mais da metade dos genes selecionados nas amostras analisadas. Embora os resultados obtidos forneçam indícios de uma provável correlação entre o aumento da expressão dos genes e o processo de tumorigênese de mama, um estudo mais amplo, com um maior número de amostras será necessário para a confirmação dos achados.

SUMMARY

Dellamano, M. [**Identification of breast tumor molecular markers by gene expression analysis**]. São Paulo; 2004. [Dissertação de Mestrado-Fundação Antonio Prudente].

Breast cancer is one of the most common malignancies and the main cause of death among women. Although there have been a considerable progress on early diagnosis and treatment, the rates of mortality due to breast cancer among women remain still high. Breast cancer is characterized by an important histoclinical heterogeneity, which makes difficult the selection of the most appropriate treatment for each case. The identification of genes differentially expressed in breast cancer has important implications in understanding the biology of breast tumorigenesis and in the identification of new potential diagnostic and prognostic markers and in the development of therapeutic targets. Expression profiling using the public serial analysis of gene expression (SAGE) database resulted in the identification of 175 genes, which were at least 4 fold overexpressed in breast cancer SAGE libraries when compared to the normal breast SAGE libraries. Considering the cell function, biological processes in which the genes are involved, cell localization and the commercial availability of antibodies, 19 candidate genes (*ALPPL2*, *BIRC5*, *CAD*, *CELSR2*, *CTSF*, *DGCR6*, *EFNA4*, *EI24*, *FN3KR3*, *GATA3*, *GATA4*, *HMG20B*, *PTPRF*, *QSCN6*, *SCARB2*, *STARD10*, *ZC3HDC1*, *KIAA0141* e *KIF22*) were selected to have the differential expression tested by Real Time RT-PCR in 22 breast cancer samples smaller than two centimeters compared to a pool of normal breast tissues. Eleven of these genes (*ALPPL2*, *CELSR2*, *FN3KR3*, *GATA3*, *GATA4*, *HMG20B*, *PTPRF*, *QSCN6*, *SCARB2*, *STARD10* and *ZC3HDC1*) showed a significant overexpression of at least 2 fold in more than 25% of the breast tumor samples tested. The genes *ALPPL2*, *CELSR2*, *GATA 3* e *HMG20B* were found overexpressed in more than 50% of the samples analyzed. The statistical analysis showed a significant correlation between the overexpression of the genes *CELSR2*, *GATA3*, *GATA4* and *KIAA0141* and the estrogen receptor immunodetection (the p values were 0.0270; 0.0223, 0.0146, and 0.0325, respectively). The association between the

overexpression of *CELSR2* and the progesterone receptor immunodetection was also statistically significant ($p=0.0278$). These preliminary results suggest that data mining of *SAGE* database seems to be an appropriate approach to identify differentially expressed genes in breast tumors. The results also suggest that these genes may be implicated in breast tumorigenesis, but further studies with larger series are needed to confirm our findings.

LISTA DE FIGURAS

Figura 1	Representação gráfica da curva de desnaturação de uma reação de amplificação por meio de RT-PCR em tempo real para a análise da expressão de um determinado gene em condições não ideais de amplificação.	31
Figura 2	Representação gráfica da curva de desnaturação de uma reação de amplificação por meio de RT-PCR em tempo real para a análise da expressão de um determinado gene em condições ideais de amplificação, resultando em um único produto, específico de reação.	31
Figura 3	Representação do ensaio para RT-PCR em tempo real da <i>Human Endogenous Control Plate (Applied Biosystems)</i> .	
	3a Placa com 96 poços para reação de RT-PCR em tempo real.	36
	3b Esquema da distribuição das sondas e <i>primers</i> nas colunas da placa.	36
Figura 4	a) Representação da reação de amplificação do gene <i>STARD10</i> por meio de reação de PCR convencional, utilizando-se como molde cDNAs provenientes de duas linhagens celulares de mama neoplásica (em condições não ideais de amplificação).	43
	b) Representação da reação de amplificação do gene <i>STARD10</i> por meio de reação de PCR convencional, utilizando-se como molde cDNAs provenientes de duas linhagens celulares de mama neoplásica (em condições ideais de amplificação).	43
Figura 5	Visualização do resultado da extração de RNA em gel de agarose a 1,5%.	47
Figura 6	Verificação da qualidade da síntese de cDNA, visualização em gel de poliacrilamida 8% corado com nitrato de prata.	48

LISTA DE TABELAS

Tabela 1	Aspectos clínico-patológicos das amostras correspondentes aos tumores iniciais selecionadas para quantificação da expressão gênica por meio de RT-PCR em tempo real.	18
Tabela 2	Origem, nome, descrição e condições de cultura apropriadas das linhagens celulares de tecido de mama utilizadas no presente estudo.	20
Tabela 3	Descrição das bibliotecas de mama disponíveis no Banco de Dados de <i>SAGE</i> do <i>CGAP</i> , em maio de 2002, que foram utilizadas no presente estudo.	24
Tabela 4	Informações sobre os <i>web-sites</i> da <i>internet</i> utilizados para o processo de seleção dos genes candidatos.	27
Tabela 5	Genes Humanos mais comumente usados como constitutivos, selecionados para a análise por <i>SAGE</i> do controle endógeno mais adequado para realização da quantificação relativa da expressão gênica por meio de RT-PCR em tempo real.	35
Tabela 6	Seqüências dos <i>primers</i> , das sondas e o tamanho esperado do fragmentos de cada gene selecionado, por meio do serviço <i>Assay By Design-Applied Biosystems</i> .	39
Tabela 7	Seqüências dos <i>primers</i> específicos de cada gene selecionado, tamanho do fragmento esperado, temperatura de <i>annealing</i> na PCR e temperatura de aquisição da fluorescência para realização das reações de amplificação por RT-PCR em tempo real.	42
Tabela 8	Massa total de RNA obtida a partir das amostras selecionadas após a extração pelo método de sedimentação por cloreto de césio.	46
Tabela 9	Genes selecionados para validação da expressão diferencial em tecido neoplásico de mama por meio de RT-PCR em tempo real.	49
Tabela 10	Avaliação das freqüências das <i>SAGEtags</i> entre as bibliotecas de <i>SAGE</i> de mama tumoral e normal por meio da utilização do <i>pipeline SAGEβBin</i> .	67
Tabela 11	Valores da diferença entre os ΔC_t s entre as amostras (mistura de tumoral e a mistura normal, de tecidos de mama) para cada um dos genes analisados no ensaio de reação de RT-PCR em tempo real <i>TaqMan Human Endogenous ControlPlate</i> .	69
Tabela 12	Porcentagem de amostras com aumento de expressão de pelo menos duas	

	vezes em relação à <i>mistura</i> de amostras normais, para os genes analisados por meio de reação de quantificação por RT-PCR em tempo real.	72
Tabela 13	Distribuição dos genes de acordo com o <i>status</i> do receptor de estrógeno e nível de expressão dos genes analisados.	74
Tabela 14	Distribuição dos genes de acordo com o <i>status</i> do receptor de progesterona e nível de expressão dos genes analisados.	76

LISTA DE ABREVIATURAS

ACTB- Homo sapiens actin, beta

ALPPL2- Homo sapiens alkaline phosphatase, placental-like 2

AJCC- the american joint committee on cancer

APC- Homo sapiens adenomatosis polyposis coli (APC), mRNA

AR- receptor de andrógeno

ATCC- american type culture collection

BIRC5- Homo sapiens baculoviral IAP repeat-containing 5 (survivin)

BRCA1- Homo sapiens breast cancer gene 1

BRCA2- Homo sapiens breast cancer gene 2

B2M- Homo sapiens beta-2-microglobulin

CAD- Homo sapiens carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase

cDNA- DNA complementar

CELSR2- Homo sapiens cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)

c-erbB2- Homo sapiens epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian) (EGFR)

cm- centímetro

cm²- centímetro quadrado

CCND1- Homo sapiens cyclin D1 (PRAD1: parathyroid adenomatosis 1)

CGAP- cancer genome anatomy project

CO₂- dióxido de carbono

Cp- *crossing point*

CsCl- cloreto de céσιο

Ct- *threshold cycle*

cTNM- classificação clínica dos tumores por meio do sistema TNM

CTSF- *Homo sapiens cathepsin F*

dbEST- *expressed sequences tags database*

DEPC- di-etil pirocarbonato

DGCR6- *Homo sapiens DiGeorge syndrome critical region gene 6*

DCIS- carcinoma Ductal *In Situ*

DMSO- dimetil sulfóxido

DNA- ácido desoxirribonucleico

dNTPs- deoxinucleotídeos fosfato

ECSM1- *endothelial cell specific molecule 1*

ECSM2- *endothelial cell specific molecule 2*

ECSM3- *endothelial cell specific molecule 3*

EFNA4- *Homo sapiens ephrin-A4*

EGFR- receptor de fator de crescimento

EI24- *Homo sapiens etoposide induced 2.4*

ER- receptor de estrógeno

EST- etiquetas de seqüências expressas, *cuja sigla do inglês é Expressed Sequence Tags*

E.U.A.- Estados Unidos da América

FLJ12171- *Homo sapiens hypothetical protein FLJ12171*

MPSS- *massively parallel signature sequencing*

mRNA- RNA mensageiro

NaOH- hidróxido de Sódio

NaCl- cloreto de sódio

NCBI- *national center for biotechnology information*

ng- nanograma

NOTCH2- *Homo sapiens Notch homolog 2 (Drosophila)*

pb- pares de bases

PCR- reação em cadeia da polimerase cuja sigla do inglês é *Polimerase Chain Reaction*

PR- receptor de progesterona

PPIA- *Homo sapiens peptidylprolyl isomerase A (cyclophilin A) (PPIA), transcript*

p53- *tumor protein p53 [Homo sapiens]*

pTNM- classificação histopatológica dos tumores por meio do sistema TNM

RPLP0- *Homo Sapiens ribosomal protein, large, P0*

RPM- rotações por minuto

PTPRF- *Homo sapiens protein tyrosine phosphatase, receptor type, F*

PGK1- *Homo sapiens phosphoglycerate kinase 1*

pH- potencial hidrogeniônico

QSCN6- *Homo sapiens quiescin Q6*

RNA- ácido ribonucléico

RT-PCR- *reverse transcriptase PCR*

RASSF1A- *Homo sapiens Ras association (RalGDS/AF-6) domain family 1 (RASSF1), transcript variant A, mRNA*



STARD10- Homo sapiens START domain containing 10

SAGE- serial analysis of gene expression

SAME- serviço de arquivo médico e estatístico

SCARB2- Homo sapiens scavenger receptor class B, member 2

STAR- steroidogenic acute regulatory protein

TAE- tris-ácido acético-EDTA

TBE- tris-ácido bórico-EDTA

TBP- Homo sapiens TATA box binding protein

TFRC- Homo sapiens transferrin receptor (p90, CD71)

T_m- melting temperature

TMA- tissue microarray

TNM- sistema de classificação de tumores malignos

TP53- Homo sapiens tumor protein p53 (Li-Fraumeni syndrome)

U- unidades

UICC- união internacional contra o câncer

uv- ultravioleta

ZC3HDC1- Homo sapiens zinc finger CCCH type domain containing 1

ÍNDICE

1	INTRODUÇÃO	1
1.1	Câncer de Mama	1
1.2	Marcadores Prognósticos	5
1.3	Aspectos Moleculares	8
2	OBJETIVOS	15
3	MATERIAIS E MÉTODOS	16
3.1	Casuística e Obtenção das Amostras de Tecidos	16
3.2	Cultura de Células	19
3.3	Extração de RNA	20
3.4	Avaliação dos RNAs	21
3.5	Síntese de cDNA	21
3.6	Avaliação da Qualidade dos cDNAs	22
3.7	Seleção dos Genes Candidatos	22
3.8	Quantificação da Expressão Gênica por Meio de RT-PCR em tempo real	27
3.9	Escolha do Gene Normalizador e Cálculo da Expressão Relativa	32
3.10	Desenho das sondas e <i>primers</i> e a Padronização das Reações para a Quantificação da Expressão Gênica por meio de RT-PCR em tempo real	37
3.11	Análise Estatística	43
4	RESULTADOS	44
4.1	Casuística e Obtenção das Amostras de Tecidos	44
4.2	Extração de RNA	44
4.3	Avaliação dos RNAs	47
4.4	Avaliação da Síntese de cDNA	47
4.5	Seleção dos Genes Candidatos	48
4.6	Escolha do Gene Normalizador	66
4.7	Quantificação da Expressão Gênica por meio de RT-PCR em tempo real	69
4.8	Análise Estatística Referente à Quantificação da Expressão Gênica por Meio de RT-PCR em tempo real	72
5	DISCUSSÃO	78
6	CONCLUSÕES	95
7	REFERÊNCIAS BIBLIOGRÁFICAS	97

ANEXOS

Anexo 1 Classificação clínica (TNM) para o câncer de mama baseada na extensão do tumor primário, na ausência ou presença e a extensão de

metástases em linfonodos regionais e a ausência ou presença de metástase à distância.

Anexo 2 Classificação patológica (pTNM) para o câncer de mama baseada na extensão do tumor primário, na ausência ou presença e a extensão de metástases em linfonodos regionais e a ausência ou presença de metástase à distância.

Anexo 3 Critérios para a graduação histológica dos carcinomas de mama, de acordo com o sistema de Elston e Ellis.

Anexo 4 Tipos histopatológicos de carcinomas mamários.

Anexo 5 Autorização para a utilização das amostras pela Comissão de Ética em Pesquisa do Hospital do Câncer A.C. Camargo (Processo nº434/02).

Anexo 6 Modelo do termo de informação e consentimento do Hospital do Câncer A.C. Camargo.

Anexo 7 Tabela com todos os genes resultantes da análise da expressão diferencial entre as bibliotecas de *SAGE* de mama tumoral e não tumoral.

Anexo 8 Medida da expressão gênica diferencial em relação a uma mistura de RNAs de tecidos de mamas normais, em todas as amostras analisadas nos genes selecionados.

1 INTRODUÇÃO

1.1 CÂNCER DE MAMA

O câncer de mama é a neoplasia mais freqüente e leva ao maior número de mortes entre as mulheres do mundo todo. A grande sensibilidade às influências endócrinas predis põe este órgão a numerosas condições patológicas. Este tipo de câncer é provavelmente o mais temido devido à sua alta incidência entre as mulheres e sobretudo pelos efeitos psicológicos que afetam a percepção de sexualidade e a própria imagem pessoal, uma vez que, em muitos casos, a forma de tratamento empregada inclui a remoção parcial ou completa da mama.

O câncer de mama representa nos países ocidentais uma das principais causas de morte entre as mulheres. Nos EUA para o ano de 2004 foram estimados 215.990 novos casos e 40.110 mortes por este tumor (JEMAL et al. 2004). No Brasil, segundo índices do Ministério da Saúde, ele se encontra entre os dez tumores de maior incidência na população feminina. Para o ano de 2003, foram estimados 41.610 novos casos e 9.335 mortes por essa doença entre as mulheres (www.inca.gov.br).

Os fatores de risco para o desenvolvimento do câncer de mama podem ser: genéticos (história familiar), hormonais, nutricionais, geográficos, ambientais e relacionados à exposição à radiação.

Os sintomas relacionados ao câncer de mama incluem a presença de nódulo na mama, acompanhado ou não de dor local. Podem também surgir alterações na pele que recobre a mama e o aparecimento de nódulos na axila e descarga papilar. Os

procedimentos básicos para o rastreamento e diagnóstico do câncer de mama compreendem o auto exame das mamas, exame clínico e mamografia. Os procedimentos auxiliares para o diagnóstico são: a ultra-sonografia, o exame citológico (punção aspirativa com agulha fina e citologia de descarga papilar) e o exame histopatológico (biópsia).

O tratamento clássico para este tipo de câncer inclui a retirada da mama comprometida e dos linfonodos regionais (gânglios linfáticos da axila próximos à mama afetada). Atualmente, quando possível, os cirurgiões têm realizado apenas a retirada do quadrante da mama em que está localizada a lesão, juntamente com o esvaziamento da axila do mesmo lado. Assim, além de resultados satisfatórios em termos de sobrevida para a paciente, a obtenção de um melhor efeito estético (uma vez que o órgão é parcialmente conservado) é alcançado, o qual exerce um importante papel sobre os efeitos psicológicos causados nas pacientes devido ao procedimento cirúrgico. Com a constatação da importância dos resultados estéticos no tratamento do câncer de mama, cada vez mais novas técnicas de cirurgia plástica que permitem a reconstrução da mama têm sido desenvolvidas.

Outros tipos de tratamentos como: quimioterapia, radioterapia, hormonioterapia e imunoterapia podem ser associados ou não ao tratamento cirúrgico, dependendo de cada caso da doença.

A classificação pelo sistema TNM (*Classification of Malignant Tumors*) desenvolvida e publicada pela União Internacional Contra o Câncer (UICC) é o método mais utilizado para a classificação de tumores malignos e a descrição de sua extensão anatômica. É de extrema importância o alcance de um consenso no registro da informação precisa da extensão da doença para cada localização anatômica, pois a

descrição clínica juntamente com a classificação histopatológica (quando possível) pode auxiliar o clínico no planejamento do tratamento, indicar prognóstico, ajudar na avaliação dos resultados do tratamento, facilitar a troca de informação entre diferentes centros de tratamento e contribuir para a pesquisa.

O sistema TNM para descrever a extensão anatômica da doença está baseado na avaliação de três componentes:

T- a extensão do tumor primário;

N- a ausência ou a presença e a extensão de metástases em linfonodos regionais;

M- a ausência ou presença de metástase à distância.

A classificação clínica designada TNM ou cTNM é baseada em evidências anteriores ao tratamento. Estas evidências surgem do exame físico, diagnóstico por imagem, endoscopia, exploração cirúrgica e outros exames relevantes. O Anexo 1 apresenta o sistema de classificação clínica TNM para o câncer de mama (SOBIN e WITTEKIND 1997). A classificação histopatológica designada pTNM (pós-cirúrgica) é baseada nas evidências obtidas do tratamento, suplementada ou modificada pela evidência adicional alcançada por meio da cirurgia e do exame histopatológico. A avaliação histopatológica do tumor primário (pT) exige a sua ressecção total ou parcial por meio de biópsia para avaliar a maior categoria pT. A avaliação histopatológica dos linfonodos regionais (pN) exige a remoção representativa de nódulos para comprovar a ausência de metástases em linfonodos regionais (pN0) e suficiente para avaliar a maior categoria pN. No Anexo 2 encontra-se representada a classificação pTNM para os tumores de mama.

Os critérios para estabelecer a graduação histológica do câncer de mama segundo o sistema de classificação de Scarff, Bloom e Richardson modificado por Elston e Ellis estão descritos no Anexo 3 (SCHMITT 2000). Neste sistema, são utilizados como parâmetros o grau de formação tubular, pleomorfismo nuclear e índice mitótico. De acordo com a intensidade de cada parâmetro (discreta, moderada ou elevada) é atribuída uma pontuação de 1 a 3. A soma obtida pelos três parâmetros corresponde ao grau da neoplasia da seguinte forma: escore 3-5: grau I; escore 6 ou 7: grau II; escore 8 ou 9: grau III. O grau histológico tem como base a combinação de parâmetros entre a razão de índice mitótico, grau nuclear e aparência morfológica celular.

Os tipos histológicos mais comuns do câncer de mama incluem o carcinoma ductal *in situ* (cuja sigla do inglês é *DCIS*) e o carcinoma ductal invasivo (POLYAK 2002). Pacientes com carcinoma ductal invasivo, na grande maioria dos casos, exibem alta incidência de comprometimento axilar e evolução clínica desfavorável, quando comparadas às pacientes com tumores menos invasivos. No Anexo 4 estão apresentados os tipos histopatológicos de carcinomas mamários (GREENE et al. 2002).

As informações hoje disponíveis em relação ao câncer de mama são ainda insuficientes para a implementação de programas de prevenção primária, ou seja, que evitem o surgimento da doença. Sendo assim, a maioria dos esforços é direcionada a ações de detecção precoce da doença, à descoberta de tumores ainda pequenos (em estágio de desenvolvimento inicial da doença), nos quais os recursos terapêuticos são mais eficazes e possibilitam a utilização de tratamentos menos mutiladores e com maior probabilidade de controle.

1.2 MARCADORES PROGNÓSTICOS

Entre os fatores prognósticos mais importantes e que podem potencialmente influenciar a abordagem terapêutica das neoplasias destacam-se: o tamanho tumoral, a idade, o grau histológico, o envolvimento de linfonodos axilares e o *status* de receptores hormonais.

O tamanho tumoral está diretamente associado com a sobrevida das pacientes (quanto mais precoce a detecção, maiores as chances de cura e a possibilidade do uso de tratamentos menos agressivos) (FISHER et al. 1983). A idade é um importante fator de predição de resposta à quimioterapia ou à terapia hormonal. Estudos têm sugerido que mulheres com idade inferior a trinta e cinco anos geralmente apresentam tumores de alto grau e possuem menor razão de sobrevida (DE LA ROCHEFORDIERE et al. 1993; NIXON et al. 1994). Diversos estudos têm demonstrado a importância do grau tumoral (principalmente em tumores de alto grau) como fator preditivo de recorrência tumoral. O grau histológico está relacionado com a invasão linfovascular, tumores que apresentam invasão linfática possuem pior prognóstico, mesmo na ausência de metástase em linfonodos (SILVESTRINI et al. 1995; TODD et al. 1987). O envolvimento de linfonodos axilares é um dos fatores de prognóstico melhores estabelecidos para o câncer de mama. Tanto o número de linfonodos envolvidos quanto o nível de envolvimento linfonodal são preditores de sobrevida. Pacientes sem o envolvimento dos linfonodos apresentam em sua grande maioria, um prognóstico favorável (FISHER et al. 1983). O *status* de receptores hormonais também é um importante fator de prognóstico para pacientes com câncer

de mama, pois estrógenos são constatados como fatores importantes na etiologia do câncer de mama. Eles são considerados como potentes agentes mitogênicos para a glândula mamária normal e também como fatores de iniciação e promoção da carcinogênese mamária (SCHMITT 1995; SOMMER e FUQUA 2001). Entretanto, a ausência de receptores de estrógeno (ER) no câncer de mama tem sido associada a um fenótipo mais agressivo e resistente à terapia hormonal. Tem se tornado bem estabelecido que pacientes com este fenótipo apresentam menor sobrevida livre de doença e menor sobrevida global, quando comparadas a pacientes portadoras do fenótipo positivo para ER (que expressam receptores de estrógeno) (NAGAI et al. 1994). Todos estes fatores têm sido utilizados para classificar os carcinomas de mama como carcinomas que provavelmente terão um prognóstico favorável ou como aqueles que se desenvolverão de forma desfavorável. No entanto, é praticamente impossível determinar o curso da doença, inclusive nas pacientes que não apresentam evidências de metástases axilares.

Atualmente, com a evolução dos métodos diagnósticos para o câncer de mama, a maioria dos tumores estão sendo diagnosticados em um estágio precoce de desenvolvimento da doença. Técnicas recentes de diagnóstico por imagem permitem a detecção de 75% de todas as pacientes que possuem pequenos carcinomas invasivos sem envolvimento de linfonodos axilares (COLPAERT et al. 2001). Entretanto, as taxas de mortalidade em estágios avançados da doença não apresentaram redução significativa, em consequência da falta de terapias efetivas. Aproximadamente 25% dessas pacientes provavelmente irão morrer em consequência da doença (PORTER et al. 2003a).

A maioria dos esforços relacionados ao controle deste tipo de câncer está dirigida às ações de detecção precoce, anterior ao comprometimento dos linfonodos axilares, o que permite, nesta fase inicial, o uso de tratamentos menos agressivos e com maior probabilidade de controle da doença. Mesmo assim, nas pacientes com linfonodos negativos, a recaída tumoral é de aproximadamente 30% em cinco anos de acompanhamento (EARLY BREAST CANCER TRIALISTS' COLLABORATIVE GROUP 1992).

Pacientes com câncer de mama em um mesmo estágio da doença possuem diferentes respostas aos tratamentos, bem como uma diferente evolução clínica. O tratamento por meio de quimioterapia e hormonioterapia reduz o risco de metástases à distância em aproximadamente um terço, entretanto, uma parcela de cerca de 70 a 80% das pacientes que receberam este tratamento provavelmente sobreviveriam sem a administração destes (VAN'T VEER et al. 2002).

Como as terapêuticas adjuvantes (por exemplo, quimioterapia e hormonioterapia) apresentam intensos efeitos colaterais indesejáveis, seria importante a utilização de marcadores moleculares que pudessem diferenciar as pacientes com bom prognóstico e que não necessitam de tratamentos complementares, daquelas com maiores probabilidades de desenvolver recorrências e/ou metástases e que necessitam de terapias adjuvantes.

Nos últimos anos, inúmeros grupos de pesquisa têm se empenhado intensamente na busca por marcadores moleculares para o câncer de mama. Vários marcadores têm sido estudados, dentre eles o oncogene *c-erbB2* e o gene supressor de tumor *p53*. Separadamente a expressão de cada um destes genes, *c-erbB2* e *p53*, parece não apresentar valor prognóstico. Entretanto, vários estudos têm demonstrado

que a presença concomitante de alterações das duas proteínas está fortemente correlacionada com um prognóstico não favorável e presença de recorrência (TAKIKAWA et al. 1994; BARBATI et al. 1997). O estudo de RUDOLPH et al. (2001) demonstrou a associação entre a detecção simultânea da expressão de c-erbB2 e p53 por imuno-histoquímica e o aumento da proliferação celular no tumor e a evolução prognóstica desfavorável.

1.3 ASPECTOS MOLECULARES

A progressão de um tecido normal para o câncer é acompanhada por uma série de alterações genéticas. Alterações na seqüência de genes, nos níveis de expressão e na estrutura de proteínas podem ser usadas como marcadores moleculares para a detecção precoce do câncer, antes de manifestação clínica, permitindo dessa forma o monitoramento da progressão da doença ou a escolha da terapia e determinação de prognóstico (SIDRANSKY 2002).

A tumorigênese de mama é um processo de múltiplos passos, iniciando-se com uma hiperplasia sem atipia, progredindo para lesão "in situ", depois para carcinomas invasivos e culminando em doença metastática (BECKMANN et al. 1997). A progressão do tumor é dirigida por alterações genéticas somáticas, que se refletem em alterações fenotípicas, bem como, em alterações nos perfis da expressão gênica (POLYAK 2001).

Diversos tipos de alterações genéticas já foram descritas em tumores, incluindo mutações pontuais (substituições, deleções e inserções) e cromossômicas (duplicações, translocações, inserções e deleções). Cerca de 5-10% dos casos de

câncer de mama e ovário podem ser associados a mutações na linhagem germinativa. Em 1990 foi identificado o gene *BRCA1*, mutações neste gene são responsáveis por aproximadamente metade de todos os casos familiares de câncer de mama e ovário (HALL et al. 1990). Em 1995, WOOSTER et al. identificaram um gene no cromossomo 13 com seis mutações diferentes em famílias com câncer de mama e o denominaram como *BRCA2*. Acredita-se que estes dois genes (*BRCA1* e *BRCA2*) se comportam como genes supressores de tumores clássicos, em que a perda de uma cópia predispõe o portador ao desenvolvimento de câncer (PARK e VOLGESTEIN 2003). Mais de 200 mutações distintas em *BRCA1* e *BRCA2* já foram descritas e grande parte destas mutações resultam em proteínas truncadas, mas uma proporção significativa corresponde a mutações do tipo *missense*, que alteram um único aminoácido e não resultam em proteína truncada (EELES 2000).

Recentemente, mecanismos epigenéticos de inativação gênica, principalmente a metilação de promotores gênicos, têm sido apontados como importantes na inativação de genes com características de supressão tumoral (JONES e BAYLIN 2002). Fatores epigenéticos são aqueles que alteram o fenótipo sem alterar o genótipo, ou seja, não representam alteração na seqüência dos nucleotídeos no DNA, mas têm consequência no fenótipo e são herdáveis durante a divisão celular (FEINBERG 2001). A metilação é um mecanismo importante para o desenvolvimento de mamíferos (JONES e TAKAI 2001). A metilação ocorre principalmente em promotores de genes que participam da supressão tumoral: genes envolvidos no sistema de reparo de DNA, no ciclo e morte celulares e genes que bloqueiam metástases, como os envolvidos em adesão celular. Recentemente um grupo de pesquisadores avaliou, de forma simultânea, a metilação em 39 genes

distintos em amostras de soro de pacientes com câncer de mama. Como resultado, foi verificado que a metilação na região promotora dos genes *RASSF1A* e *APC* pode ser utilizada como um forte indicador de pior prognóstico (MULLER et al. 2003).

Durante o processo de progressão neoplásica, o perfil de expressão gênica pode fornecer informações para a compreensão do desenvolvimento tumoral. Além disso, genes com aumento do nível de expressão no tecido neoplásico em relação ao tecido normal também podem ser utilizados como marcadores para diagnóstico ou prognóstico e podem ainda fornecer informações para o desenvolvimento de novas terapias contra o câncer. O receptor c-erbB2 foi o primeiro oncogene identificado como alvo para a terapia para o câncer de mama, com o desenvolvimento do anticorpo monoclonal humanizado anti-c-erbB2 Herceptin® (Trastuzumab) (BASELGA et al. 1996; SLAMON et al. 2001).

A classificação molecular dos tumores pela análise dos perfis de expressão gênica é uma nova maneira de investigação e diagnóstico possível em virtude dos recentes avanços tecnológicos e científicos. Em 2000 foi disponibilizado o conjunto de informações referentes às seqüências dos cromossomos humanos (LANDER et al. 2001; VENTER et al. 2001) e o Banco *dbEST* do *GenBank* conta atualmente com mais de cinco milhões de seqüências expressas nos mais diferentes tecidos humanos (www.ncbi.nlm.nih.gov/dbEST). Novas tecnologias como *SAGE* (*Serial Analysis of Gene Expression* - VELCULESCU et al. 1995) e *microarrays* de cDNA (SCHENA et al. 1998) têm sido aplicadas à análise da expressão de milhares de genes em um único experimento, permitindo a identificação de genes importantes na tumorigênese. É de extrema importância o conhecimento de genes envolvidos no crescimento e

progressão tumorais que possam representar possíveis alvos para o desenvolvimento de novas modalidades terapêuticas.

A técnica de *SAGE* permite a análise simultânea e quantitativa de um grande número de transcritos, ou seja, a técnica mostra não somente quais genes estão sendo expressos no tecido, mas também o nível de expressão dos mesmos (VELCULESCU et al. 1995). Essa técnica foi desenhada para fornecer um perfil quantitativo da expressão gênica celular, uma vez que permite uma rápida e detalhada análise da expressão de milhares de transcritos.

A técnica de *SAGE* identifica uma etiqueta (*tag*) em cada um dos transcritos que é correspondente a uma seqüência de nucleotídeos de tamanho definido, proveniente da porção codificadora dos genes e próxima ao sítio de restrição de uma endonuclease. Os dados produzidos por meio desta técnica apresentam-se como uma lista de *tags*. O nível de expressão de cada transcrito está diretamente ligado ao número de *tags* identificadas para cada gene. Atualmente, há um grande volume de dados gerados a partir de bibliotecas de *SAGE* disponível em bancos de dados públicos e que podem ser analisados através de ferramentas *on line* (www.ncbi.nlm.nih.gov/SAGE). No início deste trabalho, em agosto de 2002, neste banco de dados encontravam-se depositadas cento e sessenta e quatro bibliotecas de *SAGE*, dentre elas, vinte e duas eram de tecido mamário, sendo que sete correspondiam à glândula mamária normal (seis de tecidos e uma de linhagem celular) e quinze à glândula mamária tumoral (sete de tecidos e oito de linhagens celulares). Selecionar os dados relevantes e usá-los no planejamento de um determinado experimento para identificação de genes com expressão diferencial em tecidos normais e tumorais é um desafio. Estudos confirmatórios são longos e

dispendiosos e este problema tende a se agravar com as tecnologias de determinação de expressão gênica em larga escala. A confirmação da expressão diferencial é importante antes de se dar início a estudos de função gênica ou antes de se investigar um gene superexpresso como candidato a marcador tumoral ou alvo terapêutico (LOGING et al. 2000).

A identificação de genes diferencialmente expressos através da pesquisa em bancos de dados pode abreviar o tempo e esforços relacionados a experimentos de *microarray* realizados com o mesmo fim, por exemplo. Uma vez identificados os genes, eles podem ser validados diretamente por métodos como RT-PCR em tempo real, que podem ser realizados com pequenas quantidades de RNA e em muitas amostras simultaneamente. Isso se torna extremamente importante pelo fato de os tumores de mama atualmente serem diagnosticados na sua maioria em estágio inicial, de pequeno tamanho, havendo pequena quantidade de tecido disponível para análises.

A RT-PCR em tempo real avalia o acúmulo do produto na fase logarítmica da reação de amplificação e está diretamente relacionado à quantidade de molde existente no início da reação, sendo atualmente considerado um método bastante preciso e reprodutível para a quantificação da expressão gênica. A observação contínua da fluorescência emitida pelo DNA amplificado permite uma quantificação rápida e precisa do número de cópias de um transcrito específico presente em uma dada amostra de RNA (MORRISON et al. 1998). Desta forma, pode-se identificar genes diferencialmente expressos em dois tecidos e, em se tratando de tecido tumoral e normal, estes genes podem vir a ser utilizados como potenciais marcadores moleculares.

Diversos estudos têm sido realizados na tentativa de se compreender o perfil da expressão gênica no processo de tumorigênese de mama e novos genes candidatos a marcadores moleculares têm sido identificados (PEROU et al. 2000; SORLIE et al. 2001; VAN'T VEER et al. 2002; PORTER et al. 2003a).

PORTER et al. (2003b), desenvolveram um estudo com o objetivo de identificar genes que possam ser utilizados como marcadores moleculares para a classificação de carcinomas ductais *in situ* (DCIS) em mama, dentro de grupos biológica e clinicamente distintos. Para isso os autores compararam de forma quantitativa os transcritos presentes em células epiteliais de tecido mamário, três tipos distintos de DCIS e câncer invasivo e metastático de mama, utilizando a técnica de SAGE. Como resultado foram identificados diversos genes que poderão ser potencialmente utilizados para o diagnóstico molecular diferencial dos DCIS e tumores invasivos de mama. Os autores demonstraram que mudanças mais dramáticas no conjunto de genes transcritos em um determinado tecido (transcriptoma) ocorrem na transição do tecido normal para DCIS, ao passo que nenhuma característica molecular marcante foi encontrada entre os tumores *in situ* ou DCIS e os tumores invasivos.

JIANG et al. (2002) desenvolveram um estudo com o propósito de identificar genes específicos do tumor de mama para serem utilizados como alvo para imunoterapia. Para isto, os autores utilizaram bibliotecas de subtração de cDNA e *microarrays*. Como resultado, foram identificados quatro genes (*B709P*, *hair-specific type II keratin*, *B726P* e *gaba receptor subunit*) que têm sua expressão aumentada em tumores primários e metastáticos de mama. Como os autores enfatizaram, a identificação de genes com aumento ou diminuição da expressão em tumores de

mama pode auxiliar no melhor entendimento do processo de tumorigênese e fornecer marcadores adicionais para o tratamento e diagnóstico da doença.

Como a identificação de genes adicionais com alteração da expressão no câncer de mama é de extrema importância para a determinação prognóstica e para a escolha do tratamento mais adequado para cada caso, o presente trabalho buscou identificar genes com expressão diferencial neste tipo de tumor por meio da análise dos dados de expressão gênica disponíveis em Banco de Dados de *SAGE* e validar a expressão gênica diferencial dos genes selecionados por meio de RT-PCR em tempo real.

2 OBJETIVOS

1. Identificar genes com expressão diferencial em tumores de mama por meio da análise dos dados de expressão gênica de tecidos normais e tumorais de mama disponíveis no Banco de Dados de *SAGE* mantido pelo Instituto Nacional do Câncer dos Estados Unidos e disponibilizado no sítio <http://cgap.nci.nih.gov>.
2. Selecionar dentre os genes identificados no passo anterior candidatos a marcadores moleculares para o câncer de mama para terem a expressão diferencial validada em um conjunto de tumores iniciais de mama.
3. Confirmar a expressão gênica diferencial dos genes candidatos selecionados no passo anterior por meio de RT-PCR em tempo real em um conjunto de amostras de tumores iniciais de mama (com até dois centímetros em sua maior dimensão).

3 MATERIAIS E MÉTODOS

3.1 CASUÍSTICA E OBTENÇÃO DAS AMOSTRAS DE TECIDOS

Para obtenção das amostras utilizadas para validação dos genes candidatos foi realizado um levantamento de dados no Serviço de Arquivo Médico e Estatístico (SAME) do Hospital do Câncer A.C. Camargo por meio de análise dos prontuários de aproximadamente 200 pacientes do sexo feminino, atendidas neste hospital no período de 1999 a 2002, com tumores de mama menores do que dois centímetros e que possuíam amostras disponíveis no banco de tumores.

A utilização das amostras foi autorizada pela Comissão de Ética em Pesquisa do Hospital do Câncer A.C. Camargo (Processo nº434/02) e o termo de consentimento pós-informado foi devidamente assinado por todas as pacientes (Anexo 5 e 6, respectivamente).

As amostras foram colhidas a fresco durante o ato cirúrgico e submetidas a exame intra-operatório pelo método de congelamento para a seleção de áreas tumorais representativas e livres de necrose ou calcificações. As amostras foram submetidas a congelamento rápido em nitrogênio líquido e armazenadas a -70° C no Banco de Tumores do Hospital do Câncer A.C. Camargo/Instituto Ludwig de Pesquisa sobre o Câncer.

Os blocos foram cortados e as lâminas revisadas pelo Dr. Fernando Augusto Soares, Dr. Antônio Hugo Campos e Dra. Simone Treiger Sredni do Departamento de Anatomia Patológica do Hospital do Câncer A.C. Camargo, sendo que a última

selecionou as áreas mais representativas de tecido tumoral a serem utilizadas para posterior extração de RNA. Para o presente estudo, foram utilizadas amostras de tecidos correspondentes a carcinomas de mama com tipo histológico ductal invasivo, menores do que dois centímetros.

A Tabela 1 lista as características das amostras de neoplasia de mama selecionadas para este estudo. O estadiamento pTNM foi realizado pelo Dr. Antônio Hugo Campos do Departamento de Anatomia Patológica do Hospital do Câncer A.C. Camargo.

Tabela 1 – Aspectos clínico-patológicos das amostras correspondentes aos tumores iniciais selecionadas para quantificação da expressão gênica por meio de RT-PCR em tempo real

Amostra ¹	Tipo ²	Tamanho da Lesão (cm)	Metastases ³	ER ⁴	PR ⁴	Idade ⁶	P53 ⁷	cerbB2 ⁸	pTNM ⁹
1291	A	1,5	1/25	h	d	62	1	1	T1cN1Mx
1726	B	1,8	0/20	h	h	47	0	0	T1cN0Mx
1961	A	1,0	0/24	f	g	45	0	0	T1bN0Mx
2310	A	1,5	4/49	f	d	43	0	1	T1cN1Mx
2412	A	0,8	1/20	f	d	71	0	1	T1bN1Mx
2424	A	1,1	3/11	f	f	54	0	1	T1cN1Mx
2439	A	0,5	1/30	d	f	53	0	0	T1aN1Mx
2610	A	1,8	ND	d	d	43	1	1	T1cNxMx
2658	A	1,5	1/42	f	f	38	1	1	T1cN1Mx
2704	C	1,2	2/32	e	g	52	0	0	T1cN1Mx
2720	A	1,3	0/25-0/6	g	d	79	0	1	T1cNxMx
2734	C	1,5	0/23	d	d	45	1	1	T1cN0Mx
2789	A	1,2	0/30-0/1	g	g	32	0	1	T1cN0Mx
2933	A	2,0	1/17	d	d	54	1	1	T1cN1Mx
2937	A	0,7	0/25	f	g	47	1	1	T1bN0Mx
3143	ND	0,7	ND	f	g	45	1	1	T1bN0Mx
3177	A	1,3	0/23-0/1	g	d	64	0	0	T1cN0Mx
3364	A	1,9	0/20	d	d	43	1	0	T1cN0Mx
3448	A	1,5	0/22	h	f	55	0	0	T1cN0Mx
4110	A	1,7	0/14	h	d	59	0	1	T1cN0Mx
4111	A	2,0	0/24	d	d	30	0	0	T1cN0Mx
4379	A	1,5	1/35	h	h	78	0	0	T1cN1Mx

¹Identificação da amostra no Banco de Tumores do Hospital AC Camargo/Instituto Ludwig de Pesquisa sobre o Câncer

²Tipo Histológico:

A-carcinoma ductal invasivo

B-carcinoma lobular invasivo, com componente *in situ*

C-carcinoma infiltrante

³Presença de metastases em linfonodos axilares

⁴Status dos receptores de estrogênio (ER) e progesterona (PR) (detecção por meio de reação de imuno-histoquímica)

d-Negativo

e-Positivo em até 10% das células

f-Positivo de 10-50% das células

g-Positivo de 50-90% das células

h- positivo em mais de 90% das células

⁶Idade, em anos, da paciente ao diagnóstico

⁷Status de p53, determinado por imuno-histoquímica:

0-ausência de detecção

1-deteção

⁸Status de c-erbB-2, determinado por imuno-histoquímica:

0-ausência de detecção

1-deteção

⁹Estadiamento patológico por meio do sistema de classificação TNM

ND-não disponível

3.2 CULTURA DE CÉLULAS

O difícil acesso a tecidos tumorais faz com que a cultura de células derivadas de tumores humanos seja extensivamente usada como modelo experimental das doenças neoplásicas. Para esse projeto, foram utilizadas linhagens celulares derivadas de tecido de mama para padronização da técnica de RT-PCR em tempo real. Dessas linhagens, seis foram obtidas comercialmente da *ATCC (American Type Culture Collection - www.atcc.org)* apresentadas na Tabela 2. Além disso, foram adicionadas ao projeto duas linhagens epiteliais de tecido de mama imortalizadas cedidas de um estudo realizado por MACKAY et al. (2003). Nesse estudo, foram investigadas as mudanças no perfil de expressão gênica de células epiteliais luminiais de tecido de mama e variantes que apresentam de moderado a alto nível de expressão de *c-erbB2* usando-se cDNA *microarrays*. HB4a corresponde a uma linhagem celular epitelial luminal mamária imortalizada e HB4aC5.2 corresponde à mesma linhagem, porém transfectada com *c-erbB2*, que apresenta alta expressão desse gene.

As linhagens celulares foram cultivadas em meio apropriado, conforme descrito na Tabela 6, em estufa a 37°C e atmosfera de 5% de CO₂. Em média, as linhagens celulares foram replicadas a cada 2-4 dias e crescidas até uma confluência de aproximadamente 80% em garrafa para cultura de células com área de 75cm², que corresponde a aproximadamente 10⁶ células. Após o crescimento, as células foram tripsinizadas e centrifugadas (1.500 RPM, 4 minutos), o meio de cultura foi descartado e os sedimentos contendo as células foram estocados a -70°C.

Tabela 2 - Origem, nome, descrição e condições de cultura apropriadas das linhagens celulares de tecido de mama utilizadas no presente estudo.

Origem	Linhagem	Descrição	Meio de Cultura
ATCC	MDA-MB-468	Adenocarcinoma, glândula mamária, derrame pleural	RPMI 1640 (Invitrogen) + 10% de soro fetal bovino
	MDA-MB-435	Adenocarcinoma, glândula mamária, mama, derrame pleural	RPMI 1640 (Invitrogen) + 10% de soro fetal bovino
	MDA-MB-231	Carcinoma ductal, glândula mamária, ascite, epitélio	RPMI 1640 (Invitrogen) + 10% de soro fetal bovino
	MCF-7	Adenocarcinoma, glândula mamária, derrame pleural	McCoy's (Invitrogen) + 10% de soro fetal bovino
	SK-BR-3	Adenocarcinoma, glândula mamária, derrame pleural	Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen) + 10% de soro fetal bovino
MACKAY et al.(2003).	HB4a	Linhagem celular epitelial de tecido de mama	RPMI 1640 (Invitrogen) + 10% de soro fetal bovino + 5ug/mL de hidrocortisona e insulina
	HB4aC5.2	Linhagem celular epitelial de tecido de mama transfectada com <i>c-erbB2</i>	RPMI 1640 (Invitrogen) + 10% de soro fetal bovino + 5ug/mL de hidrocortisona e insulina

3.3 EXTRAÇÃO DE RNA

Foram extraídos RNAs das amostras de tecidos (tumoral e seu correspondente normal, quando disponível) e das linhagens celulares. A extração procedeu-se pelo Método de Sedimentação através de gradiente de Cloreto de Césio (CHRIGWIN et al. 1979).

As amostras de tecido com cerca de 0,5 mm³ e os sedimentos das linhagens celulares foram homogeneizados com auxílio de um homogeneizador *Power Gen™* em 9 mL de solução de lise (4M de Isotiocianato de Guanidina, 25mM de Citrato de Sódio, 0,1M de beta-Mercaptoetanol). O lisado foi transferido para um tubo de ultracentrífuga contendo 4mL de solução de Cloreto de Césio (5,7M CsCl, 1M Acetato de Sódio) e submetidos a centrifugação a 29.000 RPM, por 20 horas, a 20°C.

Após a centrifugação, foi removido o sobrenadante e o sedimento contendo RNA foi solubilizado em 100µL de água tratada com DEPC (di-etil pirocarbonato). A quantificação do RNA obtido foi realizada em espectrofotômetro (*GENEQUANTpro-Amersham Biosciences*).

3.4 AVALIAÇÃO DOS RNAs

Após a extração, todos os RNAs foram submetidos à análise de sua integridade em gel de agarose 1% contendo 20µg/mL de Brometo de Etídio. Para isso, foi utilizado 1µg de cada RNA, acrescentando-se igual volume de tampão desnaturante (2X TAE, 7M de uréia e 30% de glicerol). As amostras foram então incubadas a 65°C por 10 minutos e submetidas à eletroforese.

3.5 SÍNTESE DE cDNA

Após a extração e avaliação, os RNAs das amostras de tecido e das linhagens celulares, foi realizada a síntese de cDNAs. Para isso, foram utilizados 2 µg de RNA total, 0,5µg de *primer* (oligo(dT) *Invitrogen*), 0,5µM de cada dNTPs (*Invitrogen*) e água *RNase-free* (*Invitrogen*) até 12µL. A mistura foi incubada a 65°C durante 5 minutos e armazenada em gelo. Em seguida, foram adicionados 1X tampão *First Strand Buffer* (*Invitrogen*), 10mM de DTT (*Invitrogen*), 2U de *RNase OUT* (*Invitrogen*) e 10U da enzima *SuperScript II* (*Invitrogen*). A reação seguiu com incubação a 42°C por 1 hora e depois a 70°C por 15 minutos para a inativação da enzima.

3.6 AVALIAÇÃO DA QUALIDADE DOS cDNAs

Após a síntese, a qualidade dos cDNAs foi verificada por meio de RT-PCR utilizando *primers* desenhados na região 5' do gene *NOTCH2* (Direto- 5' TGTGGCCAACCAGTTCTCCT 3'; Reverso- 5' GGCAGTCATCAATATTCCTC 3') testando desta maneira a eficiência da reação de transcrição e a integridade do RNA. As seguintes condições de amplificação foram utilizadas: 1,0mM MgCl₂, 0,2μM dNTPs (*Invitrogen*), 0,4μM de cada *primer*, 1X tampão de reação 10X, 1U de *Platinum*® Taq DNA Polymerase (*Invitrogen*) e 5μL do cDNA diluído 1:10 como molde em um volume final de 25μL. O programa da PCR consistiu em 2 minutos iniciais de desnaturação a temperatura de 94°C, trinta e cinco ciclos repetidos com 1 minuto a 94°C, 1 minuto com a temperatura de 58°C e 1 minuto a 72°C, por fim, um ciclo único de 7 minutos a 72°C. O produto, um fragmento de 295 pares de base, foi visualizado por meio de eletroforese em gel de poliacrilamida 8%, com posterior coloração por prata (SANGUINETTI et al. 1994).

3.7 SELEÇÃO DOS GENES CANDIDATOS

Os genes diferencialmente expressos foram selecionados por meio do uso do banco de dados de SAGE disponível no sítio do *CGAP* (*Cancer Genome Anatomy Project* - <http://cgap.nci.nih.gov>).

Para isso, foram comparadas as frequências das *SAGEtags* depositadas no banco de dados de *SAGE* geradas a partir de tecido de mama neoplásica e não neoplásica, uma vez que o número de *SAGEtags* é diretamente proporcional ao nível

de expressão dos genes a que elas correspondem (VELCULESCU et al. 1995). Os genes com maior número de *SAGEtags* foram considerados genes mais expressos.

Esta análise foi realizada pelo Laboratório de Biologia Computacional do Instituto Ludwig sob a coordenação da Dra. Helena Brentani, utilizando-se os dados disponíveis no banco de dados de *SAGE* em agosto de 2002. Nesta época, este Banco contava com treze bibliotecas de tecido mamário, sendo sete de tecido tumoral e seis de tecido normal (Tabela 3). Para otimizar a identificação dos genes diferencialmente expressos foram adotados os critérios listados abaixo:

- a) não foram consideradas para análise as bibliotecas de *SAGE* construídas a partir de linhagens celulares;
- b) para evitar a presença de artefatos de técnica, só foram consideradas válidas as *SAGEtags* de cada transcrito que estavam representadas por no mínimo cinquenta *SAGEtags* em cada biblioteca;
- c) só foram consideradas válidas as *SAGEtags* presentes em 100% das bibliotecas;
- d) foram consideradas todas as *SAGEtags* que se apresentaram no mínimo quatro vezes mais expressas nas bibliotecas de *SAGE* de tecido tumoral do que nas bibliotecas de *SAGE* de tecido não tumoral.

A partir desta análise, foram selecionados 175 genes que apresentaram um nível de expressão mais elevado em tecido neoplásico de mama quando comparado ao tecido normal (Anexo 7).

Tabela 3 - Descrição das bibliotecas de mama disponíveis no Banco de Dados de *SAGE* do *CGAP* em maio de 2002 que foram utilizadas no presente estudo.

Nome da Biblioteca	Tecido de origem	Histologia	Informações Adicionais
SAGE_95-259	Glândula mamária	Câncer	Carcinoma ductal invasivo; paciente com 56,5 anos de idade ao diagnóstico; tumor primário com 3,0 cm; grau histológico =3; grau nuclear=3; 10 linfonodos positivos, negativo para a imunexpressão de receptor de estrógeno e progesterona
SAGE_95-260	Glândula mamária	Câncer	Tumor metastático (metástase em linfonodo de um tumor de mama primário); paciente com 56,5 anos de idade ao diagnóstico; tumor primário de 3,0 cm; grau histológico =3; grau nuclear = 3; 10 linfonodos positivos; negativo para a imunexpressão de receptor de estrógeno e progesterona
SAGE_95-347	Glândula mamária	Câncer	Carcinoma ductal invasivo; paciente com 52,8 anos de idade ao diagnóstico; tumor primário com 3,2 cm; grau histológico =3; grau nuclear =3,0; 16 linfonodos positivos; positivo para imunexpressão de receptores de estrógeno e progesterona
SAGE_95-348	Glândula mamária	Câncer	Tumor metastático (metástase em linfonodo de um tumor de mama primário); paciente com 52,8 anos de idade ao diagnóstico; tumor primário de 3,2 cm; grau histológico =3; grau nuclear = 3; 16 linfonodos positivos; negativo para a imunexpressão de receptor de estrógeno e progesterona
SAGE_DCIS	Glândula mamária	Câncer	Carcinoma ductal <i>in situ</i> , comedo; paciente com 67 anos; tumor de alto grau; paciente submetida a mastectomia parcial; positivo para imunexpressão de receptor de estrógeno e progesterona
SAGE_DCIS_2	Glândula mamária	Câncer	Carcinoma ductal <i>in situ</i> , células purificadas
SAGE_Duke_96-349	Glândula mamária	Câncer	Informação não disponível ¹
SAGE_Br_N	Glândula mamária	Normal	Epitélio mamário luminal normal, células purificadas de paciente com 36 anos de idade
SAGE_Duke_40N	Glândula mamária	Normal	Informação não disponível ¹
SAGE_Duke_48N	Glândula mamária	Normal	Informação não disponível ¹
SAGE_breast_hyperplasia	Glândula mamária	Normal	Epitélio Mamário normal de uma paciente com 31 anos de idade, portadora de câncer de mama na vigésima semana de gravidez
SAGE_breast_myoepithelial	Glândula mamária	Normal	Células mamárias mioepiteliais normais, obtidas a partir de cirurgia de redução de tecido de mama de uma paciente com 27 anos de idade
SAGE_mammary_epithelium	Glândula mamária	Normal	Células mamárias epiteliais luminais normais, obtidas do tecido de uma paciente de 43 anos de idade, submetida a mamoplastia de redução.

¹informação não disponível, biblioteca retirada do banco de dados de *SAGE* do *CGAP*.

Devido ao grande de número de genes resultantes da pesquisa por *SAGE* (175 genes), seria praticamente impossível avaliar a expressão de todos eles por meio de RT-PCR em tempo real. Sendo assim, uma análise manual deste grupo de genes foi realizada utilizando-se diversas ferramentas *online* (Tabela 4). A escolha dos genes-candidatos é um passo crucial para a execução do trabalho. Distinguir quais são os critérios mais adequados para a seleção desses genes é uma tarefa difícil, além disso, o processo de validação é demorado e dispendioso. Para se restringir o número de genes-candidatos a serem avaliados foram adotados os seguintes critérios:

- a) processo biológico que o gene está envolvido: preferencialmente, foram escolhidos genes envolvidos na proliferação, morte ou adesão celular;
- b) nível de expressão do transcrito (este valor foi atribuído conforme o número de *ESTs* do gene presentes no Banco de dados Unigene): foram escolhidos genes com maior nível de expressão avaliado pelo número de *ESTs* de cada gene depositadas no banco de dados Unigene (CAMARGO et al. 2001);
- c) Na comparação dos dados de *ESTs* e de *SAGEtags* em mama neoplásica e não neoplásica, foram escolhidos genes que apresentavam preferencialmente, uma concordância entre os dados fornecidos por meio destas duas técnicas de análise da expressão;
- d) verificação da especificidade tecidual (dados de expressão de *SAGE* em outros tecidos): esta análise foi realizada por meio de ferramentas disponíveis na interface gráfica *SAGEgenie* (www.cgap.nci.gov/SAGE) para identificar quais genes eram preferencialmente expressos em tecido

de mama e que potencialmente poderiam representar antígenos de diferenciação e úteis como potenciais alvos terapêuticos;

- e) dados de literatura: foi realizada pesquisa sobre o envolvimento do candidato com algum tipo de tumor ou no processo de tumorigênese, exceto de mama;
- f) localização celular: foram preferencialmente selecionados os genes que codificam proteínas localizadas na membrana celular por serem proteínas dessa localização mais úteis como alvos para anticorpos terapêuticos;
- g) presença de anticorpo comercial disponível para posterior validação do aumento de expressão através de ensaios de imuno-histoquímica.

Após a análise do grupo de 175 genes segundo estes critérios, foram selecionados dezessete genes para validação por meio de RT-PCR em tempo real, do aumento da expressão gênica sugerido pelos dados de *SAGE* em amostras tumorais e não tumorais de tecido de mama. Destes genes, nove fizeram parte do trabalho de iniciação científica da aluna Barbara Pereira de Mello (CNPQ-107374/2003-0), sob orientação do Dr. André Luiz Vettore e co-orientação da Dra. Otávia Caballero, como um subprojeto do presente estudo, no Laboratório de Genética do Câncer. Além disso, foram testados dois genes adicionais (*ALPPL2* e *GATA4*), dos quais já havia sondas e *primers* disponíveis no laboratório e se demonstraram como interessantes, segundo a análise realizada.

Tabela 4- Informações sobre os *web-sites* da *internet* utilizados para o processo de seleção dos genes candidatos.

Banco de Dados	Endereço do <i>web-site</i>	Principal Função
ABCAM	www.abcam.com	comercialização de anticorpos
CGAP	www.cgap.nci.nih.gov	dados de expressão de <i>SAGE</i> e <i>ESTs</i>
Dako	www.dakousa.com	comercialização de anticorpos
GeneCards	bioinfo.weizmann.ac.il/cards/index.html	informações gerais
LocusLink	www.ncbi.nlm.nih.gov/LocusLink	informações gerais
NCBI	www.ncbi.nih.gov	informações gerais
OMIM	www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM	catálogo de genes humanos e desordens genéticas
PubMed	www.ncbi.nlm.nih.gov	informações da literature científica
QED	www.qedbio.com	comercialização de anticorpos
SageGenie	http://cgap.nci.nih.gov/SAGE	dados de expressão de <i>SAGE</i> , diversas ferramentas disponíveis para análise de transcritos
SciQuest	www.sciquest.com	comercialização de anticorpos
Santa Cruz	www.scbt.com	comercialização de anticorpos
UniGene	www.ncbi.nlm.nih.gov/UniGene	informações gerais e dados de expressão em <i>ESTs</i>
AceView	www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?c=locusid&org=9606&l=5584	informações gerais
PSORT II	http://psort.nibb.ac.jp/form2.html	predição de localização do produto gênico
TMHMM	www.cbs.dtu.dk/services/TMHMM/	predição de existência de domínio transmembrana

3.8 QUANTIFICAÇÃO DA EXPRESSÃO GÊNICA POR MEIO DE RT-PCR EM TEMPO REAL

A RT-PCR em tempo real avalia o acúmulo do produto na fase logarítmica da reação de amplificação que está diretamente relacionado à quantidade de molde existente no início da reação, sendo atualmente considerado um método bastante preciso e reprodutível para a quantificação da expressão gênica. De maneira geral, a quantificação da expressão gênica por meio de RT-PCR em tempo real pode ser realizada por meio de dois sistemas de detecção de fluorescência. Um sistema refere-

se ao uso de um fluoróforo que se intercala a dupla fita de DNA resultando em emissão de fluorescência, conhecido como Sistema *SYBR Green* de detecção. O outro sistema refere-se ao uso de sondas fluorescentes de hibridização conhecidos como *Molecular beacon*, sondas de hidrólise (ou *TaqMan*) e sondas de hibridização. Para a execução deste trabalho foram escolhidos os sistemas *TaqMan* e *SYBR Green* de detecção.

A amplificação baseada no sistema *TaqMan*[®] (*Applied Biosystems*) de detecção requer a utilização de uma sonda de hibridização interna a um par de *primers*. A sonda é um oligonucleotídeo que contém na extremidade 5' uma molécula fluorescente (*reporter*) e na extremidade 3' uma outra molécula “receptora” de fluorescência (*quencher*). Enquanto a sonda está intacta, a proximidade ao *quencher* reduz, intensamente, a fluorescência emitida pelo *reporter*. Durante a fase de extensão da reação de PCR, se a seqüência alvo estiver presente, a sonda irá se parear entre os sítios dos *primers* e será clivada graças a atividade 5'-3' exonucleásica da *Taq* DNA polimerase. Essa clivagem, separa o *reporter* do *quencher* que permite a emissão de sinal livre de fluorescência sendo então detectado. A cada ciclo, moléculas adicionais de *reporter* vão sendo clivadas das suas respectivas sondas e o resultado final é um aumento da intensidade de fluorescência proporcional à quantidade do produto de PCR acumulado. O resultado é dado em *Ct* (*Threshold Cycle* ou ciclo limite) ou também chamado de *Cp* (*Crossing point* ou ponto de cruzamento), parâmetro definido como o número do ciclo em que a fluorescência gerada pela clivagem do *reporter* atinge um limite acima da fluorescência de fundo (LIVAK et al. 1995).

O sistema *TaqMan* de detecção é altamente específico. Para a utilização desta metodologia, foram sintetizadas sondas por meio do serviço *Assay By Design-Applied Biosystems*, no qual as sondas são desenhadas na junção entre dois *exons* e internas a um par de *primers* também desenhados em *exons* diferentes. Sendo assim, a alta especificidade conferida por este método de quantificação exige que apenas alguns experimentos iniciais sejam necessários para a padronização da reação de amplificação para a quantificação da expressão gênica em tempo real. Além disso, uma outra vantagem desta metodologia de detecção está na possibilidade de se realizar reações em sistema *multiplex* que permite a quantificação de mais de um produto em uma mesma reação de amplificação.

O sistema de detecção por *SYBR Green* é baseado na utilização de uma molécula fluorescente denominada *SYBR Green I* que, quando intercalada à dupla fita de DNA, passa a emitir fluorescência de forma detectável. Durante os ciclos iniciais da reação de PCR o sinal de fluorescência emitido pelo *SYBR Green I* é fraco para ser detectado (não ultrapassando o sinal de fluorescência de fundo). Entretanto, no decorrer dos ciclos da reação de PCR, há o aumento do produto amplificado e conseqüentemente o aumento do sinal de emissão de fluorescência que passa então a ser detectável. Desta forma, os valores quantificados são obtidos no ponto no qual a amplificação do produto foi detectada, diferente da reação de PCR convencional, na qual a detecção acontece depois de um número fixo de ciclos, o que representa a quantidade final de cada produto acumulado. Nesta metodologia, a quantificação da expressão gênica ocorre da mesma forma que a anterior (baseada no parâmetro de *Cts* ou *Cps*) e é tão eficiente quanto o sistema de detecção *TaqMan*. Porém, exige cuidadosa padronização, uma vez que o *SYBR Green I* se intercala a qualquer

molécula de dupla fita presente na reação. Portanto, qualquer produto inespecífico pode ser amplificado e quantificado. A especificidade da detecção de fluorescência com *SYBR Green I* pode portanto ser comprometida pela formação de dímeros de *primers*, pela concentração inadequada de *primers* e pela formação de produtos de amplificação inespecíficos. Todos estes fatores levam à formação de moléculas de dupla fita não esperadas, as quais incorporam o *SYBR Green I* e têm sua fluorescência registrada. Sendo assim, o desenho dos *primers* e a otimização da reação de PCR são passos cruciais para todo o processo de quantificação por meio de RT-PCR em tempo real.

Neste método de detecção, a confirmação da especificidade da amplificação é realizada por meio da análise da curva de desnaturação ou *melting curve*, onde são relacionadas a fluorescência e a temperatura. Esta curva consiste da monitorização da fluorescência das amostras em relação ao aumento contínuo da temperatura, o que determina a temperatura de fusão ou *melting temperature* (T_m) de cada fragmento resultante da reação de amplificação. Cada fragmento amplificado possui um T_m específico o que permite a diferenciação entre os produtos resultantes. A Figura 1 representa a curva de desnaturação para uma determinada reação de amplificação em condições não ideais, em que ocorre a formação de produtos não específicos representados graficamente pelos diversos picos na curva de desnaturação. Já a Figura 2 representa uma reação de amplificação ideal, devidamente padronizada resultando em uma reação altamente específica representada graficamente pelas curvas de desnaturação com um único pico (ou seja, um único T_m).

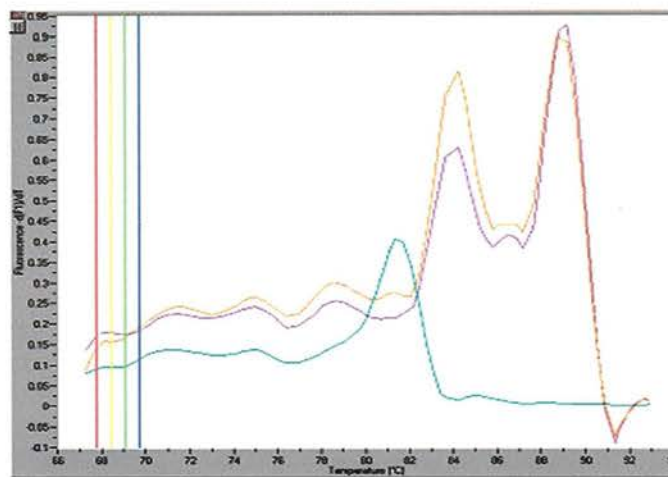


Figura 1 - Representação gráfica da curva de desnaturação de uma reação de amplificação por meio de RT-PCR em tempo real para a análise da expressão de um determinado gene em condições não ideais de amplificação, resultando em produtos não específicos de reação (representados pelos vários picos de T_m).

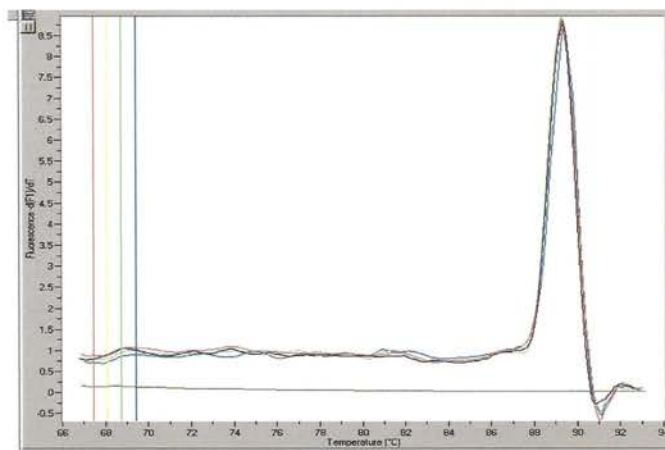


Figura 2 - Representação gráfica da curva de desnaturação de uma reação de amplificação por meio de RT-PCR em tempo real para a análise da expressão de um determinado gene em condições ideais de amplificação, resultando em um único produto, específico de reação. (representados por um único pico de T_m).

3.9 ESCOLHA DO GENE NORMALIZADOR E O CÁLCULO DA EXPRESSÃO RELATIVA

Em experimentos de quantificação da expressão gênica entre duas populações de amostras diferentes, deve-se sempre levar em consideração a existência de erros na quantificação do RNA, bem como variações na eficiência da reação de transcrição deste RNA para cDNA. Conseqüentemente, esta questão faz surgir um outro importante ponto no aspecto do desenho experimental: a escolha de um gene apropriado cuja expressão não varie ou varie muito pouco, entre as amostras, para ser usado como um gene normalizador dessas variações.

O método de escolha de um gene normalizador não é freqüentemente descrito em trabalhos de análise do perfil da expressão gênica, o que torna ainda mais complexo o processo de seleção. No presente estudo, esta escolha foi inicialmente realizada por meio da análise em bancos de *SAGE* de mama neoplásica e não neoplásica de alguns dos genes humanos descritos como apresentando expressão constitutiva (Tabela 5), com a colaboração do Laboratório de Biologia Computacional do Instituto Ludwig de Pesquisa sobre o Câncer sob a supervisão da Dra. Helena Brentani, por meio da utilização de um *pipeline* desenvolvido para a análise de comparações da expressão gênica dos dados disponíveis nas bibliotecas de *SAGE*, chamado de *SAGEβBin*. Para isso, foram comparadas as freqüências das *SAGETags* destes genes entre as bibliotecas de mama de tecidos tumorais e não tumorais e calculadas as áreas das sobreposições das distribuições destas freqüências, que correspondem ao valor de E (*bayes error rate*). Genes cujo valor obtido E fosse igual ou próximo a um foram considerados bons candidatos a normalizadores, uma

vez que não apresentaram diferenças significativas na frequência de expressão nas duas populações diferentes (no caso, tecidos neoplásicos e não neoplásicos). Um valor arbitrário de 0,3 foi usado para a escolha dos normalizadores. Genes com valores de E menores que 0,3 não foram considerados bons normalizadores, mas sim genes com expressão diferencial.

Além disso, foram analisadas as *SAGEtags* referentes a cada um dos genes quanto a sua "confiabilidade", adotando-se os seguintes critérios:

- a) utilizar sempre a primeira *SAGEtag* localizada anterior ao sítio da enzima de restrição Nla III, a partir da extremidade 3' do transcrito;
- b) Observar a presença de um sinal de poliadenilação próximo à *SAGEtag*;
- c) Observar se a *SAGEtag* está localizada fora de regiões repetitivas do genoma (por exemplo, seqüências Alu).

Dos genes analisados, apenas quatro deles apresentaram *SAGEtags* confiáveis.

Um segundo método para a escolha do gene normalizador também foi adotado. Esta outra metodologia correspondeu à utilização de um ensaio comercial para realização de reação de amplificação por RT-PCR em tempo real por meio do sistema *TaqMan* de detecção de fluorescência denominado de *TaqMan Human Endogenous Control Plate (Applied Biosystems)*, o qual consiste em uma placa de 96 poços para reação de amplificação. Nesta placa estão contidas as sondas e os *primers* correspondentes a onze genes humanos cuja expressão é comumente descrita como sendo constitutiva e um controle positivo interno de reação (*IPC- Internal Positive Control*). Cada um dos ensaios correspondentes aos genes estão dispostos em cada uma das colunas da placa em oito repetições, conforme representado na Figura 3a e b. Com exceção do 18SrRNA, todos os demais ensaios presentes nesta placa de reação

são cDNAs específicos. A coluna correspondente ao ensaio com o 18SrRNA foi descartada de nosso experimento, pois para utilizá-la teríamos que sintetizar cDNA a partir de *primers* aleatórios, o que seria impossível, pois, a síntese dos cDNAs utilizados neste estudo já havia sido realizada a partir de *primer* oligo dT. Este experimento foi realizado conforme os seguintes passos:

- a) Primeiramente foi realizada a síntese de cDNA a partir de um RNA calibrador (que faz parte dos reagentes que compõem o *kit* de reação deste ensaio), o qual fornece uma linha de base para a comparação com as outras amostras usadas na reação. Serve também como base para comparar os dados de uma amostra de múltiplos experimentos independentes (mas somente se o mesmo calibrador e as mesmas amostras forem utilizadas em todos os experimentos);
- b) Em seguida, foi realizada uma mistura de cinco cDNAs de amostras de tumores de mama a serem utilizadas neste trabalho e também cDNAs sintetizados a partir de RNA comercial obtidos de amostras normais de tecido de mama (*Clontech*), constituído de uma mistura de amostras de duas pacientes;
- c) A reação foi realizada com as seguintes condições: 25uL de *2X TaqMan Universal PCR Master Mix (Applied Biosystems)*, 4uL da mistura de cDNA em um volume final de 50uL (para cada uma das amostras a serem testadas, ou seja, para o calibrador, para a mistura de amostras tumorais e para a mistura de amostras normais). As condições de ciclagem foram: dois minutos iniciais a 50°C, seguidos de dez minutos de desnaturação

inicial a 95°C, quarenta ciclos com quinze segundos a temperatura de 95°C e um minuto a 60°C.

A análise dos dados obtidos a partir da reação de RT-PCR em tempo real deste ensaio foi realizada seguindo-se as instruções do fabricante. Para isso, foi calculado o ΔC_t na mistura referente às amostras tumorais e na mistura referente às amostras normais de tecido de mama. O ΔC_t equivale à diferença entre a média dos C_t s obtidos do gene de interesse no RNA calibrador que faz parte do *kit* e a média dos C_t s obtidos deste mesmo gene nas amostras testadas (tumoral e normal). Após definidos os valores de ΔC_t , é calculada a diferença entre a variação dos ΔC_t s de cada gene candidato a normalizador, entre as amostras analisadas. Os genes que apresentarem os valores que correspondem à menor diferença de variação são considerados bons normalizadores.

Tabela 5 - Genes humanos mais comumente usados como constitutivos, selecionados para a análise por *SAGE* do controle endógeno mais adequado para realização da quantificação relativa da expressão gênica por meio de RT-PCR em tempo real.

Símbolo do Gene	Nome do Gene	Número de Identificação*
<i>ACTB</i>	beta-actina	NM_001101
<i>B2M</i>	beta-2-microglobulina	NM_004048
<i>GAPDH</i>	gliceraldeído 3-fosfato desidrogenase	NM_002046
<i>HPRT1</i>	hipoxantina fosforibotrasferase 1	NM_000194
<i>PPIA</i>	<i>peptidylprolyl</i> isomerase A (ciclofilina A)	NM_021130
<i>PGK1</i>	fosfoglicerato quinase 1	NM_000291
<i>RPLP0</i>	proteína ribossomal, P0	NM_001002
<i>TBP</i>	<i>TATA box binding protein</i>	NM_003194
<i>TFRC</i>	receptor de transferrina	NM_003234

*Número de identificação no *GenBank* (www.ncbi.nih.gov)

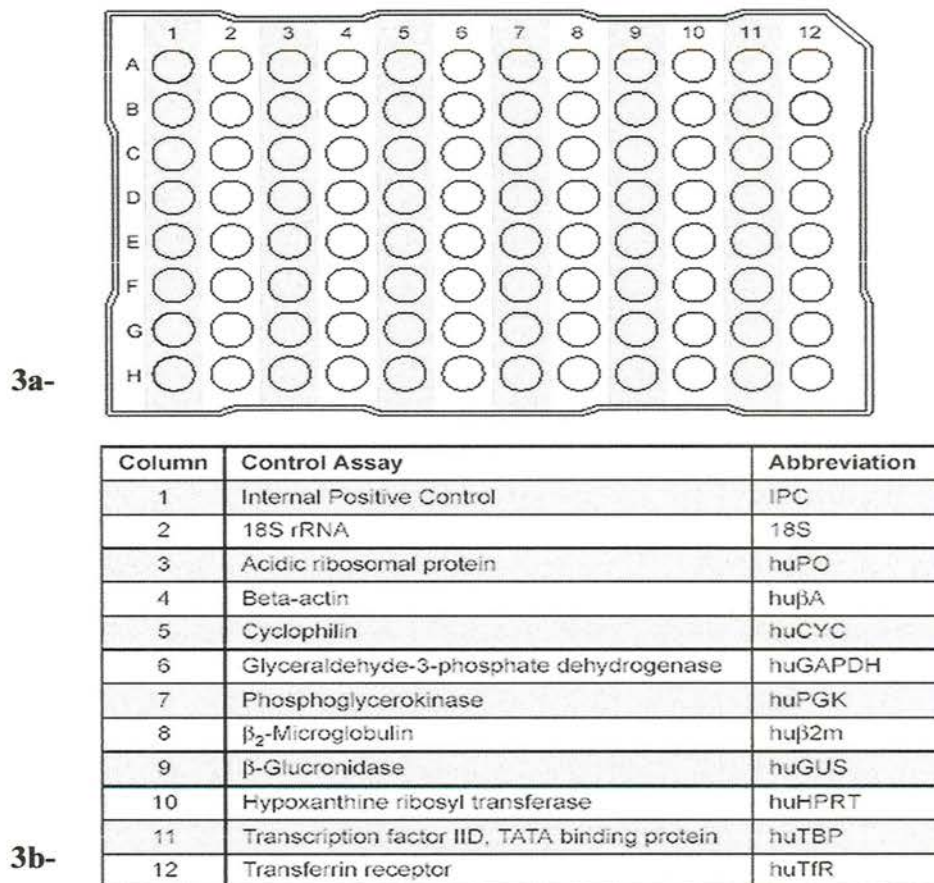


Figura 3 - Representação do ensaio para RT-PCR em tempo real da *TaqMan Human Endogenous Control Plate* (Applied Biosystems). **3a-** Placa com 96 poços para reação de RT-PCR em tempo real. **3b-** Esquema da distribuição das sondas e *primers* nas colunas da placa.

A quantificação relativa da análise da expressão gênica usando reação de amplificação por meio de RT-PCR em tempo real foi calculada pela equação $2^{-\Delta\Delta Ct}$, descrita em detalhes por LIVAK e SCHMITTGEN (2001). O valor inferido à ΔCt equivale à diferença entre a média dos *Cts* do gene de interesse e à média dos *Cts* do gene normalizador. Já o cálculo da fórmula $\Delta\Delta Ct$ envolve a subtração entre o valor de ΔCt para cada amostra no tecido neoplásico e o valor de ΔCt para o tecido não neoplásico. O resultado é expresso no número de vezes que um gene no tecido tumoral tem a sua expressão aumentada ou diminuída em relação ao tecido normal.

3.10 DESENHO DAS SONDAS E *PRIMERS* E PADRONIZAÇÃO DAS REAÇÕES PARA QUANTIFICAÇÃO DA EXPRESSÃO GÊNICA POR MEIO DE RT-PCR EM TEMPO REAL

Como já mencionado, a expressão dos genes selecionados foi avaliada por meio de dois sistemas de detecção de fluorescência. Para os genes quantificados pelo sistema sistema *TaqMan* de detecção de fluorescência foi utilizado o aparelho *ABI*[®] *7000 Sequence Detection System (Applied Biosystems)*. Para isso, foram sintetizadas sondas por meio do serviço *Assay By Design-Applied Biosystems*. As sondas foram desenhadas na junção de *exons* e internas a um par de *primers*, garantindo assim que nenhum fragmento oriundo de DNA genômico fosse amplificado e quantificado (Tabela 6). Sendo este sistema de detecção altamente específico, somente alguns experimentos para a padronização referente a quantidade de molde de cDNA a ser utilizado na reação foram necessários. As sondas para todos os genes candidatos foram marcadas na extremidade 5' com FAM como *reporter* e na extremidade 3' com *quencher* não fluorescente. Para o gene usado como normalizador (*β-actina*), fez-se uso de sonda marcada com *reporter* VIC e *quencher* TAMRA, o que permitiu a utilização de reações em sistema *multiplex*, em que um determinado gene candidato e o gene normalizador foram dosados em uma única reação de amplificação.

Existem relativamente poucos trabalhos que descrevem a utilização de reações em *multiplex* neste tipo de sistema. O sistema *TaqMan* tem sido utilizado com sucesso na quantificação de três ou mais genes alvo, sem aparente redução na sensibilidade comparada com reações individuais (SHARMA et al. 1999; SHIN et al. 1999). Para otimização das reações de PCR foi primeiramente padronizada qual seria a quantidade de molde de cDNA apropriada para os experimentos. Para isso, os

cDNAs resultantes da reação de transcrição foram diluídos dez vezes e destes foram utilizados 0,5µL, 1,0µL, 2,0µL, 4,0µL e 8,0µL em uma reação com 1X *TaqMan*[®] *Universal PCR Master Mix (2X- Applied Biosystems)*, 1X *Mix Assays-By-Design*TM *Products (20X- Applied Biosystems)* em um volume final de 50µL. As condições de ciclagem foram: 2 minutos iniciais a 50°C, seguidos de 10 minutos de desnaturação inicial a 95°C, quarenta ciclos com 15 segundos a temperatura de 95°C e um 1 a 60°C.

Após esta etapa, foi realizada uma outra reação de amplificação sob as mesmas condições de ciclagem anteriores, porém para um volume final de 25µL. Este experimento, demonstrou que mesmo utilizando-se a metade dos reagentes os resultados obtidos foram os mesmos, com isto, pode-se reduzir pela metade as quantidades dos reagentes utilizados para a reação de quantificação da expressão gênica. Em seguida foram realizados experimentos para a padronização do sistema de reação de amplificação em *multiplex*. Para isso, foram feitos experimentos nos quais foram quantificadas a expressão de um determinado gene candidato e do gene normalizador em paralelo, porém em uma mesma reação em que também foram dosadas a expressão deste mesmo gene candidato e do mesmo gene normalizador em um mesmo tubo de reação (*multiplex*). As condições de reação e de ciclagem foram as mesmas descritas acima, padronizadas para reações com volume final de 25µL. Desta forma, para os genes em que a expressão foi avaliada por meio do sistema *TaqMan* de detecção de fluorescência, as reações foram padronizadas para o volume final de 25µl em sistema *multiplex* utilizando-se 2,0µL de cDNA como molde.

Tabela 6 - Seqüências dos *primers*, das sondas, e o tamanho esperado dos fragmentos de cada gene selecionado, por meio do serviço *Assay By Design-Applied Biosystems*.

Símbolo do Gene	Seqüência do <i>primer</i> Direto (5'-3')	Seqüência do <i>primer</i> Reverso (5'-3')	Seqüência da Sonda (5'-3')	Tamanho do Fragmento (pb)
<i>ALPPL2</i>	GGAAGAATCTGGTGCAGGAATGG	GAGCTCAGTGCGGTTCCA	CCAGGGTGCCCGGTACG	71
<i>BIRC5</i>	CTGGACAGAGAAAGAGCCAAGA	CACTTTCTTCGCAGTTTCCTCAAAT	TTTCCTTTGCAATTTTG	84
<i>CAD</i>	CGCCCCGAATCCATCGA	GCTACCTTCCGAGAGGACTTCT	CCTCAGCTGGCAAACC	134
<i>CELSR 2</i>	TCTTCCACTACCTCTTTGCTACCT	AGCACCACATAGGAGAGGAAGAT	CATCCAGGGCCCCCTTC	70
<i>CTSF</i>	CAATGCCTACTCGGCCATAAAG	TGGTAGCTGTAGTCATCCTCTGT	TCCAGCCCTCCCAAATT	64
<i>EFNA 4</i>	GCTTTGAGTTCTTACCTGGAGAGA	CAAGCACTGGCCAGAACTCT	CATCTCGGTGCCCACTCC	79
<i>GATA 3</i>	GCCACGGTGCAGAGGTA	AGGGTAGGGATCCATGAAGCA	CCACCACGGGAGCCAG	79
<i>GATA 4</i>	GGAAGCCCAAGAACCTGAATAAATC	CGGGAGGAAGGCTCTCACT	CACCAGCAGCTCCTTC	69
<i>ZC3HDC</i>	TGTGACGACCATGCAAACCT	GGGCAGAGGAGTCCCAATAGT	CCCGAAGAGCATCCCAG	77
<i>KIAA0141</i>	GCTGTGAAATATCTTTGGCTTGCA	AGGCCTTTCTCATAGCAAATTCCAA	CTGCTCTGTGAGTCCCCA	84

Para a análise da expressão dos genes por meio do sistema *SYBR Green* de detecção, foi utilizado o aparelho *LightCyclerTM* (*Roche Diagnostics*) em volume final de 20 μ L de reação. Para isso foram desenhados *primers* em exons diferentes, garantindo desta maneira que nenhum produto de DNA genômico fosse amplificado (Tabela 7). Como mencionado anteriormente, no sistema *SYBR Green* existe a necessidade de otimização da reação para a obtenção de produtos de amplificação únicos e específicos. As condições da reação de amplificação foram então otimizadas para cada um dos pares de *primers* por meio de RT-PCR convencional. Para a realização das reações de amplificação foram utilizadas as seguintes condições: 1X de tampão Taq Platinum, 2,0mM de MgCl₂, 2U de *Taq Platinum DNA polimerase* (*Invitrogen*), 0,4 μ M de cada *primer*, 5% de DMSO (*Sigma Chemical Co.*), 0,2mM de dNTPs (*Invitrogen*) e 5 μ L de cDNA diluído 1:10 como molde, em volume final de 25 μ L. O programa da PCR apresenta 2 minutos iniciais de desnaturação a temperatura de 95°C, trinta e cinco ciclos repetidos com 1 minuto a 95°C, 1 minuto com a temperatura de *annealing* específica para cada par de *primer* e 1 minuto a 72°C e, para finalizar, 7 minutos a 72°C. O produto amplificado foi submetido a eletroforese em gel de agarose 1,5% para visualização dos produtos resultantes.

A Figura 4a representa a amplificação do gene *STARD10* (293pb) por meio de reação de PCR convencional, utilizando-se como molde cDNAs provenientes de duas linhagens celulares de mama neoplásica (MDA-MB-435 e MDA-MB-436). Como pode ser observado, as condições de amplificação não foram ideais, pois mais de um produto de amplificação foi detectado. Sendo assim, algumas modificações como por exemplo, aumento da temperatura de *annealing* foram realizadas e um produto específico de amplificação, foi então obtido (Figura 4b).

Após a aquisição de um produto único e específico, foram realizadas as reações de RT-PCR em tempo real. As condições destas reações foram modificadas das descritas por VOGELSTEIN e KINZLER (1999): foram utilizados, por capilar, 1X Tampão *Taq Platinum* (*Invitrogen*), 2mM de $MgCl_2$, 0,2mM de dNTPs, 0,2 μ M de cada *primer*, 0,75U de *Taq Platinum* DNA polimerase (*Invitrogen*), 5% de DMSO (*Sigma Chemical Co.*), 0,5 μ g de BSA (*Promega*), 0,1 μ L de *SYBR Green*[®] I diluído 100 vezes em água estéril, 2 μ L de cDNA diluído 1:10, em volume final de 20 μ L de reação. As condições utilizadas em um primeiro experimento foram: desnaturação inicial à 95°C por 2 minutos, amplificação com quarenta e cinco ciclos, 94°C por 15 segundos, a temperatura de *annealing* previamente determinada para cada par de *primer* por 30 segundos, 72°C por 30 segundos e temperatura de aquisição a 79°C. Para análise da curva de desnaturação, utilizou-se o seguinte programa: 95°C por 0 segundos, 65°C por 15 segundos e novamente 95°C por 0 segundos. Para término e esfriamento da reação, a temperatura chega a 40°C por 30 segundos. Os resultados foram avaliados e as temperaturas de *annealing* e de aquisição da fluorescência foram ajustadas para cada par de *primers*, conforme mostrado na Tabela 7.

Tabela 7 - Seqüências dos *primers* específicos de cada gene selecionado, tamanho do fragmento esperado, temperatura de *annealing* na PCR e temperatura de aquisição da fluorescência para realização das reações de amplificação por RT-PCR em tempo real.

Símbolo do Gene	Seqüência do primer Direto (5'-3')	Seqüência do primer Reverso (5'-3')	Tamanho do Fragmento (pb)	Temperatura de <i>Annealing</i> (°C)	Temperatura de Aquisição de Fluorescência (°C)
<i>DGCR6</i>	GGCGGGGGTGGCTGGCTTCT	GGACAGGGCTGCCTTTCTGG	182	68	86
<i>EI24</i>	GCAAAGACCCTGGCAAAGC	CAGCTTCTTCCAATCCCGCCC	229	62	82
<i>FN3KR3</i>	GGCAGGAGGACTGGGTCGTG	GCCAGCTCATATTCCGAGT	265	66	82
<i>HMG20B</i>	ACTCTGGACTTCTACATGGC	GGCTGTGAGGTTCTTCTGGG	357	66	88
<i>PTPRF</i>	TGCGTACACAGCGTCCTGC	GGGTCTGGTGGACAGAAGGG	343	62	85
<i>QSCN6</i>	CGCCGCTGTCCTCTGGCTCT	CCAGGATGATGTTGCTTGGGG	215	64	85
<i>SCARB2</i>	GGGTGTGTTCTTTGGTTTGG	CAGCCATGTCAGCCTGCTC	279	62	82
<i>STARD10</i>	CTTCTCAGTTCCTGGCTCCC	TCCGTCCCTGAAGCGGTGCG	293	68	89
<i>KIF22</i>	GGAGAAGGCTGAGGACTGCT	AAAGGCGGAGTGAGGAGACG	357	62	85
β -actina	AAATCTGGCACCACACCTTC	CACTGTGTTGCCGTACAGGT	644	64	82

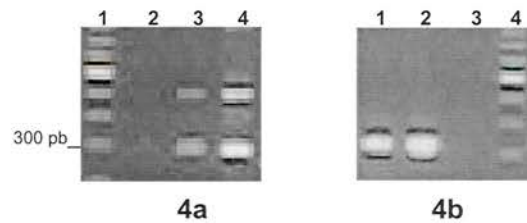


Figura 4a - Amplificação do gene *STARD10* (293 pb) utilizando as condições padrão de amplificação (conforme descritas acima). Para esta reação foi utilizado 62°C como temperatura de *annealing*. Canaleta 1- 100pb DNA Ladder (Invitrogen); canaleta 2- controle negativo (a mesma reação de PCR mas sem a adição de cDNA); canaleta 3- linhagem celular MDA-MB-435; canaleta 4- linhagem celular MDA-MB 436.

4b- Reação de amplificação do gene *STARD10* (293 pb) em condições de amplificação modificadas a partir da reação original. Para esta reação foi utilizado 68°C como temperatura de *annealing*. Canaleta 1-linhagem celular MDA-MB-435; canaleta 2- linhagem celular MDA-MB 436; canaleta 3- controle negativo (a mesma reação de PCR mas sem a adição de cDNA); canaleta 4- 100pb DNA Ladder (Invitrogen).

3.11 ANÁLISE ESTATÍSTICA

Estatísticas descritivas foram utilizadas para caracterizar a amostra. Para as variáveis numéricas ou contínuas as medidas de tendência central e de variabilidade foram aplicadas e a distribuição de frequências para as categóricas. Para a associação entre a expressão dos genes e as variáveis categóricas (*status* para receptor hormonal de estrogênio e progesterona) foi utilizado o teste não paramétrico U de Mann-Whitney devido a casuística, com nível de significância de 5%.

4 RESULTADOS

4.1 CASUÍSTICA E OBTENÇÃO DAS AMOSTRAS

Para a obtenção dos dados clínico-patológicos das amostras correspondentes aos tumores de mama em estágio inicial de desenvolvimento (menores do que dois centímetros), foi realizado um levantamento em cerca de 200 prontuários do Serviço de Arquivo Médico e Estatístico (SAME) do Hospital do Câncer A.C. Camargo. Foram anotadas as características das amostras como o tamanho da lesão, tipo histológico, grau tumoral, presença de metástases em linfonodos, presença de metástases à distância, *status* de receptores hormonais, imunoe expressão de p53 e c-erbB2. A revisão de todos os aspectos, bem como a classificação histopatológica (pTNM) foi realizada pelo Dr. Hugo Campos do Departamento de Anatomia Patológica do Hospital do Câncer A.C. Camargo .

4.2 EXTRAÇÃO DE RNA

No total foram selecionadas vinte e duas amostras de tecido de mama neoplásica para extração de RNA. Sendo que destas, as massas totais dos RNAs extraídos variaram de 0 a 112 μ g. O tamanho dos tumores variou de 0,7 a 2,0 cm. Foram selecionadas também onze amostras de tecidos mamários normais pareadas às tumorais e destas, as massas de RNA obtidas variaram de 0 a 725 μ g. Foram também obtidas onze amostras de tecidos de mama não neoplásica não pareados aos tecidos

tumorais e nestas, a variação da massa dos RNAs extraídos foi de 0 a 424 μ g. Para as linhagens celulares as massas de RNA obtidas foram de 316 a 1.936 μ g (Tabela 8).

Para a análise da expressão gênica por meio de RT-PCR em tempo real foram selecionadas vinte e duas amostras tumorais que apresentaram RNA com massa satisfatória e de boa qualidade. Para os RNAs normais, foram selecionadas cinco amostras normais pareadas aos respectivos tecidos tumorais e uma amostra normal não pareada para a elaboração de uma mistura de RNAs de tecido de mama não tumoral.

Tabela 8 - Massa total de RNA obtida a partir das amostras selecionadas (tecido tumoral e seu normal correspondente) após a extração pelo método de sedimentação por Cloreto de Césio.

Amostras*	Tamanho da lesão (cm)	Massa total de RNA dos tecidos tumorais (μg)	Massa total de RNA dos tecidos correspondentes normais (μg)
1291	1,5	21,3	0
1726	1,8	27,2	17,6
1961	1,0	5,6	34,4
2310	1,5	27,4	ND**
2412	0,8	21,2	ND
2424	1,1	15,3	ND
2439	0,5	6,4	ND
2610	1,8	50,3	ND
2658	1,5	16,9	28,8
2704	1,2	31,0	192,0
2720	1,3	20,7	ND
2734	1,5	91,4	ND
2789	1,2	10,4	ND
2933	2,0	1,3	ND
2937	0,7	54,0	ND
3143	0,7	28,7	108,0
3177	1,3	40,8	25,6
3364	1,9	32,3	725,0
3448	1,5	11,8	25,6
4110	1,7	18,4	ND
4111	2,0	112,0	ND
4379	1,5	62,4	ND

* -Código de acesso da amostra de tecido armazenada no Banco de Tumores do Hospital do Câncer A.C. Camargo.

ND** - não disponível.

4.3 AVALIAÇÃO DOS RNAS

Após a quantificação, 1 μ g de RNA de cada uma das amostras foi submetido à eletroforese em gel de agarose 1%. Foram considerados RNAs de boa qualidade aqueles que apresentaram bandas nítidas correspondentes aos RNAs ribossômicos 28S e 18S. Na Figura 5, as amostras de dois a onze apresentam RNAs íntegros e foram consideradas de boa qualidade. Já as amostras um, doze e treze apresentam degradação das moléculas de RNA e foram desconsideradas para as próximas etapas. Como controle positivo da eletroforese, foi utilizado um RNA previamente testado e sabidamente de boa qualidade (amostra 14).

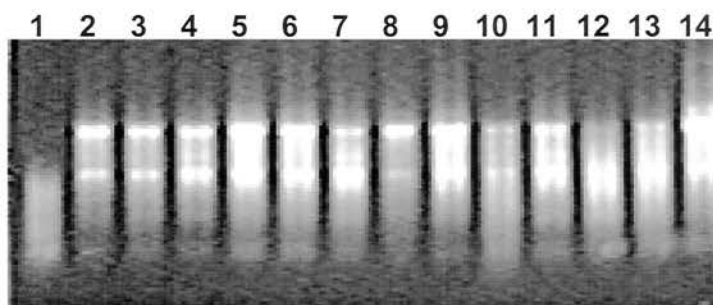


Figura 5 - Após a extração, 1 μ g de cada RNA foi submetido a eletroforese em gel de agarose 1%. Nas canaletas de 1 a 13 foram submetidas à eletroforese amostras de pacientes. Na canaleta 14 adicionou-se 1 μ g de um RNA previamente extraído e considerado de boa qualidade como controle positivo.

4.4 AVALIAÇÃO DA SÍNTESE DE cDNA

Após a extração, todas as amostras selecionadas foram utilizadas como molde para a síntese de cDNA por meio de reação de transcrição reversa. Os produtos dessa síntese foram avaliados quanto à sua qualidade por meio de RT-PCR utilizando *primers* localizados na extremidade 5' do gene *NOTCH2*, resultando em um produto com tamanho esperado de 295 pares de base. A síntese de cDNA das amostras

normais foi realizada a partir de uma mistura de 5 RNAs das amostras correspondentes disponíveis e um RNA de amostra não pareada de tecido de mama. A Figura 6 ilustra um exemplo dos produtos obtidos desta reação de amplificação.



Figura 6 - Verificação da qualidade da síntese de cDNA. Visualização em gel de poliacrilamida 8% corado com nitrato de prata. As canaletas de 2 a 14 correspondem as amostras de cDNA avaliadas, todas as amostras apresentaram uma amplificação eficiente (banda correspondente a 295 nucleotídeos), a canaleta 15 corresponde ao controle negativo da reação.

4.5 SELEÇÃO DOS GENES CANDIDATOS

A análise de bioinformática realizada com os dados disponíveis no banco público de dados de *SAGE* foi capaz de identificar 175 genes com aumento de expressão no tecido neoplásico de mama em relação ao tecido normal. Uma análise criteriosa de cada um destes 175 genes permitiu a seleção de dezessete genes candidatos a diferencialmente expressos em neoplasia de mama. Além destes, outros dois genes (*ALPPL2* e *GATA4*) também foram submetidos a este processo de análise criteriosa e selecionados para serem avaliados por meio de RT-PCR em tempo real (Tabela 9).

Tabela 9 - Genes selecionados para validação da expressão diferencial em tecido neoplásico de mama por meio de RT-PCR em tempo real.

Símbolo do Gene	Nome do Gene	ID ¹	EST ²	Razão das Freqüências ³	Localização Celular	Função celular
ALPPL2	<i>Alkaline phosphatase, placental-like 2</i>	NM_031313	20	ND ⁴	Membrana plasmática	Metabolismo, se liga a íons de magnésio
BIRC5	<i>baculoviral IAP repeat-containing 5-survivin</i>	NM_001168	409	4,0	Citoplasma	Inibidor de apoptose
CAD	<i>carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase</i>	NM_004341	317	4,2	Citoplasma	Proteína envolvida na biossíntese de pirimidinas
CELSR2	<i>cadherin, EGF LAG seven even-pass G-type receptor 2-flamingo homolog, Drosophila</i>	NM_001408	83	5,1	Membrana plasmática	Possivelmente relacionado com adesão celular
CTSF	<i>Cathepsin F</i>	NM_003793	178	7,5	Mitocôndria	Processos de proteólise
DGCR6	<i>DiGeorge syndrome critical region gene 6</i>	NM_005675	70	4,5	Provavelmente nuclear	Histogênese, organogênese e adesão celular
EFNA4	<i>Ephrin-A4</i>	NM_005227	81	20	Membrana plasmática	Sinalização célula-célula, atividade de receptor transmembrana
EI24	<i>Etoposide induced 2.4mRNA</i>	NM_004879	618	5,5	Membrana plasmática	Inibidor de apoptose
FN3KR3	<i>fructosamine-3-kinase-related protein</i>	NM_024619	218	11,0	Citoplasma	Desconhecida
GATA3	<i>GATA binding protein 3</i>	NM_002051	178	12,5	Núcleo	Fator de transcrição
GATA4	<i>GATA binding protein 4</i>	NM_002052		ND ⁴	Núcleo	Fator de transcrição
HMG20B	<i>High-mobility group 20B</i>	NM_006339	472	6,0	Núcleo	Regulação da transcrição-DNA dependente; modelamento de cromatina

¹ número de acesso do gene no *GenBank*

² número de *ESTs* depositadas no sítio do banco de dados Unigene: (www.ncbi.nlm.nih.gov/UniGene).

³ razão das freqüências das *SAGEtags* comparadas entre as bibliotecas de *SAGE* de tecidos de mama tumoral e normal.

⁴ valor não disponível, pois este gene não foi selecionado a partir da lista dos 175 genes encontrados com diferença de expressão entre as bibliotecas de *SAGE* de mama de tecidos tumorais e normais. Este gene foi incluído no trabalho devido à existência de sondas e *primers* no laboratório em que este trabalho foi desenvolvido.

Tabela 9 - Continuação dos genes selecionados para validação da expressão diferencial em tecido neoplásico de mama por meio de RT-PCR em tempo real

Símbolo do Gene	Nome do Gene	ID ¹	EST ²	Razão das Frequências ³	Localização Celular	Função celular
<i>PTPRF</i>	<i>Protein tyrosine phosphatase, receptor type, F</i>	NM_130440	665	7,8	Membrana plasmática	Adesão celular
<i>QSCN6</i>	<i>Quiescin Q6</i>	NM_002826	538	5,0	Retículo endoplasmático e membrana	Envolvido no processo de proliferação celular
<i>SCARB2</i>	<i>Scavenger receptor class B, member 2</i>	NM_005506	355	4,2	Retículo endoplasmático	Adesão celular
<i>STARD 10</i>	<i>Serologically defined colon cancer antigen 28</i>	NM_006645	520	4,5	Mitocôndria	Possível antígeno tumoral
<i>ZC3HDC1</i>	<i>zinc finger CCCH type domain containing 1</i>	NM_022750	181	17.59	Núcleo	Ligante de ácidos nucleicos
<i>KIAA0141</i>	<i>Homo sapiens gene product KIAA0141</i>	NM_014773	454	5.8	Mitocôndria	Função desconhecida (relacionada a transdução de sinal)
<i>KIF 22</i>	<i>Kinesin family member 22</i>	NM_007317	475	4.7	Nuclear	Envolvida no processo de mitose, na atividade motora dos microtúbulos

¹ número de acesso do gene no *GenBank*

² número de *ESTs* depositadas no sítio do banco de dados Unigene: (www.ncbi.nlm.nih.gov/UniGene).

³ razão das frequências das *SAGEtags* comparadas entre as bibliotecas de *SAGE* de tecidos de mama tumoral e normal.

⁴ valor não disponível, pois este gene não foi selecionado a partir da lista dos 175 genes encontrados com diferença de expressão entre as bibliotecas de *SAGE* de mama de tecidos tumorais e normais. Este gene foi incluído no trabalho devido à existência de sondas e *primers* no laboratório em que este trabalho foi desenvolvido.

Algumas das principais características dos genes selecionados são:

- *ALPPL2* (*alkaline phosphatase, placental-like 2*) → este gene está localizado no cromossomo 2, mapeado na região 2q37. Existem pelo menos quatro isoformas de fosfatases alcalinas distintas, porém relacionadas: intestinal, placentária, placentária-*like* e as denominadas como não tecido-específicas (fígado, tecido ósseo e rim). As três primeiras estão localizadas no cromossomo 2, enquanto a forma não tecido-específica é localizada no cromossomo 1. O produto resultante deste gene é uma enzima glicosilada, presente no testículo, timo e em células germinativas de determinados tumores. Segundo a predição de localização celular realizada por meio da análise do programa *PSORT II* (<http://psort.nibb.ac.jp/form2.html>), a proteína codificada por este gene possui localização citoplasmática (34,8% de probabilidade).
- *BIRC5* (*baculoviral IAP repeat-containing 5 - Survivin*) → este gene está localizado no cromossomo 17, mapeado na posição 17q25. A proteína codificada por este gene faz parte da família das proteínas inibidoras de apoptose (cuja sigla em inglês é *IAP*), expressas durante a fase G2/M do ciclo celular. *BIRC5* está associada aos microtúbulos e aos fusos mitóticos, sendo que a ruptura destas estruturas resulta em perda da atividade apoptótica e em aumento da atividade da caspase-3 durante o processo de mitose. *BIRC5* é expresso de forma abundante em células em desenvolvimento e em tecidos proliferativos, mas é pouco expresso ou ausente na maioria dos tecidos adultos (ANBROSINI et al. 1997). Entretanto, vários estudos têm demonstrado a expressão de *BIRC5* em diversos tipos tumorais, como em câncer de estômago, colo-retal, pulmão,

mama, pâncreas, próstata, melanoma e em linfomas não-Hodgkin's de alto grau (AMBROSINI et al. 1997; ADIDA et al. 1998; KAWASAKI et al. 1998; LU et al. 1998). A expressão de BIRC5 tem sido correlacionada a um pior prognóstico em neuroblastoma, câncer de bexiga, câncer colo-retal, esofágico e carcinomas gástricos (ADIDA et al. 1998; KAWASAKI et al. 1998; LACASSE et al. 1998; LU et al. 1998; SWANA et al. 1999; KATO et al. 2001). Diversos estudos têm demonstrado que a detecção da proteína por meio de imuno-histoquímica varia entre os diferentes tipos tumorais estudados. ZHU et al. (2003), desenvolveram um estudo para a detecção de BIRC5 por meio de imuno-histoquímica em amostras humanas de carcinoma gástrico e em carcinoma gástrico de ratos e demonstraram que a alta expressão desta proteína é um evento precoce no desenvolvimento da doença. KAWASAKI et al. (1998), demonstraram que a expressão de BIRC5 apresenta significado prognóstico para o câncer colo-retal em que, pacientes com expressão de BIRC5 possuem pior prognóstico em relação àqueles com ausência da expressão. Em câncer colo-retal, a expressão de *BIRC5* parece estar associada ao processo de progressão tumoral (KAWASAKI et al. 2001). TANAKA et al. (2000), em um estudo usando imuno-histoquímica, sugeriram que a inibição da apoptose por BIRC5 constitui um parâmetro significativo de pior prognóstico em pacientes com câncer de mama. O aumento da expressão de BIRC5 parece, portanto, colaborar com o processo de progressão tumoral através da inibição da apoptose.

- *CAD* (*carbonyl-phosphatase synthetase 2, aspartate transcarbamylase, and dihydroorotase*) → este gene está localizado no cromossomo 2 (2p22-

p21), codifica uma proteína de tripla função associada à atividade enzimática das três primeiras enzimas nos seis primeiros passos na via da biossíntese de pirimidinas: carbamoilfosfato sintetase, aspartato transcarbamilase e deidroorotase. Segundo a predição de localização realizada por meio da análise no programa *PSORT II*, a proteína codificada por este gene possui localização citoplasmática (39,1% de probabilidade). Esta proteína é regulada por meio da ação mitogênica da cascata de atividade da proteína quinase (MAPK), o que sugere uma ligação direta entre a ativação da cascata de MAPK e a biossíntese *de novo* de nucleotídeos de pirimidinas. O hormônio 17 β -estradiol (E2) estimula a proliferação celular da linhagem celular MCF7 e em outras linhagens celulares de câncer de mama com fenótipo positivo para receptor de estrógeno (ER+) e isto é acompanhado pela progressão do ciclo celular e pela transativação de múltiplos genes, incluindo aqueles envolvidos em resposta proliferativa. LIPPMAN et al. (1989) investigaram os efeitos de E2 em células MCF7 em diversas enzimas requeridas para a síntese de DNA, inclusive aquelas envolvidas na biossíntese de nucleotídeos. Os autores encontraram que a indução destas enzimas por E2 é acompanhada pelo aumento da síntese de DNA. Em adição, diversos genes requeridos para a biossíntese de pirimidinas, incluindo carbamoilfosfato sintetase, aspartato transcarbamilase, também são induzidas por E2. KHAN et al. (2003) demonstraram que o gene *CAD* é induzido por E2 em linhagens celulares de câncer de mama ER+ (MCF7 e ZR-75). Os dados encontrados por estes autores sugeriram que diversos genes são responsivos a E2 regulados por ER α /Sp1 (mecanismo que envolve interação proteína-

proteína) e confirmaram que o aumento da expressão hormônio induzida na atividade de enzimas associadas com a biossíntese é acompanhada pela indução da expressão do gene *CAD*.

- *CELSR2* (*cadherin, EGF LAG seven even-pass G-type receptor 2-flamingo homolog, Drosophila*) → este gene está localizado no cromossomo 1, mapeado na região 1p21. Em *Drosophila*, Flamingo (*Fmi*) é um receptor transmembrana conservado da superfamília das caderinas, que desempenha diversos papéis no processo de remodelamento neuronal e na polaridade de células planares epiteliais (CHAE et al. 1999). A proteína codificada por ele faz parte da subfamília flamingo, que consiste de caderinas “não clássicas”, uma subpopulação que não interage com cateninas (ao contrário das caderinas do tipo clássicas). Estão localizadas na membrana plasmática, possuem nove domínios caderina, sete domínios *epidermal growth factor like repeats* (*EGF-like*) e dois *repeats* laminina tipo A. Acredita-se que estas proteínas sejam receptores envolvidos no contato celular, mediando a comunicação com domínios caderina, agindo como ligantes de regiões homofílicas e domínios *EGF-like*, envolvidos na adesão celular e em interações receptor-ligante. A função específica deste membro em particular ainda não foi determinada.
- *CTSF* (*Cathepsin F*) → este gene está localizado no cromossomo 11, mapeado na região 11q13. As cisteína-proteinases são um grupo de enzimas que catalisam a hidrólise de diferentes proteínas e desempenham um papel central na degradação de proteínas intracelulares e no processo de renovação proteica (CHAPMAN et al. 1997). Estas enzimas proteolíticas podem ser divididas em mais de vinte diferentes famílias,

incluindo as da família da papaínas que compreendem uma variedade de enzimas presentes tanto em organismos eucariotos, quanto em procariotos. Em humanos incluem-se onze membros da família das catepsinas (B, C, H, F, K, L, O, S, V, W e X/Z), as quais apresentam diferentes níveis e padrões de expressão, que contribuem para os diferentes papéis fisiológicos desempenhados por estas enzimas. A análise funcional destas proteinases tem demonstrado que além da função intracelular na reciclagem proteica, elas também estão relacionadas a outros processos fisiológicos, como na apresentação de antígenos, na homeostase epidermal e no remodelamento ósseo. Além disso, tem sido sugerido que as catepsinas estejam envolvidas em uma variedade de doenças como na doença de Alzheimer, na artrite reumatóide, no câncer e no processo metastático. Em um estudo de SANTAMARIA et al. (1999), a expressão de catepsina F foi observada em uma variedade de tecidos humanos, sugerindo que ela exerça um papel no processo normal de catabolismo de proteínas. Análise do seu nível de expressão por *Northern blot* revelou um aumento significativo em algumas linhagens celulares isoladas a partir de tecidos cancerosos, o que sugere que esta enzima possa estar envolvida em processos de degradação que ocorrem durante a tumorigênese. A análise realizada com o programa *PSORT II* indicou que o produto gênico de *CTSF* possui localização preferencialmente extracelular, incluindo a parede celular (com probabilidade de 66,7%).

- *DGCR6* (*DiGeorge syndrome critical region gene 6*) → este gene foi isolado de uma região crítica para a síndrome DiGeorge que está associada com microdeleções cromossômicas na região 22q11.21. O produto gênico

de *DGCR6* apresenta homologia com a proteína gonadal de *Drosophila melanogaster* a qual participa do desenvolvimento gonadal e das células da linhagem germinativa. Essa proteína apresenta homologia com a cadeia γ -1 de laminina humana, que faz parte da formação da molécula de laminina, a qual se liga as células através de interações com receptores e é associada aos processos de adesão celular, migração e organização tecidual durante o desenvolvimento (DEM CZUK et al. 1996).

- *EFNA4* (*Ephrin-A4*) → este gene está localizado no cromossomo 1, mapeado na região 1q21-q22, codifica um membro da família das efrinas (cuja sigla em inglês é *EPH*). As EPHs e os receptores relacionados às EPHs compreendem uma vasta subfamília de receptores com atividade de tirosina quinase e parecem estar envolvidas em mediar eventos de desenvolvimento, especialmente no sistema nervoso, em coordenar processos de migração celular e no processo de eritropoiese. Os receptores tipo tirosina quinase são responsáveis pela transdução de sinais do meio extracelular para dentro da célula, desempenhando um papel crítico nos processos de proliferação, diferenciação e migração celular. Com base na estrutura e na similaridade das seqüências, as EPHs podem ser divididas em duas classes: A e B. Aquelas pertencentes à classe A estão ancoradas à membrana por meio de ligação glicofosfatidilinositol e as pertencentes à classe B correspondem às proteínas transmembrana. O gene *EFNA4* codifica uma EPH de classe A, que parece estar associada a processos de sinalização célula-célula, possui domínio conservado de atividade de tirosina quinase, o qual indica o possível envolvimento em mecanismos de transdução de sinal. Vários receptores EPHs apresentam aumento de

expressão em diversos tipos tumorais, o que sugere que eles tenham um importante papel na progressão do câncer (DODELET e PASQUALE 2000). Desde que a *EFNA1* foi isolada a partir de linhagens celulares de carcinoma hepatocelular (HIRAI et al. 1987), vários estudos têm revelado que em diversos outros tipos de carcinomas têm sido encontrada a expressão de membros da família dos receptores tipo tirosina quinase. Como é o exemplo do aumento da expressão de *EFNB2* descrito em pacientes com melanoma maligno e correlacionado com o aumento da tumorigenicidade e do potencial maligno do melanoma (VOGT et al. 1998). VARELIAS et al. (2002) desenvolveram um estudo em que foi identificada a expressão da proteína EFNB1 em amostras de osteosarcoma e correlacionada a um pior prognóstico, sugerindo que a expressão desta proteína possa indicar uma maior agressividade tumoral. Outros estudos demonstraram evidências da expressão de EFNA1 e A2 em carcinoma de mama humano e experimentos de angiogênese *in vitro* sugeriram que a expressão de EFNA1 e a interação com o receptor EFNA2 possa influenciar na neovascularização tumoral (OGAWA et al. 2000). NOREN et al. (2004), demonstraram o papel pró-angiogênico de EFNB4 no processo de progressão tumoral, em que EFNB4 promove o crescimento tumoral por estímulo à angiogênese por meio de EFNB2. O papel da subclasse tipo B de EPHs e seus receptores na neovascularização tumoral ainda não foi elucidado. Entretanto, existem diversas evidências de que estas moléculas possam ter um importante papel no processo de angiogênese durante o desenvolvimento.

- *EI24* (*Etoposide induced 2.4 mRNA*) → *EI24*, também denominado de *PIG8* está localizado na região cromossômica 11q24, que é freqüentemente alterada em cânceres humanos. A perda de heterozigotidade nesta região é observada em diversos tipos tumorais, como no câncer cervical invasivo, no câncer de mama e no melanoma maligno, sendo que os possíveis genes com função de supressão tumoral ainda não foram identificados. *EI24/PIG8* parece ter um papel importante no controle negativo do crescimento celular devido a sua função como efetor apoptótico do gene *p53*. A inativação do gene *TP53* é uma das mais comuns alterações detectadas em mais de 50% de todos os tipos tumorais. O tipo selvagem do gene *TP53* apresenta um papel chave na supressão tumoral por meio do monitoramento do dano ao DNA e pela execução de vias que controlam o crescimento celular de forma negativa, seja pelo bloqueio das células na fase G1 do ciclo celular, ou pela indução à apoptose (LEVINE 1997). *EI24* é um gene de resposta ao dano de DNA, originalmente isolado a partir de fibroblastos NIH 3T3 que foram induzidos a morte celular (LEHAR et al. 1996). POLYAK et al. (1997) isolaram *EI24* a partir de células de carcinoma de cólon em apoptose, em resposta à expressão ectópica do tipo selvagem de *TP53*. A análise da expressão gênica de mais de 7.200 transcritos revelou que *EI24* estava entre os 14 genes identificados como induzidos de maneira dependente de *TP53*. GU et al. (2000) demonstraram que *EI24* é responsivo à p53 e que o aumento da expressão de *EI24* suprime o crescimento celular por meio da indução a morte celular no processo de apoptose. Os achados deste estudo indicaram que *EI24* possa ser um provável supressor de tumor em

determinados tipos tumorais. Entretanto, esse gene encontra-se com a expressão aumentada em alguns tipos tumorais como cérebro e mama segundo dados de *SAGE* e *ESTs* (<http://cgap.nci.nih.gov/Tissues/VirtualNorthern>). A proteína codificada por *EI24* está localizada na membrana (52,5% de probabilidade, *PSORT II*) e apresenta seis domínios transmembrana (segundo a predição do banco de dados TMHMM).

- *FN3KRP* (*fructosamine-3-kinase-related protein*) → este gene está localizado no cromossomo 17, mapeado na posição 17q25.3. No início deste estudo, esse gene possuía função desconhecida (*Hypothetical protein FLJ12171*). *FN3KRP* é uma proteína relacionada à frutosamina-3 quinase (*FN3K*, membro da família das proteínas quinases), que fosforila frutosaminas ligadas à proteínas iniciando um processo que remove esses resíduos de frutosamina das proteínas. COLLARD et al. (2003), demonstraram que *FN3K* possui a capacidade exclusiva de fosforilar frutosaminas e que, ambas *FN3KRP* e *FN3K* apresentam a capacidade de fosforilar psicossaminas e ribulosaminas com afinidade similar. *FN3KRP* é encontrada em tecidos como pulmão e fígado, em que não é detectado a presença do mRNA de *FN3K*, indicando que *FN3KRP* parece ter uma ação complementar a *FN3K* no processo de retirada de açúcares ligados a proteínas. *FN3K* parece exercer o papel de enzima reparadora de proteínas, sendo assim é esperado que desempenhe um papel importante nos tecidos que contém proteínas de meia-vida longa, como é caso dos eritrócitos e do cristalino, mas isso também é verdadeiro para o cérebro, que contém mielinas que possuem uma meia-vida de mais de dez dias (DELCOUR e

PAPACONSTANTINO 1972; LAJTHA et al. 1977). Um estudo realizado em animais *knockout* para uma outra enzima com função semelhante, demonstrou que a ausência da enzima resultou na morte dos animais dentro de poucos dias (KIM et al. 1997). Este resultado ressalta a importância da ação destas enzimas de reparo no cérebro e coincidem com os níveis elevados de mRNAs de FN3K e FNK3RP encontrados no tecido cerebral, nos estudos de COLLARD et al. (2003). Todavia, a proteína codificada por este gene possui função ainda desconhecida e apresenta predição de domínio de proteína do tipo quinase. A análise realizada no banco de dados *PSORT II* indicou que a proteína resultante deste gene possui localização citoplasmática (60,9% de probabilidade).

- *GATA3* (*GATA binding protein 3*) → este gene está localizado no cromossomo 10, mapeado na região 10p15. Faz parte da família GATA de fatores de transcrição. *GATA3* é altamente expresso em células linfóides T e acredita-se em seu envolvimento no desenvolvimento destas células e na expressão das cadeias α e δ nos receptores de células T (HO et al. 1991; JOULIN et al. 1991). Este gene é também altamente expresso em áreas de cérebro embrionário e em vários outros tecidos, incluindo rim e endotélio (ORKIN 1992). Recentemente, diversos membros da família GATA têm sido relacionados a vários tipos de tumores humanos. O envolvimento de *GATA3* no câncer de mama foi demonstrado por HOCH et al. (1999) em estudos de análise da expressão gênica em linhagens celulares tumorais de mama e em quarenta e sete amostras de tumores de mama invasivos e, neste mesmo estudo, os autores correlacionaram a expressão da proteína com o fenótipo positivo para receptor de estrógeno.

- *GATA4* (*GATA binding protein 4*) → este gene está localizado no cromossomo 8, mapeado na região 8p23.1-p22. Assim como *GATA3*, *GATA4* também faz parte da família *GATA* de fatores de transcrição. Os membros desta família reconhecem o sítio *GATA*, o qual está presente na região promotora de diversos genes. Acredita-se que a proteína codificada por este gene esteja envolvida na embriogênese, diferenciação e função de células do miocárdio. Mutações neste gene têm sido associadas a defeitos no septo cardíaco. A expressão de *GATA4* em cânceres já foi associada ao adenocarcinoma de esôfago e malignidades nas gônadas (LIN et al. 2000).
- *HMG20B* (*High-mobility group 20B*) → este gene está localizado no cromossomo 19, mapeado na posição 19p13.3. Os *HMGs* (*high mobility group*) codificam um domínio conservado de ligação ao DNA em diversas proteínas e estão envolvidos no processo de regulação da transcrição gênica e na conformação da cromatina. Alguns membros desta família têm sido identificados como ligantes ao menor sulco da dupla fita de DNA, talvez se ligando a seqüências específicas e freqüentemente associados com distorção estrutural ou manipulação da estrutura do DNA. A proteína codificada por este gene apresenta provável localização nuclear (65,2%), segundo análise feita no programa *PSORT II*. Esse gene foi denominado inicialmente como *BRAF35*, e a proteína foi originalmente identificada como um componente ligante de DNA do multicomplexo de *BRCA2* (um importante gene envolvido no câncer de mama hereditário). MARMORSTEIN et al. (2001) demonstraram a associação entre *BRAF35* e *BRCA2* com a cromatina durante as fases iniciais da condensação dos cromossomos na mitose e por meio de experimentos com microinjeções de

anticorpos, sendo sugerido pelos autores o envolvimento deste complexo na regulação da progressão do ciclo celular. O complexo BRAF35/BRCA2 não confere somente um papel regulatório na progressão do ciclo celular, mas também tem sido descrito como envolvido no reparo de dano ao DNA.

- *PTPRF* (*Protein tyrosine phosphatase, receptor type, F*) → este gene está localizado no cromossomo 1, mapeado na região 1p34. Codifica uma proteína de membrana que possui um domínio citoplasmático com homologia à proteína fosfatase 1B e um domínio extracelular homólogo a uma molécula de adesão celular. A proteína codificada por este gene faz parte da família de proteínas tirosina fosfatase (cuja sigla do inglês é *PTP*), constituída por moléculas sinalizadoras que regulam uma variedade de processos celulares importantes, como a proliferação e a diferenciação celular, o ciclo mitótico e transformação oncogênica. Segundo a análise realizada para a predição de localização, a proteína codificada por este gene está localizada na membrana (33,3% de probabilidade-*PSORT II*) e apresenta um domínio transmembrana, segundo a predição realizada no banco de dados TMHMM (www.cbs.dtu.dk/services/TMHMM).
- *QSCN6* (*Quiescin Q6*) → este gene está localizado no cromossomo 1, mapeado na posição 1p24, próximo a um locus envolvido no câncer de próstata hereditário (HPC1). A proteína codificada por este gene é originada a partir da fusão de dois representantes de duas famílias gênicas, durante o processo de evolução, contendo domínios de tioredoxina (proteína oxiredutase) e *ERV1* (este gene foi identificado como necessário para a fosforilação oxidativa e vegetativa no crescimento em *S. cerevisiae*).

QSCN6 parece estar envolvido de alguma maneira na divisão celular e na manutenção do genoma mitocondrial de leveduras (LISOWSKY 1992). A expressão deste gene é induzida em fibroblastos benignos que entram na quiescência, o que sugere que este gene desempenhe um papel importante na regulação da proliferação celular (COPPOCK et al. 1993). A regulação do processo de quiescência é um passo crítico para todo o processo de regulação da proliferação celular, grande parte das falhas ocorridas durante a proliferação celular estão relacionadas ao desenvolvimento de câncer e de outras doenças proliferativas (COPPOCK et al. 1998). A proteína codificada por este gene é preferencialmente encontrada no retículo endoplasmático e na membrana (44,4% de probabilidade, segundo a predição realizada no programa *PSORT II*).

- *SCARB2* (*Scavenger receptor class B, member 2*) → este gene está localizado no cromossomo 4, mapeado na posição 4q21.21-q21.22. A proteína codificada por este gene é uma glicoproteína do tipo III localizada na membrana dos lisossomos e dos endossomos. Estudos utilizando proteínas similares em camundongos e em ratos têm sugerido que esta proteína parece participar do processo de transporte de membrana e na organização do compartimento endossomal e lisossomal. Em camundongos, a deficiência de uma proteína similar foi descrita como prejudicial aos processos de transporte de membrana celular e também por causar obstrução da junção pélvica, surdez e neuropatia periférica (GAMP et al. 2003). A proteína codificada por este gene é encontrada no retículo endoplasmático (44,4% de probabilidade, segundo a predição realizada no programa *PSORT II*). O produto gênico de *SCARB2* parece exercer

atividade de molécula de adesão, pois apresenta um domínio conservado da família de antígenos CD36, que é uma família altamente conservada durante o processo evolutivo e inclui moléculas de adesão e receptores lipídicos na superfície das células, bem como da membrana lisossomal (GAMP et al. 2003) e apresenta também domínios de receptores associados a membrana relacionados a mecanismos de transdução de sinais.

- *STARD10 (Serologically defined colon cancer antigen 28)* → este gene está localizado no cromossomo 11, mapeado na região 11q13, próximo a região correspondente aos genes *EMS* e *CCD1*, os quais são freqüentemente encontrados com aumento de expressão em câncer de mama (ORMANDY et al. 2003). A proteína codificada por este gene possui função desconhecida e parece conter domínios envolvidos no transporte lipídico, no metabolismo e possível função de antígeno tumoral. *STAR (steroidogenic acute regulatory protein)* corresponde a uma proteína mitocondrial que é sintetizada em resposta ao estímulo do hormônio luteinizante. A expressão desta proteína na ausência do estímulo hormonal é suficiente para induzir a produção de esteróides, o que sugere que essa proteína seja necessária para a regulação do processo de formação de esteróides. Conhecida também como NY-CO-28, *STARD10* foi identificada como um antígeno tumoral pelo sistema imune humoral em um trabalho de SCANLAN et al. (1998), em que foi utilizada a técnica de SEREX. Nesse trabalho, outros 48 antígenos distintos (NY-CO1 a NY-CO-48) foram identificados no soro de quatro pacientes com câncer de cólon. OLAYIOYE et al. (2004), identificaram o aumento da expressão

deste gene em diversas linhagens celulares humanas tumorais de mama e o aumento da expressão da proteína foi encontrado em 28 de 79 amostras de câncer de mama primário analisadas.

- *ZC3HDC1* (*zinc finger CCCH type domain containing 1*) → este gene está localizado no cromossomo 7, mapeado na região 7q34. No início deste trabalho este gene possuía sua função desconhecida (*hypothetical protein FLJ22693*). A proteína codificada por este gene possui domínio de dedo de zinco tipo C-x8-C-x5-C-x3-H. Acredita-se que os domínios desse tipo estejam envolvidos na ligação ao DNA e que existam diferentes tipos, dependendo das posições dos resíduos de cisteína. As proteínas que contém o domínio tipo C-x8-C-x5-C-x3-H, incluem proteínas de eucariotos envolvidas em ciclo celular e regulação da proliferação celular. A análise realizada no programa *PSORT II* indicou que a proteína resultante deste gene possui localização nuclear (39,1% de probabilidade).
- *KIAA0141* (*Homo sapiens gene product KIAA0141*) → este gene está localizado no cromossomo 5, mapeado na região 5q31.3. A proteína codificada por este gene possui função ainda não conhecida, de localização mitocondrial, segundo a predição de localização realizada com o programa *PSORT II* (60,9% de probabilidade).
- *KIF22* (*Kinesin family member 22*) → este gene está localizado no cromossomo 16, mapeado na posição 16p11.2. A proteína codificada por este gene faz parte da família de proteínas do tipo *Kinesin*. Esta família de proteínas está envolvida no processo de transporte de organelas (essencial para a morfogênese e o funcionamento celular adequado), de proteínas complexas, de mRNAs e participam também, dos movimentos dos

cromossomos e dos fusos durante os processos de mitose e meiose (HIROKAWA 1998; BRENDZA et al. 2000; SHARP et al. 2000; MIKI et al. 2001). A análise realizada no programa *PSORT II* indicou que a proteína resultante deste gene possui localização nuclear (56,5% de probabilidade).

4.6 ESCOLHA DO GENE NORMALIZADOR

Como já descrito anteriormente, genes normalizadores são genes constitutivos utilizados para corrigir variações nas quantidades iniciais de RNA nos experimentos de RT-PCR em tempo real. Estes genes devem apresentar uma variação mínima no nível de expressão de mRNA entre as diferentes populações de amostras estudadas. No presente estudo, a escolha do gene normalizador foi realizada por dois métodos.

Para a identificação por meio da análise dos dados de *SAGE*, neste estudo foram avaliados dez possíveis genes para serem usados como normalizadores (*ACTB*, *B2M*, *GAPDH*, *HPRT1*, *PPIA*, *PGK1*, *RPLP0*, *TBP*, *TFRC*). Destes genes, foram obtidas as razões entre as frequências das *SAGEtags* presentes nas bibliotecas de tecidos normais e tumorais de mama, por meio da utilização do programa *SAGEβBin* e calculadas as área de sobreposição das distribuições das frequências e atribuídos os valores de E. Nesta análise, dos dez possíveis genes normalizadores, apenas quatro deles apresentaram *SAGEtags* confiáveis e foram avaliados (Tabela 10). As *SAGEtags* que foram consideradas como não confiáveis correspondiam àquelas que estavam localizadas em regiões de repetição do genoma da família *Alu*. Portanto, os valores de expressão obtidos não poderiam ser considerados como exclusivos do gene analisado,

mas também a frequência de expressão das *SAGEtags* de outros genes poderiam estar sendo somadas aos valores de expressão do gene de interesse.

Dos quatro genes avaliados (Tabela 10), o gene *PGK1* foi o único que pode ser considerado adequado para ser utilizado como normalizador em estudos de análise de expressão gênica por RT-PCR em tempo real em amostras de tecido de mama tumoral e não-tumoral, pois possui uma *SAGEtag* confiável e valor de E próximo a um, o que significa que não existe diferença significativa entre a frequência das *SAGEtags* entre as bibliotecas de *SAGE* preparadas a partir de tecidos normais e tumorais. Entretanto, uma vez que este gene não é frequentemente utilizado em estudos de análise da expressão gênica e que os demais genes constitutivos não puderam ser incluídos nesta análise porque não possuíam *SAGEtags* confiáveis, um segundo método de escolha foi adotado.

Tabela 10 - Avaliação das frequências das *SAGEtags* entre as bibliotecas de *SAGE* de mama tumoral e normal por meio da utilização do *pipeline SAGEβBin*.

Símbolo do Gene	SAGE-Tag	E*
<i>RPLP0</i>	CTCAACATCT	0,44
<i>TBP</i>	AACCACAGTT	0,62
<i>TFRC</i>	ACCTTTACTG	0,63
<i>PGK1</i>	GAGGAAGGCT	0,83

E- valor atribuído a área de sobreposição das distribuições das frequências

Em seguida, foram realizados quatro experimentos de maneira independente com o ensaio comercial para reação de RT-PCR em tempo real com o sistema *TaqMan* de detecção da fluorescência, chamado *TaqMan Human Endogenous Control Plate (Applied Biosystems)*. Desta forma foi calculada a diferença de variação de expressão de dez genes entre as amostras normal e tumoral (conforme já descrito em material e métodos). Foram obtidas as médias e os desvios padrões de cada gene nos

quatro experimentos. Dos valores obtidos, destacaram-se os resultados referentes ao gene da β -actina (Tabela 11), o qual apresentou baixa diferença entre as amostras testadas (na mistura de tecidos tumorais e na mistura de tecidos normais de mama) em todos os experimentos realizados.

Embora a análise da frequência das *SAGEtags* por meio da utilização do *SAGE β Bin* tenha caracterizado a *SAGEtag* referente à β -actina como não confiável e portanto, não possibilitando a sua avaliação, a segunda abordagem experimental para a escolha do gene normalizador demonstrou uma baixa variação do nível de expressão deste gene entre as amostras normal e tumoral. Além disso, dados de literatura têm demonstrado a ampla utilização deste gene como normalizador em estudos de análise da expressão gênica em tumores de mama (NACHT et al. 1999; KROP et al. 2001; JIANG et al. 2002; PORTER et al. 2003a). Sendo assim, o gene selecionado para ser utilizado como gene normalizador neste estudo, foi o da β -actina, seja quando utilizado o sistema *Taqman* de quantificação ou quando o sistema utilizado foi o *SYBR Green*.

Para a análise pelo sistema *Taqman*, para a amplificação deste gene foram utilizados *primers* e sonda disponíveis comercialmente pela *Applied Biosystems (Pre developed TaqMan[®] assay reagents-Part Number: 4310881)*. Para a análise pelo sistema *SYBR Green*, foram desenhados *primers* em *exons* diferentes (*exon 3 e 5*) com seqüência direta 5'AAATCTGGCACCACACCTTC3' e reversa 5'CACTGTGTTGGCGTACAGGT 3', com produto esperado de 644 pares de bases.

Tabela 11 - Valores da diferença entre os Δ Cts entre as amostras (mistura de tumoral e a mistura normal, de tecidos de mama) para cada um dos genes analisados no ensaio de reação de RT-PCR em tempo real *TaqMan Human Endogenous Control Plate (Applied Biosystems)*.

Símbolo dos genes	Experimento 1	Experimento 2	Experimento 3	Experimento 4	Média	Desvio Padrão
<i>IPC</i> **	0,04	0,28	0,30	0,59	0,30	0,19
<i>PO</i>	1,50	1,90	4,31	1,09	2,20	1,25
<i>BA</i>	0,34	0,09	0,24	0,18	0,21	0,09
<i>CYC</i>	0,21	0,28	0,84	0,58	0,48	0,25
<i>GAPDH</i>	0,61	0,09	0,19	0,51	0,35	0,22
<i>PGK</i>	1,76	0,95	0,87	0,41	0,75	0,21
<i>B2m</i>	1,16	1,83	2,24	2,02	1,81	0,40
<i>GUS</i>	1,22	0,46	0,27	0,20	0,54	0,41
<i>HPRT</i>	0,44	0,40	0,84	0,01	0,47	0,23
<i>TBP</i>	1,32	0,04	0,58	0,40	0,59	0,47
<i>TIR</i>	2,13	0,67	0,90	1,56	1,32	0,57

* valores da diferença entre os Δ Cts entre as amostras (mistura de tumores e a mistura de tecidos normais, de mama) para cada um dos genes analisados, no experimento mencionado.

** *Internal Positive Control*

4.7 QUANTIFICAÇÃO DA EXPRESSÃO GÊNICA POR MEIO DE RT-PCR EM TEMPO REAL

A quantificação da expressão dos dezenove genes selecionados como candidatos diferencialmente expressos no câncer de mama resultantes da análise em bancos de dados de *SAGE* foi realizada em vinte e duas amostras de tecidos neoplásicos de mama, menores do que dois centímetros e em uma mistura de amostras não tumorais. Os dados obtidos da quantificação da expressão gênica estão representados na Tabela 12. Foram consideradas as amostras com aumento de expressão aquelas que apresentaram os valores de $\Delta\Delta$ Ct pelo menos duas vezes aumentado em relação ao valor de expressão da mistura de amostras normais. Destes resultados, foram considerados como tendo a expressão diferencial validada, os genes candidatos que apresentaram um aumento de expressão em pelo menos 25% das

amostras testadas (LOGING et al. 2000). Dos resultados obtidos a partir da quantificação da análise da expressão dos dezenove genes avaliados, onze deles (*ALPPL2*, *CELSR2*, *FN3KRP*, *GATA3*, *GATA4*, *HMG20B*, *PTPRF*, *QSCN6*, *SACARB2*, *STARD10* e *ZC3HDC1*) apresentaram um aumento de expressão nas amostras tumorais testadas em relação ao tecido normal, segundo os critérios adotados.

Conforme podemos observar na Tabela 12, o gene *ALPPL2* apresentou aumento da expressão em 59,0% das amostras testadas, ou seja, 13 das 22 amostras analisadas exibiram um aumento de pelo menos duas vezes, quando comparadas à mistura do tecido normal, sendo que, a variação do aumento da expressão foi de 2,01 a 29,55 vezes.

O gene *CELSR2* apresentou aumento de expressão em 81,81% das amostras testadas (18 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,3 a 48,8 vezes.

A análise da quantificação da expressão do gene *FN3KRP* revelou o aumento da expressão em 27,27% das amostras testadas (6 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,4 a 3,58 vezes.

O gene *GATA3* apresentou aumento de expressão em 77,27% das amostras testadas (17 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2 a 29,4 vezes.

A análise da quantificação da expressão do gene *GATA4* revelou o aumento da expressão em 33,33% das amostras testadas (6 das 18 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,42 a 146 vezes.

O gene *HMG20B* apresentou aumento de expressão em 54,54% das amostras (12 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,2 a 18,5 vezes.

O gene *PTPRF* apresentou aumento de expressão em 31,81% das amostras (7 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,5 a 6,3 vezes.

A análise da quantificação da expressão do gene *QSCN6* revelou o aumento da expressão em 27,27% das amostras testadas (6 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,01 a 12,95 vezes.

O gene *SCARB2* apresentou aumento de expressão em 36,36% das amostras (8 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2 a 12,95 vezes.

A análise da quantificação da expressão do gene *STARD10* revelou o aumento da expressão em 25% das amostras testadas (5 de 20 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,19 a 11,75 vezes.

O gene *ZC3HDC1* apresentou aumento de expressão em 36,36% das amostras testadas (8 das 22 amostras analisadas), sendo que a variação do aumento da expressão foi de 2,11 a 4,13 vezes.

Os demais genes (*BIRC5*, *CAD*, *CTSF*, *EFN4*, *EI24*, *KIAA0141* e *KIF22*) não apresentaram aumento de expressão (de pelo menos mais que duas vezes) em mais de 25% das amostras tumorais testadas, em relação ao tecido normal. Nenhuma das amostras testadas apresentou aumento da expressão em relação ao tecido normal quando foi analisado o gene *DGCR6*.

No Anexo 8, encontram-se os dados completos da medida da expressão gênica diferencial em todas as amostras analisadas nos genes selecionados.

Tabela 12- Porcentagem de amostras com aumento de expressão de pelo menos duas vezes em relação à mistura de amostras normais, para os genes analisados por meio de reação de quantificação por RT-PCR em tempo real

Símbolo do Gene	Porcentagem de amostras com aumento de expressão (número de amostras)	Varição do nível de aumento de expressão nas amostras (número de vezes)
<i>ALPPL2</i>	59,00 (13/22)	2,0 - 29,5
<i>BIRC</i>	22,72 (5/22)	2,7 - 17,6
<i>CAD</i>	18,18 (4/22)	2,7 - 5,4
<i>CELSR2</i>	81,81 (18/22)	2,3 - 48,8
<i>CTSF</i>	22,72 (5/22)	2,1 - 3,3
<i>DGCR6</i>	0,0 (0/22)	0
<i>EFNA 4</i>	22,72 (5/22)	2,3 - 5,0
<i>EI24</i>	4,54 (1/22)	2,2 - 2,2
<i>FN3KRP</i>	27,27 (6/22)	2,4 - 3,5
<i>GATA 3</i>	77,27 (17/22)	2,0 - 29,4
<i>GATA 4</i>	33,33 (6/18)	2,42 - 146,0
<i>HMG20B</i>	54,54 (12/22)	2,3 - 18,5
<i>PTPRF</i>	31,81(7/22)	2,3 - 14,3
<i>QSCN6</i>	27,27 (6/22)	2,3 - 7,7
<i>SCARB2</i>	36,36 (8/22)	2,7 - 12,9
<i>STARD10</i>	25,00 (5/20)	2,0 - 11,7
<i>ZC3HDC</i>	40,90 (9/22)	2,0 - 4,0
<i>KIAA0141</i>	22,72(5/22)	2,0 - 4,2
<i>KIF22</i>	4,54 (1/22)	5,4 - 5,4

4.8 ANÁLISE ESTATÍSTICA REFERENTE À QUANTIFICAÇÃO DA EXPRESSÃO GÊNICA POR MEIO DE RT-PCR EM TEMPO REAL

Frente aos resultados de análise da expressão gênica diferencial dos dezenove genes analisados, foi realizada a análise estatística para correlacionar dados clínico-patológicos disponíveis (idade, tamanho da lesão, *status* para a imunodeteção de p53 e c-erb-B2 e presença de metástase em linfonodos). Neste trabalho, foram

considerados estatisticamente significativos os dados que apresentaram um valor de p menor que 0,05.

A associação entre o aumento de expressão dos genes *CELSR2*, *GATA3*, *GATA4* e *KIAA0141* e a positividade para a imunodeteção de receptor de estrógeno se mostrou estatisticamente significativa (com valores de p iguais a: 0,0270, 0,0223, 0,0146 e 0,0325, respectivamente) e marginal em relação ao aumento da expressão do gene *CTSF* ($p=0,0600$) (Tabela13).

A associação entre a expressão do gene *CELSR2* e a positividade para a imunodeteção do receptor de progesterona se mostrou estatisticamente significativa ($p=0,0278$) (Tabela14). A associação entre o aumento de expressão do gene *GATA3* e a positividade para a imunodeteção de receptor de progesterona se mostrou marginalmente significante ($p=0,0527$).

Devido à pequena amostragem, nenhuma outra associação pode ser observada entre os dados de quantificação da expressão gênica e os demais aspectos clínico-patológicos das amostras das pacientes analisadas.

Tabela 13. Distribuição dos genes de acordo com o *status* do receptor de estrógeno e o nível de expressão dos genes analisados.

Genes	Negativo N (6)	Positivo N (16)	p-valor ¹
ALPPL2			
Min-Máx	0,48-15,08	0,40-29,54	0,9412
Mediana	2,48	2,55	
Média (dp)	5,29(5,70)	6,33(8,20)	
BIRC5			
Min-Máx	0,71-6,32	0,25-17,63	0,3020
Mediana	0,99	0,72	
Média (dp)	2,09(2,21)	2,75(4,81)	
CAD			
Min-Máx	0,21-3,05	0,22-5,46	0,2688
Mediana	1,30	0,70	
Média (dp)	1,37(0,96)	1,22(1,40)	
CELSR2			
Min-Máx	0,83-6,10	1,10-48,84	0,0270
Mediana	2,16	8,07	
Média (dp)	3,01(2,42)	13,76(15,62)	
CTSF			
Min-Máx	0,13-1,84	0,34-3,38	0,0600
Mediana	0,57	1,09	
Média (dp)	0,69(0,63)	1,37(0,88)	
DGCR6			
Min-Máx	0,05-1,45	0,04-1,45	0,7400
Mediana	0,54	0,43	
Média (dp)	0,62(0,57)	0,46(0,40)	
EFNA4			
Min-Máx	0,71-5,02	0,16-2,71	0,6583
Mediana	1,32	1,23	
Média (dp)	1,77(1,62)	1,39(0,84)	
EI24			
Min-Máx	0,27-1,84	0,25-2,21	0,6057
Mediana	0,69	0,87	
Média (dp)	0,81(0,56)	0,91(0,57)	
FN3KRP			
Min-Máx	0,27-3,58	0,26-3,30	0,8828
Mediana	0,93	0,87	
Média (dp)	1,94(1,38)	1,29(0,93)	
GATA 3			
Min-Máx	0,20-11,39	1,12-29,44	0,0223
Mediana	1,57	7,03	
Média (dp)	3,10(4,25)	10,22(8,46)	
GATA4			
Min-Máx	0,01-0,12	0,01-146,0	0,0146
Mediana	0,02	1,92	
Média (dp)	0,04(0,05)	13,2(38,4)	

¹-Valor de p obtido por meio do teste U de Mann-Whitney

Tabela 13. Continuação da distribuição dos genes de acordo com o *status* do receptor de estrógeno e o nível de expressão dos genes analisados.

Genes	Negativo N (6)	Positivo N (16)	p-valor ¹
HMG20B			
Min-Máx	0,28-18,50	0,288-10,92	0,9412
Mediana	2,75	2,20	
Média (dp)	5,47(7,05)	3,14(2,96)	
PTPRF			
Min-Máx	0,41-4,31	0,49-6,29	0,6583
Mediana	1,95	1,22	
Média (dp)	2,23(1,77)	1,70(1,45)	
QSCN6			
Min-Máx	0,42-5,35	0,50-8,84	0,7124
Mediana	1,28	1,16	
Média (dp)	1,79(1,80)	2,72(3,02)	
SCARB2			
Min-Máx	0,49-6,94	0,36-12,95	0,9999
Mediana	1,24	1,29	
Média (dp)	2,23(2,44)	2,34(3,03)	
STARD10			
Min-Máx	0,01-3,98	0,03-11,75	0,9652
Mediana	1,34	0,66	
Média (dp)	1,56(1,69)	1,69(3,00)	
ZC3HDC1			
Min-Máx	0,90-3,55	0,50-4,12	0,5553
Mediana	2,34	1,62	
Média (dp)	2,20(1,10)	1,83(1,08)	
KIAA0141			
Min-Máx	0,57-1,76	0,46-4,28	0,0325
Mediana	0,74	1,65	
Média (dp)	0,92(0,45)	1,83(1,00)	
KIF22			
Min-Máx	0,03-0,84	0,07-5,48	0,8250
Mediana	0,27	0,28	
Média (dp)	0,33(0,26)	0,70(1,33)	

¹-Valor de p obtido por meio do teste U de Mann-Whitney

Tabela 14. Distribuição dos genes de acordo com o *status* do receptor de progesterona e o nível de expressão dos genes analisados.

Genes	Negativo N (6)	Positivo N (16)	p-valor ¹
ALPL2			
Min-Máx	0,40-17,08	0,48-29,54	0,2786
Mediana	3,30	1,77	
Média (dp)	6,59(5,98)	5,51(9,01)	
BIRC5			
Min-Máx	0,31-17,63	0,25-7,01	0,3088
Mediana	1,23	0,75	
Média (dp)	3,38(5,48)	1,30(1,95)	
CAD			
Min-Máx	0,21-5,46	0,30-3,28	0,4905
Mediana	5,46	3,28	
Média (dp)	1,41(1,54)	1,11(1,00)	
CELSR2			
Min-Máx	0,83-30,80	2,67-48,8	0,0278
Mediana	2,47	9,38	
Média (dp)	6,38(9,06)	15,2(17,14)	
CTSF			
Min-Máx	0,13-3,38	0,34-2,14	0,9476
Mediana	0,68	0,89	
Média (dp)	1,26(1,07)	1,11(0,64)	
DGCR6			
Min-Máx	0,04-1,45	0,08-1,03	0,5766
Mediana	0,24	0,53	
Média (dp)	0,52(0,57)	0,48(0,29)	
EFNA4			
Min-Máx	0,69-2,71	0,16-5,02	0,6224
Mediana	1,26	1,34	
Média (dp)	1,22(0,57)	1,77(1,39)	
E124			
Min-Máx	0,25-2,11	0,33-1,95	0,3578
Mediana	0,63	0,90	
Média (dp)	0,80 (0,59)	0,96(0,52)	
FN3KRP			
Min-Máx	0,27-3,58	0,26-3,30	0,5767
Mediana	0,77	0,97	
Média (dp)	1,29(1,33)	1,39(1,00)	
GATA 3			
Min-Máx	0,20-19,76	1,74-29,44	0,0527
Mediana	2,30	9,35	
Média (dp)	5,71(7,10)	10,85(8,58)	
GATA4			
Min-Máx	0,0-146,0	0,03-12,55	0,1826
Mediana	0,20	1,92	
Média (dp)	15,3(45,92)	4,06(4,87)	

¹-Valor de p obtido por meio do teste U de Mann-Whitney

Tabela 14. Continuação da distribuição dos genes de acordo com o *status* do receptor de progesterona e o nível de expressão dos genes analisados.

Genes	Negativo N (6)	Positivo N (16)	p-valor¹
HMG20B			
Min-Máx	0,28-8,19	0,45-18,50	0,1228
Mediana	1,37	2,80	
Média (dp)	2,40(5,48)	5,14(5,51)	
PTPRF			
Min-Máx	0,41-4,27	0,49-6,29	0,6224
Mediana	1,24	1,45	
Média (dp)	1,62(1,19)	2,07(1,82)	
QSCN6			
Min-Máx	0,42-8,84	0,70-7,72	0,1783
Mediana	0,97	1,29	
Média (dp)	2,16(2,85)	2,78(2,70)	
SCARB2			
Min-Máx	0,36-12,95	0,44-6,94	0,2786
Mediana	1,19	1,35	
Média (dp)	2,29(3,65)	2,33(1,91)	
STARD10			
Min-Máx	0,03-11,75	0,31-4,61	0,3823
Mediana	0,66	0,90	
Média (dp)	1,91(3,77)	1,44(1,50)	
ZC3HDC1			
Min-Máx	0,504-1,6	0,54-3,21	0,6224
Mediana	1,14	1,91	
Média (dp)	1,09(1,28)	1,96(0,87)	
KIA0141			
Min-Máx	0,46-4,28	0,77-3,07	0,1077
Mediana	1,12	1,75	
Média (dp)	1,43(1,16)	1,74(0,74)	
KIF22			
Min-Máx	0,03-5,48	0,07-1,67	0,3088
Mediana	0,30	0,25	
Média (dp)	0,80(1,56)	0,40(0,49)	

¹-Valor de p obtido por meio do teste U de Mann-Whitney

5 DISCUSSÃO

O câncer é uma doença genética caracterizada por alterações em genes que estão envolvidos no processo de proliferação, morte e diferenciação celular. A progressão de uma pré-neoplasia para o câncer é um processo composto por diversas etapas, acompanhado pelo acúmulo de alterações genéticas que resultam na ausência ou em padrões de expressão inadequados de genes e modificações na estrutura e função de proteínas.

O câncer de mama é a neoplasia mais freqüente que acomete as mulheres do mundo todo. Embora exista um considerável progresso no diagnóstico e no tratamento precoce para esse tipo de tumor, a mortalidade por essa doença entre as mulheres ainda permanece relativamente alta. O câncer de mama é uma doença complexa, caracterizada pelo acúmulo de diversas alterações moleculares que resultam em uma importante heterogeneidade clínica e histopatológica, que faz com que seja difícil selecionar qual estratégia terapêutica se adequará da melhor forma para cada caso (BERTUCCI et al. 2003). Com o passar dos anos, têm sido aplicadas variações no tratamento do câncer de mama, com o uso de terapias sistêmicas adjuvantes (quimioterapia e hormonioterapia) tendo a finalidade de controlar a doença. Inicialmente, a quimioterapia foi amplamente aplicada em pacientes portadoras de doença avançada. Entretanto, diversos estudos têm revelado que a quimioterapia sistêmica e a terapia hormonal também reduzem o risco de recorrência e mortalidade em pacientes portadoras de tumores precoces. Devido ao avanço das técnicas de diagnóstico e da maior divulgação de campanhas de educação e saúde pública, o diagnóstico de pacientes com câncer de mama em estágio precoce da doença sem o

envolvimento de linfonodos (que confere pior prognóstico) tem aumentado. A maior parte destas pacientes, cerca de 75%, não apresentarão recidivas. Portanto, é inapropriada a administração de terapia sistêmica para todas as pacientes com câncer de mama sem o envolvimento de linfonodos (FISHER et al. 1983; BLACK et al. 1997; CHIA et al. 2004). De acordo com essa questão, é importante definir quais pacientes irão se beneficiar do uso deste tipo de terapia daquelas que não se beneficiariam, uma vez que, ela envolve um tratamento de alto custo e com intensos efeitos colaterais.

Tumores de mama com aparentemente o mesmo tipo histológico e estadiamento variam amplamente em sua resposta às terapias. Novos marcadores de prognóstico e fatores preditivos são necessários para permitir a individualização da terapia (BERTUCCI et al. 2000).

A análise do perfil da expressão gênica permite avaliar a heterogeneidade do câncer de mama e identificar tumores com pior prognóstico e os subtipos tumorais que não seriam reconhecidos apenas por parâmetros histoclínicos usuais. A identificação de marcadores moleculares pode ser importante não somente para estabelecimento do prognóstico, mas também para a detecção da doença em estágios precoces, uma vez que nesta fase a doença é mais passível de cura. Também podem ser úteis no acompanhamento de pacientes, denunciando a recidiva da doença ou o seu desaparecimento (SIDRANSKY 1997).

A detecção precoce e a identificação de marcadores moleculares como fatores de prognóstico úteis para o câncer de mama permitirão a escolha da terapia mais adequada e eficaz para cada tumor e como consequência poderá melhorar a sobrevida global da população portadora deste tipo de câncer. A escolha do tratamento poderia diminuir a agressividade da evolução clínica. Seletivamente, poderiam ser aplicadas

terapias radicais mais precocemente, ajustadas às características moleculares dos tumores mais agressivos e também, terapias mais brandas poderiam ser instituídas às pacientes com câncer de mama com características moleculares mais favoráveis. O resultado de um tratamento individualizado pode possibilitar maior sobrevida e melhor qualidade de vida às pacientes portadoras desta neoplasia.

O perfil de expressão gênica pode contribuir para o desenvolvimento de novos agentes terapêuticos, assim como para a identificação de marcadores para o câncer de mama para monitorar tanto a progressão, quanto o tratamento da doença (JIANG et al. 2002). Alterações genéticas, resultando em níveis de expressão alterados de mRNA e proteínas têm sido amplamente descritas no processo de tumorigênese de mama, como por exemplo a ativação ou amplificação de oncogenes ou a perda de genes supressores de tumor (JONES et al. 1995; WALKER et al. 1997).

Recentemente, diversos genes têm sido descritos como tendo o aumento da expressão durante o desenvolvimento do câncer de mama (JIANG et al. 2002; MANNI et al. 2001; KAURANIEME et al. 2004). A maior parte destes genes está associada a mecanismos de controle de proliferação celular e podem representar potenciais marcadores prognósticos ou alvos para a terapia do câncer de mama.

Com o desenvolvimento de técnicas de biologia molecular de alta demanda, foram criadas novas abordagens para identificar e reconhecer o perfil da expressão gênica e um melhor entendimento do complexo processo biológico que é a tumorigênese. Técnicas como cDNA *microarray*, *SAGE* (VELCULESCU et al. 1995) e *MPSS* (*Massively Parallel Signature Sequencing*) (JONGENEEL et al. 2003) têm permitido a rápida análise no perfil da expressão gênica de um determinado tipo tumoral.

Além disso, a análise no nível de expressão de proteínas é também de grande importância para a identificação de possíveis marcadores moleculares para serem utilizados como alvos terapêuticos. A técnica de *Tissue Microarray (TMA)* (KONONEN et al. 1998) permite que centenas de amostras de tecidos sejam fixadas em uma lâmina e analisadas em um único experimento. Esta técnica possibilita a utilização de espécimes teciduais de arquivo clínico fixadas e emblocadas em parafina com o uso mínimo de reagentes, o que resulta também em economia de amostras de tecido com potencial para gerar um grande volume de informação para posterior correlação clínica. Grande parte da utilização da tecnologia de *TMA* tem surgido a partir de pesquisas correlacionadas ao câncer, como por exemplo a análise da frequência de alterações moleculares em um amplo número de tumores, na exploração da progressão tumoral, na identificação de fatores prognósticos e na validação de novos genes descobertos como potenciais marcadores de diagnóstico e alvos para terapia (KALLIONIEMI et al. 2001).

A disponibilidade dos dados de expressão gênica de diversos tecidos via *internet* é uma abordagem econômica para fornecer subsídios para a realização de inúmeros experimentos. Atualmente, há um grande volume de dados que foram gerados a partir de bibliotecas de *SAGE* que se encontram disponíveis em bancos de dados públicos e podem ser analisados por meio de ferramentas *on line*. O número de *SAGETags* reflete de forma direta a abundância dos transcritos em um determinado tipo de tecido, sendo portanto, os dados de *SAGE* altamente acurados e quantitativos. As informações depositadas nestes bancos de dados são de grande potencial como o ponto inicial de estudos de análise da expressão gênica. Entretanto, "vasculhar" estes bancos de dados, definir quais são os critérios mais relevantes e adaptá-los para uma determinada aplicação não é uma tarefa simples. Trabalhar com todos estes dados em

um curto período de tempo parece ser um desafio. Diversas ferramentas de bioinformática têm sido desenvolvidas com o objetivo de auxiliar e facilitar todo este processo de busca por genes diferencialmente expressos para posterior verificação experimental. Como exemplo, podemos citar o desenvolvimento do *website SAGEGenie* (<http://cgap.nci.nih.gov/SAGE>), que se trata de uma interface gráfica mantida pelo *CGAP (Cancer Genome Anatomy Project)* para a apresentação e análise dos dados de *SAGE*, na qual é fornecida uma série de ferramentas para análise da expressão gênica. Nesse *website*, podem ser realizadas comparações horizontais da expressão gênica (por exemplo, entre tecido de cérebro normal *versus* tecido de cérebro tumoral), bem como comparações verticais (diferentes tipos de tecidos ou tipos de órgãos) (BOON et al. 2002; LIANG 2002).

Diversos estudos têm utilizado a técnica de *SAGE* para análise da expressão gênica. LOGING et al. (2000) utilizaram os dados de *SAGE* para a identificação de marcadores moleculares para o glioblastoma multiforme (GBM) e a identificação de potenciais alvos para imunoterapia. Neste trabalho, os autores realizaram uma análise "*in silico*" a partir das bibliotecas de *SAGE* disponíveis *on line* e selecionaram dezessete genes candidatos a diferencialmente expressos para serem avaliados por meio de RT-PCR em tempo real. Destes genes, sete demonstraram-se potenciais candidatos a marcadores tumorais para o GBM.

HUMINIECKI e BICKNELL (2000) identificaram novos genes específicos do endotélio por meio da análise de bancos de dados. Os autores utilizaram duas diferentes fontes de dados de expressão: *Unigene/EST* (utilizando as bibliotecas de *ESTs* do *dbEST*) e *SAGE/CGAP* (utilizando as bibliotecas de *SAGE*). Após a identificação de um grupo de genes alvo do estudo, os autores partiram para a verificação experimental por meio de RT-PCR. Como resultado foram identificados

quatro novos genes específicos do endotélio (*ECSM1*, *ECSM2*, *ECSM3* e *Magic roundabout*).

CHEN et al. (2002) também fizeram uso dos dados de *SAGEtags* disponíveis *on line* para a identificação de novos genes a partir de *SAGEtags* que não mapeavam em genes conhecidos para testar a hipótese de que estas *SAGEtags* corresponderiam a novos transcritos ou novos genes que não haviam sido identificados no genoma humano. Os dados provenientes deste estudo forneceram evidências de que parte destas *SAGEtags* são de fato oriundas de novos transcritos e representam novos genes em potencial.

Em um outro trabalho, HOUGH et al. (2000) utilizaram os dados de *SAGE* para a identificação de marcadores moleculares para o câncer de ovário. Neste trabalho, os autores comparam o perfil da expressão gênica entre amostras de tecido de ovário tumoral e não tumoral a partir da construção de bibliotecas de *SAGE*, em que foram identificados diversos genes com diferença de expressão. Quatro dos genes encontrados com aumento de expressão no tecido tumoral foram então avaliados por meio de imuno-histoquímica. Destes, foram identificados três genes (*ApoJ*, *Claudin 3* e *Claudin 4*) que nunca haviam sido descritos em trabalhos prévios como relacionados ao câncer de ovário.

A comparação da expressão gênica por meio de *SAGE* foi realizada por St CROIX et al. (2000) para identificar genes endoteliais de tecido colo-retal normal e tumoral e determinar quais seriam diferencialmente expressos. Neste trabalho, os autores encontraram 46 transcritos com alto nível de expressão (mais de 10 vezes) no endotélio tumoral do que no normal e 33 transcritos foram identificados como expressos em baixos níveis no tecido tumoral em relação ao normal.

Sendo assim, há evidências de que a utilização dos dados gerados a partir da técnica de *SAGE* tem se mostrado como uma abordagem bastante eficaz para estudos de análise da expressão gênica. Com o propósito de contribuir para a identificação de possíveis alterações na expressão de genes envolvidos no processo inicial de tumorigênese de mama, o presente estudo buscou avaliar a expressão diferencial entre tecidos de mama neoplásica e não neoplásica de genes selecionados a partir de bancos de dados públicos de *SAGE*.

A confirmação da expressão diferencial observada *in silico* ou por outros métodos como por exemplo experimentos de cDNA *microarray* é necessária previamente à identificação de potenciais marcadores tumorais. A RT-PCR em tempo real é uma potente técnica para se quantificar de forma rápida e precisa a expressão gênica em diversas amostras, o que a torna extremamente útil para a quantificação de genes escolhidos por meio de pesquisa em bancos de dados públicos disponibilizados *on-line*.

No presente estudo, foi inicialmente gerada uma lista contendo 175 genes potencialmente mais expressos no câncer de mama em relação aos tecidos de mama normal, utilizando os dados das bibliotecas de *SAGE* disponíveis. A partir desta lista de genes, foi realizado um estudo desses candidatos através de uma análise da função, localização celular, dados de literatura, entre outras informações por meio de ferramentas disponíveis *on line*, o que resultou na escolha de dezenove genes para serem então validados quanto à expressão gênica diferencial por meio da técnica de RT-PCR em tempo real em amostras de tumores iniciais de mama. Dentre os dezenove genes inicialmente testados, onze genes apresentaram aumento de expressão em mais de 25% das amostras analisadas. Os genes *ALPPL2*, *CELSR2*, *GATA3* e *HMG20B* apresentaram aumento de expressão em mais de 50% das amostras

analisadas. Esses resultados demonstram o potencial da abordagem utilizada neste estudo para a escolha e validação de genes diferencialmente expressos.

O gene *ALPPL2* apresentou aumento de expressão em 59% das amostras tumorais testadas, quando comparado ao tecido normal. A proteína codificada por este gene é uma proteína semelhante à fosfatase alcalina placentária. As fosfatases alcalinas correspondem a um grupo de enzimas responsáveis pela hidrólise de compostos que contêm fosfatos, em pH alcalino. A fonte primária das fosfatases alcalinas placentárias humanas (cuja sigla do inglês é *PALP*) é a placenta, na qual a síntese destas enzimas ocorre durante o segundo e o terceiro trimestre da gravidez (SHE et al. 2000). É sugerido que *PALP* esteja envolvida na transferência da IgG maternal para o feto (STEFANER et al. 1997) e um acúmulo de evidências indica um possível envolvimento de *PALP* no processo de divisão de células normais e transformadas (DABARE et al. 1999). A expressão de *PALP* também é observada em diversos tipos tumorais, mas sua contribuição para o processo de tumorigênese ainda não foi elucidada. Desde que o crescimento fetal e tumoral apresentam características comuns, é provável que o aumento da expressão de *PALP* altere tanto a proliferação do tecido fetal, quanto de tecidos cancerosos. SHE et al. (2000) estudaram os efeitos da expressão de *PALP* na proliferação de fibroblastos de camundongos e em fibroblastos de fetos humanos não transformados e transformados por *H-ras*. Neste trabalho, os autores demonstraram que nestas linhagens celulares, a proteína *PALP* altamente purificada seria responsável por estimular a síntese de DNA e a proliferação celular através de várias enzimas regulatórias do processo de proliferação celular, já conhecidas. A isoenzima *ALPPL2* é encontrada em testículo e timo, sendo que o nível desta enzima é elevado no soro de pacientes portadores de tumores de células germinativas de testículo, principalmente seminomas.

O gene *HMG20B* apresentou um aumento da expressão em 54,5% das amostras testadas, quando comparado ao tecido normal. A proteína codificada por este gene é muito semelhante à Hmg20b murina e pode ser a subunidade central do complexo ativador transcricional SWI/SNF. O grupo de proteínas *HMG* (*high mobility group*) é composto de proteínas cromossomais não-histonas que contêm um domínio conservado de ligação com o DNA chamado de *HMG box* e estão envolvidos no processo de regulação da transcrição gênica e no modelamento da cromatina. *HMG20B* é também denominado como *BRAF35*, cuja proteína foi originalmente identificada como um componente ligante de DNA do multicomplexo de *BRCA2* (importante gene envolvido no câncer de mama hereditário). MARMORSTEIN et al. (2001), demonstraram a associação entre *BRAF35* e *BRCA2* com a cromatina durante as fases iniciais da condensação dos cromossomos na mitose e por meio de experimentos com microinjeções de anticorpos, sugerindo o envolvimento deste complexo na regulação da progressão do ciclo celular. De forma similar aos outros complexos regulatórios, o complexo *BRAF35/BRCA2* não confere somente um papel regulatório na progressão do ciclo celular, mas também tem sido descrito como envolvido no reparo de dano ao DNA.

O gene *CELSR2* apresentou-se com aumento de expressão em 81,8% das amostras de tumores testadas, quando comparado ao tecido normal. Apesar da função dessa proteína ainda não ter sido esclarecida, há evidências por estudos em outros organismos que ela possa estar envolvida na mediação da comunicação célula-célula, que é um mecanismo importante no que diz respeito ao envio de sinais para processos como crescimento, proliferação e diferenciação celulares. Nosso estudo forneceu indícios de que o aumento da expressão de *CELSR2* é um evento freqüente no início

do processo de tumorigênese da mama, uma vez que o aumento significativo de sua expressão foi observado na maioria das amostras testadas.

Um outro resultado importante obtido no presente estudo, foi o referente à expressão do gene *GATA3*. Este gene apresentou o aumento da expressão em 77,2% das amostras tumorais testadas. O gene *GATA3* faz parte da família GATA de fatores de transcrição do tipo dedos de zinco, na qual existem pelo menos seis membros já identificados, os quais estão divididos em duas subfamílias que são importantes na regulação de diversos genes e são expressos em uma variedade de tecidos durante o desenvolvimento e em adultos (SIMON 1995). De maneira geral, as proteínas da família GATA têm sido associadas a papéis críticos dentro do processo de desenvolvimento, incluindo a regulação da diferenciação celular, o controle da proliferação e do movimento celular (LACROIX e LECLERCQ 2004). A subfamília composta por *GATA 1, 2 e 3* é predominantemente expressa em células sangüíneas hematopoiéticas e endoteliais (WEISS e ORKIN 1995). *GATA 4, 5 e 6* são os membros da segunda subfamília e são expressos principalmente em tecidos do sistema cardiovascular e em tecidos derivados do endoderma, incluindo fígado, pulmão, pâncreas e intestino (SUZUKI et al. 1996; LACROIX e LECLERCQ 2004). *GATA3* é altamente expresso em células linfóides T e acredita-se que esteja envolvido na diferenciação destas células (HO et al. 1991; JOULIN et al 1991). Este gene é também altamente expresso em áreas de cérebro embrionário, sistema nervoso central e periférico de adultos e em vários outros tecidos, incluindo rim e endotélio (ORKIN 1992). A presença de *GATA3* é observada em tecido de mama e próstata normais, que são dois tecidos cuja proliferação é dependente de hormônios esteróides. Uma correlação estatisticamente significativa entre a expressão do receptor de estrógeno e a expressão de *GATA3* foi descrita por HOCH et al. (1999), em um trabalho em que

utilizaram cDNA *microarray* para comparar o padrão da expressão gênica entre linhagens celulares derivadas de câncer de mama discordantes quanto à expressão de receptores de estrógeno (ER). Estes experimentos indicaram a abundante expressão de *GATA3* nas linhagens celulares ER+ e a mínima ou nenhuma expressão deste gene nas linhagens celulares ER-. Um painel composto de 47 amostras de tumores primários de mama invasivos foi usado para avaliar a expressão de PR, ER e *GATA3* por meio de ensaio por imunoperoxidase, o qual estabeleceu uma correlação estatisticamente significativa ($p < 0,0001$) entre a expressão de *GATA3* e ER. BERTUCCI et al. (2000), desenvolveram um estudo utilizando cDNA *microarray* com 1276 genes candidatos em 34 amostras de tumores primários em mama (três amostras TI, 20 amostras TII, 10 amostras TIII e 1 amostra não avaliável) e observaram a alta expressão de *GATA3* em tumores com fenótipo positivo para o ER, sugerindo a importância deste gene em tumor de mama e seu envolvimento com ER.

A análise estatística no nosso estudo revelou associações estatisticamente significativas entre a expressão dos genes *CELSR2*, *GATA3*, *GATA4* e *KIAA0141* e a positividade para a imunodeteção do receptor de estrógeno (ER) e marginalmente significativa entre a expressão do gene *CTSF* e a positividade para a imunodeteção de ER. Uma correlação estatisticamente significativa foi encontrada entre a expressão do gene *CELSR2* e positividade para a imunodeteção do receptor de progesterona (PR) e marginalmente significativa entre a expressão do gene *GATA3* e a positividade para a imunodeteção do receptor de progesterona.

A glândula mamária, o ovário e o útero nas mulheres são os principais tecidos-alvo dos hormônios sexuais que incluem o estrógeno, progesterona e andrógeno. Estrógeno é um hormônio importante para a proliferação, diferenciação e função nos tecidos reprodutivos (CURTIS et al. 2000). Evidências têm se acumulado de que

estrógeno apresenta efeito mitogênico em tumores de mama. Os mecanismos que causam esse efeito ainda não são totalmente entendidos, mas provavelmente resultam do complexo mecanismo de modulação que os receptores de estrógeno exercem sobre a transcrição gênica (PLATET et al. 2004). A progesterona é um regulador essencial nos eventos reprodutivos associados à gravidez, incluindo a ovulação, o desenvolvimento do útero e da glândula mamária (CONNELLY et al. 2002). O andrógeno é indispensável para o desenvolvimento e a função normal dos tecidos reprodutivos. Estes hormônios exercem suas funções em seus tecidos-alvo, por meio de receptores intracelulares específicos: os receptores de estrógeno (cuja sigla do inglês é ER), os receptores de progesterona (cuja sigla do inglês é PR) e os receptores de andrógeno (cuja sigla do inglês é AR), os quais fazem parte da superfamília de receptores nucleares (cuja sigla do inglês é NR) e funcionam como fatores de transcrição que regulam a expressão de genes alvo. Existem dois receptores para estrógeno, ESR1 (alfa, ER- α) e ESR2 (beta, ER- β). Embora essas duas proteínas sejam muito semelhantes em seus domínios de ligação ao DNA e aos esteróides, ESR2 está localizado em um diferente cromossomo e codifica uma proteína mais curta (ANDERSON 2002). O padrão de expressão dessas duas proteínas também difere. A porcentagem de células ESR1 positivas é geralmente baixa nas células epiteliais de glândulas mamárias fora da lactação (20 a 30%), mas aumenta em doenças proliferativas benignas, particularmente quando associadas com atipia celular e em carcinomas ductais *in situ* de baixo grau, sugerindo que uma elevada expressão desse gene possa estar associada ao maior risco de tumorigênese. Ao contrário, os níveis de ESR2 diminuem na progressão de hiperplasia ductal a DCIS, enquanto que em DCIS de alto grau a expressão de ambos receptores diminui. A expressão de ESR1 é encontrada em 50 a 80% dos tumores invasivos, enquanto vários estudos indicam que

a expressão de *ESR2* diminui em tumores de mama invasivos em relação ao tecido adjacente normal (PLATET et al. 2004).

O conhecimento do *status* de receptores de estrógeno e também de progesterona é essencial para decisões terapêuticas em relação à hormonioterapia.

Cerca de um terço dos carcinomas de mama são hormônio dependentes, ou seja, expressam receptores de estrógenos e respondem à abordagem terapêutica com o uso de anti-estrógenos, como por exemplo o tamoxifeno (JORDAN 1999). Em células mamárias sensíveis, o tamoxifeno liga-se a ER e este, por sua vez, não ativa a transcrição gênica. Em células mamárias que se tornam resistentes à ação dos moduladores seletivos de receptor de estrógeno, como é caso do tamoxifeno, é observada a maior expressão de receptores de fator de crescimento EGFR e ERBB2, por meio da ativação da via de sinalização da MAPK (quinase ativada por mitógeno), a qual pode fosforilar ER, facilitando a ligação do receptor com proteínas co-ativadoras, o que restabelece sua atividade transcricional e o estímulo da proliferação celular. A ação de anti-estrógenos, não somente exclui co-ativadores ligantes, mas também pode facilitar o recrutamento de co-repressores e resultar em ação repressiva da expressão basal de certos genes (FOLGUEIRA e BRENTANI 2004).

Tumores com o fenótipo ER negativo são resistentes ao tratamento adjuvante. Pacientes portadoras de tumores ER negativos não apresentam sempre a mesma resposta à quimioterapia. A ausência da expressão de ER em tumores de mama está associada também a um fenótipo mais agressivo e tem se tornado bem estabelecido que pacientes com câncer de mama ER- (hormônio-independente) apresentam menor tempo de sobrevida livre de doença e menor sobrevida global quando comparadas com pacientes portadoras de tumores com o fenótipo ER+ (McGUIRE e CLARK 1992; NAGAI et al. 1994). Dentre os tumores ER positivos, aqueles que são

concomitantemente PR positivos têm maior probabilidade de responder à terapia hormonal que os que são negativos para o receptor de progesterona (JACOBSEN et al. 2003). Diversos estudos têm investigado as diferenças no perfil de expressão gênica entre os tumores de mama ER+ e ER-.

Entretanto, a relação entre resposta ao tratamento hormonal e o *status* de ER não é perfeita e isso torna possível que outros genes relacionados à expressão de ER possam ser mais importantes do que o próprio ER para caracterizar o fenótipo hormônio-responsivo. Tais genes podem ser úteis como fatores preditivos na terapia e estimular a pesquisa levando à identificação de novos alvos terapêuticos (BERTUCCI et al. 2000).

Um estudo recente realizado por LACROIX e LECLERCQ (2004) avaliou a co-expressão de três genes (*GATA3*, *HNF3A* e *XBPI*) com o gene de receptor de estrógeno (*ESR1*) em tumores de mama e em linhagens celulares derivadas de câncer de mama. Como resultado deste trabalho os autores sugeriram que estes três fatores de transcrição exercem um papel importante no acompanhamento e controle dos efeitos mediados por ER α no câncer de mama.

A indução da proliferação celular é o maior efeito mediado por ER- α em células de câncer de mama. Para que este processo ocorra é necessário energia e supostamente uma grande síntese de proteínas. Uma outra forma de controlar a atividade de ER- α é modulando a atividade de seus ligantes. As seqüências ligantes para GATA3 são freqüentemente encontradas entre outros elementos comuns do DNA, sugerindo que os fatores GATA podem exercer uma função essencial em cooperação com outras proteínas ligantes de DNA. A forte correlação entre a expressão de ER α e GATA3 sugere que GATA3 possa cooperar com esse receptor esteróide para regular genes hormônio-responsivos, específicos para o tecido de mama

(LACROIX e LECLERCQ 2004). O fato de não ser a resposta à hormonioterapia função somente da positividade dos receptores de estrógenos, a identificação de genes adicionais envolvidos nesse processo é essencial. O nosso estudo além de ter confirmado a correlação positiva entre expressão de GATA3 e ER, revelou outros genes cujo aumento de expressão está associado à positividade do receptor de estrógenos. A extensão desse trabalho com análise de um grande número de casos com longo tempo de acompanhamento pode indicar se o aumento de expressão desses genes pode ser usado como um fator preditivo, adicional à expressão dos receptores de estrógeno, da resposta à hormonioterapia.

Outro fato que demonstra a utilidade da abordagem adotada neste estudo para a identificação de genes candidatos a marcadores em câncer de mama foi a recente publicação de um trabalho por OLAYIOYE et al. (2004), que estudaram *STARD10*, um dos genes escolhidos para validação neste estudo, e este apresentou-se com aumento de expressão de 2 a 12 vezes em 25% das amostras testadas. Nesse trabalho, os autores demonstraram que a fosfoproteína *STARD10* apresenta aumento de expressão em tumores de mama em humanos e em camundongos. Os autores encontraram o aumento da expressão em 35% dos carcinomas de mama primários e em 64% das linhagens celulares de câncer de mama, sendo mais freqüente a expressão em tecidos com aumento de expressão de *ERBB2*. A coexpressão de *STARD10* com *ERBB1* em fibroblastos de murinos aumentou o crescimento independente de ancoragem no ensaio de *soft* ágar, evidenciando uma cooperação funcional entre *STARD10* e *ERBB1*. Juntos, estes dados sugerem que o aumento da expressão desta proteína pode contribuir para a oncogênese da mama e nossos dados sugerem que o aumento de expressão desse gene ocorre em uma porcentagem de tumores iniciais.

Os resultados apresentados no presente estudo demonstraram que a estratégia de escolha de genes a partir da análise de banco de dados públicos disponíveis *on line* foi adequada, por terem sido validados por meio de RT-PCR em tempo real mais da metade dos genes selecionados. Além disso, essa estratégia mostrou-se com um potencial de rápida execução, uma vez que experimentos são necessários apenas para a validação dos candidatos escolhidos e não para a escolha dos mesmos. A escolha da técnica de RT-PCR em tempo real como método de validação foi bastante efetiva, permitindo que, com o uso de pequena quantidade de RNA de cada amostra, a análise da expressão dos múltiplos genes selecionados na primeira etapa do estudo de uma maneira rápida e precisa. Outro fator que demonstra a utilidade da abordagem escolhida é a obtenção, mesmo com um pequeno número de amostras, da correlação positiva e estatisticamente significativa da expressão de alguns dos genes testados com a positividade dos receptores de estrógeno, provavelmente demonstrando a importância desses genes na tumorigênese da mama. Além disso, após o início do estudo foram publicados trabalhos, como o desenvolvido por OLAYIOYE et al. (2004), que mostraram o papel de um dos genes escolhidos nesse estudo, como o gene *STARD10* na transformação celular e na tumorigênese da mama. Entretanto, o que entendemos que seja uma etapa crucial nesse tipo de trabalho é o estudo detalhado de cada gene e estabelecimento dos critérios de escolha dos candidatos para validação. Essa etapa, provavelmente a mais laboriosa e que requer estudos detalhados, é decisiva para o sucesso desse tipo de estudo e aparentemente os critérios usados neste estudo foram adequados. Embora os resultados obtidos no presente estudo forneçam indícios de uma provável correlação entre o aumento da expressão dos genes e o processo de tumorigênese de mama, um estudo mais amplo, com um maior número de amostras e um maior tempo e acompanhamento das pacientes é imprescindível para

confirmar os resultados obtidos e para demonstrar se os achados moleculares aqui mostrados poderiam ser utilizados como marcadores para o acompanhamento de pacientes portadoras do câncer de mama ou como alvos terapêuticos.

6 CONCLUSÕES

- A abordagem utilizada neste estudo para a identificação de genes diferencialmente expressos potencialmente úteis como marcadores prognósticos, diagnósticos ou alvos terapêuticos, que envolveu a análise dos dados de expressão gênica de tecidos normais e tumorais de mama disponíveis no Banco de Dados de *SAGE* mantido pelo Instituto Nacional do Câncer dos Estados Unidos, disponibilizado no sítio <http://cgap.nci.nih.gov>, foi validada pois, dos 175 genes inicialmente identificados, 19 foram escolhidos para avaliação e destes foram encontrados onze genes (*ALPPL2*, *CELSR2*, *GATA3*, *GATA4*, *FN3KRP*, *HMG20B*, *PTPRF*, *QSCN6*, *SCARB2*, *START10* e *ZC3HDC1*) com aumento da expressão em pelo menos mais de 25% das amostras tumorais analisadas em comparação ao tecido normal. Desses, o aumento de expressão de quatro deles (*ALPPL2*, *CELSR2*, *GATA3*, *HMG20B*) parece ser um evento freqüente na tumorigênese da mama por ter sido constatado em mais de 50% das amostras tumorais analisadas.
- Este trabalho confirmou que a identificação de genes com diferença de expressão através da utilização de bancos de dados disponíveis publicamente e posterior validação dos genes resultantes desta seleção por meio de RT-PCR em tempo real é uma estratégia capaz de identificar potenciais marcadores para o câncer.
- Embora os resultados obtidos neste estudo indiquem haver uma correlação entre o aumento da expressão dos genes e o processo de tumorigênese de mama, um estudo mais amplo, com um maior número de amostras e com

maior tempo de acompanhamento das pacientes é necessário para demonstrar, de forma irrefutável, que estes achados moleculares possam ser utilizados como marcadores para o diagnóstico e acompanhamento de pacientes portadoras desta neoplasia.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Adida C, Crotty PL, McGrath J, Berrebi D, Diebold J, Altieri DC. Developmentally regulated expression of the novel cancer anti-apoptosis gene survivin in human and mouse differentiation. **Am J Pathol** 1998; 152:43-9.

Anderson E. The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis. **Breast Cancer Res** 2002; 4:197-201.

Ambrosini G, Adida C, Altieri DC. A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. **Nat Med** 1997; 3:917-21.

Barbati A, Cosmi EV, Sidoni A, et al. Value of c-erbB-2 and p53 oncoprotein co-overexpression in human breast cancer. **Anticancer Res** 1997; 17:401-405.

Baselga J, Tripathy D, Mendelsohn J, et al. Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with HER2/neu-overexpressing metastatic breast cancer. **J Clin Oncol** 1996; 14:737-744.

Beckmann MW, Niederacher D, Schnurch HG, Gusterson BA, Bender HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. **J Mol Med** 1997; 75:429-39.

Bertucci F, Houlgatte R, Benziene A, et al. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. **Hum Mol Genet** 2000; 9:2981-91.

Bertucci F, Loriod B, Nasser V, et al. Gene expression profiling of breast carcinomas using nylon DNA arrays. **C R Biol** 2003; 326:1031-9.

Black RJ, Bray F, Ferlay J, Parkin DM. Cancer incidence and mortality in the European Union: cancer registry data and estimates of national incidence for 1990. **Eur J Cancer** 1997;1075-107. Erratum in: **Eur J Cancer** 1997;33:2440.

Boon K, Osorio EC, Greenhut SF, et al. An anatomy of normal and malignant gene expression. **Proc Natl Acad Sci U S A** 2002; 99:11287-92.

Brendza RP, Serbus LR, Duffy JB, Saxton WM. A function for kinesin I in the posterior transport of oskar mRNA and Stauf protein. **Science** 2000; 289:2120-2.

Camargo AA, Samaia HP, Dias-Neto E, et al. The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. **Proc Natl Acad Sci U S A** 2001; 98:12103-8. Erratum in: **Proc Natl Acad Sci USA** 2004; 101:414. Melo M B [corrected to Barbosa de Melo M].

Chae J, Kim MJ, Goo JH, et al. The Drosophila tissue polarity gene starry night encodes a member of the protocadherin family. **Development** 1999; 126:5421-9.

Chapman HA, Riese RJ, Shi GP. Emerging roles for cysteine proteases in human biology. **Annu Rev Physiol** 1997; 59:63-88.

Chen J, Sun M, Lee S, Zhou G, Rowley JD, Wang SM. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. **Proc Natl Acad Sci U S A** 2002; 99:12257-62.

Chia SK, Speers CH, Bryce CJ, Hayes MM, Olivotto IA. Ten-year outcomes in a population-based cohort of node-negative, lymphatic, and vascular invasion-negative early breast cancers without adjuvant systemic therapies. **J Clin Oncol** 2004; 22:1630-7.

Chrigwin MJ, Przybyla EA, MacDonald R and Rutter J. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. **Biochemistry** 1979; 18:5294-5299.

Collard F, Delpierre G, Stroobant V, Matthijs G, Van Schaftingen E. A mammalian protein homologous to fructosamine-3-kinase is a ketosamine-3-kinase acting on psicossamines and ribulosamines but not on fructosamines. **Diabetes** 2003; 52:2888-95.

Colpaert C, Vermeulen P, Jeuris W, et al. Early distant relapse in "nodo negative" breast cancer patients is not predicted by occult axillary lymph node metastases, but by the features of the primary tumor. **J Pathol** 2001; 193:442-9.

Conneely OM, Mulac-Jericevic B, DeMayo F, Lydon JP, O'Malley BW. Reproductive functions of progesterone receptors. **Recent Prog Horm Res** 2002; 57:339-55.

Coppock DL, Kopman C, Scandalis S, Gilleran S. Preferential gene expression in quiescent human lung fibroblasts. **Cell Growth Differ** 1993; 4:483-93.

Coppock DL, Cina-Poppe D, Gilleran S. The quiescin Q6 gene (QSCN6) is a fusion of two ancient gene families: thioredoxin and ERV1. **Genomics** 1998; 54:460-8.

Curtis Hewitt S, Couse JF, Korach KS. Estrogen receptor transcription and transactivation: estrogen receptor knockout mice: what their phenotypes reveal about mechanisms of estrogen action. **Breast Cancer Res** 2000; 2:345-52. Epub 2000 Jul 12.

Dabare AA, Nouri AM, Cannell H, Moss T, Nigam AK, Oliver RT. Profile of placental alkaline phosphatase expression in human malignancies: effect of tumour cell activation on alkaline phosphatase expression. **Urol Int** 1999; 63:168-74.

de la Rochefordiere A, Asselain B, Campana F, et al. Age as prognostic factor in premenopausal breast carcinoma. **Lancet** 1993; 341:1039-43.

Delcour J, Papaconstantinou J. Biochemistry of bovine lens proteins. IV. Synthesis and aggregation of -crystallin subunits in differentiating lens cells. **J Biol Chem** 1972; 247:3289-95.

Demczuk S, Thomas G, Aurias A. Isolation of a novel gene from the DiGeorge syndrome critical region with homology to *Drosophila* *gdl* and to human LAMC1 genes. **Hum Mol Genet** 1996; 5:633-8.

Dodelet VC, Pasquale EB. Eph receptors and ephrin ligands: embryogenesis to tumorigenesis. **Oncogene** 2000; 19:5614-9.

Early Breast Cancer Trialist's Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. **Lancet** 1992; 339:71-85.

Eeles RA. Future possibilities in the prevention of breast cancer: intervention strategies in BRCA1 and BRCA2 mutation carriers. **Breast Cancer Res** 2000; 2:283-90.

Feinberg AP. Cancer epigenetics takes center stage. **Proc Natl Acad Sci USA** 2001; 98:392-4.

Fisher B, Bauer M, Wickerham DL, et al. Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. **Cancer** 1983; 52:1551-7.

Folgueira MAAK, Brentani MM. Câncer de mama. In: Ferreira CG, Rocha JCC, editores. **Oncologia molecular**. São Paulo: Atheneu; 2004. p.135-44.

Gamp AC, Tanaka Y, Lullmann-Rauch R, et al. LIMP-2/LGP85 deficiency causes ureteric pelvic junction obstruction, deafness and peripheral neuropathy in mice. **Hum Mol Genet** 2003; 12:631-46.

Greene FL, Page DL, Fleming ID, et al. **AJCC cancer staging manual**. 6th ed. New York: Springer Verlag; 2002.

Gu Z, Flemington C, Chittenden T, Zambetti GP. ei24, a p53 response gene involved in growth suppression and apoptosis. **Mol Cell Biol** 2000; 20:233-41.

Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. **Science** 1990; 250:1684-9.

Hirai H, Maru Y, Hagiwara K, Nishida J, Takaku F. A novel putative tyrosine kinase receptor encoded by the eph gene. **Science** 1987; 238:1717-20.

Hirokawa N. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. **Science** 1998; 279:519-26.

Ho IC, Vorhees P, Marin N, et al. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. **EMBO J** 1991; 10:1187-92.

Hoch RV, Thompson DA, Baker RJ, Weigel RJ. GATA-3 is expressed in association with estrogen receptor in breast cancer. **Int J Cancer** 1999; 84:122-8.

Hough CD, Sherman-Baust CA, Pizer ES, et al. Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. **Cancer Res** 2000; 60:6281-7.

Huminiecki L, Bicknell R. In silico cloning of novel endothelial-specific genes. **Genome Res** 2000; 10:1796-806.

Jacobsen BM, Richer JK, Sartorius CA, Horwitz KB. Expression profiling of human breast cancers and gene regulation by progesterone receptors. **J Mammary Gland Biol Neoplasia** 2003; 8:257-68.

Jemal A, Tiwari RC, Murray T, et al. Cancer statistics, 2004. **CA Cancer J Clin** 2004; 54:8-29.

Jiang Y, Harlocker SL, Molesh DA, et al. Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays. **Oncogene** 2002; 21:2270-82.

Jongeneel CV, Iseli C, Stevenson BJ, et al. Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing. **Proc Natl Acad Sci U S A** 2003; 100:4702-5.

Jones KA, Brown MA and Solomon E. Molecular genetics of sporadic and familial breast cancer. **Cancer Surv** 1995; 25:315-34.

Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. **Science** 2001; 293:1068-70.

Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. **Nat Rev Genet** 2002; 3:415-28.

Jordan VC. Estrogen receptor as a target for the prevention of breast cancer. **J Lab Clin Med** 1999; 133:408-14.

Joulin V, Bories D, Eleouet JF, Labastie MC, Chretien S, Mattei MG, Romeo PH. A T-cell specific TCR delta DNA binding protein is a member of the human GATA family. **EMBO J** 1991; 10:1809-16.

Kallioniemi OP, Wagner U, Kononen J, Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. **Hum Mol Genet** 2001; 10:657-62.

Kato J, Kuwabara Y, Mitani M, et al. Expression of survivin in esophageal cancer: correlation with the prognosis and response to chemotherapy. **Int J Cancer** 2001; 95:92-5.

Kauraniemi P, Hautaniemi S, Autio R, et al. Effects of Herceptin treatment on global gene expression patterns in HER2-amplified and nonamplified breast cancer cell lines. **Oncogene** 2004; 23:1010-3.

Kawasaki H, Altieri DC, Lu CD, Toyoda M, Tenjo T, Tanigawa N. Inhibition of apoptosis by survivin predicts shorter survival rates in colorectal cancer. **Cancer Res** 1998; 58:5071-4.

Kawasaki H, Toyoda M, Shinohara H, et al. Expression of survivin correlates with apoptosis, proliferation, and angiogenesis during human colorectal tumorigenesis. **Cancer** 2001; 91:2026-32.

Khan S, Abdelrahim M, Samudio I, Safe S. Estrogen receptor/Sp1 complexes are required for induction of cad gene expression by 17beta-estradiol in breast cancer cells. **Endocrinology** 2003; 144:2325-35.

Kim E, Lowenson JD, MacLaren DC, Clarke S, Young SG. Deficiency of a protein-repair enzyme results in the accumulation of altered proteins, retardation of growth, and fatal seizures in mice. **Proc Natl Acad Sci U S A** 1997; 94:6132-7.

King L, Half JM, Lee MK, Newman B. Linkage of early onset breast cancer to chromosome 17q21. **Science** 1990; 250:1684-89.

Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. **Nat Med** 1998; 4:844-7.

Krop IE, Sgroi D, Porter DA, et al. HIN-1, a putative cytokine highly expressed in normal but not cancerous mammary epithelial cells. **Proc Natl Acad Sci U S A** 2001; 98:9796-801.

LaCasse EC, Baird S, Korneluk RG & MacKenzie AE. The inhibitors of apoptosis (IAPs) and their emerging role in cancer. **Oncogene** 1998; 17:3247-59.

Lacroix M, Leclercq G. About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. **Mol Cell Endocrinol** 2004; 219:1-7.

Loging WT, Lal A, Siu IM, et al. Identifying potential tumor markers and antigens by database mining and rapid expression screening. **Genome Res** 2000; 10:1393-1402.

Lu CD, Altieri DC, Tanigawa N. Expression of a novel antiapoptosis gene, survivin, correlated with tumor cell apoptosis and p53 accumulation in gastric carcinomas. **Cancer Res** 1998; 58:1808-12.

Mackay A, Jones C, Dexter T, et al. cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. **Oncogene** 2003; 22:2680-8.

Manni A, Trout D, Verderame MF, Washington S, Mauger D, Demers L. Effect of alpha-difluoromethyl-ornithine on the expression and function of the epidermal growth factor receptor in human breast epithelial cells in culture. **Breast Cancer Res Treat** 2001; 68:139-46.

Marmorstein LY, Kinev AV, Chan GK, et al. A human BRCA2 complex containing a structural DNA binding component influences cell cycle progression. **Cell** 2001;104:247-57.

McGuire WL, Clark GM. Prognostic factors and treatment decisions in axillary-node-negative breast cancer. **N Engl J Med** 1992;326:1756-61.

Miki H, Setou M, Kaneshiro K, Hirokawa N. All kinesin superfamily protein, KIF, genes in mouse and human. **Proc Natl Acad Sci U S A** 2001; 98:7004-11.

Morrison TB, Weiss JJ, Wittwer CT. Quantification of low-copy transcripts by continuous SYBR green I monitoring during amplification. **Biotechniques** 1998; 24:954-958,960,962.

Muller HM, Widschwendter A, Fiegl H, et al. DNA methylation in serum of breast cancer patients: an independent prognostic. **Cancer Res** 2003; 63:7641-5.

Nagai MA, Marques LA, Yamamoto L, Fujiyama CT, Brentani MM. Estrogen and progesterone receptor mRNA levels in primary breast cancer: association with patient survival and other clinical and tumor features. **Int J Cancer** 1994;59:351-6.

Nacht M, Ferguson AT, Zhang W, et al. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. **Cancer Res** 1999; 59:5464-70.

Nixon AJ, Neuberg D, Hayes DF, et al. Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer. **J Clin Oncol** 1994; 12:888-94.

Noren NK, Lu M, Freeman AL, Koolpe M, Pasquale EB. Interplay between EphB4 on tumor cells and vascular ephrin-B2 regulates tumor growth. **Proc Natl Acad Sci U S A** 2004; 101:5583-8.

Ogawa K, Pasqualini R, Lindberg RA, Kain R, Freeman AL, Pasquale EB. The ephrin-A1 ligand and its receptor, EphA2, are expressed during tumor neovascularization. **Oncogene** 2000; 19:6043-52.

Olayioye MA, Hoffmann P, Pomorski T, et al. The phosphoprotein StarD10 is overexpressed in breast cancer and cooperates with ErbB receptors in cellular transformation. **Cancer Res** 2004; 64:3538-44.

Orkin SH. GATA-binding transcription factors in hematopoietic cells. **Blood** 1992; 80:575-81.

Ormandy CJ, Musgrove EA, Hui R, Daly RJ, Sutherland RL. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. **Breast Cancer Res Treat** 2003; 78:323-35.

Overgaard J, Yilmaz M, Guldborg P, Hansen LL, Alsner J. TP 53 mutation is an independent prognostic marker for poor outcome in both node-negative and node-positive breast cancer. **Acta Oncol** 2000; 39:327-33.

Park BH, Vogelstein B. Tumor-suppressor genes: BRCA1 and BRCA2 genes.. In: Kurfe DW, Pollock RE, Weishelbaum RR, Bast RC, Gansler T, Rolland J, editors. **Holland Frei cancer medicine**. 6th ed. Canadá: Feri Emil; 2003. Available from <URL:<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowSection&rid=cmed6.chapter.1566>> [2004 Ago 3]

Platet N, Cathiard AM, Gleizes M, Garcia M. Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. **Crit Rev Oncol Hematol** 2004; 51:55-67.

Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. **Nature** 2000; 406:747-52.

Polyak K, Xia Y, Zweier JL, Kinzler KW, Vogelstein B. A model for p53-induced apoptosis. **Nature** 1997; 389:300-5.

Polyak K. On the birth of breast cancer. **Biochim Biophys Acta** 2001; 1552:1-13.

Polyak K. Molecular alterations in ductal carcinoma in situ of the breast. **Curr Opin Oncol** 2002; 14:92-6.

Porter D, Weremowicz S, Chin K, et al. A neural survival factor is a candidate oncogene in breast cancer. **Proc Natl Acad Sci U S A** 2003a;100:10931-6.

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

Rudolph P, Alm P, Olsson H, et al. Concurrent overexpression of p53 and c-erbB-2 correlates with accelerated cycling and concomitant poor prognosis in node-negative breast cancer. **Hum Pathol** 2001; 32:311-9.

Sanguinetti CJ, Dias Neto E, Simpson AJ. Rapid silver staining and recovery of PCR products separated on polyacrilamide gels. **Biotechniques** 1994; 17:914-21.

Santamaria I, Velasco G, Pendas AM, Paz A, Lopez-Otin C. Molecular cloning and structural and functional characterization of human cathepsin F, a new cysteine proteinase of the papain family with a long propeptide domain. **J Biol Chem** 1999; 274:13800-9.

Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW. Microarrays: biotechnology's discovery platform for functional genomics. **Trends Biotechnol** 1998; 16:301-6.

Schmitt F. Mama. In: Brasileiro filho G, editor. **Bogliolo patologia**. 6 ed. Rio de Janeiro: Guanabara Koogan; 2000. p.558.

Schmitt FC. Multistep progression from an oestrogen-dependent growth towards an autonomous growth in breast carcinogenesis. **Eur J Cancer** 1995; 31A:2049-52.

Scanlan MJ, Chen YT, Williamson B, et al. Characterization of human colon cancer antigens recognized by autologous antibodies. **Int J Cancer** 1998; 76:652-8.

Sharma VK, Dean-Nystrom EA, Casey TA. Semiautomated fluorogenic PCR assays (TaqMan) for rapid detection of Escherichia coli O157: H7 and other Shiga toxigenic E. coli. **Mol Cel Probes** 1999; 13:291-302.

Sharp DJ, Rogers GC, Scholey JM. Microtubule motors in mitosis. **Nature** 2000; 407:41-7.

She QB, Mukherjee JJ, Huang JS, Crilly KS, Kiss Z. Growth factor-like effects of placental alkaline phosphatase in human fetus and mouse embryo fibroblasts. **FEBS Lett** 2000; 469:163-7.

Shin JH, Nolte FS, Holloway BP, Morrison CJ. Rapid identification of up three Candida species in a single reaction tube by a 5' exonuclease assay using fluorescent DNA probes. **J Clin Microbiology** 1999; 37:165-70.

Sidransky D. Nucleic acid-based methods for the detection of cancer. **Science** 1997; 278(5340):1054-9.

Sidransky D. Emerging molecular markers of cancer. **Nature Rev Cancer** 2002; 2:210-9.

Silvestrini R, Daidone MG, Luisi A, et al. Biologic and clinicopathologic factors as indicators of specific relapse types in node-negative breast cancer. **J Clin Oncol** 1995; 13:697-704.

Simon MC. Gotta have GATA. **Nat Genet** 1995; 11:9-11.

Slamon DJ, Leyland-Jones B, Shak S. et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. **N Engl J Med** 2001; 344:783-92.

Sobin LH, Wittekind Ch. **TNM classificação dos tumores malignos**. Trad. De Ministério da Saúde. 5 ed. Rio de Janeiro: Coordenação de Programas do Controle de Câncer; 1997. Tumores da mama; p.128-36.

Sommer S, Fuqua SA. Estrogen receptor and breast cancer. **Semin Cancer Biol** 2001; 11:339-52.

Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. **Proc Natl Acad Sci U S A** 2001; 98:10869-74.

St Croix B, Rago C, Velculescu V, et al. Genes expressed in human tumor endothelium. **Science** 2000; 289:1197-202.

Stefaner I, Stefanescu A, Hunziker W, Fuchs R. Expression of placental alkaline phosphatase does not correlate with IgG binding, internalization and transcytosis. **Biochem J** 1997; 327:585-92.

Suzuki E, Evans T, Lowry J, et al. The human GATA-6 gene: structure, chromosomal location, and regulation of expression by tissue-specific and mitogen-responsive signals. **Genomics** 1996; 38:283-90.

Swana HS, Grossman D, Anthony JN, Weiss RM, Altieri DC. Tumor content of the antiapoptosis molecule survivin and recurrence of bladder cancer. **N Engl J Med** 1999; 341:452-3.

Tanaka K, Iwamoto S, Gon G, Nohara T, Iwamoto M, Tanigawa N. Expression of survivin and its relationship to loss of apoptosis in breast carcinomas. **Clin Cancer Res** 2000; 6:127-34.

Takikawa Y, Noguchi M, Kitagawa H, Thomas M. Immunohistochemical detection of p53 and c-erbB-2 proteins: prognostic significance in operable breast cancer. **Breast Cancer** 1994; 30:17-23.

Todd JH, Dowle C, Williams MR, et al. Confirmation of a prognostic index in primary breast cancer. **Br J Cancer** 1987; 56:489-92.

Varelias A, Koblar SA, Cowled PA, Carter CD, Clayer M. Human osteosarcoma expresses specific ephrin profiles: implications for tumorigenicity and prognosis. **Cancer** 2002; 95:862-9.

van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. **Nature** 2002; 415:530-6.

Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. **Science** 2001; 291:1304-51.

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

Vogelstein B, Kinzler KW. Digital PCR. **Proc Natl Acad Sci USA** 1999; 96:9236-41.

Vogt T, Stolz W, Welsh J, et al. Overexpression of Lerk-5/Eplg5 messenger RNA: a novel marker for increased tumorigenicity and metastatic potential in human malignant melanomas. **Clin Cancer Res** 1998; 4:791-7.

Walker RA, Jones JL, Chappell S, Walsh T, Shaw JA. Molecular pathology of breast cancer and its application to clinical management. **Cancer Metastasis Rev** 1997; 16:5-27.

Weiss MJ, Orkin SH. GATA transcription factors: key regulators of hematopoiesis. **Exp Hematol** 1995; 23:99-107.

Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. **Nature** 1995; 378:789-92.

Zhu XD, Lin GJ, Qian LP, Chen ZQ. Expression of survivin in human gastric carcinoma and gastric carcinoma model of rats. **World J Gastroenterol** 2003; 9:1435-8.

ANEXOS

ANEXO 1 - Classificação clínica (TNM) para o câncer de mama baseada na extensão do tumor primário, na ausência ou presença e a extensão de metástases em linfonodos regionais e a ausência ou presença de metástase à distância.

Extensão do Tumor Primário		Linfonodos Regionais		Metástase à Distância	
T	O tumor primário não pode ser avaliado	NX	Os linfonodos regionais não podem ser avaliados	MX	A presença de metástase à distância não pode ser avaliada
T0	Não há evidência de tumor primário	N0	Ausência de metástase em linfonodos regionais	M0	Ausência de metástase à distância
Tis	Carcinoma <i>in situ</i> : carcinoma intraductal, ou carcinoma lobular <i>in situ</i> , ou doença de Paget do mamilo sem tumor	N1	Metástase em linfonodo(s) axilar(es) homolateral(ais), móvel (eis)	M1	Metástase à distância
T1	Tumor de 2,0cm ou menos em sua maior dimensão	N2	Metástase em linfonodo(s) axilar(es) homolateral(ais) fixo(s) uns aos outros ou a outras estruturas		
T1mic	Microinvasão de 0,1cm ou menos em sua maior dimensão	N3	Metástase em Linfonodo(s) mamário(s) interno(s) homolateral(ais)		
T1a	Mais de 0,1cm, porém não mais de 0,5 cm em sua maior dimensão				
T1b	Mais de 0,5cm, porém não mais de 1,0cm em sua maior dimensão				
T1c	Mais de 1,0cm, porém não mais de 2,0cm em sua maior dimensão				
T2	Tumor de mais de 2,0cm, porém não mais de 5,0cm em sua maior dimensão				
T3	Tumor de mais de 5,0cm em sua maior dimensão				
T4	Tumor de qualquer dimensão, com extensão direta à parede torácica ou à pele				
T4a	Extensão à parede torácica				
T4b	Edema, ou ulceração da pele mamária, ou nódulos cutâneos satélite confinados à mesma mama				
T4c	Ambos T4a e T4b				
T4d	Carcinoma inflamatório				

ANEXO 2 - Classificação patológica (pTNM) para o câncer de mama baseada na extensão do tumor primário, na ausência ou presença e a extensão de metástases em linfonodos regionais e a ausência ou presença de metástase à distância.

Extensão do Tumor Primário (pT)		Linfonodos Regionais (pN)		Metástase à Distância (pM)
pT	As categorias pT correspondem às categorias T	pNX	Os linfonodos regionais não podem ser avaliados	As categorias pM correspondem às categorias M
T0	Não há evidência de tumor primário	pN0	Ausência de metástase em linfonodo(s) axilar(es) homolateral(ais), móvel(eis).	
Tis	Carcinoma <i>in situ</i> : carcinoma intraductal ou carcinoma lobular <i>in situ</i> ou doença de Paget do mamilo sem tumor	pN1a	Somente micrometástase (nenhum linfonodo maior que 0,2cm)	
T1	Tumor de 2,0cm ou menos em sua maior dimensão	pNbi	Metástase em 1 a 3 linfonodos, com um maior que 0,2cm e todos os outros menores do que 2,0cm em sua maior dimensão	
T1mic	Microinvasão de 0,1cm ou menos em sua maior dimensão	pNbii	Metástase em 4 ou mais linfonodos, com um maior que 0,2cm e todos os outros menores do que 2,0cm em sua maior dimensão.	
T1a	Mais de 0,1cm, porém não mais de 0,5 cm em sua maior dimensão	pNbiii	Extensão do tumor além da cápsula de um linfonodo metastático, com menos de 2,0cm em sua maior dimensão	
T1b	Mais de 0,5cm, porém não mais de 1,0cm em sua maior dimensão	pN1binv	Metástase em um linfonodo com 2,0cm ou mais, em sua maior dimensão	
T1c	Mais de 1,0cm, porém não mais de 2,0cm em sua maior dimensão	pN2	Metástase em linfonodos axilares homolaterais, fixos uns aos outros ou a outras estruturas	
T2	Tumor de mais de 2,0cm, porém não mais de 5,0cm em sua maior dimensão	pN3	Metástase em linfonodo(s) mamário(s) interno(s) homolateral(ais)	
T3	Tumor de mais de 5,0cm em sua maior dimensão			
T4	Tumor de qualquer dimensão, com extensão direta à parede torácica ou à pele			
T4a	Extensão à parede torácica			
T4b	Edema ou ulceração da pele mamária ou nódulos cutâneos satélite confinados à mesma mama			
T4c	Ambos T4a e T4b			
T4d	Carcinoma inflamatório			

ANEXO 3 - Critérios para a graduação histológica dos carcinomas de mama, de acordo com o sistema de Elston e Ellis.

Formação tubular	>75%=1	10-75%=2	<10%=3
Pleomorfismo nuclear	discreto=1	moderado=2	elevado=3
Índice mitótico em 10	1 ponto	2 pontos	3 pontos
Campos de grande aumento			
Objetiva 25X(Leitz Ortholux)	0-9	10-19	>20
Objetiva 40X (Nikon Labophot)	0-5	6-10	>11
Objetiva 40X (Leitz Diaplan)	0-11	12-2	>23
Grau I: 3-5 pontos			
Grau II: 6-7 pontos			
Grau III: 8-9 pontos			

ANEXO 4 - Tipos Histopatológicos de carcinomas mamários.

Carcinomas sem outras especificações

Ductal

- Intraductal (*in situ*)
- Invasivo com predominância do componente intraductal
- Invasivo
- Comedocarcinoma
- Inflamatório
- Medular
- Mucinoso ou colóide
- Papilar
- Esquitroso
- Tubular
- Outros

Lobular

- In situ*
- Invasivo com componente *in situ* predominante
- Invasivo

Mamilar

- Doença de Paget
- Doença de Paget com carcinoma intraductal
- Doença de Paget com carcinoma ductal invasivo

Carcinoma Indiferenciado

Outros

São Paulo, 31 de Julho de 2002

À

Dra. Márcia Dellamano

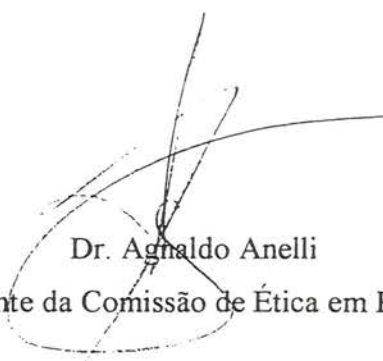
Ref. Comunicação da Comissão de Ética em Pesquisa

Projeto: “Identificação de marcadores moleculares para o câncer de mama por meio da análise da expressão gênica”. (CEP – n.º 434/02)

Prezada Dra. Márcia,

Seu projeto de pesquisa, acima mencionado, foi apreciado pela Comissão de Ética em Pesquisa (CEP) do Hospital do Câncer em sua última reunião de 30/07/2002. Os membros desta comissão **aprovaram** a realização deste estudo. Informações a respeito do andamento do referido projeto deverão ser encaminhados à secretaria do CEP dentro de 12 meses.

Atenciosamente,



Dr. Agnaldo Anelli
Presidente da Comissão de Ética em Pesquisa

C/c: Dra. Otávia Luisa Silva Damas de Caballero

Dr. André Luiz Vettore

Dra. Simone Treiger Sredni

TERMO DE INFORMAÇÃO E CONSENTIMENTO

Para obter um maior conhecimento clínico e científico do câncer, o Corpo Clínico deste Hospital (médicos e pesquisadores) desenvolve pesquisa clínica científica. Através dessa pesquisa é possível conhecer melhor os mecanismos da doença e, portanto, oferecer novas possibilidades de diagnóstico e tratamento. Ainda mais, este trabalho envolve a busca, em tumores, de novos genes ou de lesões em genes já existentes. Estes estudos são realizados em fragmentos de tumores removidos em cirurgias, ou em material biológico colhido.

Voce está sendo admitido (a) neste Hospital para estabelecimento de diagnóstico e/ou tratamento de alguma forma de tumor. Para fins de diagnóstico, fator prognóstico e /ou como parte de seu tratamento, há a necessidade da remoção do tumor e/ou material biológico relacionado à enfermidade. É prática rotineira deste Hospital usar parte do tumor e/ou outro material biológico retirado, para exames clínicos laboratoriais, necessários para um diagnóstico definitivo. O restante do tumor que é retirado , ou do material biológico, não é utilizado, sendo então congelado e armazenado para novos exames se necessários. Caso contrário, são descartados, conforme Legislação Sanitária regulamentar sobre o assunto.

A obtenção e o estudo dos referidos fragmento de tumor e material biológico não implicarão em riscos adicionais no seu tratamento ou na sua cirurgia, nem tampouco, em aumento no tempo da operação ou na extensão da mesma. O projeto de pesquisa proposto que vier utilizar este material será previamente apresentado à apreciação da Comissão de Ética em Pesquisa do Hospital.

O fragmento de tumor e /ou material biológico será identificado no laboratório por um código formado por números e letras e, portanto, sua privacidade e identidade serão sempre preservadas.A eventual inclusão dos resultados em publicação científica será feita de modo a manter o anonimato do paciente.

Concordando com o uso do material para os fins acima descritos, é necessário esclarece-lo(a) que não existem quaisquer benefícios ou direitos financeiros a receber sobre os eventuais resultados decorrentes da pesquisa. Se voce não concordar em doar o material para pesquisa, sua decisão não influenciará, de nenhum modo, no seu tratamento.

Caso voce ainda tenha questões a fazer sobre este Termo de Consentimento ou alguma dúvida que não tenha sido esclarecida pelo seu médico, por gentileza, entre em contato com a Comissão de Ética Médica, pelo telefone 3272-5085.

Voce receberá uma cópia deste documento e o original será arquivado em seu prontuário.Somente assine este Termo, se consentir.

D E C L A R A Ç Ã O

Declaro estar ciente das informações ora prestadas, tendo lido atentamente e concordado com todo o teor.

São Paulo,.....de.....de

.....
Responsável ou Paciente

Nome:.....

RG:.....

RGH:.....



ANEXO 7 - Genes selecionados por meio da análise das bibliotecas de tecido de mama tumoral e não tumoral do banco de dados de *SAGE*.

Razão da frequência ¹	Identificação do gene ²	Nome do gene
91,63690202	M11730	<i>v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)</i>
33,27504301	NM_000251	<i>mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)</i>
24,74196355	BC021299	<i>prostate epithelium-specific Ets transcription factor</i>
24,19214213	BC033143	<i>Homo sapiens, hypothetical gene LOC126695, clone MGC:45763 IMAGE:3960220, mRNA, complete cds</i>
23,36332807	BC001618	<i>phosphoserine aminotransferase</i>
23,36332807	NM_003038	<i>solute carrier family 1 (glutamate/neutral amino acid transporter), member 4</i>
19,82342988	AJ006352	<i>ephrin-A4</i>
19,06047562	NM_003748	<i>aldehyde dehydrogenase 4 family, member A1</i>
17,69949096	NM_022750	<i>hypothetical protein FLJ22693</i>
17,59428519	AK026152	<i>cathepsin H</i>
17,59428519	NM_004308	<i>Rho GTPase activating protein 1</i>
16,28353169	M97252	<i>Kallmann syndrome 1 sequence</i>
16,12809476	AK023594	<i>hypothetical protein FLJ21080</i>
16,12809476	NM_004663	<i>RAB11A, member RAS oncogene family</i>
15,39499954	AK000600	<i>sialyltransferase 7D ((alpha-N-acetylneuraminyl-2,3-beta-galactosyl-1,3)-N-acetyl galactosaminide alpha-2,6-sialyltransferase)</i>
15,39499954	NM_003104	<i>sorbitol dehydrogenase</i>
14,84517813	NM_053056	<i>cyclin D1 (PRAD1: parathyroid adenomatosis 1)</i>
13,45161313	AY007096	<i>Homo sapiens clone TCCCIA00164 mRNA sequence</i>
13,19571389	U38847	<i>TAR (HIV) RNA binding protein 1</i>
12,64589248	BC006793	<i>GATA binding protein 3</i>
12,46261868	NM_005496	<i>SMC4 structural maintenance of chromosomes 4-like 1 (yeast)</i>
12,46261868	BC031406	<i>transducer of ERBB2, 1</i>
11,72952346	NM_007040	<i>E1B-55kDa-associated protein 5</i>
11,72952346	AF246239	<i>GL004 protein</i>
11,72952346	AK024216	<i>hypothetical protein FLJ14154</i>

11,72952346	AF459094	<i>testis intracellular mediator protein</i>
11,27133895	AK026581	<i>solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 1</i>
10,99642824	AK024268	<i>Homo sapiens cDNA FLJ14206 fis, clone NT2RP3003157</i>
10,62988063	AJ488506	<i>Homo sapiens mRNA for protein phosphatase 3 catalytic subunit beta3 (PPP3CB gene)</i>
10,62988063	AL136631	<i>hypothetical protein FLJ12171</i>
10,62988063	BC006983	<i>NME7</i>
10,62988063	BC028049	<i>protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform (calcineurin A beta)</i>
10,61969458	J05500	<i>spectrin, beta, erythrocytic (includes spherocytosis, clinical type I)</i>
10,26333303	AK057983	<i>Homo sapiens, clone IMAGE:4430514, mRNA, partial cds</i>
10,26333303	BF033739	<i>hypothetical transmembrane protein SBB154</i>
10,26333303	AY013288	<i>RNA helicase family</i>
9,911714939	AL713764	<i>tissue inhibitor of metalloproteinase 2</i>
9,911714939	AK027024	<i>transcription elongation factor A (SII), 3</i>
9,896785418	Y17979	<i>fucosyltransferase 8 (alpha (1,6) fucosyltransferase)</i>
9,896785418	AK025209	<i>Homo sapiens cDNA: FLJ21556 fis, clone COL06353</i>
9,53023781	AK001135	<i>Sec23-interacting protein p125</i>
9,285872738	AB011117	<i>KIAA0545 protein</i>
9,203735301	AF355402	<i>BTB (POZ) domain containing 1</i>
9,163690202	AK001950	<i>hypothetical protein FLJ11088</i>
8,870452116	X80199	<i>MLN51 protein</i>
8,797142594	AK054598	<i>Homo sapiens cDNA FLJ30036 fis, clone 3NB692001496</i>
8,797142594	NM_032376	<i>hypothetical protein MGC4251</i>
8,797142594	NM_006330	<i>lysophospholipase I</i>
8,797142594	BC030556	<i>protein inhibitor of activated STAT3</i>
8,797142594	AF201385	<i>thioredoxin reductase 2</i>
8,674960058	NM_007293	<i>complement component 4A</i>
8,674960058	NM_000592	<i>complement component 4B</i>
8,30841245	AK022030	<i>Homo sapiens cDNA FLJ11968 fis, clone HEMBB1001133</i>
8,30841245	BC020965	<i>integral membrane protein 1</i>

8,064047378	AF072928	<i>myotubularin related protein 6</i>
7,787776024	AJ131891	<i>polymerase (DNA directed), mu</i>
7,787776024	NM_002840	<i>protein tyrosine phosphatase, receptor type, F</i>
7,69749977	AK023329	<i>Homo sapiens cDNA FLJ13267 fis, clone OVARC1000964</i>
7,645135826	NM_032999	<i>general transcription factor II, i</i>
7,645135826	AF036613	<i>general transcription factor II, i, pseudogene 1</i>
7,645135826	AK057750	<i>Homo sapiens cDNA FLJ25021 fis, clone CBL01740</i>
7,575317234	NM_014335	<i>CREBBP/EP300 inhibitory protein 1</i>
7,575317234	AK056917	<i>split hand/foot malformation (ectrodactyly) type 3</i>
7,330952162	BC011682	<i>cathepsin F</i>
7,330952162	AK027154	<i>hypothetical protein FLJ10111</i>
7,330952162	AB033073	<i>similar to glucosamine-6-sulfatases</i>
7,147678358	AK057157	<i>RAB6A, member RAS oncogene family</i>
7,147678358	AL136727	<i>RAB6C, member RAS oncogene family</i>
7,08658709	NM_006767	<i>leucine-zipper-like transcriptional regulator, 1</i>
7,08658709	BC007301	<i>phosphatidylinositol glycan, class S</i>
6,964404554	AK074092	<i>melanoma antigen, family D, 1</i>
6,964404554	AK022725	<i>protein phosphatase methylesterase-1</i>
6,597856946	BC007493	<i>galactosidase, beta 1</i>
6,597856946	AL080126	<i>KIAA0683 gene product</i>
6,597856946	AK054823	<i>O-sialoglycoprotein endopeptidase</i>
6,464566906	BC000426	<i>H1 histone family, member X</i>
6,451237902	BC013032	<i>similar to HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X</i>
6,414583142	AK025526	<i>hypothetical protein FLJ20275</i>
6,371816747	BC013590	<i>eukaryotic translation initiation factor 2B, subunit 5 (epsilon, 82kD)</i>
6,353491874	NM_001677	<i>ATPase, Na⁺/K⁺ transporting, beta 1 polypeptide</i>
6,336037226	AK056727	<i>HCGIV-6 protein</i>
6,336037226	AF036144	<i>meningioma expressed antigen 5 (hyaluronidase)</i>
6,323871865	BC034378	<i>Homo sapiens cDNA: FLJ21547 fis, clone COL06206</i>

6,231309338	AK024466	<i>cation-chloride cotransporter-interacting protein</i>
6,231309338	AK025398	<i>HLA-B associated transcript 5</i>
6,109126802	BC000504	<i>succinate-CoA ligase, GDP-forming, alpha subunit</i>
6,048035533	BM920374	<i>hypothetical protein IMAGE3455200</i>
5,864761729	NM_013286	<i>chromosome 3p21.1 gene sequence</i>
5,864761729	NM_002028	<i>farnesyltransferase, CAAx box, beta</i>
5,864761729	BC003505	<i>high-mobility group 20B</i>
5,864761729	BC006428	<i>hypothetical protein HSPC195</i>
5,864761729	BC000854	<i>hypothetical protein MGC5576</i>
5,864761729	BC007855	<i>KIAA0141 gene product</i>
5,864761729	NM_003634	<i>nipsnap homolog 1 (C. elegans)</i>
5,864761729	U83993	<i>purinergic receptor P2X, ligand-gated ion channel, 4</i>
5,864761729	NM_004592	<i>splicing factor, arginine/serine-rich 8 (suppressor-of-white-apricot homolog, Drosophila)</i>
5,742579193	AL117420	<i>CGI-86 protein</i>
5,663837108	NM_016024	<i>CGI-79 protein</i>
5,498214121	AK026642	<i>chromosome 20 open reading frame 35</i>
5,498214121	AB020694	<i>KIAA0887 protein</i>
5,470018152	NM_004879	<i>etoposide-induced mRNA</i>
5,457486609	AK024892	<i>SH3 domain binding glutamic acid-rich protein like</i>
5,376031585	NM_005766	<i>FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)</i>
5,376031585	NM_022575	<i>vacuolar protein sorting 16 (yeast)</i>
5,351595078	BC016329	<i>Homo sapiens, clone IMAGE:4103364, mRNA</i>
5,314940317	AL442086	<i>chromosome 20 open reading frame 81</i>
5,204976035	L29376	<i>Homo sapiens (clone 3.8-1) MHC class I mRNA fragment</i>
5,131666513	AF234887	<i>cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)</i>
5,131666513	AK023964	<i>hypothetical protein FLJ13902</i>
5,009483977	NM_002826	<i>quiescin Q6</i>
4,98504747	NM_018320	<i>hypothetical protein FLJ11099</i>
4,948392709	BC031301	<i>KIAA1185 protein</i>

4,887301441	BM906096	<i>hypothetical protein BC000919</i>
4,887301441	NM_032366	<i>hypothetical protein MGC13114</i>
4,887301441	AB037745	<i>KIAA1324 protein</i>
4,887301441	AF070523	<i>vitamin A responsive; cytoskeleton related</i>
4,838428427	AB014536	<i>copine III</i>
4,829803777	AK055515	<i>Homo sapiens cDNA: FLJ21260 fis, clone COL01441</i>
4,805846417	AK056322	<i>hypothetical protein MGC3180</i>
4,805846417	AF226732	<i>NPD007 protein</i>
4,765118905	AF372839	<i>A20-binding inhibitor of NF-kappaB activation-2</i>
4,765118905	AL137262	<i>Homo sapiens mRNA; cDNA DKFZp434J1313 (from clone DKFZp434J1313)</i>
4,765118905	NM_015701	<i>hypothetical protein CL25084</i>
4,765118905	AF218006	<i>hypothetical protein FLJ13612</i>
4,765118905	BC004352	<i>kinesin-like 4</i>
4,765118905	AL110209	<i>lysophospholipase 3 (lysosomal phospholipase A2)</i>
4,765118905	AF135168	<i>N-ethylmaleimide-sensitive factor</i>
4,765118905	AF247703	<i>zinc finger, DHHC domain containing 3</i>
4,642936369	NM_014634	<i>partner of PIX 2</i>
4,642936369	AK023249	<i>phafin 2</i>
4,581845101	AF246238	<i>hypothetical protein FLJ20886</i>
4,54519034	AK056156	<i>Homo sapiens cDNA FLJ13553 fis, clone PLACE1007454</i>
4,490208199	NM_033084	<i>Fanconi anemia, complementation group D2</i>
4,490208199	AK021913	<i>Homo sapiens cDNA FLJ11851 fis, clone HEMBA1006744</i>
4,490208199	AK025297	<i>Homo sapiens cDNA: FLJ21644 fis, clone COL08411</i>
4,490208199	AF217749	<i>protocadherin beta 9</i>
4,490208199	AF151810	<i>serologically defined colon cancer antigen 28</i>
4,437155256	NM_006184	<i>nucleobindin 1</i>
4,398571297	NM_005675	<i>DiGeorge syndrome critical region gene 6</i>
4,398571297	BC020981	<i>general transcription factor IIC, polypeptide 2 (beta subunit, 110kD)</i>
4,398571297	BC011365	<i>glyoxalase I</i>

4,398571297	AK055070	<i>Homo sapiens cDNA FLJ20167 fis, clone COL09512</i>
4,398571297	AF400044	<i>Homo sapiens clone IMAGE:1628386 ST7OT3 mRNA, non-coding transcript</i>
4,398571297	AK056357	<i>hypothetical protein FLJ31795</i>
4,398571297	NM_018728	<i>myosin 5C</i>
4,398571297	BM909365	<i>polyglutamine binding protein 1</i>
4,398571297	NM_003365	<i>ubiquinol-cytochrome c reductase core protein 1</i>
4,312324801	NM_002168	<i>isocitrate dehydrogenase 2 (NADP+), mitochondrial</i>
4,306934395	BC002917	<i>hypothetical protein FLJ11856</i>
4,306934395	BC008426	<i>palmitoyl-protein thioesterase 1 (ceroid-lipofuscinosis, neuronal 1, infantile)</i>
4,293843409	NM_016030	<i>CGI-87 protein</i>
4,276388761	NM_005506	<i>CD36 antigen (collagen type 1 receptor, thrombospondin receptor)-like 2 (lysosomal integral membrane protein II)</i>
4,276388761	NM_021823	<i>hypothetical protein MDS018</i>
4,251952254	BC001381	<i>hypothetical protein FLJ22940</i>
4,251952254	BC012814	<i>zinc finger protein-like 1</i>
4,215297493	BC006540	<i>c6.1A</i>
4,215297493	NM_004341	<i>carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase</i>
4,154206225	NM_015339	<i>activity-dependent neuroprotector</i>
4,154206225	AK055117	<i>Homo sapiens cDNA FLJ30555 fis, clone BRAWH12003818</i>
4,154206225	AB018335	<i>KIAA0792 gene product</i>
4,154206225	NM_002979	<i>sterol carrier protein 2</i>
4,086896085	AK055071	<i>Homo sapiens cDNA FLJ30509 fis, clone BRAWH12000595</i>
4,086896085	NM_005080	<i>X-box binding protein 1</i>
4,072751201	NM_002887	<i>arginyl-tRNA synthetase</i>
4,032023689	BC014210	<i>arylsulfatase A</i>
4,032023689	NM_012106	<i>binder of Arl Two</i>
4,032023689	BC013149	<i>Homo sapiens, clone IMAGE:3917549, mRNA, partial cds</i>
4,032023689	BC008573	<i>Homo sapiens, clone MGC:17005 IMAGE:4182067, mRNA, complete cds</i>
4,032023689	BC031610	<i>Homo sapiens, membrane-spanning 4-domains, subfamily A, member 12 4-domains, subfamily A, member 7, clone MGC:35295 IMAGE:5176452, mRNA, complete cds</i>

4,032023689	NM_013332	<i>hypoxia-inducible protein 2</i>
4,032023689	BC032528	<i>leukotriene A4 hydrolase</i>
4,032023689	NM_007121	<i>nuclear receptor subfamily 1, group II, member 2</i>
4,032023689	AL713766	<i>RAP2B, member of RAS oncogene family</i>

¹ razão das frequências das *SAGTags* comparadas entre as bibliotecas de tecidos de mama tumoral e normal

² número de identificação no *GenBank*

ANEXO 8 - Medida da expressão diferencial em relação a uma mistura de RNAs de tecido de mama não tumoral, normalizadas pela expressão do gene da β -actina.

Amostras	<i>APPL2</i>	<i>BIRC5</i>	<i>CAD</i>	<i>CELSR2</i>	<i>CTSF</i>	<i>DGCR6</i>	<i>EFNA4</i>	<i>EI24</i>	<i>FN3KR3</i>	<i>GATA3</i>	<i>GATA4</i>	<i>HMG20B</i>	<i>PTPRF</i>	<i>QSCN6</i>	<i>SCARB2</i>	<i>SATARD10</i>	<i>ZC3HDC</i>	<i>KIAA0141</i>	<i>KIF22</i>
1291 ¹	12,25	0,32	0,22	2,48	0,68	0,12	0,69	0,43	0,62	2,31	0,01	0,29	1,25	0,98	1,19	0,04	0,82	1,59	0,17
1726	29,55	0,70	0,80	20,97	1,08	0,16	2,69	0,42	0,27	1,74	0,03	2,81	0,49	0,92	1,14	0,55	1,91	1,76	0,08
1961	2,01	0,37	0,53	3,71	2,14	0,55	1,34	0,92	0,76	4,94	ND ²	2,86	0,99	1,15	2,80	0,93	0,96	1,91	0,16
2310	0,40	17,63	5,46	30,80	3,39	1,46	1,03	2,21	3,00	19,77	146,02	3,82	2,51	8,85	2,53	11,75	4,13	4,29	5,48
2412	7,19	0,34	0,71	1,11	1,32	0,25	0,80	1,20	1,09	1,13	0,29	1,80	0,85	0,76	1,32	0,77	0,98	1,24	0,56
2424	4,61	0,26	0,38	10,09	0,66	0,37	0,50	0,49	1,38	10,97	2,26	1,48	1,45	2,31	0,45	0,35	0,55	0,82	0,09
2439	0,48	0,76	1,59	5,94	0,90	1,04	5,03	1,84	2,80	11,39	ND	18,51	4,32	5,35	6,94	3,99	2,98	1,77	0,34
2610	2,31	0,72	1,48	0,88	0,66	1,46	1,26	0,99	1,36	0,21	0,02	4,14	2,56	1,72	1,64	1,35	0,90	0,79	0,28
2658	0,86	1,98	0,50	9,38	0,82	0,51	1,89	0,95	0,98	3,06	0,14	2,15	0,69	1,18	3,32	0,54	1,33	1,03	0,28
2704	1,27	0,29	3,28	48,84	0,57	0,84	2,58	0,81	3,31	29,45	ND	8,51	6,30	7,46	2,67	2,19	1,61	3,07	1,67
2720	17,09	9,92	0,70	14,62	2,48	0,10	1,37	0,33	0,77	6,00	3,48	0,85	0,93	0,51	0,43	0,66	1,14	1,51	0,36
2734	1,89	2,79	1,13	6,11	0,13	0,83	1,39	0,76	3,58	2,06	0,12	8,20	4,27	1,46	2,78	2,46	3,56	0,59	0,84
2789	1,78	0,75	0,66	47,84	1,86	0,08	0,16	0,91	0,68	14,22	12,55	5,64	3,26	0,70	1,27	4,61	3,22	1,72	1,00
2933	15,08	1,24	0,21	1,68	0,48	0,05	1,49	0,27	0,27	3,58	0,02	1,37	0,50	0,42	0,86	0,03	1,74	0,70	0,25
2937	0,67	0,81	2,73	3,62	0,35	0,63	1,20	1,19	1,33	22,78	5,72	2,27	1,62	1,48	1,35	0,91	2,17	0,77	0,28
3143	15,19	7,01	0,31	4,10	2,14	0,54	1,16	0,84	0,71	4,44	1,60	1,02	0,66	1,03	0,65	0,56	3,11	1,38	0,13
3177	1,05	0,33	0,84	6,77	2,20	1,09	2,71	1,33	2,19	18,83	1,06	4,10	2,52	6,77	12,95	ND	3,16	2,87	0,30
3364	9,32	0,72	3,05	2,65	1,85	0,26	0,77	0,64	0,52	1,09	0,02	0,35	1,36	1,12	0,72	ND	1,09	1,12	0,27
3448	3,11	0,46	0,78	2,68	0,63	0,54	2,53	1,96	2,45	9,35	0,18	10,93	1,82	7,73	3,96	1,01	2,11	2,06	0,26
4110	3,31	1,91	0,98	2,35	0,58	0,05	1,27	0,26	0,39	7,59	2,42	1,27	0,70	0,51	0,37	0,19	0,50	0,46	0,34
4111	2,66	6,32	0,80	0,84	0,13	0,10	0,72	0,40	0,46	0,29	ND	0,29	0,42	0,71	0,50	0,02	2,96	0,58	0,04
4379	1,10	0,98	0,70	10,85	1,11	0,11	0,46	0,33	0,72	7,01	10,02	0,45	1,21	1,29	1,11	0,32	1,64	2,85	0,18

¹Código das amostras do Banco de tumores do Hospital do Câncer AC Camargo/ Instituto Ludwig de Pesquisa sobre o Câncer

²ND-não detectado.

