



*Caracterização de genes
diferencialmente expressos em
linhagens celulares de mama
que expressam diferentes níveis
de c-erbB2*

ANA PAULA MEDEIROS SILVA

**Dissertação de Doutorado apresentada à Fundação
Antônio Prudente para obtenção do Grau de
Doutor em Ciências**

Área de concentração: Oncologia

Orientadora: Dra. Anamaria Aranha Camargo

Co-orientadora: Dra. Dirce Maria Carraro



São Paulo

2006

FICHA CATALOGRÁFICA

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

Silva, Ana Paula Medeiros

**Caracterização de genes diferencialmente expressos em linhagens
celulares de mama que expressam diferentes níveis de c-erbB2** / Ana Paula
Medeiros Silva - São Paulo, 2006.

163p.

Tese (doutorado) – Fundação Antônio Prudente.

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

Orientadora: Anamaria Aranha Camargo.

Descritores: 1. CÂNCER DA MAMA 2. CROMOSSOMO 17. 3. PERFIL DA
EXPRESSÃO GÊNICA. 4. GENE c-erbB2. 5. ONCOGENES.

*“Mais importante do que
acreditar no que fazemos é fazer
aquilo que acreditamos com
amor, paciência e
dedicação”.*
(Deise Maria)

**Aos meus pais,
com todo meu amor e admiração**

AGRADECIMENTOS

À **Dra. Anamaria A. Camargo**, por me receber com tanta disponibilidade e carinho. É quase impossível, apenas com palavras, agradecer toda amizade, apoio e incentivo que tenho recebido durante esses anos. Mais que orientadora, para mim, você é um exemplo de pessoa e de profissional. Muito obrigada pela oportunidade, pela confiança, pela paciência, pela preocupação, por todos os ensinamentos enfim, obrigada pela chance de poder trabalhar e conviver com a pessoa maravilhosa que você é.

Ao **Ricardo Moura**, pelo convívio diário, pelo incentivo e apoio constantes, pelas dicas de trabalho e também pela amizade que é sempre o mais importante. Você é um amigo especial.

À **Valéria Paixão**, Valerie, pelo importante e indispensável apoio pessoal e técnico sempre com muita disponibilidade, competência, amizade, carinho e alto astral. Obrigada também pelo incentivo e por todos os conselhos sempre muito sinceros. Você é uma grande amiga.

Ao **Érico**, pelos ensinamentos, por sempre se preocupar em oferecer ajuda, pela paciência, pelo apoio e pela convivência.

Ao **Fabrcício**, que mesmo não estando mais presente em nosso laboratório, se faz presente não só por e-mails mas também pelas recordações que nos deixou.

À **Jane Kaiano** e **Daniela Ierardi** pelos momentos de descontração dentro e fora do laboratório.

À **Dirce** pelas importantes sugestões, conselhos e pela disponibilidade em ler este trabalho.

À toda **turma do laboratório**, Maria Cristina, Murilo, Anna Christina, Daniel, Ricardo, Érico, Fabrício, Daniela, Lílian, Fabiana, Raphael, Newton, Felícia, Elisa, Mônica, Lilianzinha, Maria Isabel Valéria e Fernando pela convivência e pelos bons momentos vividos dentro e fora do laboratório. Sem vocês, cada um a seu modo, certamente tudo teria sido mais difícil.

Ao **Dr. San Ming Wang e Dr. Jianjun Chen**, da Universidade de Chicago, por me receberem no laboratório e colaborarem com a padronização do GLGI-MPSS.

Ao **Dr. Sandro de Souza** e toda sua equipe do Laboratório de Bioinformática pela disponibilidade, sugestões e auxílio com o Banco de Dados do MPSS.

Ao **Jorge Estefano e Pedro Valente** pela boa vontade e excelente ajuda com os bancos de dados.

À **Dra. Luiza Lina Vila** e sua equipe, em especial a **Dra. Andréa Trevisan**, pela disponibilidade e boa vontade em ensinar como utilizar e também por dividir a utilização do Real Time durante o período em que não tínhamos o nosso próprio equipamento.

À **Prof. Dra. Mari Sogayar** e toda sua equipe, em especial a **Rita Figueira**, pela amizade, alto astral e boa vontade com que sempre me receberam e se dispuseram a ajudar. Obrigada também pelo importante auxílio com as amostras de RNA dos pacientes.

Ao **Dr. Fernando Lopes Alberto e Dra. Eloísa de Sá Moreira**, do Laboratório Fleury, pela importante ajuda na correta interpretação dos dados gerados pelo Real Time, pela incansável boa vontade com que sempre nos receberam e esclareceram nossas dúvidas e também por possibilitarem a utilização do equipamento Rotor Gene.

Ao **Prof. Dr. Sérgio Verjowski** e sua equipe, em especial ao **Rodrigo Loro**, pela ajuda inicial com o Real Time.

A todos os **funcionários do ILPC** pelo suporte técnico e administrativo e também pela convivência durante este período. Em especial à **Ana Claudia**, pela boa vontade e eficiência em resolver os problemas do dia-a-dia.

Aos demais colegas do ILPC pela convivência durante este período.

À **Suely Francisco** e demais funcionários da biblioteca da Fundação Antônio Prudente, em especial à **Janaína Beltrame dos Santos**, pela boa vontade e auxílio na obtenção e organização do material bibliográfico.

À **Ana Maria Kuninari e Luciana Pitombeira** pelo excelente trabalho desenvolvido junto a secretaria de pós graduação e pelo carinho com que sempre me receberam.

Ao **Dr. Luiz Fernando Lima Reis** pela excelente coordenação do curso de Pós Graduação, pela disponibilidade e também pelos importantes conselhos.

Ao **Dr. Humberto Torloni e Prof. Dr. Fernando Augusto Soares** pela fundação e administração do Banco de Tumores, respectivamente, que foi importante fonte de material para a realização deste trabalho.

Ao Departamento de Anatomia Patológica, em especial ao **Prof. Dr. Fernando Soares e a Dra. Isabela Werneck** pela ajuda com as análises de imunohistoquímica.

À **Prof Dra. Marisis Aranha Plessmann de Camargo**, pela importante ajuda na revisão do inglês (summary).

Ao **Dr. Adalberto Abrão Siufi**, do Hospital do Câncer Alfredo Abrão, pela coleta de parte das amostras de tumores de mama utilizadas neste estudo. Ao **Dr. Ismael Dale Cotrim Guerreiro da Silva** da UNIFESP por disponibilizar as amostras.

Ao **André Leme**, pela ajuda com os problemas de informática.

Ao **Prof. Dr. Ricardo Renzo Brentani** que através da administração do Hospital do Câncer e do Instituto Ludwig, possibilitou o desenvolvimento desse trabalho em um local conceituado e bem estruturado.

Aos meus tios, primos, avós e amigos que sempre torceram por mim. Ao Walter, Silvana, Juliana, Renata, Wilson, Simone, Paula e Gabriela por serem mais que amigos. Obrigada pela torcida e incentivo sempre.

Ao **meu irmão**, que sempre sem dizer muito, sei que se preocupa, me apóia e torce por mim.

Aos **meus pais** pelo amor, amizade, dedicação, paciência, incentivo, confiança e equilíbrio. Vocês são meus grandes exemplos.

À **FAPESP** (Fundação de Amparo à Pesquisa do Estado de São Paulo), pela bolsa de estudo concedida.

RESUMO

Silva APM. **Caracterização de genes diferencialmente expressos em linhagens celulares de mama que expressam diferentes níveis de c-erbB2.** São Paulo, 2006. [Dissertação de Doutorado - Fundação Antônio Prudente].

O oncogene c-erbB2 é um receptor de membrana com atividade de tirosina quinase e pertence a família de receptores de fatores de crescimento epidermal. A super-expressão do oncogene c-erbB2 é observada em 25-30% dos tumores de mama e é também um fator de pior prognóstico. Para estudar os mecanismos moleculares de atuação desse oncogene, Harris et al (1999) desenvolveram um modelo de super-expressão de c-erbB2 em células epiteliais luminais imortalizadas da mama. Duas linhagens celulares, HB4a-C3.6 e HB4a-C5.2, expressando diferentes níveis de c-erbB2, foram derivadas da linhagem imortalizada HB4a.

Com o objetivo de identificar transcritos diferencialmente expressos entre as linhagens celulares HB4a e HB4a-C5.2, foi utilizada a técnica de MPSS (*Massive Parallel Signature Sequencing*). A partir de ambas as linhagens celulares foi gerado um total de 24.521.236 *tags*, correspondendo a 24.065 *tags* distintas e válidas. Depois de estabelecidas as correlações entre *tags* e transcritos humanos conhecidos, 9% das *tags* não foram associadas a transcritos conhecidos (*tags* órfãs). Entre as *tags* órfãs, muitas apresentaram expressão diferencial entre as linhagens HB4a and HB4a-C5.2 e, provavelmente, correspondem a novos transcritos regulados por c-erbB2. Para caracterizar alguns desses novos transcritos, foi desenvolvida, em nosso laboratório, a técnica de GLGI-MPSS (Generation of Longer cDNA fragments for Gene Identification) e 83 *tags* órfãs foram convertidas em fragmentos 3' de cDNA. Para 41 das 83 amplificações de GLGI-MPSS, foi possível obter um fragmento predominante e de forte intensidade quando analisado em gel de agarose.

Esses 41 fragmentos foram clonados e seqüenciados e as seqüências geradas foram analisadas através do programa BLAST, para verificar a presença de similaridade com transcritos humanos depositados em bancos de dados públicos. A análise desses fragmentos permitiu a identificação de 10 novos transcritos

possivelmente regulados por c-erbB2, 3 transcritos polimórficos nos quais a presença de *SNPs* gerou *tags* de MPSS alternativas, 2 formas alternativas de poliadenilação de transcritos já conhecidos e *tags* de MPSS artefatuais. A utilização do GLGI-MPSS nos permitiu ainda identificar 5 transcritos *antisense*, dos quais 4 foram validados através de RT-PCR fita específica.

A expressão diferencial dos 10 transcritos possivelmente regulados por c-erbB2 e dos 3 transcritos correspondentes às *tags* alternativas associadas a *SNPs* foi avaliada através de PCR em Tempo Real nas linhagens HB4a e HB4a-C52. Para 5 dos 13 transcritos analisados, foi possível confirmar a expressão diferencial entre as duas linhagens. Dos 5 transcritos validados, 4 se mostraram mais expressos na linhagem HB4a-C5.2. Em seguida, o padrão de expressão dos 4 transcritos super-expressos na linhagem HB4a-C5.2 foi analisado por PCR em Tempo Real em linhagens tumorais e em amostras de tumores de mama que apresentavam ou não super-expressão do oncogene c-erbB2. A super-expressão dos transcritos correspondentes às *tags* 6 e 28 foi observada na maioria das linhagens tumorais com super-expressão de c-erbB2. Já nos experimentos que avaliaram a expressão desses transcritos em amostras tumorais, foi possível verificar que a média do nível de expressão desses transcritos em relação ao tecido normal de mama era maior no grupo de tumores c-erbB2 positivos e, para as *tags* 7 e 28, a diferença observada entre os dois grupos foi estatisticamente significativa. Esses resultados confirmam a expressão diferencial desses 4 transcritos em tumores c-erbB2 positivos e a caracterização funcional dos mesmos nos permitirá compreender melhor os mecanismos moleculares de atuação do c-erbB2, abrindo novas perspectivas para o acompanhamento e tratamento de pacientes com câncer de mama.

SUMMARY

Silva APM. **Caracterização de genes diferencialmente expressos em linhagens celulares de mama que expressam diferentes níveis de c-erbB2** [Characterization of genes differentially expressed in breast cell lines, expressing different levels of c-erbB2]. São Paulo; 2006. [Dissertação de Doutorado – Fundação Antonio Prudente].

The c-erbB2 oncogene is a membrane receptor with tyrosine kinase activity and belongs to the epidermal growth factor receptor family. C-erbB2 over-expression is observed in 25-30% of breast tumors and is an adverse prognostic factor. To study the molecular mechanisms of c-erbB2 over-expression, Harris et al. (1999) developed a model of c-erbB2 over-expression in conditionally immortalized mammary luminal epithelial cells. Two new cell lines, HB4a-C3.6 and HB4a-C5.2, expressing different levels of c-erbB2 were derived from the immortalized cell line HB4a.

In order to identify transcripts differentially expressed between HB4a and HB4a-C5.2 cell lines, the MPSS (Massively Parallel Signature Sequencing) technique was used. A total of 24.521.236 MPSS tags, representing 24.065 unique reliable tags, was generated from both cell lines. After establishing reliable correlations between tags and known human transcripts, 9% of the tags could not be associated with a known transcripts (orphan tags). Among the orphan tags there were several that showed differential expression between HB4a and HB4a-C5.2 cell lines and probably represent novel transcripts regulated by c-erbB2. To further characterize some of these novel transcripts, the GLGI-MPSS (Generation of Longer cDNA fragments for Gene Identification) technique was developed in our laboratory and 83 orphan tags were converted into 3' cDNA fragments. The amplification of a dominant band was observed for 41 of 83 GLGI-MPSS amplifications, when analyzed in agarose gel.

These 41 fragments were cloned and sequenced and the sequences were analyzed using BLAST to identify similarities with human transcripts submitted to public databases. The analysis of these fragments allowed the identification of 10

novel transcripts, putatively regulated by c-erbB2, 3 polymorphic transcripts in which the presence of a *SNP* generated an MPSS alternative tag, 2 alternative polyadenylation isoforms of known transcripts and artefactual MPSS tags. GLGI-MPSS also allowed us to identify 5 antisense transcripts from which 4 were further validated by strand specific RT-PCR.

Differential expression of the 10 transcripts putatively regulated by c-erbB2, and of the 3 transcripts in which the presence of a *SNP* generated an MPSS alternative tag was evaluated by Real Time PCR in the HB4a and HB4a-C5.2 cell lines. Differential expression between the two cell lines was confirmed for 5 out of the 13 transcripts analyzed. Four out of the 5 validated transcripts were shown to be over-expressed in the HB4a-C5.2 cell line. Next, we examined, by Real Time PCR, the expression pattern of the 4 transcripts over-expressed in the HB4a-C5.2 cell line in breast tumor cell lines and breast tumor samples, over -expressing or not c-erbB2. Over-expression of the transcripts corresponding to tags 6 and 28 was observed in the majority of the breast tumor cell lines over-expressing c-erbB2. On the other hand, in the experiments carried out using tumor samples, we observed that the average expression level of these 4 transcripts, in relation to normal breast tissue, was higher in the c-erbB2 positive group and that the difference observed for tags 7 and 28 was statistically significant. These results confirm the differential expression of these 4 transcripts in c-erbB2 positive tumors and their functional characterization will allow us to better understand the molecular mechanisms behind c-erbB2 over-expression and will eventually open new perspectives in the management and treatment of breast cancer patients.

LISTA DE FIGURAS

Figura 1	Representação esquemática de um corte de mama	5
Figura 2	Representação esquemática da progressão tumoral	6
Figura 3	Representação esquemática do cromossomo 17	13
Figura 4	Estrutura de um receptor erbB ativado	14
Figura 5	Domínio extracelular dos receptores c-erbB2 e c-erbB3	15
Figura 6	Representação esquemática da família de receptores erbB e seus ligantes específicos	19
Figura 7	Vias de sinalização celular e a ativação do receptor c-erbB2	21
Figura 8	Interações entre o Trastuzumab e as células tumorais	28
Figura 9	Modelo de células luminais desenvolvido por Harris et al. (1999)	34
Figura 10	Representação esquemática da técnica de GLGI	83
Figura 11	Resultados obtidos com o protocolo modificado de GLGI	86
Figura 12	Resultados obtidos com o protocolo de GLGI para as <i>tags</i> controles	88
Figura 13	Amplificação GLGI-MPSS	90

Figura 14	Reação de GLGI-MPSS utilizando um maior número de ciclos para a amplificação dos fragmentos	91
Figura 15	Alinhamento entre o transcrito <i>antisense</i> correspondente à <i>tag</i> 12 e a seqüência de mRNA do transcrito <i>sense</i>	96
Figura 16	Validação experimental do transcrito polimórfico correspondente à <i>tag</i> 60	97
Figura 17	Alinhamento entre o transcrito correspondente à <i>tag</i> 60 e a seqüência de mRNA do transcrito polimórfico NASP	98
Figura 18	Alinhamento entre o transcrito correspondente à <i>tag</i> 12 e a seqüência de mRNA do transcrito conhecido	101
Figura 19	Alinhamento entre o transcrito correspondente à <i>tag</i> 65 e a seqüência de mRNA do transcrito conhecido	102
Figura 20	Representação esquemática dos resultados obtidos a partir da amplificação GLGI para as 83 <i>tags</i> órfãs de MPSS	103
Figura 21	Padronização da reação de amplificação por PCR em Tempo Real para o oncogene c-erbB2	106
Figura 22	Representação gráfica das análises de expressão diferencial dos transcritos correspondentes às <i>tags</i> 7, 28, 44 e 6 e do oncogene c-erbB2 em amostras tumorais de mama c-erbB2 negativas e c-erbB2 positivas	121

Figura 23	Geração da seqüência completa do transcrito representado pela <i>tag</i> 44 através da técnica de RACE	129
Figura 24	Seqüência consenso do transcrito correspondente à <i>tag</i> 44	130
Figura 25	Análise dos fragmentos gerados por MPSS-GLGI que correspondem à transcritos <i>antisenses</i>	135
Figura 26	RT-PCR fita específica	137

LISTA DE TABELAS

Tabela 1	Dados de incidência do câncer de mama no Brasil	3
Tabela 2	Meios de cultura utilizados no cultivo celular das linhagens celulares de mama cedidas pelo Dr. Robert Whitehead e pelo Dr. Michael O'Hare	47
Tabela 3	<i>Tags</i> associadas à transcritos conhecidos	60
Tabela 4	Seqüência dos iniciadores específicos correspondentes às <i>tags</i> órfãs	60
Tabela 5	Iniciadores para PCR em Tempo Real	68
Tabela 6	Iniciadores para RACE	75
Tabela 7	Iniciadores <i>Antisense</i>	76
Tabela 8	Dado de expressão diferencial do oncogene c-erbB2 entre as linhagens celulares HB4a e HB4a-C5.2	80
Tabela 9	Categorização das <i>tags</i> .de MPSS	81
Tabela 10	Resultados obtidos através do protocolo GLGI-MPSS para as <i>tags</i> órfãs	93
Tabela 11	Concentração dos iniciadores e concentração de cDNA utilizadas nas reações de PCR em Tempo Real	105

Tabela 12	Eficiência de amplificação dos iniciadores utilizados nas reações de PCR em Tempo Real	107
Tabela 13	Dados de expressão relativa dos novos transcritos determinadas por PCR em Tempo Real e a frequência com que cada tag foi encontrada nas bibliotecas HB4a e HB4a-C5.2 de acordo com o MPSS	108
Tabela 14	Dados de expressão e amplificação do oncogene c-erbB2 em linhagens celulares tumorais de mama	110
Tabela 15	PCR em Tempo Real para avaliar o nível de expressão diferencial do oncogene c-erbB2 e dos novos transcritos em linhagens tumorais de Mama	112
Tabela 16	Dados referentes a expressão diferencial dos novos transcritos e do oncogene c-erbB2 obtidos através de PCR em Tempo Real	119
Tabela 17	Caracterização dos transcritos <i>antisenses</i> quanto a presença de sinal de poliadenilação e cauda de poli A	133

LISTA DE ABREVIATURAS

μg	do inglês <i>microgram</i>
μl	do inglês <i>microliter</i>
μM	do inglês <i>micromolar</i>
AJCC	do inglês <i>The American Joint Committee on Cancer</i>
AR	do inglês <i>amphiregulin</i>
ATP	do inglês <i>adenosine 5'-triphosphate</i>
BCRP	do inglês <i>Breast Cancer Resistance Protein</i>
BLAST	do inglês <i>Basic Local Alignment Search Tool</i>
BRCA 1	do inglês <i>Breast Cancer Gene 1</i>
BRCA 2	do inglês <i>Breast Cancer Gene 2</i>
BTC	do inglês <i>betacellulin</i>
cDNA	do inglês <i>complementary DNA</i>
CDK	do inglês <i>Cyclin-Dependent kinase</i>
c-erbB	do inglês <i>Human Epidermal Growth factor Receptor</i>
CXCR4	do inglês <i>CXC chemokine receptor 4</i>
CpG	do inglês <i>cytosine phosphate guanine</i>
C_T	do inglês <i>Cycle Threshold</i>
DNA	do inglês <i>deoxyribonucleic acid</i>
DNase	do inglês <i>desoxyribonuclease</i>
dNTP	do inglês <i>deoxynucleoside 5' triphosphate</i>
DEPC	do inglês <i>Di-etil pirocarbonato</i>
DTT	do inglês <i>1,4-dithiothreitol</i>
EDTA	do inglês <i>ethylenediamine tetraacetate</i>
EGF	do inglês <i>Epidermal Growth Factor</i>
EGFR	do inglês <i>Epidermal Growth Factor Receptor</i>
EPR	do inglês <i>epiregulin</i>
ERK	do inglês <i>Extracellular Signal Regulated Kinase</i>
ESTs	do inglês <i>Expressed Sequence Tags</i>

FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FISH	do inglês <i>Fluorescence in Situ Hybridization</i>
GAPDH	do inglês <i>glyceraldehyde-3-phosphate dehydrogenase</i>
GLGI	do inglês <i>Generation of Longer cDNA fragments from SAGE tags for Gene Identificatio</i>
Grb2	do inglês <i>Growth Factor Receptor-Bound Protein 2</i>
HB-EGF	do inglês <i>Heparin-Binding Epidermal Growth Factor</i>
HAT	do inglês <i>Histone Acetyltransferase</i>
HCl	do inglês <i>chloridric acid</i>
HDAC	do inglês <i>Histone Deacetylase</i>
HPRT	do inglês <i>Hypoxanthine Phosphoribosyl Transferase</i>
Ig	do inglês <i>Immunoglobulin</i>
IPTG	do inglês <i>isopropyl-β-D-thiogalactoside</i>
IRAK	do inglês <i>IL-1R- associated quinase</i>
Kda	do inglês <i>kilodalton</i>
LPS	do inglês <i>lipopolysaccharide</i>
Lix1	do inglês <i>Limb Expression 1</i>
MAPK	do inglês <i>Mitogen-Activated Protein Kinase</i>
MgCl₂	do inglês <i>magnesium chloride</i>
mM	do inglês <i>millimolar</i>
MPSS	do inglês <i>Massively Parallel Signature Sequencing</i>
mRNA	do inglês <i>messenger RNA</i>
NaCl	do inglês <i>sodium chloride</i>
NCBI	do inglês <i>National Center for Biotechnology Information</i>
NH₄OAc	do inglês <i>ammonium acetate</i>
NIH	do inglês <i>National Institutes of Health</i>
ng	do inglês <i>nanogram</i>
nm	do inglês <i>nanometre</i>
NRG	do inglês <i>neuregulin</i>
OD	do inglês <i>optical density</i>
ORF	do inglês <i>Open Reading Frame</i>

p16	do inglês <i>16 KDa protein</i>
p21	do inglês <i>21 KDa protein</i>
p27	do inglês <i>27 KDa protein</i>
p53	do inglês <i>53 KDa protein</i>
p85	do inglês <i>85 KDa protein</i>
pb	do inglês <i>base pair</i>
PCR	do inglês <i>polimerase chain reaction</i>
PDGF	do inglês <i>Platelet-Derived Growth Factor</i>
pH	do inglês <i>hydrogen ionic potential</i>
PI3K	do inglês <i>Phosphatidyl-Inositol-3-Kinase</i>
PKC	do inglês <i>Protein Kinase C</i>
Poli-A	do inglês <i>polyadenylate</i>
RACE	do inglês <i>Rapid Amplification of cDNA ends</i>
RNA	do inglês <i>acid ribonucleic</i>
RNAse	do inglês <i>ribonuclease</i>
rpm	do inglês <i>rotation per minute</i>
RT-PCR	do inglês <i>Reverse transcriptase – polymerase chain reaction</i>
SAGE	do inglês <i>Serial Analysis of Gene Expression</i>
SH2	do inglês <i>Src Homology 2 Domain</i>
SIGIRR	do inglês <i>Single Immunoglobulin IL-1R-related</i>
SNPs	do inglês <i>Single Nucleotide Polymorphisms</i>
TIR	do inglês <i>Toll-interleukin-1 receptor</i>
TLR	do inglês <i>Toll-like receptor</i>
TNM	do inglês <i>Tumor Node Metastasis</i>
TRAF6	do inglês <i>Tumor Necrosis Factor (TNF) receptor-associated factor 6</i>
Tris	do inglês <i>hydroxymethyl aminomethane</i>
UICC	do francês <i>Union Internationale Contre le Cancer</i>
VEGF	do inglês <i>Vascular Endothelial Growth Factor</i>
X-GAL	do inglês <i>5-bromo-4-chloro-3-indolyl-β-D-galactoside</i>

ÍNDICE

1	INTRODUÇÃO	2
1.1	Aspectos epidemiológicos e clínicos do câncer de mama	2
1.2	Aspectos genéticos e moleculares do câncer de mama	9
1.2.1	Receptores tirosina quinase e câncer	11
1.2.1.1	O oncogene c-erbB2	12
1.2.1.2	Ativação do oncogene c-erbB2	17
1.2.1.3	Vias de sinalização celular ativadas pelo oncogene c-erbB2	20
1.2.1.4	Significado clínico do oncogene c-erbB2 no câncer de mama	24
1.2.1.5	O oncogene c-erbB2 como alvo terapêutico	27
1.3	Modelo HB4a / HB4a-C5.2	33
1.4	MPSS e a análise do perfil de expressão nas linhagens HB4a e HB4a-C5.2	36
1.5	GLGI	40
2	OBJETIVOS	43
2.1	Objetivo geral	43
2.2	Objetivos específicos	43
3	MATERIAIS E MÉTODOS	46
3.1	Linhagens celulares e amostras tumorais	46
3.2	Extração deRNA total e controle de contaminação com DNA genômico	48
3.3	Tratamento do RNA total com DNase I	50
3.4	GLGI	51
3.4.1	Extração de RNA total e isolamento de mRNA	51
3.4.2	Síntese da primeira e segunda fita de cDNA	52
3.4.3	Digestão da dupla fita de cDNA	54
3.4.4	Recuperação do fragmento 3' de cDNA	54
3.4.5	Ligação dos adaptadores ao fragmento 3' de cDNA	55
3.4.6	Padronização da amplificação do fragmento 3' de cDNA	56
3.4.7	Amplificação dos fragmentos 3' de cDNA em larga escala	57
3.4.8	Extensão do fragmento 3' de cDNA a partir da <i>tag</i> de MPSS	58
3.4.9	Purificação do produto obtido na reação de extensão do fragmento 3' de cDNA	62

3.5	Clonagem e transformação bacteriana	63
3.6	Seqüenciamento automático	64
3.7	Análise das seqüências de GLGI-MPSS	64
3.8	Análise do nível de expressão por PCR em Tempo Real	66
3.8.1	Síntese de cDNA	66
3.8.2	PCR em Tempo Real	66
3.8.3	Padronização das concentrações dos iniciadores e da quantidade de cDNA	68
3.8.4	Análise dos dados de expressão	69
3.8.5	Análise estatística dos dados de expressão	72
3.9	RACE (<i>Rapid Amplification of cDNA Ends</i>)	73
3.10	RT-PCR fita específica	75
4	RESULTADOS E DISCUSSÃO	79
4.1	Cultura de células e expressão diferencial de c-erbB2 nas linhagens HB4a e HB4a-C5.2	79
4.2	Análise das linhagens HB4a e HB4a-C5.2 por meio de MPSS	80
4.3	Padronização da técnica de GLGI-MPSS	82
4.3.1	<i>Tags</i> controles	87
4.3.2	<i>Tags</i> órfãs	88
4.4	Quantificação da expressão dos fragmentos 3' específicos e dos transcritos correspondentes às <i>tags</i> alternativas associadas a <i>SNPs</i> por meio de PCR em Tempo Real nas linhagens celulares HB4a e HB4a-C5.2	104
4.4.1	Validação da expressão diferencial dos transcritos regulados por c-erbB2 em diferentes linhagens tumorais de mama por meio de PCR em Tempo Real	109
4.4.2	Quantificação da expressão dos transcritos diferencialmente expressos nas linhagens celulares HB4a e HB4a-C5.2 por meio de PCR em Tempo Real em amostras tumorais de mama	113
4.5	Caracterização dos novos transcritos regulados por c-erbB2 e validados por PCR em Tempo Real	122
4.6	Caracterização dos transcritos <i>antisense</i> regulados por c-erbB2	132

5	CONSIDERAÇÕES FINAIS	142
6	CONCLUSÕES	147
7	REFERÊNCIAS BIBLIOGRÁFICAS	150

ANEXOS

Anexo 1 Generation of longer 3' cDNA fragments from massively parallel signature sequencing tags.

Anexo 2 Conversion of MPSS Orphan Tags into 3' ESTs.

Anexo 3 The impact of SNPs on the interpretation of SAGE and MPSS experiments data.

Anexo 4 Curriculum Vitae

Introdução

1 INTRODUÇÃO

1.1 ASPECTOS EPIDEMIOLÓGICOS E CLÍNICOS DO CÂNCER DE MAMA

O câncer de mama representa nos países ocidentais uma das principais causas de morte entre as mulheres. As estatísticas indicam o aumento de sua incidência tanto nos países desenvolvidos quanto nos países em desenvolvimento (RADICE e REDAELLI 2003; JEMAL et al. 2004). Segundo a Organização Mundial da Saúde (OMS), nas décadas de 60 e 70 houve um aumento de 10 vezes nas taxas de incidência, para esse tipo de câncer, nos registros de diversos continentes. Entretanto, as taxas de mortalidade associadas a esse tipo de neoplasia vêm diminuindo consideravelmente, o que pode ser atribuído, principalmente, ao avanço nos métodos de diagnóstico e de tratamento (DUMITRESCU e COTARLA 2005). No Brasil, o câncer de mama é o que mais causa mortes entre as mulheres. Estima-se que para o ano de 2006 sejam diagnosticados, no país, 48.930 novos casos (Tabela 1) e que ocorram 9000 óbitos decorrentes da doença por ano (Ministério da Saúde 2006).

Tabela 1 - Dados de incidência do câncer de mama no Brasil.

Localização Primária Neoplasia Maligna	Casos	Taxa Bruta
Mama Feminina	48.930	51,66
Traquéia, Brônquio e Pulmão	9.320	9,82
Estômago	8.230	8,65
Colo do Útero	19.260	20,31
Cólon e Reto	13.970	14,73
Esôfago	2.610	2,74
Leucemias	4.220	4,45
Cavidade Oral	3.410	3,58
Pele Melanoma	3.050	3,16
Outras Localizações	63.320	66,78
Subtotal	176.320	185,95
Pele não Melanoma	61.160	64,53
Todas as Neoplasias	237.480	250,45

Fonte: Modificado de Ministério da Saúde (2006)

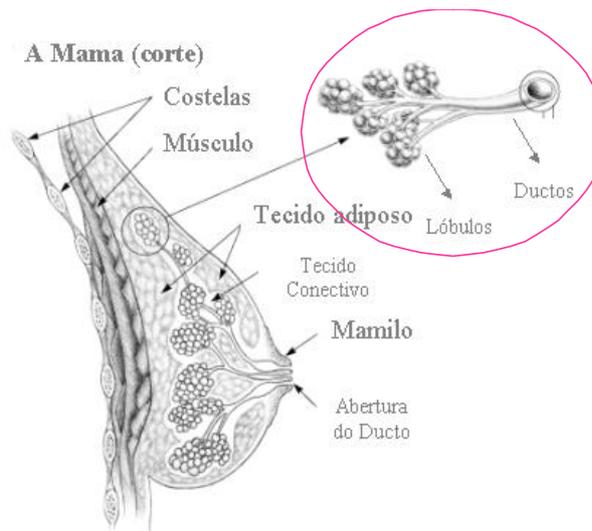
Legenda: Estimativas para o ano de 2006 das taxas brutas de incidência por 100.000 mulheres e do número de casos novos por câncer, em mulheres, segundo localização primária.

O câncer de mama é uma doença bastante heterogênea, apresentando uma grande variabilidade clínica e histopatológica. Essa heterogeneidade reflete a sua complexa etiologia, que sofre a influência de fatores exógenos (o tipo de dieta e o uso de contraceptivos orais – o que ainda é alvo de muita controvérsia) e de fatores endógenos (as taxas hormonais e a predisposição genética – considerado o fator de risco mais importante) (BIÉCHE e LIDEREAU 1995). Outros fatores que levam a uma maior suscetibilidade a essa doença incluem a menarca precoce, a menopausa tardia (além dos 50 anos, em média) e a nuliparidade, ou a primeira gravidez após os 30 anos de idade. Além desses fatores, o consumo habitual de álcool também parece estar associado ao risco aumentado de desenvolvimento de câncer de mama (SHERMAN e HOSSFELD 1993; RAHMAN e STRATTON 1998).

O diagnóstico precoce dos tumores de mama favorece a eficácia do tratamento e a possibilidade de cura da doença, que se mostra superior a 90% quando

o tumor é detectado ainda em estágios iniciais. Os métodos utilizados para a detecção do câncer de mama incluem o auto-exame, o exame clínico-palpatório, a mamografia, a ultra-sonografia e os exames cito e histopatológicos, conferindo uma acuidade de até 95% para o diagnóstico da doença. Porém, o elevado custo desses exames torna o diagnóstico tardio, além de dificultar o tratamento (VAN POZNAK e SEIDMAN 2002; SUNPAWERAVONH e SUNPAWERAVONG 2004). No Brasil, cerca de 70% dos casos de câncer de mama são diagnosticados tardiamente em estágios avançados (<http://www.hcanc.org.br>) e, embora tenham ocorrido alguns avanços no que diz respeito ao diagnóstico precoce da doença, métodos diagnósticos mais sensíveis e acessíveis precisam ser desenvolvidos.

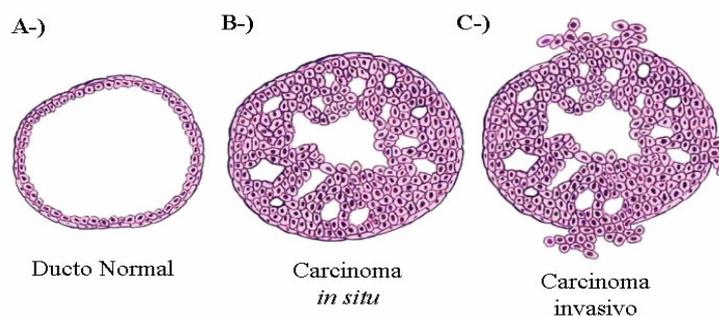
A apresentação clínica do câncer de mama pode variar desde formas não invasivas até outras altamente metastáticas. A maioria dos cânceres de mama, mais de 95%, pertence à classe dos carcinomas, originando-se do epitélio dos lóbulos e ductos da glândula, mais especificamente das células luminais (Figura 1) (HARRIS et al. 1999; PAGE et al. 1999).



Fonte: modificado de <http://www.cancervic.org.au/cancer1/prevent/breasthealth/dcis.gif>

Figura 1 - Representação esquemática de um corte de mama. Em destaque estão representados os lóbulos e ductos da glândula.

O carcinoma invasivo encontra-se frequentemente associado a focos de carcinoma *in situ* e a lesões hiperplásicas, sugerindo que essas lesões pré-malignas possam representar lesões precursoras do câncer. Em estádios mais avançados, as células tumorais invadem a membrana basal bem como o tecido adjacente. Já nas formas altamente metastáticas, as células tumorais invadem outros tecidos como ossos, fígado, cérebro e pulmões (Figura 2) (LAKHANI et al. 1999; SHERMAN e HOSSFELD 1993).



Fonte: modificado de http://www.breastcancer.org/is_cancer_invasive.html

Figura 2 - Representação esquemática da progressão tumoral. Ducto normal (A); Carcinoma ductal *in situ* (B) e Carcinoma ductal invasivo (C).

O carcinoma mamário dissemina-se por via linfática, pela corrente sanguínea e por extensão direta. Na disseminação por via linfática, os linfonodos agem como uma espécie de filtros das células que se desprendem do tumor primário. Depois de atravessar o filtro ganglionar, geralmente, as células tumorais invadem os linfonodos supraclaviculares e alcançam a circulação venosa (SHERMAN e HOSSFELD 1993).

A determinação do estadiamento clínico do câncer de mama teve início no século 20, considerando especificamente o tamanho do tumor e a sua localização, o acometimento de linfonodos regionais e a presença de metástases distantes, culminando no desenvolvimento do sistema de classificação TNM criado pelo AJCC (*The American Joint Committee on Cancer*) e pelo UICC (*Union Internationale Contre le Cancer*). Esse sistema baseia-se no tamanho do tumor (T), no envolvimento de linfonodos (N) e na presença de metástases distantes (M) (STYBLO e WOOD 1998).

A principal forma de tratamento para o câncer de mama é a ressecção cirúrgica podendo ou não ser seguida de terapia adjuvante (quimioterapia, terapia hormonal e radioterapia). As formas mais comuns de ressecção são a cirurgia

conservadora com esvaziamento de linfonodos, ou a mastectomia radical (VAN POZNAK e SEIDMAN 2002; SHERMAN e HOSSFELD 1993). A ressecção cirúrgica é curativa em mais de 50% dos casos, os quais apresentam pouca ou nenhuma doença micrometastática. A terapia adjuvante aumenta a expectativa de vida, provavelmente devido à erradicação ou à inibição da progressão da doença micrometastática. A terapia adjuvante deve ser adotada em pacientes com tumores maiores que 1 cm ou com metástases nos linfonodos axilares. Entretanto, cerca de 30% das pacientes que não apresentam acometimento de linfonodos morrem em decorrência da doença. Essas pacientes poderiam beneficiar-se da administração de terapia adjuvante, porém a dificuldade em se identificar tais pacientes permanece um desafio (DOWSETT 1998). Por outro lado, submeter todas as pacientes à terapia adjuvante não é viável, devido à alta toxicidade ocasionada pelo tratamento e também ao elevado custo (THOMSEN 2003).

Os fatores prognósticos tradicionais para o câncer de mama incluem o acometimento de linfonodos, o tamanho do tumor primário, seu grau histológico e seu subtipo histológico. Entre os fatores prognósticos tradicionais, a presença de metástases em linfonodos axilares e o tamanho do tumor são os dois fatores prognósticos mais importantes. Assim, a necessidade de se determinar quais pacientes não apresentam comprometimento de linfonodos e possuem um risco aumentado de recorrência da doença tem conduzido diversos estudos a respeito da identificação de marcadores moleculares (STYBLO e WOOD 1998; JARDINES e BERGER 1998).

A identificação de marcadores moleculares tem despertado grande interesse uma vez que os mesmos podem ser utilizados no diagnóstico precoce da doença, na

avaliação do prognóstico da paciente e também na determinação da conduta terapêutica (DOWSETT 1998; SRIVASTAVA e GOPAL-SRIVASTAVA 2002; KEEN e DAVIDSON 2003). As características neoplásicas e metastáticas adquiridas pela célula tumoral associam-se à ocorrência de alterações genéticas e epigenéticas que afetam genes envolvidos nos processos de proliferação celular, apoptose, reparo de DNA, angiogênese e adesão celular. Os genes afetados, assim como as alterações genéticas e epigenéticas relacionadas, podem ser utilizados na prática clínica como marcadores moleculares da doença (SRINIVAS et al. 2001 e NEGM et al. 2002).

A utilização de marcadores moleculares no diagnóstico precoce do câncer de mama ainda é bastante limitada. Porém, esses marcadores moleculares estão sendo cada vez mais utilizados na prática clínica, tanto para avaliar a evolução da doença quanto a resposta ao tratamento (ESTEVA et al. 2002; ARCIERO et al. 2003; ESTEVA e HORTOBAGYI 2004). Como exemplo de marcadores moleculares, com reconhecido valor prognóstico em câncer de mama, pode-se mencionar os receptores de estrógeno e de progesterona, o antígeno de proliferação Ki67 e o oncogene c-erbB2. Os receptores de estrógeno e de progesterona e o oncogene c-erbB2 são utilizados também como marcadores moleculares para avaliar respectivamente a resposta à terapia hormonal e ao tratamento com o anticorpo monoclonal Trastuzumab. A utilização na prática clínica de outros marcadores moleculares de prognóstico, por exemplo, o gene supressor de tumor p53, o receptor de ácido retinóico β , o fator de crescimento derivado de plaquetas (PDGF) e o fator de crescimento vascular endotelial (VEGF), entre outros, ainda está sob investigação (ESTEVA et al. 2002; ARCIERO et al. 2003; KEEN e DAVIDSON 2003; ESTEVA e HORTOBAGYI 2004).

1.2 ASPECTOS GENÉTICOS E MOLECULARES DO CÂNCER DE MAMA

O estudo das bases moleculares do desenvolvimento de tumores gerou o conceito de que o câncer é uma doença causada pelo acúmulo progressivo de alterações genéticas (BAYLIN e OHM 2006). Essas alterações genéticas afetam diferentes etapas nas vias que regulam os processos de proliferação, diferenciação e sobrevivência celular. Acredita-se que as células presentes em um tumor sejam descendentes de uma célula ancestral comum que em algum momento, geralmente décadas antes do tumor tornar-se palpável, desencadeou um processo de proliferação exacerbada (WEINBERG 1996). Essas células, então, devido à maior capacidade de proliferação, sofrem uma expansão clonal transmitindo geneticamente as alterações para as células que dela se originaram (KING 2000).

É importante destacar que o processo de tumorigênese também pode ser desencadeado por alterações epigenéticas (mecanismos que alteram o padrão de expressão gênica sem alterar a seqüência de DNA). As alterações epigenéticas envolvem tanto ganhos e perdas de metilação no DNA como modificações de histonas. Embora os mecanismos moleculares que envolvem essas alterações estejam apenas começando a ser elucidados, sabe-se que a repressão transcricional de genes supressores de tumor representa um dos principais mecanismos de alteração epigenética. A supressão transcricional pode estar associada à metilação aberrante do DNA, em regiões denominadas ilhas de CpG que se localizam na região promotora dos genes (BAYLIN e OHM 2006). A acetilação e deacetilação de histonas representam um outro tipo de alteração epigenética que é responsável por modular a

estrutura da cromatina e regular a expressão gênica nas células eucarióticas (KOURAKLIS et al. 2006). As histonas acetiltransferases (HATs) e as histonas deacetilases (HDACs) são duas classes de enzimas que devem atuar em equilíbrio para que ocorra o controle da acetilação de histonas. A presença de mutações ou alterações na expressão de genes que codificam as histonas acetiltransferases ou deacetilases têm sido associadas a carcinogênese e progressão tumoral (MEI et al. 2004; MAI et al. 2005).

Os genes envolvidos na tumorigênese podem ser funcionalmente divididos em duas classes: os proto-oncogenes e os genes supressores de tumor. Os genes supressores de tumor, atuam restringindo a proliferação celular e, geralmente, as alterações nesses genes estão associadas a perda de função. Os proto-oncogenes, ao contrário, geralmente estão envolvidos na regulação positiva da proliferação celular e, quando alterados ou amplificados no genoma, passam a ser denominados oncogenes. As alterações em oncogenes estão freqüentemente associadas a um ganho de função e, por essa razão, são ditas dominantes em relação à tumorigênese (MENDELSON et al. 1995).

No câncer de mama, a seqüência de eventos genéticos que conduzem à formação de um tumor palpável ainda não está claramente compreendida (KING 2000). Entretanto, algumas alterações comumente encontradas nesses tumores já foram descritas. Nesse processo de malignização, ocorre o envolvimento de diversos proto-oncogenes (c-MYC, c-erbB2, c-erbB1 e ciclina D1) e genes supressores de tumor (BRCA1, BRCA2, TP53, p16, E-caderina entre outros) (VOLGELSTEIN e KINZLE 1998; FOLGUEIRA e BRENTANI 2004).

Uma classe de oncogenes, já caracterizada, corresponde aos receptores para fatores de crescimento presentes na superfície das células. Esses receptores são proteínas transmembrânicas que possuem um domínio extracelular, ao qual se liga o fator de crescimento, e um domínio citoplasmático, que é capaz de desencadear uma cascata de sinalização intracelular. A superprodução desses receptores tem sido identificada em alguns tipos de tumores, entre eles os tumores de mama, resultando na ativação da proliferação celular, sem que haja necessidade da ocorrência de um estímulo extracelular. Um exemplo clássico de receptores para fatores de crescimento que atuam como oncogenes, são os receptores de membrana com atividade de tirosina quinase pertencentes a família de receptores de crescimento epidérmico erbB (CAMARGO e COSTA 2003).

1.2.1 Receptores Tirosina Quinase e câncer

Os receptores com atividade de tirosina quinase são enzimas que atuam como mediadores primários no processo de sinalização celular, regulando a progressão do ciclo celular, rearranjo do citoesqueleto, diferenciação celular e apoptose. (CASALINI et al. 2003; ROSKOSKI 2004). A sub-classe I da super-família de receptores tirosina quinase é constituída por receptores de fatores de crescimento epidérmico e compreende quatro membros: EGFR/c-erbB1; c-erbB2; c-erbB3 e c-erbB4. Todos os membros dessa família de receptores possuem um domínio de ligação extracelular rico em cisteína, contendo sítios que podem ser glicosilados, um domínio transmembrana e um domínio citoplasmático contendo resíduos de tirosina (XIE et al. 2000; WITTON 2003; HYNES e LANE 2005).

A porção extracelular desses receptores é constituída por quatro domínios, sendo os domínios I e III (também chamados de L1 e L2 e constituídos por aproximadamente 190 aminoácidos), os envolvidos na ligação dos receptores aos fatores de crescimento. Os domínios II e IV (também chamados de S1 e S2) são ricos em resíduos de cisteína, os quais são mantidos unidos por pontes dissulfeto. (WITTON 2003; MARMOR et al. 2004; ROSKOSKI 2004; HYNES e LANE 2005). Quando os receptores erbB estão no estado inativo, o domínio II, envolvido no mecanismo de dimerização entre receptores, torna-se bloqueado por uma interação intramolecular com o domínio IV. Entre os membros da família de receptores tirosina quinase, os que estão freqüentemente associados a diversos tipos de câncer são o c-erbB1 e o c-erbB2. Nos tumores de mama, as alterações moleculares encontradas ocorrem freqüentemente no receptor c-erbB2 e geralmente associam-se a super-expressão do mesmo (KRAUS et al. 1987; STERN 2000; HYNES e LANE 2005).

1.2.1.1 O oncogene c-erbB2

Em 1981, um grupo de pesquisadores transfectaram células 3T3 (fibroblastos de camundongo) utilizando material genético obtido a partir de células de neuroblastoma de rato e identificaram uma nova seqüência de DNA com potencial transformante. Este oncogene foi mapeado no cromossomo 17 e foi denominado “neu” (neuroblastoma). Após alguns anos, o homólogo humano do oncogene “neu” foi clonado a partir de uma biblioteca de cDNA, sendo denominado “HER2”. Em seguida, a seqüência genômica desse gene foi clonada e denominada c-erbB2 (PLACIDO et al. 1998; XIE et al. 2000).

Experimentos realizados em glândula mamária de camundongos transgênicos indicaram que a super-expressão de c-erbB2 induz o câncer de mama metastático. Esse modelo animal revelou que a carcinogênese desencadeada pela super-expressão de c-erbB2 é um processo que envolve múltiplas etapas. A tumorigenicidade requer alterações estruturais adicionais no oncogene, mutações em outros oncogenes ou ainda uma excessiva produção de agonistas de EGF (STERN 2000).

O oncogene c-erbB2 localiza-se no braço longo do cromossomo 17 (17q11.2 – q12) (Figura 3), sendo composto por 4.530 pares de bases, distribuídos ao longo de 27 exons. Esse oncogene codifica uma glicoproteína transmembrânica (receptor) de 185-KDa com atividade tirosina-quinase em sua porção carboxi-terminal. Essa proteína, constituída por 1.255 aminoácidos, apresenta um alto grau de similaridade com outros membros da família de receptores tirosina-quinase (c-erbB1, c-erbB3 e c-erbB4) (ISHII et al. 1987; XIE et al. 2000; RABINDRAN 2005).

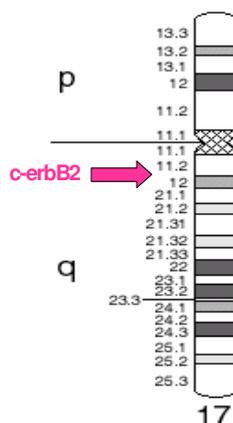
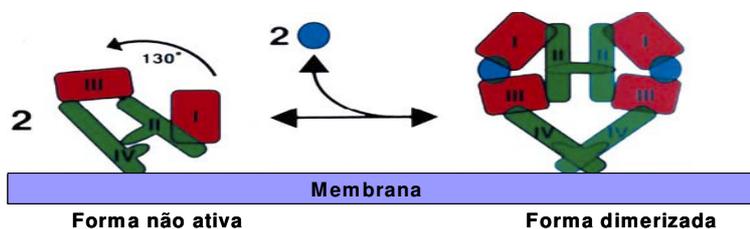


Figura 3 - Representação esquemática do cromossomo 17. Em destaque está representado o oncogene c-erbB2, cuja localização citogenética ocorre no braço longo do cromossomo (17q) entre as bandas cromossômicas 11.2 e 12.

A estrutura da porção extracelular do receptor c-erbB2 não se assemelha à estrutura dos demais membros de sua família. Dados cristalográficos sobre a estrutura do domínio extracelular de c-erbB2, na ausência de um ligante, revelaram que sua conformação estrutural lembra a forma ativada dos receptores c-erbB1 e c-erbB3 (Figura 4), sugerindo que esse receptor seja autoativado (ROSKOSKI 2004; HYNES e LANE 2005).

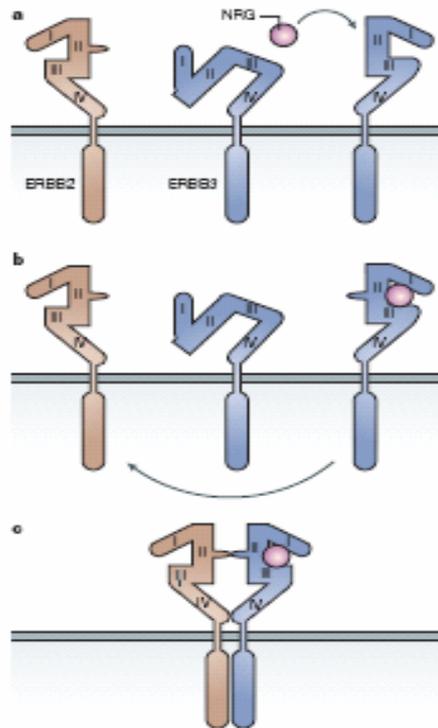


Fonte: Modificado de ROSKOSKI (2004).

Figura 4 - Estrutura de um receptor erbB ativado. Representação esquemática das alterações estruturais que ocorrem nos receptores erbB induzidas pela presença de um ligante.

Para o receptor c-erbB2, a ocorrência de interação entre os domínios II e IV é inexistente (Figura 5) (ROSKOSKI 2004; HYNES e LANE 2005). Além disso, diversos resíduos que desempenham importante papel na interação dos demais receptores erbB com os ligantes, são substituídos no c-erbB2 por resíduos (Met10, Arg13 e Pro15) que impedem a interação do receptor com o ligante. O fato do receptor c-erbB2 apresentar a região que contém o sítio de ligação para fatores de crescimento mutada e obstruída, explica a ausência de um ligante específico para esse receptor (ROSKOSKI 2004). O receptor c-erbB2, em sua conformação nativa, apresenta o domínio II completamente exposto, indicando sua constante disponibilidade para formar homodímeros ou formar heterodímeros com outros

membros ativos de sua família. Essa estrutura peculiar do c-erbB2, explica a razão pela qual esse receptor é o parceiro preferencial de outros membros de sua família. Por essa razão, o oncogene c-erbB2 tornou-se um importante alvo terapêutico (HYNES e LANE 2005).



Fonte: Modificado de HYNES e LANE (2005).

Legenda: (a) domínios I e III estão envolvidos na ligação do peptídeo heregulina ao receptor c-erbB3, (b) fazendo com que o domínio II do receptor se torne exposto para a dimerização com outro receptor erbB. (c) o receptor c-erbB2 possui uma estrutura nativa que lembra a forma ativada dos receptores c-erbB1 e c-erbB3.

Figura 5 - Domínio extracelular dos receptores c-erbB2 e c-erbB3.

Em 1986, BARGMANN et al. identificaram uma mutação pontual, no códon 664 (V664G), na região transmembrana do gene neu (gene de rato homólogo ao c-erbB2 humano), que foi capaz de aumentar o potencial transformante desse gene. Camundongos transgênicos portadores do gene neu ativado pela mutação, ou que apresentavam super-expressão do gene normal, freqüentemente, desenvolvem

adenocarcinoma de mama (XIE et al. 2000). Embora nenhuma mutação ainda tenha sido identificada no gene humano, análises de clones de cDNA do gene c-erbB2 permitiram a identificação de um polimorfismo no domínio transmembrana, no códon 655. Esse polimorfismo codifica uma isoleucina (Ile; ATC) em vez de uma valina (Val; GTC), o que poderia resultar em um aumento da atividade tirosina quinase da proteína (XIE et al. 2000; BENUSIGLIO et al. 2005).

Nos tumores de mama humanos, o principal mecanismo de ativação do c-erbB2 é a amplificação gênica, que é freqüentemente associada à super-expressão da proteína. Entretanto, vale ressaltar que existem diversos tumores de mama que apresentam super-expressão da proteína c-erbB2 mesmo sem a ocorrência de amplificação do gene (KARUNAGARAN et al. 1996; RÉVILLION et al. 1998). Amplificação do gene c-erbB2 e a concomitante super-expressão da proteína têm sido associadas à progressão tumoral, estando presentes em todos os estágios dos carcinomas mamários incluindo as lesões metastáticas (LIU et al. 1992; KARUNAGARAN et al. 1996).

A amplificação do gene c-erbB2, ou a super-expressão de seu produto protéico, ocorre em 25-30% dos carcinomas ductais invasivos de mama e está associado a um pior prognóstico (STYBLO e WOOD 1998; YAMAUCHI et al. 2001; TIMMS et al. 2002). Essa associação deve-se ao fato de que a super-expressão de c-erbB2 é responsável por aumentar a proliferação celular do tumor e favorecer a formação de metástases e desenvolvimento de quimioresistência (STYBLO e WOOD 1998; YAMAUCHI et al. 2001; TIMMS et al. 2002).

Atualmente, acredita-se que o papel desempenhado pelo oncogene c-erbB2 durante a progressão tumoral ocorra ainda nos estágios iniciais do processo de

transformação, sendo que a super-expressão do oncogene confere às células tumorais uma vantagem proliferativa que garante a sobrevivência das mesmas durante a seleção clonal. Essa hipótese é sustentada pela alta frequência com que a super-expressão de c-erbB2 ocorre nos carcinomas ductais *in situ* quando comparada aos carcinomas mais avançados (CASALINI et al. 2004).

1.2.1.2 Ativação do oncogene c-erbB2

Os membros da família de receptores tirosina quinase são expressos em tecidos de origem epitelial, mesenquimal e neuronal. Sob condições fisiológicas, a ativação dos receptores erbB é controlada pela expressão temporal de seus ligantes (membros da família de fator de crescimento epidérmico – EGF). Todos os ligantes específicos dos receptores erbB compartilham um domínio composto por 45-55 aminoácidos e seis resíduos de cisteína que interagem covalentemente formando três loops (STERN 2000; ROSKOSKI 2004; CASALINI et al. 2004; HYNES e LANE 2005). Essa região confere a especificidade de ligação aos receptores e permite a divisão dos ligantes em três grupos distintos:

- o primeiro grupo inclui o fator de crescimento epidérmico (EGF), fator de crescimento transformante alfa ($TGF-\alpha$) e a anfiregulina (AR), que se ligam ao receptor c-erbB1;

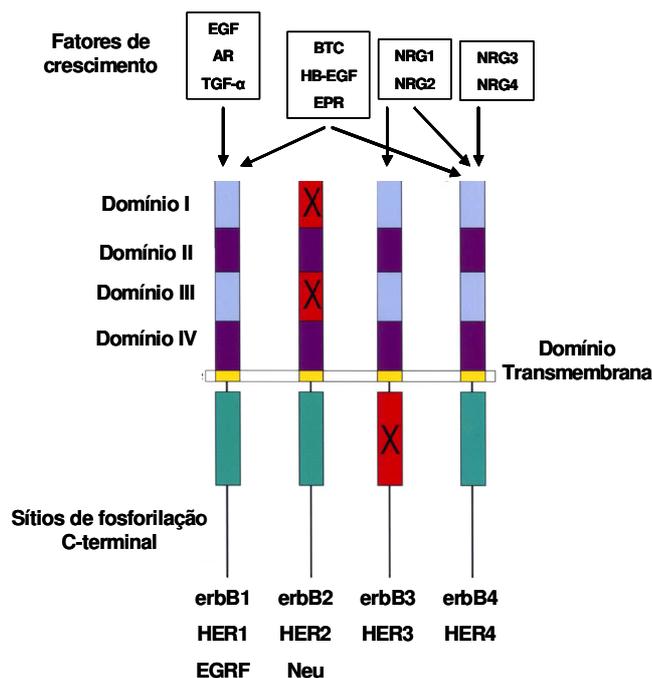
- o segundo grupo inclui a betacelulina (BTC), o fator de crescimento epidérmico ligado à heparina (HB-EGF) e a epiregulina (EPR), que se ligam aos receptores c-erbB1 e c-erbB4;

- o terceiro grupo inclui as neuregulinas (NRG), também conhecidas como heregulinas, ou fator de diferenciação neu, que subdividem-se em dois outros grupos:

as neuregulinas 1 e 2 se ligam aos receptores c-erbB3 e c-erbB4, e as neuregulinas 3 e 4 se ligam apenas ao receptor c-erbB4 (STERN 2000; OLAYIOYE et al. 2001; HYNES e LANE 2005).

Os ligantes são peptídeos produzidos como precursores transmembrana, cujos ectodomínios são processados por proteólise tornando-os fatores de crescimento solúveis (CASALINI et al. 2004; HYNES e LANE 2005). Esse processo de clivagem proteolítica, extremamente importante para que haja controle entre a disponibilidade dos ligantes e a ativação do receptor, conta com o envolvimento de proteases pertencentes à família de metaloproteinases e metaloproteinases de matrix (HYNES e LANE 2005). Além da ocorrência de amplificação gênica, a presença constante de ligantes EGF nos tumores também pode ser responsável pela estimulação constitutiva dos receptores erbB.

Os ligantes apresentam preferências pelos diferentes tipos de dímeros formados pelos receptores. Do mesmo modo, cada receptor dimerizado possui um conjunto diferente de resíduos tirosina que, quando fosforilados, atuam como sítios de ligação para proteínas adaptadoras específicas que recrutam combinações diferentes de moléculas de sinalização (MONTROYAMA et al. 2002; CASALINI et al. 2004). A resposta das células a diferentes estímulos depende de quais receptores são ativados. Por essa razão, a regulação diferencial dos receptores é importante, uma vez que cada receptor possui propriedades únicas de sinalização (STERN 2000). Na figura 6, podemos observar a representação esquemática dos receptores da família erbB e seus ligantes específicos.



Fonte: Modificado de ROSKOSKI (2004).

Legenda: Os domínios de ligação inativos do receptor erbB2 estão destacados em vermelho assim como o domínio quinase inativo do receptor erbB3.

Figura 6 - Representação esquemática da família de receptores erbB e seus ligantes específicos.

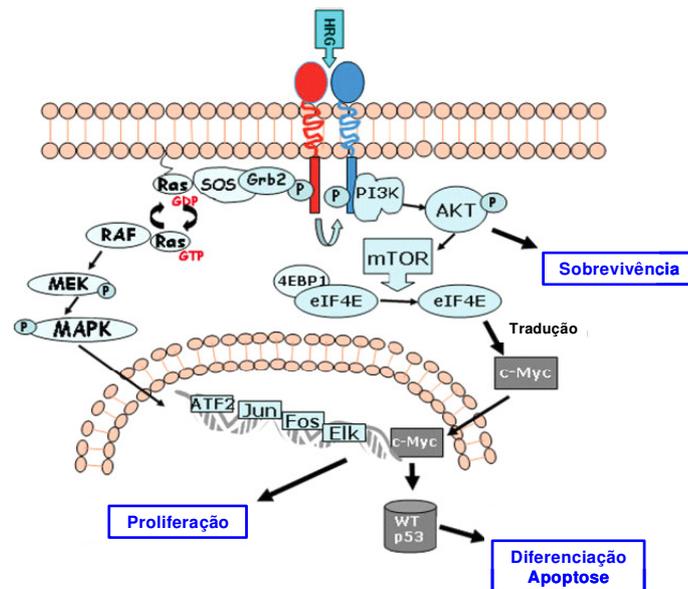
Os heterodímeros formados pelo c-erbB2 possuem um potencial de sinalização extremamente elevado devido à dramática redução na taxa de dissociação do ligante, promovendo a ativação prolongada das vias de sinalização downstream ao receptor (CASALINI et al. 2004). No entanto, é importante lembrar que quando o receptor c-erbB2 encontra-se superexpresso, a presença de um ligante para que a sua heterodimerização com os outros receptores da sua família ou sua homodimerização ocorra não se faz necessária (OLAYIOYE et al. 2001). Isto explica a capacidade desse receptor em desencadear a ativação constante de uma via de transdução de sinal mesmo quando nenhum ligante específico está presente (ROSKOSKI 2004).

A importância desses receptores durante o desenvolvimento normal pode ser demonstrada por meio da letalidade embrionária ou perinatal de camundongos knockout ou portadores de mutações. Esses animais apresentaram desenvolvimento neural e cardiovascular deficientes (c-erbB2, c-erbB3 e c-erbB4), e também apresentam deficiência na formação do epitélio, do trato respiratório e gastrointestinal (c-erbB1) (MARMOR et al. 2004; CASALINI et al. 2004; HYNES e LANE 2005). Além disso, esses receptores estão envolvidos no desenvolvimento normal da glândula mamária durante a puberdade e a gravidez (STERN 2000; CASALINI et al. 2004).

1.2.1.3 Vias de sinalização celular ativadas pelo oncogene c-erbB2

A homodimerização ou heterodimerização entre os receptores tirosina quinase torna-os ativos, promovendo a estimulação de uma rede de sinalização celular bastante complexa que é capaz de ativar diferentes moléculas de sinalização *downstream* ao receptor (MÉNARD et al. 2004; HYNES e LANE 2005).

As principais vias de sinalização desencadeadas pela ativação do receptor c-erbB2 são as vias das proteínas-quinases ativadas por mitógenos (MAPKs), da fosfatidil-inositol 3-quinase (PI3K), e a via da proteína quinase C (PKC) (TIMMS et al. 2002; ZHOU e HUNG 2003). As diferentes vias mencionadas estão esquematizadas na Figura 7.



Fonte: Modificado de CASALINI et al. (2004).

Legenda: Representação esquemática das principais vias de sinalização celular desencadeadas pela ativação do receptor c-erbB2.

Figura 7 - Vias de sinalização celular e a ativação do receptor c-erbB2.

A super-expressão de c-erbB2 induz a ativação prolongada da via de sinalização MAPK, conferindo às células que super-expressam esse receptor um maior potencial transformante. Isso pode ser explicado pelo fato de que a ativação do receptor c-erbB2 promove a fosforilação de resíduos tirosina específicos na porção intracelular do receptor, que atuam como sítios de ligação para domínios SH2 presentes em moléculas de sinalização intracelular como, por exemplo, a proteína citoplasmática (adaptadora) Grb2. A proteína Grb2 recruta a proteína SOS que, em seguida, ativa Ras, pela conversão GDP-GTP. Ras, por sua vez, se liga e ativa Raf quinase, ocasionando a fosforilação de MEK1/2 (MAPKK) e de ERK1/2 (MAPK) (LIU et al. 1992; RÉVILLION et al. 1998; ORR et al. 2000; XIE et al. 2000; YAMAUCHI et al. 2001). Em seguida, ERK fosforila diversas proteínas citoplasmáticas e de citoesqueleto, incluindo a quinase p70-S6. Além disso, ERK

também sofre uma rápida translocação para o núcleo, onde promove a fosforilação e consequente ativação de diversos fatores de transcrição, como por exemplo, Elk-1 e E2F (OLAYIOYE et al. 2001; TIMMS et al. 2002; MARMOR et al. 2004).

A via de sinalização PI3K-AKT também é modulada pela ativação do receptor c-erbB2 e parece estar associada com inibição de apoptose e com invasão das células epiteliais de mama (TIMMS et al. 2002; CASALINI et al. 2004). Essa via de sinalização, basicamente, pode ser estimulada por meio do recrutamento da subunidade adaptadora p85 de PI3K pelo receptor c-erbB2 ativado. Pela estimulação da via PI3K-AKT a proteína mTOR, que atua como um sensor regulando a disponibilidade de energia e nutrientes para a célula, também pode ser modulada fazendo com que ocorra um aumento de síntese protéica, por exemplo, do receptor de quimiocina CXCR4, envolvido na formação de metástases do tumor de mama para o pulmão (LI et al. 2004; HYNES e LANE 2005).

Outra via de sinalização que também pode ser ativada por meio de interações entre fatores extracelulares e receptores tirosina quinase corresponde à via de PKC. Após ser ativada, a proteína de membrana fosfolipase C degrada fosfolipídeos de membrana, produzindo mensageiros intracelulares secundários (DAG e IP3). Esses mensageiros secundários ativam uma classe de proteínas denominadas PKC (proteína quinase dependente de Ca^{+2}). Uma vez ativada, PKC pode regular, por meio de fosforilação, diversas proteínas que controlam o desenvolvimento celular (MARMOR et al. 2004).

Um dos principais processos celulares afetados pela sinalização de c-erbB2 é a progressão do ciclo celular (LANE et al. 2000; TIMMS et al. 2002; MARMOR et al. 2004). A existência de conexão entre a super-expressão de c-erbB2 e a proteína

nuclear p21 foi descrita pela primeira vez em 1998, porém o mecanismo por meio do qual essas proteínas estariam interagindo não pôde ser elucidado nessa ocasião. Evidências experimentais posteriores, indicaram que uma possível explicação para esse fato seria a maior estabilidade de p21 no citoplasma, onde essa proteína atua como inibidor de apoptose. Esses dados pareceram controversos uma vez que a localização nuclear de p21 é responsável pelo efeito inibitório exercido sobre o crescimento celular, sendo regulada pelo sinal de localização nuclear presente na porção carboxi terminal da proteína. Entretanto, trabalhos posteriores revelaram que a super-expressão de c-erbB2, em tumores de mama, pode induzir a localização citoplasmática de p21 através da ativação de AKT. AKT fosforila p21 no resíduo de treonina 145, presente no sinal de localização nuclear, resultando na permanência da proteína no citoplasma e, conseqüentemente, alterando a atividade inibitória de crescimento celular exercida por p21 (ZHOU e HUNG 2003).

Outro importante regulador negativo do ciclo celular que possui expressão reduzida em tumores de mama que super-expressam c-erbB2 é a proteína nuclear p27. Essa proteína é responsável pela parada do ciclo celular em G1 e sua regulação ocorre principalmente por ubiquitinação. A ativação da via MAPK nos tumores de mama que super-expressam o c-erbB2 resulta em aumentada degradação de p27 pelo proteassoma, ainda no citoplasma conduzindo à perda de atividade inibitória de crescimento exercida pela proteína (ZHOU e HUNG 2003).

A super-expressão da ciclina D1 também tem sido encontrada em muitos dos tumores de mama cuja transformação maligna é mediada pelo oncogene c-erbB2. Em um trabalho realizado por YU et al. (2001), utilizando camundongos com deficiência de ciclina D1, os resultados revelaram que esses animais eram resistentes ao

desenvolvimento de tumores de mama induzidos pelos oncogenes neu e ras. Esses achados reforçam a ligação entre a super-expressão de c-erbB2, a ativação da via MAPK e a super-expressão da ciclina D1 no processo de transformação das células epiteliais de mama (TIMMS et al. 2002). A ciclina D1 ativa as quinases dependentes de ciclina CDK4 e CDK6, e assim, promove a progressão do ciclo celular da fase G1 para a fase S. A ativação da via de sinalização das MAPKs induzida por c-erbB2 resulta na super-expressão da ciclina D1 por meio dos fatores de transcrição SP1 e E2F e também por meio da estabilização da proteína através de sua fosforilação por AKT. Entretanto, a exata contribuição de cada uma dessas vias para o aumento da ciclina D1, assim como os mecanismos exatos por meio dos quais o oncogene c-erbB2 atua sobre a progressão do ciclo celular ainda permanecem pouco compreendidos (LENFERINK et al. 2001; MARMOR et al. 2004).

Depois de disparadas as vias de sinalização celular, o sinal de ativação recebido pelo receptor começa a ser atenuado. A atenuação do sinal ocorre por meio da *down*-regulação nos níveis do receptor, presentes na superfície celular, através de sua endocitose. Os receptores dimerizados são internalizados em regiões da membrana plasmática, revestidas por clatrina que invaginam formando vesículas endocíticas. A etapa subsequente determina se o receptor será reciclado de volta à membrana, ou se sofrerá degradação lisossomal. (STERN 2000; MARMOR et al. 2004).

1.2.1.4 Significado clínico do oncogene c-erbB2 no câncer de mama

Diversos estudos têm mostrado a associação entre a super-expressão da proteína c-erbB2 e o desenvolvimento do câncer de mama. O produto protéico do

oncogene c-erbB2 encontra-se superexpresso em aproximadamente 25-30% dos carcinomas invasivos, geralmente devido à amplificação do gene (KEEN e DAVIDSON 2002). Entretanto, o valor prognóstico e preditivo do c-erbB2 ainda necessita ser melhor investigado, uma vez que existem muitas discrepâncias entre os achados dos diferentes estudos. Essas discrepâncias provavelmente estão relacionadas a fatores tais como: o número de pacientes incluídos em cada estudo, as características do tumor, o método utilizado para avaliar a positividade de c-erbB2 e até mesmo o estabelecimento de um nível mínimo para a positividade do oncogene. Apesar disso, algumas associações puderam ser notadas.

- Receptores esteróides: a grande maioria dos estudos clínicos revelaram uma relação inversa entre a positividade de c-erbB2 e a presença de receptores esteróides. Em geral, a positividade de c-erbB2 é duas vezes maior nos tumores que são receptores de estrógeno e progesterona negativos do que nos tumores que apresentam positividade para esses receptores.

- Tipo histológico do tumor: Entre os carcinomas, a positividade de c-erbB2 apresenta maior incidência no tipo ductal em relação ao lobular. Nos carcinomas ductais invasivos, a positividade do oncogene é duas vezes maior nos carcinomas *in situ* de alto grau, sub-tipo comedoniano, em relação aos carcinomas invasivos. Essa aparente discrepância, considerando que a super-expressão desse oncogene contribui com a maior capacidade de invasão tumoral, pode ser explicada pelo fato de que o carcinoma ductal *in situ* comedoniano pode representar uma lesão pré-invasiva, enquanto os que apresentam negatividade do oncogene devem derivar diretamente de uma hiperplasia atípica (LIU et al. 1992).

Também foram relatadas associações entre positividade de c-erbB2 e tumores com menor grau de diferenciação histológica, tumores aneuplóides e tumores com alta taxa de proliferação celular. Por outro lado, parâmetros como idade ao diagnóstico, tamanho do tumor, menopausa, envolvimento linfonodal e associação a outros receptores da família erbB não foram claramente associados à positividade de c-erbB2 (RÉVILLION et al. 1998).

Evidências indicam que, em pacientes linfonodo positivo, a positividade para c-erbB2 está associada com a diminuição de sobrevida livre de doença e/ou sobrevida global em análise univariada. Esses dados indicam que c-erbB2 é um marcador biológico de uma forma mais agressiva de câncer de mama. No caso de pacientes linfonodo negativo, a maioria dos estudos realizados não considerou o c-erbB2 como um indicador de prognóstico. Uma das possíveis explicações para esse achado é que o poder estatístico necessário para que se possa demonstrar um efeito na sobrevida deve ser diretamente proporcional ao número de eventos (recaídas ou mortes) no estudo. Uma vez que o risco de recorrência da doença em pacientes linfonodo negativo é pequeno, seriam necessários muitos casos para que um número de eventos estatisticamente significativo fosse obtido.

Quando o valor prognóstico de c-erbB2 é considerado em relação ao tratamento, o efeito favorável, tanto da hormonioterapia como da quimioterapia na sobrevida foi restrito aos pacientes c-erbB2 negativos. Isso sugere que o oncogene c-erbB2 pode estar envolvido em mecanismos de resistência a drogas, ou, ao menos, ser considerado um marcador de resistência (RÉVILLION et al. 1998).

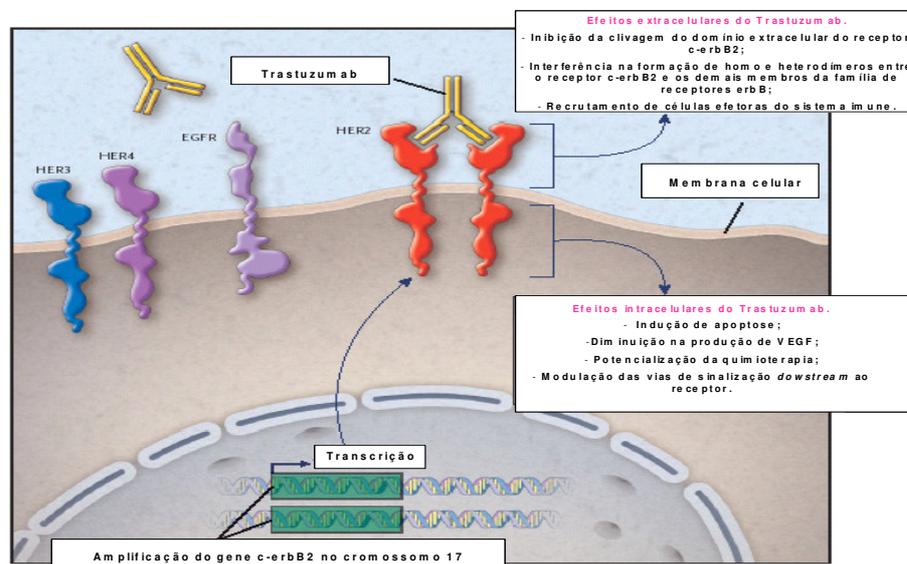
1.2.1.5 O oncogene c-erbB2 como alvo terapêutico

A super-expressão de c-erbB2 caracteriza uma fração bastante significativa dos carcinomas de mama e, por essa razão, esse oncogene tem sido alvo para o desenvolvimento de novas abordagens terapêuticas. Diferentes linhas de pesquisa têm sugerido que pacientes com câncer de mama que apresentam super-expressão de c-erbB2 não são responsivos aos tratamentos convencionais disponíveis. Assim, novas abordagens terapêuticas que visam combater as células que super-expressam c-erbB2 estão sendo desenvolvidas (RÉVILLION et al. 1998; ZHOU e HUNG 2003). Dois tipos principais de inibidores de c-erbB2 estão em uso clínico: os anticorpos humanizados, que agem diretamente sobre o domínio extracelular do c-erbB2; e moléculas inibidoras de atividade de tirosina quinase, que competem com o ATP no domínio de ligação tirosina-quinase do receptor impedindo a fosforilação do mesmo (HORTOBAGYI 2001; HYNES e LANE 2005; RABINDRAN 2005).

A primeira geração de anticorpos monoclonais avaliados clinicamente foi baseada em anticorpos monoclonais murino, cujo alvo é o domínio extracelular do receptor c-erbB2. A utilização desses anticorpos murino, em linhagens celulares de tumor de mama humano e também em modelo animal (xenograft), mostrou-se capaz de inibir o crescimento das células tumorais. Entretanto, o problema da imunogenicidade e a limitada capacidade desses anticorpos em recrutar mecanismos efetores do sistema imune levaram ao desenvolvimento de um anticorpo monoclonal anti-erbB2 humanizado (RÉVILLION et al. 1998; ADAMS e WEINER 2005). A versão humanizada do anticorpo murino (mAB4D5) foi produzida a partir da inserção de porções do sítio de ligação ao antígeno, do anticorpo murino, em um anticorpo monoclonal humano (IgG1) e recebeu o nome de Trastuzumab (Herceptin;

Genentech, Inc, South San Francisco, CA) (ARTEAGA 2004; ADAMS e WEINER 2005; BURSTEIN 2005; HORTOBAGYI 2005).

Desde 1998, o trastuzumab foi aprovado para o tratamento de tumores de mama metastáticos que super-expressam c-erbB2 (HORTOBAGYI 2001; HYNES e LANE 2005). O trastuzumab previne parcialmente a formação de heterodímeros, previne a clivagem do domínio extracelular do receptor, promove a parada do ciclo celular na fase G1 e induz p27, cuja atuação reduz o número de células na fase S (Figura 8) (LOHRISCH e PICCART 2001; KUTE et al. 2004). Além disso, ao contrário dos inibidores tirosina quinase, os anticorpos têm a capacidade inerente de recrutar células efetoras do sistema imune (macrófagos e monócitos) para o tumor por meio da ligação entre o domínio constante do anticorpo e os receptores específicos dessas células (HYNES e LANE 2005).



Fonte: Modificado de BURSTEIN (2005).

Legenda: O Trastuzumab é um anticorpo monoclonal humanizado que se liga ao receptor c-erbB2 inibindo o crescimento da célula tumoral por meio de uma variedade de mecanismos intracelulares e extracelulares.

Figura 8 - Interações entre o Trastuzumab e as células tumorais.

Estudos pré-clínicos demonstraram que o trastuzumab bloqueia apenas as células tumorais que super-expressam c-erbB2. Assim, esse agente terapêutico é prescrito apenas para as pacientes com câncer de mama que apresentam elevados níveis de c-erbB2. Esse anticorpo induz a regressão tumoral em aproximadamente 20-30% dos tumores metastáticos de mama (ARTEAGA 2004; HORTOBAGYI 2005). Embora ainda não esteja claro qual a melhor forma de administração do trastuzumab, testes clínicos utilizando a combinação desse anticorpo a outros agentes citotóxicos comumente utilizados no tratamento do câncer de mama metastático (vinorelbine, docetaxel, cisplatina e paclitaxel) têm revelado resultados satisfatórios (60-80% de taxa de resposta) com relação ao aumento na taxa de sobrevida das pacientes (TANNER et al. 2004; ADAMS e WEINER 2005; HORTOBAGYI 2005). Dois testes clínicos de fase III fizeram a comparação entre quimioterapia como tratamento de primeira linha e a combinação entre a quimioterapia e o Trastuzumab em pacientes com câncer de mama metastático com super-expressão de c-erbB2. No grupo que recebeu tratamento combinado, houve uma melhora significativa no tempo de progressão da doença, na taxa de resposta, na duração da resposta ao tratamento e aumento na sobrevida em até 29 meses. (HORTOBAGYI 2005). Diversos estudos mostram que o trastuzumab é bem tolerado apresentando uma baixa incidência de efeitos adversos. Entretanto, testes clínicos revelaram a ocorrência de disfunção cardíaca, reversível após o término do tratamento, em aproximadamente 16% das pacientes que receberam tratamento combinado trastuzumab-quimioterapia. Esse efeito cardiotóxico pode ser explicado pelo fato de que o tecido cardíaco normal expressa uma pequena quantidade da proteína e sua supressão prolongada afetaria a

manutenção do músculo cardíaco (HYNES e LANE 2005; RABINDRAN 2005; HORTOBAGYI 2005).

Como mencionado anteriormente, apenas 30% dos tumores metastáticos que super-expressam c-erbB2 respondem ao tratamento com trastuzumab e teorias como, por exemplo, a existência de vias de sinalização aberrantes downstream ao receptor podem explicar essa aparente resistência tumoral ao trastuzumab. A ativação constante da via de sinalização PI3K-AKT pode resultar no desenvolvimento de resistência das células tumorais ao trastuzumab. A principal função da fosfatase citoplasmática PTEN é defosforilar PI3K, inibindo a proteína-quinase AKT, que por sua vez é responsável pela transdução de sinais de sobrevivência e proliferação celular e inibição de apoptose. Assim, torna-se claro que a perda ou a presença de mutações em PTEN, favorecem a ativação constante de AKT tornando as células resistentes ao trastuzumab (HYNES e LANE 2005; RABINDRAN 2005). Além disso, os ligantes dos receptores erbB podem facilitar o desenvolvimento de resistência ao trastuzumab, por meio da ativação de heterodímeros alternativos. Essa hipótese foi testada experimentalmente, e os resultados revelaram que o trastuzumab não é capaz de bloquear a proliferação de células tumorais que têm ativação autócrina do receptor c-erbB1. Nesse caso, o anticorpo não é capaz de bloquear a formação de heterodímeros contendo o c-erbB2, ou bloquear a ativação das vias de sinalização *downstream* ao receptor, pois o trastuzumab se liga ao domínio IV do receptor, o qual não está envolvido na dimerização do mesmo. Por outro lado, o pertuzumab, outro anticorpo monoclonal cujo alvo é o receptor c-erbB2, se liga ao domínio II do receptor, o qual é responsável pela dimerização do mesmo. Esta característica pode explicar parcialmente porque esse anticorpo, ao contrário do

trastuzumab, é capaz de inibir o crescimento de tumores que expressam baixos níveis de c-erbB2 (BADACHE e HYNES 2004; HYNES e LANE 2005).

Uma outra classe de agentes capazes de inibir a sinalização celular desencadeada pela super-expressão do oncogene c-erbB2, representando uma alternativa terapêutica para o câncer de mama, são as moléculas inibidoras de atividade tirosina quinase. Células tumorais tratadas com esses agentes sofrem alterações em muitas de suas vias intracelulares (PI3K-AKT e MAPK), essenciais para o desenvolvimento e progressão tumoral (HORTOBAGYI 2001; HYNES e LANE 2005; RABINDRAN 2005).

A fosforilação dos resíduos tirosina dos receptores erbB é uma etapa bastante crítica para que as vias de sinalização reguladas por esses receptores sejam disparadas, culminando no desencadeamento de diversos eventos intracelulares. Os dois grandes desafios a serem superados com relação à identificação dos primeiros inibidores tirosina quinase foram os seguintes: seletividade e toxicidade. Os receptores erbB compõem uma, entre muitas, família de receptores com domínio tirosina quinase extremamente semelhantes o que poderia dificultar a identificação de inibidores específicos para cada um dos receptores. Além disso, esses receptores apresentam expressão não somente em diversos tipos de tumores sólidos como também em muitos tecidos normais, o que poderia representar um problema com relação ao nível de toxicidade desencadeado pelos inibidores. Entretanto, diversos testes realizados com preparações enzimáticas obtidas a partir de linhagens celulares com super-expressão de receptores erbB permitiram a identificação de uma classe de inibidores tirosina quinase, anilinoquinazoline específicos para o receptor c-erbB1 (GROSS et al. 2004; WAKELING 2005). A descoberta da classe de inibidores

anilinoquinazoline, representou a primeira etapa do programa que selecionou o Gefinitib (Iressa; ZD1839 - AstraZeneca) como candidato a testes clínicos. O gefinitib é um inibidor tirosina quinase que interage com o domínio intracelular do receptor c-erbB1, já aprovado para utilização no tratamento clínico. Já o inibidor tirosina quinase Lapatinib (GW572016 - GlaxoSmithKline), específico para os receptores c-erbB1 e c-erbB2, está em fase III dos testes clínicos para utilização em pacientes com câncer de mama não responsivos a Herceptina e quimioterapia (WAKELING 2005; HYNES e LANE 2005).

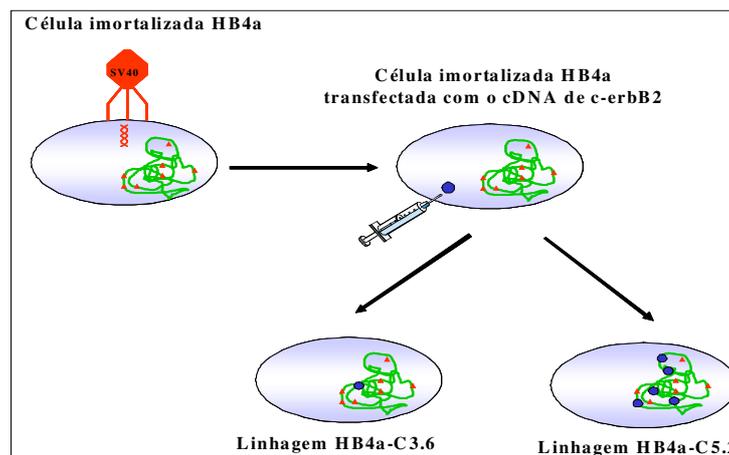
O conhecimento dos mecanismos de sinalização que envolvem a super-expressão do receptor c-erbB2 tem avançado e permitido importante progresso no conhecimento da tumorigênese do câncer de mama. Novas abordagens terapêuticas que atuam sobre as células que super-expressam c-erbB2 têm sido desenvolvidas e os resultados preliminares, revelam que essas abordagens, em geral, têm ocasionado a regressão dos tumores de mama (RABINDRAN 2005). No entanto, os mecanismos moleculares de atuação desse oncogene ainda não foram completamente elucidados e o desenvolvimento de novas abordagens que permitam a análise do perfil de expressão gênica das células tumorais que super-expressam o oncogene c-erbB2, cada vez mais, se faz necessário.

Nesse contexto, o presente trabalho, que teve por objetivo identificar genes diferencialmente expressos em linhagens celulares de mama, que expressam diferentes níveis do oncogene c-erbB2, poderá contribuir não somente para o melhor entendimento do papel biológico desempenhado por esse oncogene no câncer de mama como também para a identificação de novos alvos terapêuticos e/ou marcadores para diagnóstico.

1.3 MODELO HB4a / HB4a-C5.2

Para compreender o papel desempenhado pelo oncogene c-erbB2 no desenvolvimento do câncer de mama é preciso estudar as interações entre esse receptor, os demais membros de sua família e seus ligantes. Diversos modelos experimentais, incluindo células não humanas e não epiteliais, além de diferentes linhagens celulares de câncer de mama têm sido utilizadas para investigar o papel da super-expressão de c-erbB2 no desenvolvimento do câncer de mama (HARRIS et al. 1999).

HARRIS et al. (1999) geraram um painel de células luminais de mama imortalizadas e transfectadas com o cDNA completo de c-erbB2 humano (derivado de uma linhagem celular de câncer de mama já estabelecida – BT474) para estudar o potencial transformante de c-erbB2 e a sua interação com outros receptores de sua família. As células imortalizadas utilizadas nesse estudo, denominadas HB4a, foram obtidas por meio da infecção de células luminais de mama com o vetor retroviral que expressa o antígeno T do vírus SV40 (“Simian Virus” 40). Duas novas linhagens celulares, HB4a-C3.6 e HB4a-C5.2, expressando diferentes níveis de c-erbB2 foram derivadas da linhagem não transformada HB4a (HARRIS et al. 1999). Após a caracterização dessas linhagens utilizando citometria de fluxo e comparações com linhagens celulares de mama que expressam níveis conhecidos de c-erbB2, a linhagem HB4a-C3.6 foi considerada como apresentando um nível moderado de expressão de c-erbB2, enquanto que as células HB4a-C5.2 foram consideradas como tendo um elevado nível de expressão (HARRIS et al. 1999). O modelo de células luminais desenvolvido por HARRIS et al. (1999) está esquematizado na Figura 9.



Legenda: A célula imortalizada HB4a foi transfectada com o cDNA de c-erbB2, gerando duas linhagens celulares HB4a-C3.6 e HB4a-C5.2. A caracterização dessas linhagens utilizando citometria de fluxo e comparações com linhagens celulares de mama que expressam níveis conhecidos de c-erbB2 revelaram que a linhagem HB4a-C3.6 apresenta um nível moderado de expressão de c-erbB2, enquanto as células HB4a-C5.2 apresentam um elevado nível de expressão de c-erbB2.

Figura 9 - Modelo de células luminais desenvolvido por HARRIS et al. (1999).

Os resultados obtidos a partir dos ensaios de transformação *in vitro*, mostraram que a super-expressão de c-erbB2 é capaz de ocasionar a transformação das células HB4a, sendo que a extensão deste evento está correlacionada com o nível de expressão desse receptor. A expressão moderada de c-erbB2 não levou a nenhuma alteração morfológica nas células HB4a-C3.6, enquanto as células HB4a-C5.2, cujo nível de expressão de c-erbB2 é aumentado, apresentaram crescimento empilhado, indicando a perda de inibição por contato. Observou-se também que a ativação constitutiva do receptor c-erbB2 pode ser desencadeada por um alto nível de expressão, independentemente dos ligantes EGF e HRG. Nas células HB4a-C5.2, a proteína adaptadora Shc encontra-se constitutivamente associada ao c-erbB2 e à proteína Grb2, o que desencadeia a ativação da via de sinalização de RAS. Entretanto, apesar dos efeitos observados *in vitro*, a super-expressão do c-erbB2 não

causou tumorigênese *in vivo*. Esse fato foi observado por meio da inoculação de camundongos atímicos, sugerindo a necessidade da ocorrência de outros eventos posteriores até a completa malignização (HARRIS et al. 1999).

Um recente trabalho desenvolvido por MACKAY et al. (2003), utilizando as linhagens celulares HB4a-C3.6, HB4a-C5.2 e HB4a, teve por objetivo investigar as alterações associadas a super-expressão de c-erbB2. A abordagem do grupo foi baseada na utilização de um *microarray* de cDNA constituído por 9.930 clones representando aproximadamente 6.000 genes distintos. Utilizando o dado normalizado dos 9.930 clones de cDNA, foi gerada uma lista de genes diferencialmente expressos, quando as 3 linhagens acima mencionadas foram comparadas entre si. Esse trabalho possibilitou a identificação de um total de 61 genes diferencialmente expressos associados a super-expressão de c-erbB2.

Alguns dos genes que apresentaram expressão aumentada incluem membros da família de proteínas ligadoras de cálcio (S100P e S100A4), que foram previamente associadas a tumorigênese de mama. O gene mais significativamente super-expresso e que não havia sido associado a expressão de c-erbB2 foi o LGALS1 (galectina 1 – GBP). As galectinas são uma família de proteínas que estão envolvidas na modulação das interações célula-célula e célula-matriz. Ainda nesse trabalho, a LGALS3 (galectina 3) apresentou uma expressão reduzida nas linhagens HB4a-C3.6 e HB4a-C5.2 em relação à linhagem HB4a, o que está em concordância com dados já publicados que indicam a existência de uma correlação entre altos níveis de LGALS1 e baixos níveis de LGALS3 em carcinomas de mama com potencial metastático. Um dos genes identificado nesse trabalho, e que apresentou redução significativa em seu nível de expressão, foi o gene IGFBP3 (maior determinante dos níveis circulantes de

IGFs, as quais são potentes agentes mitogênicos nas células tumorais de mama). A relevância clínica de alguns desses genes, HSPB1, LGALS3, FN1, ERBB2, MYC e IGFBP3, foi demonstrada em nível de proteína através de imunohistoquímica em um *tissue array* composto por 48 carcinomas c-erbB2 positivos e 47 carcinomas c-erbB2 negativos.

Em dois outros trabalhos desenvolvidos por VAN'T VEER et al (2002) e VAN DE VIJVER et al (2002), baseados em análises de *microarray*, foi possível identificar um perfil de expressão gênica associado a um pior prognóstico para pacientes com câncer de mama. Nesses estudos, foram utilizados 70 genes para classificar 295 pacientes com carcinoma de mama primário como tendo uma assinatura de expressão gênica associada a um pior ou melhor prognóstico. Os resultados obtidos revelaram que entre os 295 pacientes, 180 possuíam um perfil de expressão gênica associado a um pior prognóstico, apresentando uma sobrevida global de 54,6% e sobrevida livre de metástases de 50,6% em 10 anos. Por outro lado, 115 pacientes apresentaram um perfil de expressão gênica associado a um melhor prognóstico, apresentando uma sobrevida global de 94,5% e sobrevida livre de metástases de 85,2% em 10 anos.

A exemplo do trabalho realizado por MACKAY et al. (2003), esse trabalho teve como objetivo isolar e caracterizar transcritos diferencialmente expressos em linhagens celulares de mama que expressam diferentes níveis de c-erbB2. Para tanto, o nível de expressão dos diferentes transcritos expressos nas linhagens HB4a e HB4a-C5.2 foi quantificado por meio de uma nova metodologia para análise da expressão gênica, denominada MPSS.

1.4 MPSS E A ANÁLISE DO PERFIL DE EXPRESSÃO NAS LINHAGENS HB4a e HB4a-C5.2

Massively parallel signature sequencing (MPSS) é uma técnica desenvolvida pela empresa *Lynx Therapeutics*, e que se baseia nos mesmos princípios de SAGE (*Serial Analysis of Gene Expression*), ou seja, na produção e quantificação de *tags* próximas à extremidade 3' dos transcritos. No entanto, ao contrário do SAGE, o MPSS utiliza a clonagem *in vitro* de fragmentos de DNA em *microbeads* e no seqüenciamento em larga escala dessas partículas sem a necessidade de separação física dos fragmentos a serem seqüenciados. A associação dessas duas tecnologias permite a produção de um número 100 vezes maior de *tags* em relação à técnica de SAGE e a avaliação da expressão de transcritos pouco abundantes (BRENNER et al. 2000a). Vale destacar, ainda, que o grande volume de dados gerados permite a realização de análises estatísticas sólidas e a caracterização de transcritos diferencialmente expressos, independentemente do fato de serem ou não transcritos conhecidos, e independente do nível de expressão desses transcritos ser alto ou baixo (BRENNER et al. 2000b).

Em um experimento realizado pela *Lynx Therapeutics*, o padrão de expressão das duas linhagens celulares de mama (HB4a e HB4a-C5.2) foi comparado com o objetivo de estudar as conseqüências moleculares relacionadas à super-expressão de *c-erbB2*. Um total de 24.521.236 *tags* foram geradas a partir de ambas as linhagens celulares correspondendo a 24.065 *tags* distintas. As seqüências das *tags* geraram um banco de dados experimental (JONGENEEL et al. 2003). Os dados gerados, foram analisados por dois grupos de bioinformática (Dr. Sandro de Souza e Dr. Victor

Jongeneel) com o objetivo de detectar *tags* diferencialmente expressas e também estabelecer a correlação entre *tags* e transcritos. Cada grupo de bioinformática fez a associação *tag*-transcrito utilizando estratégias de mapeamento distintas.

O grupo do Dr. Victor Jongeneel utilizou a estratégia de gerar uma seqüência consenso dos transcritos humanos, baseando-se nos alinhamentos entre seqüências expressas (mRNAs e *ESTs*) e a seqüência genômica. Os sítios de poliadenilação alternativos também foram considerados na montagem das seqüências consenso. As seqüências consenso de todos os transcritos foram então, analisadas quanto à presença do sítio de restrição *DpnII*. Depois de identificado o sítio da enzima localizado mais próximo à extremidade 3' de cada transcrito, uma *tag* virtual (constituída por 13 nucleotídeos) adjacente ao sítio de restrição foi extraída. O conjunto dessas *tags* virtuais gerou um banco de dados de referência. Por fim, os bancos de dados experimental e de referência foram cruzados e, então, parte das *tags* geradas foi associada a transcritos conhecidos (JONGENEEL et al. 2003).

O grupo do Dr. Sandro de Souza utilizou uma estratégia de mapeamento, baseando-se em todas as seqüências de mRNAs disponíveis nos bancos de dados públicos. Cada seqüência de mRNA foi analisada quanto à presença do sítio de restrição *DpnII*, e, então, a *tag* virtual adjacente ao sítio localizado mais próximo da extremidade 3' do transcrito foi extraída. As seqüências das *tags* virtuais geraram um banco de dados de referência em que também foram armazenadas informações a respeito de cada transcrito a partir do qual a *tag* foi extraída. Finalmente, as informações contidas nos bancos de dados experimental e de referência foram cruzadas e os transcritos a partir dos quais as *tags* experimentais de MPSS foram geradas puderam ser identificados (SILVA et al. 2004b) (Anexo 3).

Os resultados das análises realizadas pelos grupos de bioinformática indicaram que mais de 90% das *tags* de MPSS foram associadas a genes conhecidos e vários genes diferencialmente expressos foram identificados. Uma fração menor (9%), porém significativa, de *tags* (incluindo várias diferencialmente expressas entre as duas linhagens celulares) não foi associada a transcritos conhecidos e a origem delas permanece incerta. Uma parte dessas *tags* órfãs deve representar novos transcritos humanos. Entretanto, uma fração ainda não calculada dessas *tags* pode representar artefatos gerados pela técnica de MPSS. As *tags* artefatuais podem ser geradas basicamente por duas razões: digestão parcial da enzima de restrição *DpnII*, ou anelamento interno do oligo d(T) no momento da síntese de cDNA.

No entanto, devido ao pequeno tamanho das *tags* (17 nt), torna-se difícil usar sua informação para uma posterior validação experimental e caracterização dos transcritos correspondentes. Para solucionar esse problema, a técnica GLGI (*Generation of longer cDNA fragments from SAGE tags for gene identification*), desenvolvida por CHEN et al (2002a), foi desenvolvida com o objetivo de estender as *tags* gerando fragmentos 3' de cDNA constituídos por algumas centenas de pares de bases. Devido à similaridade existente entre as metodologias de SAGE e MPSS, foi possível utilizar a técnica de GLGI, com algumas modificações, para a extensão de *tags* geradas por meio de MPSS. As adaptações feitas no protocolo original da técnica de GLGI foram desenvolvidas como parte deste trabalho e em colaboração com o grupo do Dr. San Ming Wang, na Universidade de Chicago. O protocolo adaptado foi nomeado GLGI-MPSS.

1.5 GLGI

O protocolo do GLGI foi desenvolvido para converter uma seqüência curta de nucleotídeos correspondente a uma *tag* de SAGE em um fragmento 3' de cDNA com algumas centenas de pares de bases (CHEN et al. 2000). Por meio dessa conversão e com base na seqüência do fragmento 3' de cDNA, é possível caracterizar de forma mais precisa e específica o transcrito representado pela *tag* de SAGE. A técnica de GLGI também pode ser utilizada com sucesso na caracterização de transcritos diferencialmente expressos, nos casos em que uma única *tag* de SAGE está associada a vários transcritos.

Um trabalho realizado pelo mesmo grupo que desenvolveu a técnica do GLGI, avaliou o impacto causado pelo tamanho das seqüências das *tags* de SAGE na identificação de transcritos. Os resultados desse trabalho revelaram que as *tags* de SAGE, em geral, não possuem um tamanho adequado para permitir a ocorrência de um alinhamento específico entre a *tag* e um transcrito. Entretanto, quando essas seqüências curtas são estendidas a um fragmento 3' de cDNA, através do GLGI, os alinhamentos entre as *tags* e os genes a partir dos quais essas *tags* foram geradas ocorrem de forma bastante específica (LEE et al. 2002).

Em um outro estudo, semelhante ao presente trabalho, esse mesmo grupo, avaliou o perfil de expressão gênica em células humanas CD15+. Para tanto, a técnica de SAGE foi utilizada, e a informação quantitativa a cerca da expressão de 37.519 *tags* de SAGE foi gerada. Desse total de *tags*, 34% apresentaram alinhamento com mais de um transcrito conhecido, e 47% das *tags* não apresentaram nenhum alinhamento sendo, portanto, consideradas possíveis novos transcritos. A técnica de

GLGI foi, então, utilizada para estender as *tags* que apresentaram alinhamentos múltiplos e que apresentaram um elevado número de cópias, em um fragmento 3' de cDNA. A combinação das técnicas de SAGE e GLGI permitiu a identificação de um conjunto de genes importantes na diferenciação mielóide que apresentaram diferentes níveis de expressão e também formas alternativas de *splicing* (LEE et al. 2001). Esses trabalhos, revelaram o potencial da técnica de GLGI na identificação de transcritos em larga escala.

No protocolo de GLGI, o cDNA é sintetizado utilizando um iniciador oligo d(T) biotilado, importante para a recuperação do fragmento 3' de cDNA após a digestão do mesmo com a enzima de restrição utilizada para a produção da *tag*. Em seguida, são ligados ao fragmento 3' de cDNA adaptadores, que são importantes para a amplificação por PCR, em larga escala, do fragmento acima mencionado. Por fim, a extensão do fragmento 3' de cDNA a partir da *tag* é realizada utilizando o cDNA amplificado, o iniciador *sense* correspondente à *tag* e um iniciador *antisense* incorporado à seqüência do oligo dT utilizado na síntese do cDNA dupla fita (CHEN et al. 2002a e b).

Objetivos

2 OBJETIVOS

2.1 OBJETIVO GERAL

Este projeto teve como objetivo geral, identificar e caracterizar novos transcritos humanos regulados por c-erbB2 e que apresentassem expressão diferencial entre as linhagens celulares HB4a e HB4a-C5.2.

2.2 OBJETIVOS ESPECÍFICOS

- Selecionar *tags* de MPSS diferencialmente expressas entre as linhagens HB4a e HB4a-C5.2 que não foram associadas a nenhum transcrito conhecido (*tags* órfãs);
- Utilizar a metodologia GLGI para estender as *tags* órfãs diferencialmente expressas;
- Utilizar a informação gerada pelo GLGI para validar experimentalmente, por meio de PCR em Tempo Real, a expressão diferencial dos novos transcritos nas linhagens HB4a e HB4a-C5.2;
- Gerar a seqüência completa dos novos transcritos diferencialmente expressos nas duas linhagens celulares;

- Analisar, por meio de PCR em Tempo Real, a expressão dos transcritos diferencialmente expressos em linhagens tumorais e em amostras de tumores de mama c-erbB2 positivas e c-erbB2 negativas.

Materiais e Métodos

3 MATERIAS E MÉTODOS

3.1 LINHAGENS CELULARES E AMOSTRAS TUMORAIS

A linhagem HB4a foi obtida a partir da imortalização de células luminais de mama com um vetor retroviral que expressa o antígeno T do vírus SV40 (STAMPS et al. 1994). A linhagem HB4a-C5.2 foi gerada por meio da transfecção de células HB4a com o plasmídeo Pj5E c-erbB2. Esse plasmídeo contém toda a região codificadora do gene c-erbB2 humano sob o controle da região promotora de Mouse Mammary Tumor Virus (HARRIS et al. 1999). Essas linhagens celulares foram gentilmente cedidas pelo Dr. Michael O'Hare (Instituto Ludwig de Londres) e as linhagens celulares SKBR3, MCF7, PMC42, MDA-MB-435 foram cedidas pelo Dr. Robert Whitehead (Vanderbilt University Medical Center). Todas as linhagens foram cultivadas até a aquisição de massa suficiente, correspondendo a 3×10^6 células, para a obtenção do RNA. Os meios de cultura utilizados no cultivo de cada linhagem celular estão listados na Tabela 2.

Tabela 2 - Meios de cultura utilizados no cultivo celular das linhagens celulares de mama cedidas pelo Dr. Robert Whitehead e pelo Dr. Michael O'Hare.

Linhagem Celular	Meio de cultura
HB4a	RPMI – 10% soro fetal bovino e 5µg/ml de insulina e 5µg/ml de hidrocortisona
SKBR3	RPMI – 10% soro fetal bovino
MCF7	McCoy's – 10% soro fetal bovino
PMC42	RPMI – 15% soro fetal bovino e 5µg/ml de insulina
MDA-MB-435	RPMI – 10% soro fetal bovino

As linhagens celulares SKBR7, BT20, MDA-MB-453 e MDA-MB-468, não foram cultivadas em nosso laboratório. O RNA total dessas linhagens foi cedido pelo Dr. Michael Ohare (Instituto Ludwig de Londres).

Todas as amostras tumorais utilizadas neste estudo foram obtidas junto ao Banco de Tumores do Hospital A.C.Camargo/Instituto Ludwig de Pesquisas sobre Câncer e do Hospital do Câncer Alfredo Abrão. As amostras foram obtidas mediante consentimento informado dos pacientes e aprovação do comitê de ética da Instituição. Para a realização deste estudo, foram selecionadas 49 amostras de carcinomas ductais invasivos de mama. Dentre as 49 amostras tumorais, 27 amostras apresentaram super-expressão da proteína c-erbB2. Os dados referentes a positividade das amostras para a proteína c-erbB2 foram obtidos por meio de imunohistoquímica realizada no Departamento de Patologia da Instituição e extraídos diretamente dos prontuários dos pacientes. Todas as amostras tiveram os níveis de mRNA de c-erbB2 quantificados por PCR em Tempo Real e foram classificadas como c-erbB2 negativas quando o nível de expressão do oncogene foi menor ou igual a 3 e as amostras foram consideradas c-erbB2 positivas quando o nível de expressão

do oncogene foi maior ou igual a 20 em relação a um *pool* de amostras de mama normal.

3.2 EXTRAÇÃO DE RNA TOTAL E CONTROLE DE CONTAMINAÇÃO COM DNA GENÔMICO

A extração de RNA foi feita a partir das linhagens celulares e dos tumores acima descritos foi realizada por meio da técnica de Sedimentação em Cloreto de Césio. As células foram homogeneizadas em 9 ml de solução de lise (4M de isotiocianato de guanidina / 25 mM citrato – pH 7.0 / 0.1M β -mercaptoetanol). O lisado foi, em seguida, aplicado sobre 4ml de um gradiente de cloreto de césio (5.7M CsCl e 25mM de NaAc) e centrifugado a 29.000 rpm durante 17 horas a 20°C em ultracentrífuga Beckman, utilizando o rotor SW40Ti. Após a centrifugação, o RNA formou um precipitado no fundo do tubo, que foi solubilizado em 50 a 200 μ l de água biodestilada tratada com DEPC (Di-etil pirocarbonato).

A integridade dos RNAs foi visualizada aplicando-se 1 μ g de RNA total em gel de agarose 1%. Antes de ser aplicado no gel, o RNA foi desnaturado a 65°C durante 5 minutos, sendo mantido em condição desnaturante em tampão de amostra contendo uréia (7M uréia e 30% glicerol). Todos os procedimentos foram realizados com materiais próprios para uso exclusivo de RNA. A coloração do gel foi feita utilizando brometo de etídio e o mesmo foi visualizado em luz ultravioleta. Foram considerados íntegros os RNAs que apresentaram as bandas correspondentes aos

RNAs ribossômicos 28S e 18S evidentes, sendo a primeira duas vezes mais intensa que a segunda.

A quantificação do RNA foi feita pela leitura em espectrofotômetro com comprimento de onda equivalente a 260 nm, considerando-se que 1DO260 nm equivale a 40µg/ml de RNA. A relação entre as leituras realizadas a 260 e 280 nm foi utilizada como parâmetro na estimativa do grau de contaminação do RNA por proteínas.

Para se ter a garantia de que o RNA extraído não estava contaminado com DNA genômico, 100ng de RNA total foi amplificado com iniciadores específicos para regiões intrônicas do gene MLH1 (Forward - 5'TGG TGT CTC TAG TTC TGG3' e Reverse - 5'CAT TGT TGT AGT AGC TCT GC3'). A amplificação de um fragmento de aproximadamente 250 pares de bases indica a existência de contaminação da amostra de RNA por DNA genômico.

A PCR para amplificação do fragmento genômico correspondente ao gene MLH1 teve 35 ciclos com uma temperatura de desnaturação de 94°C por 1 minuto, uma temperatura de anelamento de 55°C por 1 minuto e uma temperatura de extensão de 72°C por 45 segundos. Como controle positivo da reação foram utilizadas 100ng de DNA de cérebro normal.

Além do teste de hMLH1, também foi realizado o teste de contaminação, utilizando iniciadores específicos para os fragmentos em estudo e 100ng de RNA total tratado. Como controle positivo da reação utilizou-se 1µl cDNA da linhagem HB4a-C5.2.

3.3 TRATAMENTO DO RNA TOTAL COM DNASE I

O RNA total das linhagens e amostras tumorais foi tratado com DNase livre de RNase (1U/ μ l, Promega), utilizando 1 unidade da enzima para cada 2,5 μ g de RNA total de linhagem celular e 1 unidade da enzima para cada 1,5 μ g de RNA total de tumor. As reações, para tratar 50 μ g de RNA total, foram realizadas em um volume final de 300 μ l contendo 1X tampão da enzima DNaseI (GibcoBRL), 4U de inibidor de Rnase (Invitrogen), 1mM de DTT (Gibco), 50 μ g de RNA total e 33U da enzima DnaseI (Gibco) e água q.s.p. 300 μ l. As reações foram mantidas a 37°C durante 90 minutos. Em seguida, foram adicionados às amostras 1 vez o volume de fenol/clorofórmio e as mesmas foram centrifugadas a 4°C durante 10 minutos a 12000rpm. Após o término da centrifugação, o sobrenadante foi recolhido e transferido para um tubo eppendorf novo. Ao sobrenadante foi adicionado 1 vez o volume clorofórmio e, então, a solução foi homogeneizada com o auxílio do vórtex. Novamente as amostras foram centrifugadas a 4°C durante 10 minutos a 12.000rpm e, após a centrifugação, o sobrenadante foi transferido para um tubo eppendorf novo. Com o objetivo de precipitar o RNA tratado, foram adicionados ao sobrenadante 1/10 do volume final de acetato de sódio 3M pH5.2 e 3 volumes de etanol 100%. Os RNAs foram mantidos a -20°C durante toda a noite.

As amostras foram centrifugadas a 4°C durante 15 minutos a 12.000rpm. O sobrenadante foi retirado com pipeta Pasteur e desprezado. Foram adicionados ao pellet de RNA 1ml de etanol 70% e, novamente, as amostras foram centrifugadas a 4°C durante 15 minutos a 12000rpm. O sobrenadante foi desprezado por inversão do

tubo eppendorf e mantido invertido sobre um papel limpo para secar o pellet de RNA. Depois de seco, o pellet foi solubilizado em 20-30µl de água DEPC. A quantificação foi feita pela leitura de uma alíquota da amostra no espectrofotômetro a 260 e 280 nm conforme descrito anteriormente no item 3.2 dos materiais e métodos.

3.4 GLGI

3.4.1 Extração de RNA total e isolamento de mRNA

A extração do RNA total foi realizada como descrita anteriormente no item 3.2. Para a purificação do mRNA, foi utilizado um sistema de purificação magnética disponibilizado pela Dynal (Dynabeads Oligo (dT)₂₅ e Dynal MPC). As partículas magnéticas acopladas ao oligo d(T) foram inicialmente lavadas em 1ml de 2X tampão de ligação (20mM Tris-HCL pH 7.5, 1M LiCl, 2mM EDTA). Para tanto, o tubo contendo as partículas magnéticas foi colocado na MPC (*Magnet Particle Concentrator*) e o sobrenadante foi removido. Em seguida, as partículas magnéticas foram ressuspensas em 200µl de 2X tampão de ligação e misturadas as 180µg de mRNA, previamente aquecido a 65°C durante 2 minutos a fim de que as estruturas secundárias fossem rompidas. Essa mistura foi mantida a temperatura ambiente durante 5 minutos de forma a permitir a ligação do mRNA ao oligo d(T) acoplado às partículas magnéticas. Em seguida, o tubo foi colocado novamente na MPC durante 1 minuto. O sobrenadante foi removido, e o mRNA imobilizado nas partículas foi lavado, 3 vezes, com 800µl de tampão de lavagem (10mM Tris-HCl pH8.0, 150mM LiCl, 1mM EDTA). Após a lavagem do mRNA, as partículas foram ressuspensas em 15µl de água DEPC e, em seguida, foram aquecidas a 65°C durante 2 minutos. O

tubo foi colocado na MPC, e o sobrenadante contendo o mRNA foi removido e transferido para um tubo novo. As partículas foram novamente ressuspensas em 15µl de água DEPC e, mais uma vez, foram aquecidas a 65°C durante 2 minutos. O tubo foi colocado na MPC, e o sobrenadante foi recolhido e combinado aos 15µl recolhidos na etapa anterior. Com o objetivo de avaliar a eficiência de purificação do mRNA, foram retirados 2µl de cada amostra, HB4a-C5.2 e HB4a, que foram analisados em um gel de agarose 1% corado com brometo de etídeo.

3.4.2 Síntese da primeira e segunda fita de cDNA

A síntese da primeira e da segunda fita de cDNA foram realizadas segundo especificações do fornecedor da enzima MMLV RT – Moloney Murine Leukemia Vírus Reverse Transcriptase (Gibco cDNA synthesis kit 18267-013).

Para a síntese da primeira fita de cDNA foi utilizada uma mistura de oligos dT biotinizados que possuíam em suas extremidades 5' uma sequência adaptadora adicional destacada em negrito:

5'biotin-**ACT ATC TAG AGC GGC CGC**-T16-R onde R=A/G

5'biotin-**ACT ATC TAG AGC GGC CGC**-T16-C-V onde V=A/G/C

As concentrações de ambos os oligos foram ajustadas para 0,5µg/µl e os volumes equivalentes dos dois oligos foram misturados. A primeira fita de cDNA foi sintetizada utilizando 2µl da mistura de oligos dT biotinizados, 28µl de mRNA, 1X tampão de obtenção da primeira fita, 0,5mM dNTPs, 40U de RNasin, 10mM DTT, 500U da enzima MMLV RT e água q.s.p 50µl. A reação de síntese foi mantida a 37°C durante 40 minutos. Após esse intervalo, a reação foi aquecida a 65°C durante 2 minutos. Em seguida, a reação foi mantida a 37°C durante 30 minutos, sendo

adicionados mais 2 μ l da enzima MMLV RT. Essa etapa foi repetida duas vezes para a adição da enzima. Para avaliar a qualidade de síntese da primeira fita de cDNA, foram retirados 2 μ l da reação que foram analisados em gel de agarose 1% corado com brometo de etídeo.

Para a síntese da segunda fita de cDNA, foram adicionados, aos 48 μ l da reação da primeira fita, 0,25mM de dNTPs , 1X tampão de obtenção da segunda fita, 100U de *E. coli* DNA polimerase I (GibcoBRL), 12,5U de *E. coli* DNA ligase (GibcoBRL) e água q.s.p 291 μ l. A reação de síntese foi mantida a 16°C durante 2 horas. Em seguida, foram adicionados 10U de RNaseH (Promega) e a reação foi mantida a 37°C durante 30 minutos. A amostra foi, então, tratada com 400 μ l de fenol/clorofórmio. A fase aquosa foi recuperada e foi dividida em dois tubos eppendorfs novos. Em cada um dos tubos, foram adicionados 60 μ g de glicogênio (Roche Diagnostics; Indianápolis, IN), 150 μ l de 7,5M NH₄OAC e 1ml de 100% etanol gelado. A reação foi homogeneizada por inversão dos tubos e, em seguida, os mesmos foram centrifugados a 14.000 rpm (Eppendorf 5417 R – rotor F 45-30-11) durante 15 minutos a temperatura ambiente. Após a centrifugação, os precipitados foram novamente combinados em um mesmo tubo e foram lavados com 1ml de 70% etanol. Os precipitados secaram a temperatura ambiente e, depois, foram dissolvidos em 82 μ l de LoTE (3mM Tris-HCl pH 7.5, 0.2mM de EDTA pH 7.5). Para avaliar a eficiência de síntese da segunda fita de cDNA foram retirados 2 μ l da amostra para análise em gel de agarose 1% corado com brometo de etídeo.

3.4.3 Digestão da dupla fita de cDNA

Após a síntese da dupla fita de cDNA, a mesma foi digerida com a enzima de restrição *DpnII* (New England Biolabs). A digestão da dupla fita foi realizada a partir de 80µl de amostra (cDNA dupla fita), 1X tampão para a enzima *DpnII*, 150U da enzima *DpnII* (Fermentas) e água q.s.p 180µl. A amostra foi mantida a 37°C durante 3 horas. Para avaliar a eficiência de digestão do cDNA dupla fita, foi retirada uma alíquota de 4µl da reação para análise em gel de agarose 1% corado com brometo de etídeo.

3.4.4 Recuperação do fragmento 3' de cDNA

Após a digestão do cDNA dupla fita, os fragmentos 3' de cDNA foram recuperados utilizando um sistema de purificação magnética disponibilizado pela *Dynal* (*Dynal MPC e Dynabeads M-280 Streptavidin*). As partículas magnéticas acopladas a streptavidina foram inicialmente lavadas com 300µl de 2X tampão de ligação/lavagem (10mM Tris-HCl pH7.5, 1mM EDTA e 2M de NaCl). Para tanto, o tubo contendo as partículas foi colocado na MPC durante 1 minuto e o sobrenadante foi removido. As partículas foram ressuspensas em 180µl de 2X tampão de ligação/lavagem. Após a lavagem das partículas, as mesmas foram misturadas aos 180µl de cDNA digerido com a enzima *DpnII* para favorecer a interação entre a streptavidina acoplada à partícula magnética e à biotina ligada à extremidade 5' do oligo d(T). O tubo contendo o cDNA digerido e as partículas foi mantido em constante agitação, durante 60 minutos a temperatura ambiente. Ao final dos 60 minutos, o tubo foi colocado na MPC durante 1 minuto e o sobrenadante foi removido. Em seguida, as partículas foram lavadas em 1ml de 1X tampão de

ligação/lavagem (5mM Tris-HCl pH7.5, 0.5mM EDTA e 1M NaCl) por 3 vezes e com 1ml de TE (10mM Tris e 1mM EDTA). Após as lavagens das partículas, o tubo foi colocado na MPC e o sobrenadante foi removido e descartado. O tubo contendo as partículas ligadas ao fragmento 3' de cDNA foi mantido no gelo até o momento do uso.

3.4.5 Ligação dos adaptadores ao fragmento 3' de cDNA

Para a posterior amplificação do fragmento 3' de cDNA foram ligados à extremidade 5' do fragmento de interesse uma seqüência adaptadora. A importância da ligação dessa seqüência ao fragmento de cDNA que se deseja amplificar, se deve ao fato de que existe uma complementaridade, de 20 bases, entre a seqüência desses adaptadores e a seqüência de um dos iniciadores (SAGE P1) utilizado na reação de amplificação do fragmento 3' de cDNA. A seqüência adaptadora foi gerada a partir do anelamento de dois iniciadores complementares. Para tanto, as concentrações de ambos os iniciadores foram ajustadas para 100pmoles/ μ l através da utilização de Low TE e 24,84 μ l do iniciador 1A foi misturado a 25,5 μ l do iniciador 1B. Para a reação de anelamento dos iniciadores foram adicionados aos 50 μ l da mistura acima descrita, 50 μ l de 10X tampão 2 (New England Biolabs) e 400 μ l de Low TE pH7,5. A reação foi aquecida a 95°C durante 2 minutos e, em seguida, foi mantida a temperatura ambiente por aproximadamente 2 horas.

As partículas ligadas ao fragmento 3' de cDNA (descritas no final da etapa 3.4.4) foram lavadas por 3 vezes com 250 μ l de 1X tampão da ligase T4. Em seguida, o tubo contendo as partículas foi colocado durante 1 minuto na MPC e o sobrenadante foi removido e descartado. Foi adicionado ao tubo contendo as

partículas 19µl de água, 1X tampão da ligase T4 e 1,4µl da sequência adaptadora. Essa reação foi homogeneizada e aquecida a 50°C durante 2 minutos e, posteriormente, mantida a temperatura ambiente durante 15 minutos. Em seguida foram adicionados 18U de T4 DNA ligase (GibcoBRL). A reação de ligação foi homogeneizada e mantida a 16°C por 24 horas

As seqüências dos iniciadores, utilizadas para a produção da seqüência adaptadora, estão listadas abaixo e, em vermelho, estão destacadas as bases complementares entre as duas seqüências.

Primer 1A:

5' TTTGGA TTTGCTGGTGCAGTACA ACTAGGCTTAATAGGGAGATC 3'

Primer 1B:

5'-p TCCCTATTAAGCCTAGTTGTACTGCACCAGCAAATCC [amino-modificadoC7]-3'

3.4.6 Padronização da amplificação do fragmento 3' de cDNA

Para a amplificação do fragmento 3' de cDNA ligado a seqüência adaptadora, foram utilizados dois iniciadores: SAGE P1 (cuja seqüência é complementar a seqüência adaptadora ligada ao fragmento de cDNA) e SAGE-*antisense* (complementar a seqüência adicional do iniciador oligo d(T)). O principal objetivo dessa padronização é determinar a quantidade ideal de cDNA, e também o número de ciclos a serem utilizados na reação de amplificação do fragmento 3' de cDNA em larga escala. Antes da amplificação, as partículas (descritas no item 3.4.5) foram lavadas por 3 vezes com 500µl de 1X tampão de ligação/lavagem. Ao final de cada

lavagem, o tubo contendo as partículas era colocado na MPC e o sobrenadante era removido e descartado. Após a última lavagem as partículas foram ressuspensas em 100µl de Low TE pH7,5 e mantidas no gelo até o momento de uso. As seqüências dos dois iniciadores utilizados na reação de amplificação estão listadas abaixo:

SAGE-P1: 5'-GGATTTGCTGGTGCAGTACA-3'

SAGE-*antisense*: 5'-ACTATCTAGAGCGGCCGCTT-3'

Foram utilizados, na reação de amplificação, 1X tampão da enzima, 2mM de MgCl₂, 0,1mM dNTPs, 100ng do iniciador SAGE-P1, 100ng do iniciador SAGE-*antisense*, 2,5U da enzima Taq Platinum (Invitrogen), 1µl do cDNA 3' ligado ao adaptador para um volume final de 50µl. A reação de padronização foi realizada sob as seguintes condições de amplificação: 2 minutos de denaturação inicial a 95°C, seguida de 16/18/20, ou 25 ciclos de denaturação a 94°C durante 30 segundos, anelamento dos iniciadores a 55°C durante 30 segundos e extensão a 72°C durante 35 segundos. As reações foram mantidas a 72°C durante 7 minutos para a extensão da moléculas durante o último ciclo. Em cada um dos ciclos mencionados acima, foram retiradas alíquotas de 5µl da reação para que o número ideal de ciclos para a amplificação fosse determinado. As amostras foram analisadas em gel de agarose 1% corado com brometo de etídeo.

3.4.7 Amplificação dos fragmentos 3' de cDNA em larga escala

Após as condições utilizadas na reação de amplificação do fragmento 3' de cDNA terem sido padronizadas, a reação de amplificação foi realizada em larga escala. Para tanto, foram realizadas 48 reações de amplificação do fragmento 3' de cDNA para cada uma das linhagens utilizadas (HB4a-C5.2 e HB4a). Foram

utilizados, na reação, 1X tampão da enzima, 2mM MgCl₂, 10mM dNTPs, 60ng de SAGE-P1, 60ng de SAGE-*antisense*, 2,5U da enzima Taq Platinum (Invitrogen), 1,5µl do cDNA 3' ligado ao adaptador para um volume final de 50µl. A reação foi realizada sob as seguintes condições de amplificação: 2 minutos de desnaturação inicial a 95°C seguida de 20 ciclos de desnaturação a 94°C durante 30 segundos, anelamento dos iniciadores a 55°C durante 30 segundos e extensão a 72°C durante 35 segundos. As reações foram mantidas a 72°C durante 7 minutos. Todo o conteúdo das 48 reações foi recolhido e dividido em 4 tubos eppendorfs de 1,5ml contendo aproximadamente 600µl de amostra em cada tubo. Em seguida, foram adicionados 600µl de fenol/clorofórmio/álcool isoamílico em cada tubo. Os tubos foram agitados vigorosamente e em seguida foram centrifugados a 14.000 rpm (Eppendorf 5417 R – rotor F 45-30-11) durante 15 minutos a temperatura ambiente. Para a precipitação do material amplificado, foram adicionados 50µg de glicogênio (Roche Diagnostics; Indianápolis, IN), 133µl de 7,5M de NH₄OAc e 1ml de 100% etanol em cada tubo. Os tubos foram mantidos a -80°C durante 30 minutos e, em seguida, foram centrifugados a 14.000 rpm (Eppendorf 5417 R – rotor F 45-30-11) durante 15 minutos a temperatura ambiente. Por fim, as amostras foram lavadas com 1 ml de 70% etanol. Após a lavagem, os precipitados secaram a temperatura ambiente e foram dissolvidos em 400µl de Low TE pH7,5.

3.4.8 Extensão do fragmento 3' de cDNA a partir da tag de MPSS

Nessa reação foram utilizados os iniciadores *sense* específicos para as *tags* previamente selecionadas, o iniciador SAGE-*antisense* (5'-ACTATCTAGAGCGGCCGCTT-3'), que é complementar à sequência do oligo d(T)

e 1µl de cDNA 3' amplificado, como descrito no item 3.4.6. Além das 83 *tags* órfãs selecionadas para serem estendidas, 2 *tags* associadas a transcritos conhecidos também foram selecionadas e tiveram suas seqüências estendidas como um controle para a reação de amplificação. As seqüências dos iniciadores específicos, correspondentes tanto às *tags* associadas quanto às *tags* órfãs, estão listadas nas tabelas 3 e 4 respectivamente. É importante destacar que cada iniciador *sense* específico possui em sua seqüência, além dos nucleotídeos que compõe cada uma das *tags*, 4 nucleotídeos (GATC) que correspondem ao sítio de corte da enzima de restrição *DpnII* e 6 nucleotídeos (CAGGGA) que correspondem a algumas das bases da seqüência adaptadora ligada ao fragmento 3' de cDNA (descrito no item 3.4.5). A adição dessas 6 bases que correspondem aos adaptadores é importante para que haja um aumento na especificidade de anelamento do iniciador *sense* ao seu transcrito correspondente.

Tabela 3 – *Tags* associadas a transcritos conhecidos. Sequências dos iniciadores específicos, correspondentes às *tags* associadas, número de acesso no GenBank e tamanho esperado do fragmento GLGI-MPSS.

	Sequência da <i>tag</i>	Número de acesso no GenBank	Transcrito	Tamanho do fragmento
CT1	CAGGGA GATC CGA GAA CGC TCT A	NM_000900	Matrix Gla protein (MGP)	383bp
CT2	CAGGGA GATC TTG CTG GAT GTG A	NM_005557	Keratina 16	407bp

Tabela 4 – Sequências dos iniciadores específicos, correspondentes às *tags* órfãs.

TAGS ÓRFAS	SEQÜÊNCIA
1	5'-CAGGGA GATC CTCCAGTTGCTGC -3'
2	5'-CAGGGA GATC AAGTAATCAAGTA -3'
3	5'-CAGGGA GATC TTGGTGTTGCAA -3'
4	5'-CAGGGA GATC TAAGAGACTCAGA -3'
5	5'-CAGGGA GATC CGATGCAGGGGAG -3'
6	5'-CAGGGA GATC TTTCTAGGGATGT -3'
7	5'-CAGGGA GATC TGTCCACTTCTGG -3'
8	5'-CAGGGA GATC TAAGAAGAAAATA -3'
9	5'-CAGGGA GATC GTTCTTTCTTTCT -3'
10	5'-CAGGGA GATC ATTGGGAGCTTTT -3'
11	5'-CAGGGA GATC TCGCCAACAGACA -3'
12	5'-CAGGGA GATC TCTCAGTAAACTT -3'
13	5'-CAGGGA GATC CGGGATTCTCAG -3'
14	5'-CAGGGA GATC GGGGGAGGGTGGG -3'
15	5'-CAGGGA GATC AAAGCTTGAAAAA -3'
16	5'-CAGGGA GATC CTGTAGTTCCTA -3'
17	5'-CAGGGA GATC TTTTGCTAATTA -3'
18	5'-CAGGGA GATC GGCGGCGGGCGGG -3'
19	5'-CAGGGA GATC AAAAACTCAAAAA -3'
20	5'-CAGGGA GATC TTCCAAAACAGAG -3'
21	5'-CAGGGA GATC CAAGTGATGATAC -3'
22	5'-CAGGGA GATC CCACTCTTCTTCC -3'
23	5'-CAGGGA GATC TGTACCAGGTAGG -3'
24	5'-CAGGGA GATC AGTTAAGGAGCTA -3'
25	5'-CAGGGA GATC CAAGGGCACAAAGG -3'
26	5'-CAGGGA GATC GCTCGCTCGTCCG -3'
27	5'-CAGGGA GATC CATGAAGTCAGTC -3'
28	5'-CAGGGA GATC ATAGGGACAGCGG -3'
29	5'-CAGGGA GATC ATTGTCAACCAGG -3'
30	5'-CAGGGA GATC ATGAGGGAGCGAG -3'
31	5'-CAGGGA GATC TTCAGCCAGCCAG -3'
32	5'-CAGGGA GATC TAAAAGGCTGGCT -3'
33	5'-CAGGGA GATC AAGGATGCAAGCA -3'
34	5'-CAGGGA GATC AGGAGGAGGAGGA -3'
35	5'-CAGGGA GATC TTTATCAACTGGC -3'
36	5'-CAGGGA GATC AGAGTTGTAGCCA -3'

37	5'-CAGGGA GATC TTTTCACAGGGCA-3'
38	5'-CAGGGA GATC TAAAACTCACT-3'
39	5'-CAGGGA GATC TAAATCAAGCTGC-3'
40	5'-CAGGGA GATC GTGACATACAAAT-3'
41	5'-CAGGGA GATC AATGGTGGAGAGA-3'
42	5'-CAGGGA GATC AGATGCTTGAAG-3'
43	5'-CAGGGA GATC CAGGCACATTGA-3'
44	5'-CAGGGA GATC CAGACATGCCTGA-3'
45	5'-CAGGGA GATC TGGTTGTCAAGTTG-3'
46	5'-CAGGGA GATC CAAAAATTCTTTA-3'
47	5'-CAGGGA GATC TAGTAAAAAGAGA-3'
48	5'-CAGGGA GATC TTGCTACATAGTT-3'
49	5'-CAGGGA GATC CACTGGAAGTCCA-3'
50	5'-CAGGGA GATC CGTGACAAACAAA-3'
51	5'-CAGGGA GATC CGGGACTTCAGGT-3'
52	5'-CAGGGA GATC CACATTTGTAAAG-3'
53	5'-CAGGGA GATC AACGACAAAAAGA-3'
54	5'-CAGGGA GATC TAAGTTTCCGGAG-3'
55	5'-CAGGGA GATC CTATACTCATCGC-3'
56	5'-CAGGGA GATC TGGCTTGGGAACA-3'
57	5'-CAGGGA GATC CTAGAATCGGTGA-3'
58	5'-CAGGGA GATC TCTGCGTTTTTTT-3'
59	5'-CAGGGA GATC CAAAAGTTCAGTT-3'
60	5'-CAGGGA GATC TCTGGTTTGAAG-3'
61	5'-CAGGGA GATC TCCTTGAATGAAT-3'
62	5'-CAGGGA GATC TGCCAAGAAAAGC-3'
63	5'-CAGGGA GATC TGGCAAGGTCACC-3'
64	5'-CAGGGA GATC CGACCAGCCAGAA-3'
65	5'-CAGGGA GATC AGAGTTCTGTCAA-3'
66	5'-CAGGGA GATC TGCAGTCTAAGTG-3'
67	5'-CAGGGA GATC AATGCGCCTACTG-3'
68	5'-CAGGGA GATC GGGAAGGCAGATT-3'
69	5'-CAGGGA GATC AGTTCTTGACAGTC-3'
70	5'-CAGGGA GATC TGACAGTGGTCAC-3'
71	5'-CAGGGA GATC TCTGCTCGCTGCG-3'
72	5'-CAGGGA GATC CTCTGGACCATGA-3'
73	5'-CAGGGA GATC AGCGCTCAAAAA-3'
74	5'-CAGGGA GATC CGTATCGGTTCCC-3'
75	5'-CAGGGA GATC CGCCTGAAGCAGA-3'
76	5'-CAGGGA GATC ATTCTTCAATTAG-3'
77	5'-CAGGGA GATC GCCGTTCTGGACT-3'
78	5'-CAGGGA GATC TTGCCAGCTTGCT-3'
79	5'-CAGGGA GATC ACGCACAGCTCTA-3'
80	5'-CAGGGA GATC TGGCCATTTGTTT-3'
81	5'-CAGGGA GATC ACCAGGGTCAGTT-3'
82	5'-CAGGGA GATC AAAAAAGGCAGAA-3'
83	5'-CAGGGA GATC CTCTGCAGTCCAG-3'

Foram utilizados na reação de amplificação, 1X tampão da enzima, 1,2mM de MgCl₂, 0,08mM dNTPs, 50ng de iniciador *sense* específico, 50ng de iniciador

SAGE-*antisense*, 1,5U da enzima Taq Platinum (Invitrogen), 1µl do cDNA 3' amplificado para um volume final de 30µl. A reação foi realizada sob as seguintes condições de amplificação: 2 minutos de denaturação inicial a 94°C, seguida de 30 ciclos de denaturação a 94°C durante 30 segundos, anelamento dos iniciadores a 64°C durante 30 segundos e extensão a 72°C durante 35 segundos. As reações foram mantidas a 72°C durante 7 minutos para o último ciclo. Foram checados em gel de agarose 1% corado com brometo de etídeo 3µl de cada amostra.

3.4.9 Purificação do produto obtido na reação de extensão do fragmento 3'de cDNA

Foram adicionados a cada amostra 10µg de glicogênio (Roche Diagnostics; Indianápolis, IN), 15µl de 7,5M de NH₄OAc e 85µl de 100% etanol. Em seguida as amostras foram mantidas a temperatura ambiente durante 15 minutos e foram centrifugadas a 4.000 rpm durante 35 minutos (centrífuga SORVALL RC5C plus; rotor: SH3000). O sobrenadante foi removido, e cada uma das amostras foi lavada com 150µl de 70% etanol. As amostras foram centrifugadas a 4.000 rpm durante 15 minutos, o sobrenadante foi removido e cada precipitado, após secar naturalmente, foi dissolvido em 5µl de água deionizada. Após a purificação do produto obtido na reação de extensão do fragmento 3' de cDNA, o mesmo pode ser clonado e seqüenciado como descrito a seguir.

3.5 CLONAGEM E TRANSFORMAÇÃO BACTERIANA

Os fragmentos de cDNA estendidos a partir da *tag* específica foram clonados utilizando o pGEM[®] - T Easy Vector (Promega) segundo especificações do fornecedor. A reação de ligação foi preparada em um volume final de 10 μ l, contendo 50ng do vetor pGEM[®] - T Easy , 1X tampão da enzima, 5U de T4 DNA ligase e 2 μ l do produto purificado a partir da reação de GLGI. A ligação foi mantida a 4°C por 16 horas. A transformação bacteriana foi realizada utilizando 2 μ l da ligação e 100 μ l de bactérias DH10B eletrocompetentes. Após a ligação do produto à bactéria, foi dado o choque elétrico (resistência de 200 ohms, capacitância de 25 microfaradays e voltagem de 1.8 Kv em eletroporador Gene Pulser - BioRad*). Após o choque, as bactérias foram incubadas durante 1 hora em 1mL de meio LB sem antibiótico (1% de triptona; 0,5% de extrato de levedura; 1% de NaCl; pH 7,5), a 37°C sob agitação constante (200rpm). Em seguida, as bactérias foram sedimentadas por centrifugação, ressuspendidas em 100ml de meio LB e plaqueadas sobre meio LB-ágar contendo 50 μ g/ml de ampicilina, 20 μ g/ml de IPTG (isopropiltio- β -D-galactosídeo) e 4 μ g/ml de X-gal (5-bromo-4-cloro-3indolil- β -D-galactosídeo). As placas foram incubadas a 37°C por 18 horas. Para verificar a presença de inserto e selecionar clones para o seqüenciamento, foi realizada uma seleção de 8 colônias de cada placa por meio de PCR, em que em cada tubo, uma colônia selecionada foi palitada. As reações foram realizadas em um volume final de 20 μ l contendo 1X tampão da enzima Taq DNA polimerase (GibcoBRL), 1,5mM MgCl₂, 0,2mM dNTPs, 1U Taq DNA polimerase (GibcoBRL) e 10pmoles de iniciadores específicos para o plasmídeo (T7: 5'TAA

TAC GAC TCA CTA TAG GG3' e SP6: 5'ATT TAG GTG ACA CTA TAG3'). As condições da reação foram: desnaturação durante 4 minutos a 94°C, seguida de 30 ciclos de desnaturação a 94°C durante 45 segundos anelamento a 55°C durante 45 segundos e extensão a 72°C durante 1 minuto. Ao final da ciclagem, as moléculas foram estendidas durante 7 minutos a 72°C.

3.6 SEQÜENCIAMENTO AUTOMÁTICO

De 300 a 500 ng dos produtos que foram gerados a partir do PCR das colônias selecionadas foram seqüenciados, utilizando o *Kit Big Dye terminator* (*Applied Biosystems*), seguindo as instruções do fornecedor. Os produtos do seqüenciamento foram submetidos à eletroforese por capilaridade no seqüenciador automático de DNA ABI 3100 (*Applied Biosystems- Perkin Elmer*). Os iniciadores utilizados foram T7: 5'TAA TAC GAC TCA CTA TAG GG3' e SP6: 5'ATT TAG GTG ACA CTA TAG3'.

3.7 ANÁLISE DAS SEQÜÊNCIAS DE GLGI-MPSS

Todas as seqüências que não apresentaram as 13 bases correspondentes à seqüência da *tag* de MPSS (excluindo o sítio *DpnII*) foram excluídas das nossas análises. As seqüências selecionadas foram analisadas contra o banco de dados GenBank (NR e dbEST) através da utilização do programa BLAST N. A seqüência gerada a partir de uma *tag* órfã de MPSS foi classificada como novo transcrito se,

após as análises, não houve nenhum alinhamento entre a seqüência em questão e seqüências de mRNA ou ESTs. Um fragmento GLGI-MPSS foi considerado como um transcrito conhecido se houve um alinhamento entre a seqüência analisada e a seqüência de mRNA com mais de 95% de similaridade na mesma orientação, incluindo as 17 bases da seqüência da *tag* de MPSS. Entretanto, se a seqüência alinhada apresentou similaridade apenas com ESTs, os fragmentos GLGI-MPSS foram classificados como derivados a partir de transcritos parcialmente representados por ESTs. Em ambos os casos, moléculas de cDNA completas e ESTs, esses transcritos foram considerados como novos transcritos regulados por c-erbB2. Se o alinhamento observado foi parcial na região correspondente a seqüência de *tag* de MPSS, os fragmentos de GLGI-MPSS foram classificados como inespecíficos ou como *tags* alternativas derivadas da presença de *SNPs*. Fragmentos que apresentaram uma alta similaridade com transcritos conhecidos, porém na orientação oposta, foram classificados como transcritos *antisense*. Finalmente, fragmentos GLGI-MPSS foram considerados como formas de poliadenilação alternativa, ou como seqüências artefatuais derivadas de anelamento interno, se elas alinharam com a parte mediana de moléculas completas de cDNA na orientação 5' - 3'.

3.8 ANÁLISE DO NÍVEL DE EXPRESSÃO POR PCR EM TEMPO REAL

3.8.1 Síntese de cDNA

A síntese da primeira fita de cDNA foi realizada segundo especificações do fornecedor da enzima SuperScript II (Gibco BRL). Uma micrograma de RNA total livre de DNA genômico foi misturado a 25ng de oligo (dT)12-18, 0,5mM de dNTP e o volume foi acertado para 12µl. Esta mistura foi aquecida a 65°C por 5 minutos e rapidamente resfriada em gelo. Em seguida, a reação foi processada em presença de 1X tampão de obtenção da primeira fita, 10mM de DTT, 10U de Transcriptase Reversa (SuperScript II, Gibco BRL) e 2U de RNase OUT (Invitrogen). A reação foi processada a 42°C durante 60 minutos e inativada por meio do aquecimento a 70°C durante 15 minutos.

3.8.2 PCR em Tempo Real

A diferença no nível de expressão dos 13 novos transcritos nas linhagens HB4a e HB4a-C5.2 foi confirmada por meio de PCR em Tempo Real. Nessa reação, a fluorescência emitida a cada ciclo de amplificação é monitorada permitindo, desta forma, uma quantificação precisa do produto amplificado ciclo a ciclo.

A metodologia utilizada para a análise foi o SYBR[®] Green (Applied Biosystems) (MORRISON et al. 1998). Nessa abordagem o SYBR[®] Green liga-se à moléculas de fita dupla emitindo fluorescência. Uma vez que a cada ciclo da reação novas moléculas são formadas, conseqüentemente, o nível de fluorescência emitida

aumenta ao longo da reação. Assim, quanto menor o ciclo em que a fluorescência atinge o limite determinado pelo aparelho, maior é o nível de expressão do gene analisado (MORRISON et al. 1998).

Valores quantitativos foram obtidos, durante a ciclagem, no ponto no qual a amplificação do produto foi detectada, em vez da quantidade do produto acumulado ao final de um determinado número de ciclos. O parâmetro C_T (*Cycle Threshold*) foi definido como uma fração do número de ciclos na qual a fluorescência emitida ultrapassou um padrão estabelecido, sendo os dados de quantificação analisados pelo *software* de análise do aparelho (MORRISON et al. 1998).

Foram desenhados iniciadores específicos para cada transcrito por meio da utilização de um *software*, primer Express 2.0, do equipamento ABI Prism[®] 7700 Sequence Detection System (Applied Biosystems). As seqüências de cada um dos iniciadores estão listadas na Tabela 5. O gene constitutivo utilizado para corrigir variações nas quantidades iniciais de cDNA foi a β -actina (FW 5' GGC ACC CAG CAC AAT GAA G 3' e REV 5' CCG ATC CAC ACG GAG TAC TTG 3'). No caso da amostras tumorais, foram ainda utilizados três outros genes constitutivos: GAPDH (FW 5' ACC CAC TCC TCC ACC TTT GA 3' e REV 5' CTG TTG CTG TAG CCA AAT TCG T 3'); HPRT (FW 5' GAA CGT CTT GCT CGA GAT GTG A 3' e REV 5' TCC AGC AGG TCA GCA AAG AAT 3') e BCRP (FW 5' CCT TCG ACG TCA ATA ACA AGG AT 3' e REV 5' CCT GCG ATG GCG TTC AC 3').

Tabela 5 - Iniciadores para PCR em Tempo Real.

Categoria	Tags específicas	Seqüência dos iniciadores
<i>Up-regulada</i>	6	FW – 5' CCA CAG TAG CAT CAC CAG AAA CCT 3' REV – 5' CTG AAA GCC TGC CAT TGT TGT 3'
	7	FW – 5' TTA AGG CCA TGA GGG AAC GA 3' REV – 5' GCC CGG TGC GAA TAA TGA 3'
	10	FW – 5' CAC TGA TCA TTG GGA GCT TTT G 3' REV – 5' GAG GCC ACA CAT GCT TGG T 3'
	13	FW – 5' TGC CAA ATA GGC AGA CTG GAA 3' REV – 5' GCT CGT TCT TCT CCC TTG GA 3'
	28	FW – 5' CTC GGC TCG CGA AAC TAC A 3' REV – 5' GGG AGC TAC ATA TCA TCC TTG GA3'
	39	FW – 5' CTG CCC TGA GTC GCT TTA GG 3' REV – 5' GCG TGT CAT CCT CAG GAA ATG 3'
	44	FW – 5' GAC AAC CGG AGG CGA TTG 3' REV – 5' TCT CTG CCT CTC CCG CTA AC 3'
<i>Down-regulada</i>	16	FW – 5' GGA GAT CTG AAG CAC AAA CT 3' REV – 5' TAC CCT GTG AGA TTG TAT GG 3'
	52	FW – 5' AGT TCT AAG AAG CCA GCA GT 3' REV – 5' CTG GGG TTG CAT TTC TCT TT 3'
	76	FW – 5' AAA TGT CAT CCG GAA TGT GC 3' REV – 5' TGT TAG CAA GCC CAT TGT TC 3'
	81	FW – 5' CAC ACA GAA CCC CTC ATC ATC TC 3' REV – 5' CAA TCC TTG CCA ACC TGA TGT 3'
	60	FW – 5' TGA AAC CAG CTG TCT TGC AA 3' REV – 5' GTT TTT GTA GCC CCC ATC CA 3'
	63	FW – 5' TGC TGC TGG TGT TGG TGA A 3' REV – 5' TCT CGG GTC TGC CCA TTC T 3'

Legenda: Seqüências dos iniciadores específicos utilizados para avaliar a expressão diferencial de cada um dos transcritos nas linhagens celulares HB4a e HB4a-C5.2, por meio de PCR em Tempo Real.

3.8.3 Padronização das concentrações dos iniciadores e da quantidade de cDNA

As reações foram realizadas no aparelho ABI Prism[®] 7700 Sequence Detection System (Applied Biosystems) em um volume final de 20µl. Cada reação foi realizada em duplicata na presença de 10µl de SYBR[®] Green PCR Master Mix (Applied Biosystems), 100nM a 400nM de cada iniciador e 5ng a 50ng de cada amostra de cDNA. As reações foram realizadas em placa de 96 wells - MicroAmp[®] Optical (Applied Biosystems).

As concentrações dos iniciadores foram padronizadas considerando dois fatores: - a utilização, sempre que possível, das menores concentrações dos iniciadores a fim de evitar a formação de dímeros; e - a obtenção dos menores valores de CT possíveis. No caso das amostras de cDNA, as quantidades utilizadas foram padronizadas de modo a garantir a obtenção dos menores C_{TS} . As condições de amplificação consistiram em 2 minutos a 50°C para a ativação da enzima AmpErase (quando presente na reação), 10 minutos a 95°C para a ativação da AmpliTaq Gold DNA Polymerase, seguidos de 40 ciclos de denaturação a 95°C durante 15 segundos e 1 minuto a 60°C de anelamento dos iniciadores e extensão. Ao final da amplificação, adicionou-se uma etapa de 20 minutos de duração, na qual a temperatura aumenta gradualmente de 60°C para 95°C a 0,2°C por segundo com a contínua aquisição da fluorescência (MORRISON et al. 1998). Esta última etapa de 20 minutos, tem por objetivo submeter os produtos amplificados à análise da curva de dissociação. Esta análise por sua vez, serve para confirmar a especificidade da amplificação. A especificidade dos fragmentos amplificados também foi analisada em gel de agarose 1%.

3.8.4 Análise dos dados de expressão

A expressão diferencial dos transcritos de interesse (transcritos alvo) foi determinada pelo método de quantificação relativa em relação a um gene normalizador. A quantificação entre as linhagens HB4a e HB4a-C5.2 foi realizada utilizando a linhagem HB4a como amostra referência. No caso de quantificação entre as amostras de tumores de mama, um pool de 3 amostras de tecido de mama normal foi utilizado como amostra referência. Essas amostras foram obtidas comercialmente

da Clontech. Foram considerados como diferencialmente expressos transcritos que apresentaram diferença de expressão acima de 3 vezes.

Para o cálculo da expressão diferencial foi utilizando o modelo matemático proposto por PFAFFL (2001).

$$R = (E \text{ alvo})^{\Delta C_T \text{ alvo}} / (E \text{ normalizador})^{\Delta C_T \text{ normalizador}}$$

em que ***E alvo*** corresponde a eficiência de amplificação dos iniciadores específicos para o transcrito alvo, ***E normalizador*** corresponde a eficiência de amplificação dos iniciadores específicos para o gene normalizador, **$\Delta C_T \text{ alvo}$** corresponde a diferença entre o C_T do transcrito alvo obtido para a amostra em análise e o C_T do transcrito alvo obtido para a amostra referência, **$\Delta C_T \text{ normalizador}$** corresponde a diferença entre o C_T do gene endógeno obtido para a amostra em análise e o C_T do gene endógeno obtido para a amostra referência.

As eficiências de amplificação dos iniciadores específicos para os transcritos alvos e o gene normalizador foram calculadas de acordo com a equação: $E 10^{(-1/slope)}$. Em que, o ***slope*** corresponde a inclinação da reta obtida quando se analisa a variação do C_T dos transcritos alvo e normalizador em função do *log* de diferentes quantidades de cDNA. Foram utilizadas 4 diferentes quantidades de cDNA obtidas por meio da diluição seriada da amostra de cDNA original, que foi sintetizada a partir de 2 μ g de RNA total. As quantidades utilizadas foram 100, 10, 1 e 0,1ng.

Para a análise dos dados de expressão das amostras tumorais foram utilizados 4 genes normalizadores e para as linhagens celulares foi utilizado apenas 1 gene normalizador. Os normalizadores utilizados foram a β -actina, GAPDH, HPRT e BCR. Esses genes foram selecionados com base em dados da literatura (VANDESOMPELE et al. 2002; KOK et al. 2004; ZHANG et al. 2005; HUGGETT

et al. 2005). Para tanto, foi utilizada uma ferramenta, disponível *on line* (<http://medgen.ugent.be/~jvdesomp/genorm/>), desenvolvida para o *Microsoft Excel* e que foi denominada *geNorm* (VANDESOMPELE et al. 2002). O *geNorm* determina quais são os genes normalizadores mais estáveis em um conjunto de amostras de cDNA e calcula um fator de normalização da expressão gênica para cada amostra com base na média geométrica dos valores obtidos para os diferentes normalizadores. A estabilidade dos genes normalizadores é determinada partindo do princípio de que dois genes normalizadores ideais possuem razões de expressão idênticas em todas as amostras de cDNA a serem analisadas, independente de quaisquer condições experimentais ou tipos celulares. O *geNorm* avalia a razão de expressão de cada gene normalizador em relação aos demais normalizadores sempre aos pares, o que permite a exclusão dos genes menos estáveis. A cada normalizador excluído, o programa automaticamente atribui um novo valor referente à estabilidade dos demais normalizadores nas amostras de cDNA. Esse processo ocorre diversas vezes até que permaneçam apenas os genes cujos níveis de expressão sejam os mais estáveis determinando, assim, a combinação adequada de normalizadores a serem utilizados nas análises de PCR em Tempo Real. Preconiza-se a utilização mínima de três genes normalizadores estáveis para obter-se o fator de normalização confiável. Para determinar a necessidade de utilizar mais que três genes normalizadores na análise dos dados de PCR em Tempo Real, é recomendável fazer a inclusão seqüencial de outros genes normalizadores e avaliar o impacto dessa inclusão no fator de normalização. Se a inclusão do novo normalizador não variar o fator de normalização, a inclusão do mesmo não é necessária. Um valor de 0,15 foi estabelecido como um *cut off*, acima do qual a inclusão de um novo normalizador

seria adequada para a obtenção dos fatores de normalização (VANDESOMPELE et al. 2002).

3.8.5 Análise estatística dos dados de expressão das amostras tumorais

A análise estatística da associação entre a super-expressão de c-erbB2 e a super-expressão dos novos transcritos, nas amostras tumorais de mama, foi realizada utilizando o programa computacional GraphPad Prism versão 4.0 (GraphPad Software Inc., San Diego, CA). As diferenças entre as médias das amostras c-erbB2 positivas e c-erbB2 negativas foi testada pelo teste-t de Student não pareado. Para todos os testes estabeleceu-se um erro $\alpha=5\%$, ou seja, os resultados foram considerados estatisticamente significativos quando $p<0,05$.

3.9 RACE (RAPID AMPLIFICATION OF CDNA ENDS)

Com o objetivo de gerar a seqüência completa dos novos transcritos diferencialmente expressos entre as linhagens celulares HB4a e HB4a-C5.2, utilizamos a técnica de RACE. Para tanto, utilizou-se Smart Amplification cDNA Kit (Clontech® n°: K1811-1). A primeira fita de cDNA foi sintetizada utilizando um iniciador oligo dT modificado, denominado 5'-RACE CDS primer. Esse iniciador possui dois nucleotídeos degenerados na posição 3'terminal, que auxiliam no posicionamento do iniciador no início da cauda poli A. Após a enzima transcriptase reversa alcançar o final da molécula de mRNA, são adicionados de 3-5 resíduos dC na porção 3' da primeira fita de cDNA devido à atividade terminal transferase da enzima. Em seguida o oligonucleotídeo SMART II A, cuja porção terminal apresenta uma série de resíduos dG, se liga a cauda de cDNA rica em resíduos dC servindo como um molde para a Transcriptase Reversa PowerScript. Dessa forma, esse processo garante a síntese de moléculas de cDNA completas. A síntese da primeira fita de cDNA foi realizada utilizando 1µg de RNA total livre de DNA genômico, 1µM de 5'-CDS primer, 1µM de SMART II A oligo, e o volume foi acertado para 5µl. Essa mistura foi aquecida a 70°C durante 2 minutos e resfriada em gelo durante 2 minutos. Em seguida a reação foi processada em presença de 1X tampão de obtenção da primeira fita, 2mM de DTT, 1mM de dNTP e 100U de Transcriptase Reversa PowerScript (CLONTECH Laboratories, INC). A reação foi mantida a 42°C durante 90 minutos. Em seguida o produto da reação de síntese da primeira fita de cDNA foi diluído em 100µl de tampão Tricina-EDTA e aquecido a 72°C durante 7 minutos.

Finalmente, a primeira fita de cDNA pode então ser utilizada diretamente na reação de 5'RACE, pela utilização de iniciadores específicos para a amplificação do transcrito de interesse. A reação foi realizada utilizando 2,5µl de cDNA 5'RACE, 0,2mM de dNTP, 1X Advantage 2 PCR Buffer, 1X Advantage 2 Polimerase Mix (Clontech), 1X UPM, 0,2µM de cada iniciador externo específico, para um volume final de reação de 50µl. Todas as reações foram realizadas em uma GeneAmp PCR System 9700 (Applied Biosystems), utilizando as seguintes condições para a ciclagem: 25 ciclos de 30 segundos a 94°C, 30 segundos a 68°C e 3 minutos a 72°C.

Para aumentar a especificidade das reações a estratégia de Nested-PCR sugerida pelo fornecedor foi adotada. Os produtos obtidos a partir da primeira reação foram utilizados como molde para a segunda reação de RACE, agora utilizando o iniciador NUP e o iniciador interno específico para o transcrito em questão. Os produtos obtidos foram analisados em gel de agarose 1%, purificados utilizando o QIAquick Gel Extraction Kit (Qiagen) e clonados utilizando o pGEM[®]-T Vector System I (Promega), de acordo com as especificações do fabricante. Os clones foram seqüenciados como descrito acima, o que permitiu a extensão da seqüência do transcritos. O transcrito que teve sua seqüência completamente gerada por meio da estratégia de RACE é correspondente à *tag* 44. Na tabela 6, estão representadas as seqüências dos iniciadores que foram construídos para as reações.

Tabela 6 - Iniciadores para RACE.

Iniciadores	Seqüência dos iniciadores
44-GSP	5' GAG AGT TTA TTG GCT CAT TCA ACG GAT A 3'
44-GSPN	5' CAA CGG ATA AGT CCA GTG GTA AAA CTA G 3'
UPM (Universal Primer Mix)	Long – 5' CTA ATA CGA CTC ACT ATA GGG CAA GCA GTG GTA TCA ACG CAG AGT 3' Short – 5' CTA ATA CGA CTC ACT ATA GGG C 3'
NUP (Nested Universal Primer)	5' AAG CAG TGG TAT CAA CGC AGA GT 3'
5'RACE CDS Primer	5'-(T) ₂₅ N ₁ N-3' (N = A, C, G ou T; N ₁ = A, G ou C)
SMART II A oligo	5' AAG CAG TGG TAT CAA CGC AGA GTA CGC GGG 3'

Legenda: Seqüências dos iniciadores específicos (GSP-Gene Specific Primer) e dos iniciadores *nested* (GSPN-Gene Specific Primer Nested) utilizados para gerar a seqüência completa do transcrito correspondente à *tag* 44. Seqüências dos iniciadores utilizados universalmente nas reações de RACE.

3.10 RT-PCR FITA ESPECÍFICA

Para a validação experimental dos transcritos *antisenses* realizamos uma RT-PCR fita específica. A orientação relativa a transcrição foi determinada por meio do iniciador presente durante a ação da enzima transcriptase reversa na síntese da primeira fita de cDNA. Assim, foram realizadas quatro reações de RT-PCR para a validação de cada transcrito *antisense*. Na primeira reação, foi utilizado somente o iniciador complementar ao transcrito *sense*. Na segunda reação, foi utilizado apenas o iniciador complementar ao transcrito *antisense*. A terceira reação foi realizada na presença do iniciador complementar ao transcrito *antisense*, mas na ausência de transcriptase reversa, e a quarta reação foi realizada na presença da enzima transcriptase reversa, porém na ausência do iniciador complementar ao transcrito *antisense*. A primeira reação foi realizada para avaliar a expressão do transcrito *sense*. A segunda reação foi realizada para avaliar a expressão e comprovar a

existência do transcrito *antisense*. As duas últimas reações foram realizadas com o objetivo de verificar respectivamente a presença de contaminação com DNA genômico e a ausência de formação de estruturas secundárias durante a síntese da primeira fita de cDNA, as quais poderiam servir como iniciadores para a transcriptase reversa.

As reações foram realizadas utilizando 1µg de RNA total livre de DNA genômico, 0,5mM de dNTPs, 0,9µM de cada iniciador, e o volume foi acertado para 33µl. Essa mistura foi aquecida a 65°C durante 5 minutos e resfriada a 50°C. Em seguida a reação foi processada em presença de 1X tampão de obtenção da primeira fita, 0,01M de DTT, 40U de RNase OUT e 200U de Transcriptase Reversa (Superscript II, Gibco BRL). A reação foi mantida a 50°C durante 60 minutos e inativada pelo aquecimento a 70°C durante 15 minutos. Na tabela 7, estão listadas as seqüências dos iniciadores utilizados tanto na síntese da primeira fita de cDNA quanto nas reações de amplificação por PCR.

Tabela 7 - Iniciadores *Antisense*. Seqüências dos iniciadores específicos utilizados para a validação dos transcritos *antisense*.

Transcritos correspondentes às tags:	Seqüências dos iniciadores
12	Fw - 5'GAT CTC TCA GTA AAC TTC A 3' Rev - 5'ATT GGA TGA TAA TTT CTT GGA A 3'
46	Fw - 5'TCT TTA CTC AGT CAC ACA AC 3' Rev - 5'TTC AAG TGG AAT TAG TTG GTT 3'
57	Fw - 5'GAT CCT AGA ATC GGT GAC 3' Rev - 5'CTC TGC AGG AAG ACA CTA 3'
69	Fw - 5'GAT CAG TTC TTG CAG TCC 3' Rev - 5'GAA TCC GAT ATG GGC AAA C 3'
70	Fw - 5'CTT GAT ACA AAA ATG ATA GC 3' Rev - 5'TGT ACT ATT TAT GGT GGT GG 3'

Para verificar a expressão dos transcritos *sense* e *antisense*, 2µl do cDNA fita específica foram amplificados por PCR. As reações foram realizadas em um volume final de 25 µl contendo 1X tampão da enzima Taq DNA polimerase (GibcoBRL), 0,1 mM dNTPs, 1,4 mM MgCl₂, 0,4 µM de cada um dos iniciadores e 1U Taq DNA polimerase (GibcoBRL). As condições das reações foram: desnaturação durante 4 minutos a 94°C, seguida de 40 ciclos de desnaturação a 94°C durante 40 segundos, anelamento a 58°C durante 40 segundos e extensão a 72°C durante 1 minuto. Ao final da ciclagem, as moléculas foram estendidas durante 6 minutos a 72°C. Os fragmentos obtidos foram analisados em gel de poliacrilamida 8% e corados com prata.

Resultados e Discussão

4 RESULTADOS E DISCUSSÃO

4.1 CULTURA DE CÉLULAS E EXPRESSÃO DIFERENCIAL DE C-ERBB2 NAS LINHAGENS HB4a E HB4a-C5.2

As linhagens celulares HB4a e HB4a-C5.2 foram cedidas pelo Dr. Michael O'Hare do Instituto Ludwig de Londres e mantidas em cultura em nosso laboratório até atingirem confluência de 80%. A linhagem celular HB4a, apresentou características morfológicas típicas de células luminais epiteliais de mama normal. Essas células mantiveram a arquitetura normal do tecido, uma vez que todas as células apresentaram tamanhos similares e formato cubóide. Por outro lado, as células HB4a-C5.2 apresentaram tamanhos irregulares e também um formato mais alongado. Como relatado anteriormente (HARRIS et al. 1999), as células HB4a-C5.2 mostraram uma tendência de crescimento empilhado devido à perda de inibição por contato.

Para confirmar a expressão diferencial de c-erbB2 entre as duas linhagens celulares, o nível de expressão do oncogene foi quantificado por meio de PCR em Tempo Real. O gene constitutivo β -actina foi utilizado como normalizador da reação e a linhagem HB4a foi utilizada como referência. As concentrações dos iniciadores bem como as quantidades de cDNA utilizadas, foram padronizadas como descrito em Materiais e Métodos e o cálculo de expressão diferencial foi realizado segundo o modelo matemático proposto por PFAFFL (2001). Esse modelo leva em

consideração a eficiência de amplificação do transcrito alvo e do normalizador. A eficiência de amplificação dos iniciadores utilizados para o gene normalizador β -actina foi de 97%, e de 100% para o oncogene c-erbB2. Com base no resultado obtido por meio do PCR em Tempo Real, pôde-se verificar que o nível de expressão do oncogene c-erbB2 é 145 vezes maior na linhagem HB4a-C5.2 em relação a linhagem celular HB4a, o que demonstra haver concordância entre os nossos resultados e os dados já descritos na literatura (HARRIS et al. 1999). Os dados referentes a quantificação da expressão do oncogene c-erbB2 nas duas linhagens celulares estão relacionados na Tabela 8.

Tabela 8 - Dado de expressão diferencial do oncogene c-erbB2 entre as linhagens celulares HB4a e HB4a-C5.2.

Transcrito	Linhagens	C _T amostra	C _T controle endógeno: β -actina	$R = (E_{\text{alvo}})^{DCT_{\text{alvo}}}/(E_{\text{referência}})^{DCT_{\text{referência}}}$
c-erbB2	HB4a	23,69±0,09	15,08±0,08	145
	HB4a-C5.2	17,56±0,02	16,15±0,16	

4.2 ANÁLISE DAS LINHAGENS HB4a E HB4a-C5.2 ATRAVÉS DE MPSS

O experimento realizado pela empresa Lynx Therapeutics teve por objetivo, ao comparar o padrão de expressão das duas linhagens celulares HB4a e HB4a-C5.2, analisar as conseqüências moleculares relacionadas com a super-expressão do oncogene c-erbB2. Foram geradas um total de 24.521.236 *tags* a partir de ambas as linhagens celulares correspondendo a 24.065 *tags* distintas (Lynx Therapeutics). Os dados gerados foram analisados por dois grupos de bioinformática (Dr. Sandro de

Souza e Dr. Victor Jongeneel) com o objetivo de detectar *tags* diferencialmente expressas e também estabelecer a correlação entre *tags* e transcritos.

Por meio dessas análises, mais de 90% das *tags* de MPSS foram associadas a transcritos conhecidos e vários transcritos diferencialmente expressos foram identificados. Entretanto, 9% das *tags*, que incluíram *tags* diferencialmente expressas entre as duas linhagens, não foram associadas a transcritos conhecidos, podendo representar novos transcritos humanos. Essas *tags*, sem nenhuma associação com transcrito conhecidos foram denominadas *tags* órfãs. Esses dados estão relacionados na Tabela 9.

Tabela 9 – Categorização das *tags* de MPSS.

Categoria das Tags	Número de Tags	Tags associadas	Tags órfãs
Sem expressão diferencial	21.791 (100%)	20.111 (92%)	1.680 (8%)
5 ou mais vezes up-reguladas	873 (100%)	644 (74%)	229 (26%)
5 ou mais vezes down-reguladas	1.401 (100%)	1.155 (83%)	246 (17%)
Total	24.065	21.910 (91%)	2.155 (9%)

Legenda: Na tabela estão representadas as três categorias em que as *tags* foram divididas, o número de *tags* em cada categoria e as informações referentes ao número de *tags* que foram associadas a transcritos conhecidos e o número de *tags* que não foram associadas a transcritos conhecidos – *tags* órfãs.

As 2.155 *tags* órfãs, foram divididas em duas categorias: *Tags up-reguladas* (são *tags* com frequência 5 vezes maior na biblioteca HB4a-C5.2 em relação a HB4a) e *Tags down-reguladas* (são *tags* com frequência 5 vezes menor na biblioteca HB4a-

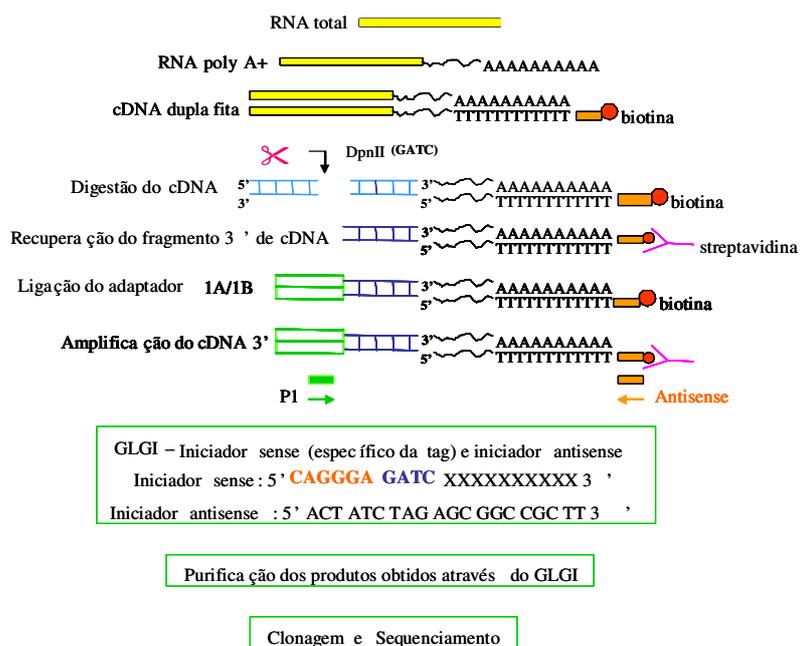
C5.2 em relação a HB4a). Foram selecionadas para as nossas análises, apenas as *tags* que apresentaram um nível de expressão diferencial entre as duas linhagens celulares maior que 5.

Devido ao pequeno tamanho das *tags* geradas pelo MPSS (17 nucleotídeos), torna-se difícil utilizar sua informação para a validação experimental das mesmas e posterior caracterização dos transcritos correspondentes. Essa limitação da metodologia foi solucionada por meio da adaptação da técnica de GLGI (Generation of Longer cDNA fragments from SAGE *tags* for Gene Identification) para a extensão de *tags* de MPSS.

4.3 PADRONIZAÇÃO DA TÉCNICA DE GLGI-MPSS

A adaptação do protocolo de GLGI, para a extensão de *tags* de MPSS, foi realizada na Universidade de Chicago em colaboração com o grupo do Dr. San Ming Wang e sob orientação do Dr. Jianjun Chen. Como mencionado anteriormente, o protocolo de GLGI foi desenvolvido inicialmente para estender *tags* geradas a partir de SAGE. Por essa razão, o protocolo original sofreu algumas modificações e o protocolo adaptado, recebeu o nome de GLGI-MPSS (CAMARGO et al. 2004; SILVA et al. 2004a e b) (Anexos 1 e 2). As modificações incluíram alterações na seqüência do adaptador utilizado para a amplificação do fragmento 3' de cDNA, aumento no número de ciclos da PCR durante a amplificação do fragmento de GLGI-MPSS e seleção de 8 colônias para cada fragmento gerado pelo GLGI a ser seqüenciado. A alteração na seqüência do adaptador foi necessária, devido ao uso de uma enzima de restrição diferente, (*DpnII*), da utilizada no protocolo de SAGE para a

construção da biblioteca de MPSS. O aumento do número de ciclos na PCR e de colônias a serem seqüenciadas foi necessário, devido ao baixo nível de expressão dos transcritos correspondentes às *tags* de MPSS. Na Figura 10, pode - se visualizar a representação esquemática do protocolo do GLGI.



Legenda: A partir do RNA total do tecido de interesse obtém-se o mRNA utilizado para a síntese da dupla fita de cDNA. O oligonucleotídeo d(T) utilizado na síntese da dupla fita de cDNA é marcado com biotina. Em seguida, a dupla fita de cDNA é digerida com a enzima de restrição *DpnII*. Os fragmentos 3' de cDNA resultantes da digestão são, então, recuperados por meio de um sistema de partículas magnéticas acopladas a streptavidina. Após a recuperação dos fragmentos 3' de cDNA, são ligados aos mesmos adaptadores importantes para a subsequente amplificação dos fragmentos. Após a amplificação dos fragmentos 3' de cDNA, ocorre a realização da reação de GLGI, purificação dos fragmentos gerados, clonagem e seqüenciamento dos mesmos.

Figura 10 - Representação esquemática da técnica de GLGI.

O início aos experimentos se deu partindo de aproximadamente 180 µg de RNA total de cada uma das linhagens celulares. Depois de confirmar a integridade dos RNAs (figura 11A), a etapa seguinte consistiu na purificação do mRNA realizada por meio da utilização de um sistema constituído de partículas magnéticas acopladas

ao oligo d(T). Após a purificação do mRNA uma alíquota desse material foi analisada em um gel de agarose 1% corado com brometo de etídeo. Com base no gel, pôde-se observar que, embora houvesse uma pequena contaminação da amostra com RNA ribossômico (que pôde ser visualizada através das bandas 28S e 18S), havia ocorrido uma recuperação de mRNA bastante eficiente (Figura 11B) e suficiente para prosseguir com o protocolo.

Em seguida foi realizada a síntese da primeira e segunda fitas de cDNA por meio da utilização de uma mistura de oligos d(T) biotinilados que possuíam na sua extremidade 5' uma seqüência adicional adaptadora (complementar a seqüência do iniciador SAGE-*antisense* posteriormente utilizado para a amplificação em larga escala do fragmento *DpnII* 3' do cDNA). Para avaliar a eficiência da síntese tanto da primeira quanto da segunda fita de cDNA, foram retiradas alíquotas de cada uma das etapas de síntese que foram analisadas em gel de agarose 1% corado com brometo de etídeo. Pôde-se observar um rastro que nos indicou que a síntese de ambas as fitas de cDNA das linhagens celulares haviam sido eficientes (Figura 11C).

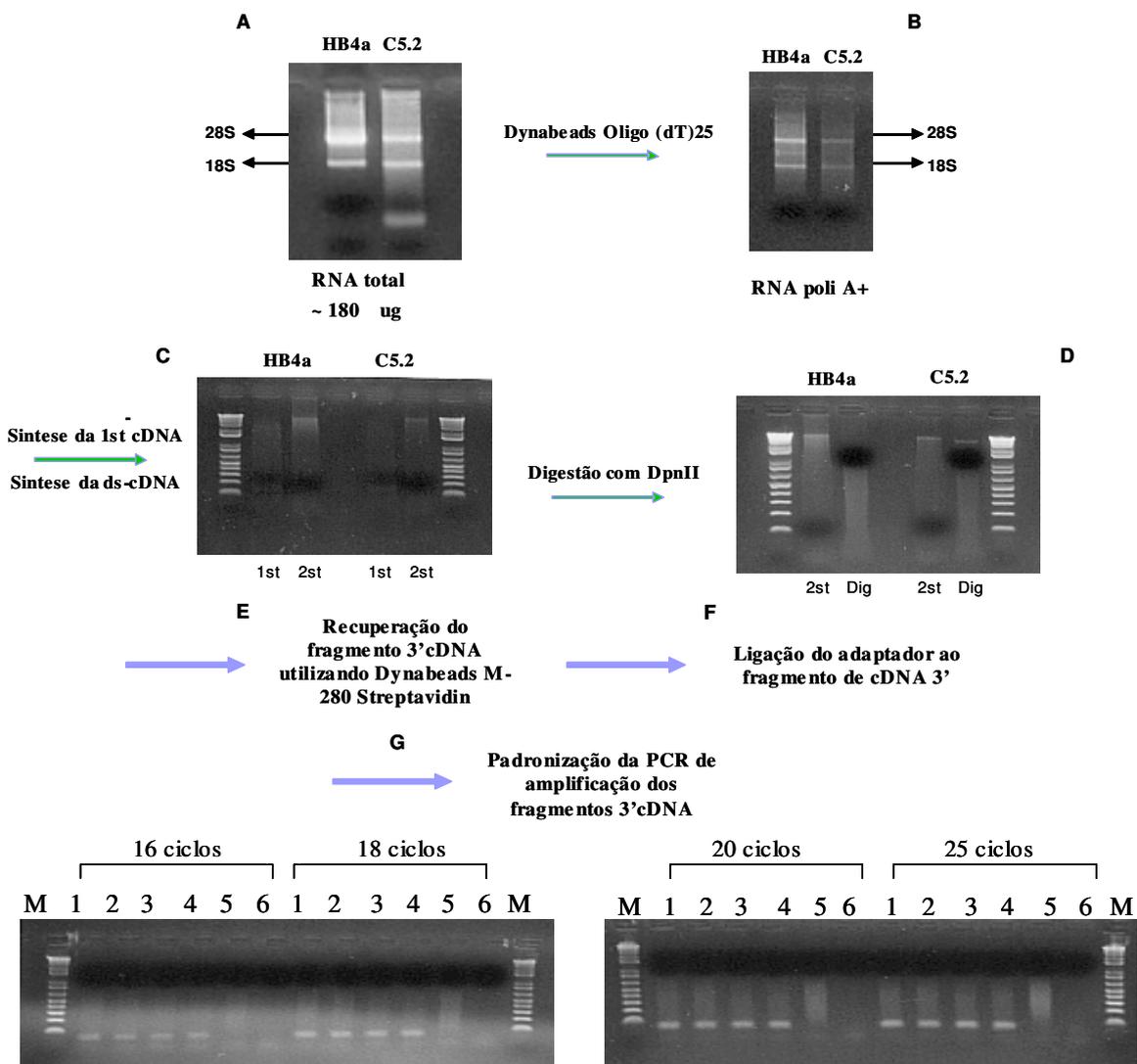
Após a dupla fita de cDNA ter sido sintetizada, a mesma foi digerida com a enzima de restrição *DpnII* cujo sítio de reconhecimento e corte são as bases GATC. Para avaliar a eficiência de digestão, uma alíquota da reação foi aplicada, em conjunto com uma alíquota de cDNA não digerido, em gel de agarose 1% corado com brometo de etídeo. De acordo com o esperado, pôde-se verificar que houve uma diminuição no tamanho do fragmento de cDNA dupla fita digerido quando comparado a amostra não digerida (Figura 11D).

Após a digestão, os fragmentos 3' de cDNA foram recuperados por meio da utilização de um sistema de partículas magnéticas acopladas a streptavidina. A

streptavidina interage com a biotina ligada a extremidade 5' do oligo d(T) utilizado durante a síntese do cDNA permitindo a purificação dos fragmentos (Figura 11E).

Após a purificação dos fragmentos 3' *DpnII*, foi ligada aos mesmos uma seqüência adaptadora (Figura 11F) complementar a seqüência do iniciador SAGE P1, utilizado em combinação com o iniciador SAGE-*antisense* na reação de amplificação do fragmento 3' de cDNA.

Durante a padronização da técnica, o número de ciclos a ser utilizado na reação de amplificação em larga escala foi testado. Para tanto, foram retiradas alíquotas de 5µl em cada um dos seguintes ciclos: 16, 18, 20 e 25. Essas amostras foram aplicadas em gel de agarose 1% corado com brometo de etídeo e com base no gel, pôde-se concluir que as amostras estavam adequadas para serem utilizadas na reação de amplificação em larga escala e que 20 ciclos seriam suficientes para que a amplificação do fragmento desejado ocorresse (Figura 11G).



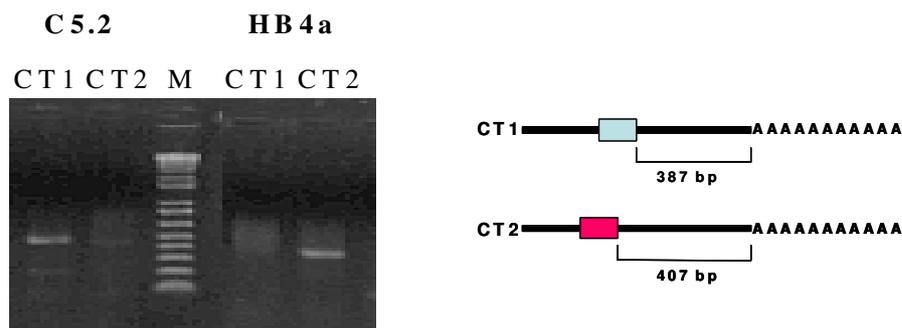
Legenda: (A) extração de RNA total das linhagens celulares HB4a e HB4a-C5.2 (B) purificação de RNA poli A+ (C) Síntese da primeira e segunda fita de cDNA das linhagens celulares (D) comparação entre a dupla fita de cDNA sintetizada pré e pós digestão com a enzima de restrição *DpnII* (E) etapa de recuperação do fragmento 3' de cDNA gerado a partir da digestão do cDNA dupla fita (F) ligação dos adaptadores ao fragmento de cDNA a ser amplificado (G) reação de padronização da amplificação dos fragmentos 3' de cDNA. As amostras de cDNA foram utilizadas em duplicatas onde para cada ciclo os números 1 e 2 representam o cDNA da linhagem HB4a; 3 e 4 representam o cDNA da linhagem HB4a-C5.2; 5 representa o cDNA utilizado como controle positivo da reação e 6 representa o controle negativo da reação (amostra sem nenhum cDNA).

Figura 11 - Resultados obtidos com o protocolo modificado de GLGI.

Após as condições utilizadas na reação de amplificação do fragmento 3' de cDNA terem sido padronizadas, a reação de amplificação foi realizada em larga escala. Para tanto, as reações foram realizadas em uma placa com capacidade para 96 reações. A placa foi dividida ao meio, de modo que foram realizadas 48 reações de amplificação do fragmento 3' de cDNA para cada uma das linhagens utilizadas (HB4a e HB4a-C5.2).

4.3.1 Tags controles

Para a padronização das reações de extensão das *tags* de MPSS, foram utilizadas 2 *tags* controles, CT1 (NM_000900) e CT2 (NM_005557), correspondentes a transcritos conhecidos. Essas 2 *tags* associadas a transcritos conhecidos representaram os controles, pois os tamanhos dos fragmentos 3' de cDNA estendidos a partir do GLGI foram calculados, considerando a posição da *tag* de MPSS em relação a extremidade 3' do transcrito. Os fragmentos amplificados a partir de cada uma das *tags* apresentaram o tamanho esperado e além disso, os fragmentos obtidos eram bem definidos e únicos. Pôde-se observar também que a *tag* controle CT2, pertencente à categoria de *tags* com menor expressão na linhagem celular HB4a-C5.2, apresentou amplificação específica apenas na linhagem HB4a, e a *tag* controle CT1, pertencente à categoria de *tags* com maior expressão na linhagem celular HB4a-C5.2, foi estendida apenas a partir do cDNA proveniente da mesma (figura 12). A especificidade dos fragmentos foi confirmada após a clonagem e o seqüenciamento dos mesmos. As seqüências de ambos os fragmentos obtidos foram analisadas por meio da utilização do programa BLAST.



Legenda: Resultado da extensão dos fragmentos 3' de cDNA a partir das *tags* controles de MPSS representadas como CT1 e CT2. O padrão de peso molecular utilizado foi o 100 bp (M). O tamanho molecular dos fragmentos amplificados estão indicados em pares de bases (pb) à direita da figura, na representação esquemática da distância entre a *tag* de cada um dos transcritos controle e a cauda poliA. O retângulo verde representa a *tag* do CT1 e o retângulo rosa representa a *tag* do CT2.

Figura 12 - Resultados obtidos com o protocolo de GLGI para as *tags* controles.

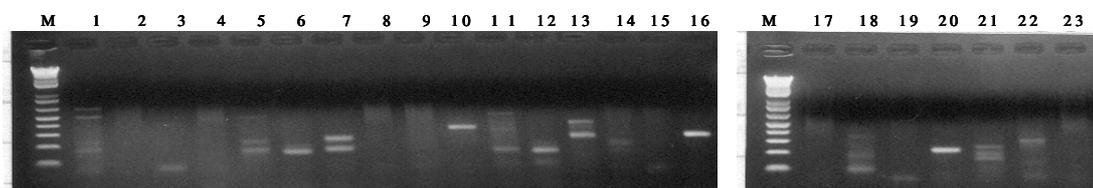
4.3.2 Tags órfãs

Uma vez que os resultados obtidos na reação de extensão dos fragmentos 3' de cDNA para as *tags* controles de MPSS foram positivos, partiu-se para a realização da extensão dos fragmentos correspondentes às *tags* órfãs. Para dar início às ampliações através do GLGI-MPSS, foram selecionadas 83 *tags* órfãs sendo que 40 *tags* pertenciam à categoria *up* regulada e 43 *tags* pertenciam à categoria *down* regulada.

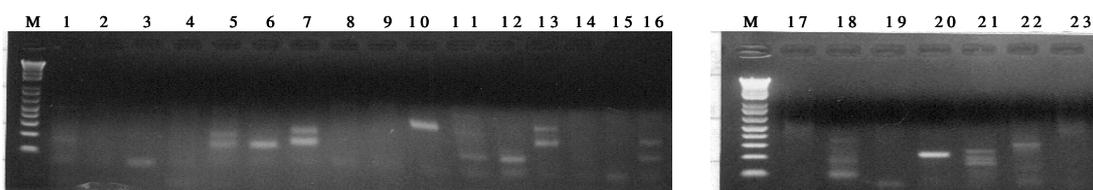
Após a realização das reações de amplificação das 83 *tags* selecionadas, observou-se que para 42 das 83 reações de GLGI-MPSS não foi possível visualizar em gel de agarose, um fragmento de cDNA bem definido, após 30 ciclos de amplificação (figuras 13 A1;2 ,B1;2 e C1;2). Este fato pode ser explicado devido ao baixo nível de expressão dos transcritos a partir dos quais as *tags* de MPSS foram geradas. Dados da literatura mostram que a eficiência de amplificação do GLGI-SAGE é proporcional à abundância das *tag*. A amplificação de *tags* de SAGE com

um número acima de 50 cópias, geralmente, produz um fragmento representado, em gel de agarose, por uma única e intensa banda. Em contrapartida, a amplificação de *tags* de SAGE com um número inferior a 50 cópias, em geral, resulta em fragmentos de menor intensidade podendo haver amplificação de bandas adicionais (CHEN et al. 2002a). As *tags* órfãs selecionadas apresentam, em média, 12.1 cópias na biblioteca HB4a e 13,8 cópias na biblioteca HB4a-C5.2.

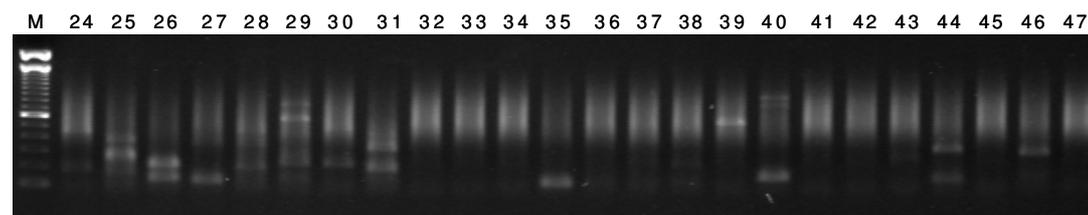
A1



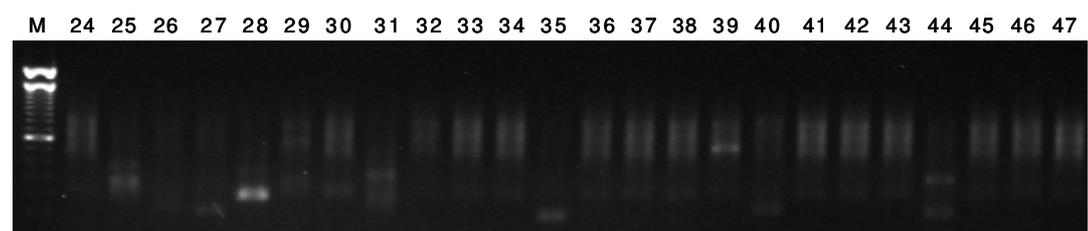
A2



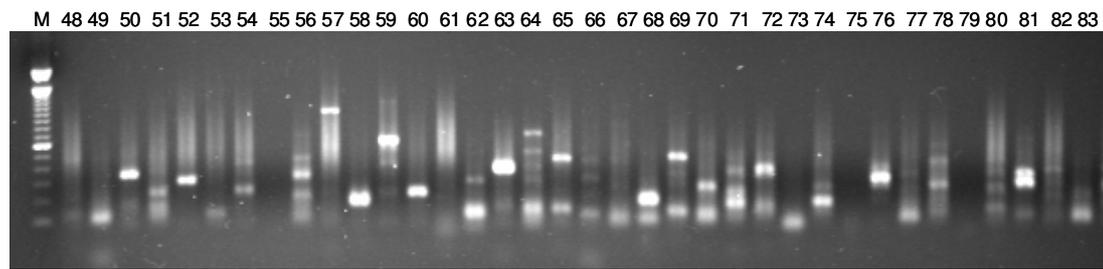
B1



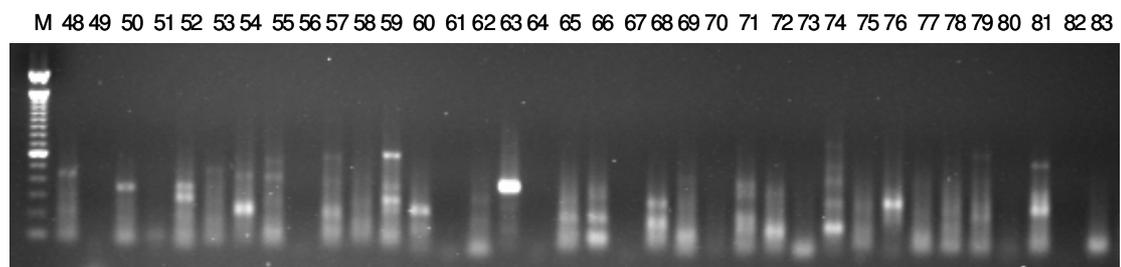
B2



C1



C2



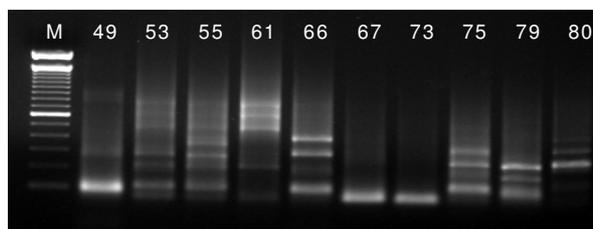
Legenda: Nas figuras estão representadas as reações de amplificação GLGI realizadas para as *tags* órfãs de MPSS (1-83) em que A1, B1 e C1 correspondem às amplificações que utilizaram cDNA da linhagem HB4a e A2, B2 e C2 correspondem às amplificações que utilizaram cDNA da linhagem HB4a-C5.2. Os fragmentos foram analisados em gel de agarose 1% corado com brometo de etídeo. O marcador de peso molecular utilizado foi o 100 bp (M).

Figura 13 - Amplificação GLGI-MPSS.

Em teoria, modificações no protocolo GLGI-MPSS, especialmente o aumento do número de ciclos poderiam resolver o problema de amplificação de *tags* com um número de cópias inferior a 50. Entretanto, foi relatado que no caso de GLGI-SAGE a amplificação com um elevado número de ciclos pode resultar em amplificação inespecífica, devido ao anelamento parcial do iniciador *sense* com outros transcritos (CHEN et al. 2002a). Baseando-se nessas informações, um total de 10 *tags* de MPSS, para as quais não havia ocorrido a amplificação de um fragmento de cDNA bem definido, foram re-amplificadas utilizando 35 ciclos. Os resultados obtidos a partir

desse experimento foram similares aos descritos na literatura para as *tags* de SAGE. Como pode ser visto na figura 14, fragmentos bem definidos com tamanho acima de 100 pares de bases foram obtidos para 2 das 10 *tags* testadas (*tags* 66 e 80). Esses fragmentos foram, então, clonados, seqüenciados, e mostraram-se inespecíficos após análise das seqüências por meio do programa BLAST.

Outra possibilidade para explicar a baixa eficiência de amplificação observada para o GLGI-MPSS é o fato de que, nem sempre, a *tag* de MPSS possui uma seqüência adequada para o desenho dos iniciadores e eficiente amplificação durante a PCR. Muitos iniciadores *sense* utilizados na amplificação GLGI-MPSS formaram estruturas secundárias. Entretanto, essa hipótese provou ser apenas parcialmente correta uma vez que pôde-se obter fragmentos GLGI-MPSS específicos para algumas *tags* que apresentavam formação de estruturas secundárias em seus iniciadores.



Legenda: Dez *tags* órfãs (49, 53, 55, 61, 66, 67, 73, 75, 79 e 81) que não produziram bandas bem definidas na reação padrão de GLGI-MPSS foram re-amplificadas utilizando 35 ciclos. Os fragmentos foram analisados em gel de agarose 1% e foram corados com brometo de etídeo. O padrão de peso molecular utilizado foi o 100 bp (M).

Figura 14 - Reação de GLGI-MPSS utilizando um maior número de ciclos para a amplificação dos fragmentos.

A amplificação de um fragmento bem definido foi observada para 41 das 83 (49,4%) reações de GLGI-MPSS (figura 13). A distribuição de tamanho dos fragmentos amplificados variou de 29 a 588 pares de bases com uma média de tamanho de 195 pares de bases. Mais da metade dos fragmentos 3' de cDNA gerados apresentaram tamanho acima de 120 pares de bases, o que representa cerca de 10 vezes o tamanho inicial da *tag* de MPSS. Sendo assim, os dados gerados proporcionam informação adicional para a posterior caracterização dos transcritos correspondentes.

Os 41 fragmentos obtidos foram clonados e 8 colônias de cada fragmento foram seqüenciadas e analisadas quanto à similaridade com outros transcritos por meio do programa BLAST N contra o banco de dados do GenBank (NR e dEST). Uma alta similaridade com seqüências transcritas foi encontrada para todos os fragmentos analisados e, na maioria dos casos, as seqüências com as quais os fragmentos apresentaram similaridade, foram submetidas ao GenBank depois da construção do nosso banco de dados de referência de *tags* de MPSS (Tabela 10).

Tabela 10 - Resultados obtidos por meio do protocolo GLGI-MPSS para as *tags* órfãs.

Tag	Sequência	Tamanho	Similaridade
3	GATC TTGGTGTTCGCAA	29 bp	Inespecífica
5	GATC CGATGCAGGGGAG	323 bp	Inespecífica
6	GATC TTTCTAGGGATGT	108 bp	SNP
7	GATC TGTCCTACTTCTGG	122 bp	Novo gene regulado por c-erbB2 (FL)
10	GATC ATTGGGAGCTTTT	271 bp	Novo gene regulado por c-erbB2 EST
12	GATC TCTCAGTAAACTT	105 bp	Antisense
13	GATC CGGGATTCTCTCAG	201 bp	Novo gene regulado por c-erbB2 (FL)
16	GATC CTTGTAGTTCCTA	219 bp	Novo gene regulado por c-erbB2 EST
20	GATC TTCAAACACAGAG	588 bp	Anelamento interno
25	GATC CAAGGGCACAAGG	269 bp	Inespecífico
26	GATC GCTCGCTCGCTCG	101 bp	Inespecífico
27	GATC CATGAAGTCAGTC	92 bp	Inespecífico
28	GATC ATAGGGACAGCGG	173 bp	Novo gene regulado por c-erbB2 (FL)
31	GATC TTCAGCCAGCCAG	118 bp	Inespecífico
35	GATC TTTATCACTGGC	72 bp	Anelamento interno
39	GATC TAAATCAAGCTGC	509 bp	Novo gene regulado por c-erbB2 EST
40	GATC GTGACATACAAAT	102 bp	Anelamento interno
44	GATC CAGACATGCCTGA	95 bp	Novo gene regulado por c-erbB2 EST
46	GATC CAAAAATTCTTTA	253 bp	Antisense
50	GATC CGTGACAAACAAA	127 bp	Inespecífico
51	GATC CGGGACTTCAGGT	211 bp	Inespecífico
52	GATC CACATTTGTTAAG	266 bp	Novo gene regulado por c-erbB2 EST
54	GATC TAAGTTCCGGAG	168 bp	Inespecífico
56	GATC TGGCTTGGGAACA	285 bp	Inespecífico
57	GATC CTAGAATCGGTGA	280 bp	Antisense
58	GATC TCTGCGTTTTTTT	150 bp	Inespecífico
59	GATC CAAAAGTTCACTT	393 bp	Anelamento interno
60	GATC TCTGGTTTGAAG	200 bp	SNP
62	GATC TGCCAAGAAAAGC	75 bp	Poliadenilação alternativa
63	GATC TGGCAAGGTCACC	344 bp	SNP
64	GATC CGACCAGCCAGAA	108 bp	Inespecífico
65	GATC AGAGTTCTGTCAA	110 bp	Poliadenilação alternativa
68	GATC GGGAAAGGCAGATT	157 bp	Anelamento interno
69	GATC AGTTCCTGCAGTC	294 bp	Antisense
70	GATC TGACAGTGGTCAC	236 bp	Antisense
71	GATC TCTGCTCGCTGCG	141 bp	Inespecífico
72	GATC CTCTGGACCATGA	92 bp	Inespecífico
74	GATC CGTATCGGTTCCC	109 bp	Inespecífico
76	GATC TTCTTCAATTAG	279 bp	Novo gene regulado por c-erbB2 (FL)
78	GATC TTGCCAGCTTGCT	91 bp	Inespecífico
81	GATC ACCAGGGTCAGTT	121 bp	Novo gene regulado por c-erbB2 EST

Legenda: Na tabela, estão representadas as 41 *tags* de MPSS e o tamanho do fragmento gerado por meio do GLGI-MPSS. Os resultados das análises de similaridade podem ser visualizados em diferentes cores.

Para 16 dos 41 fragmentos analisados (39%), observou-se, após análise por meio do BLAST, uma alta similaridade com transcritos já conhecidos (tabela 9). Entretanto, esta similaridade foi parcial na região correspondente ao iniciador *sense* (em geral as últimas 4 a 7 bases do iniciador *sense*). Esse dado sugere que o iniciador *sense*, correspondente à *tag* de MPSS, anelou inespecificamente ao transcrito conhecido durante a amplificação GLGI-MPSS.

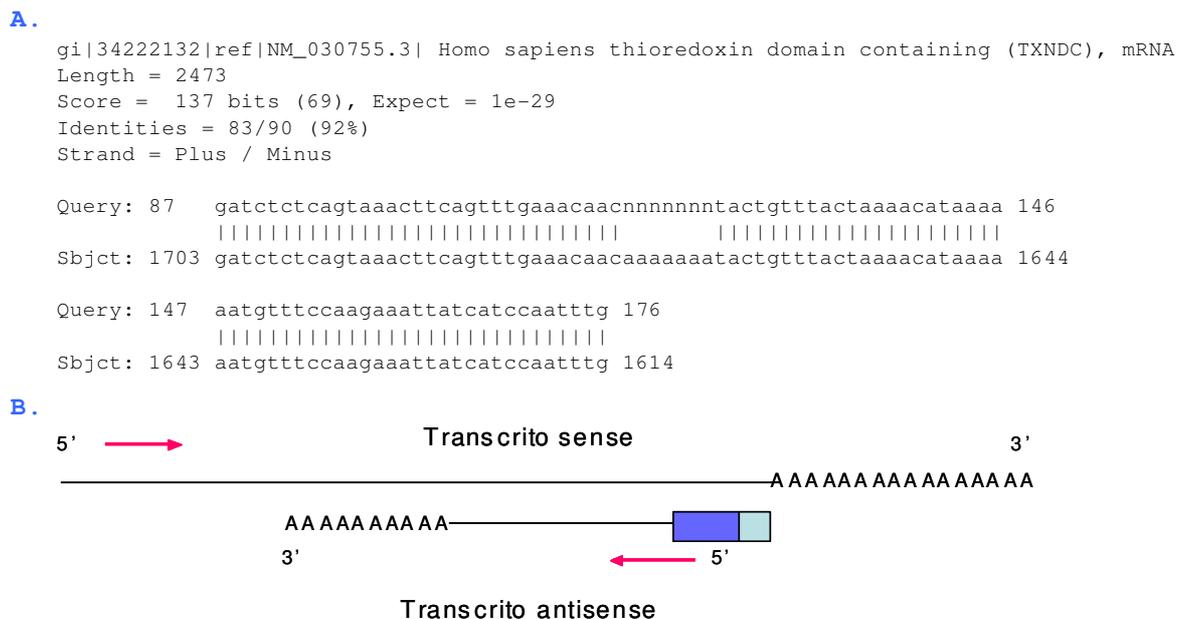
Ao contrário de uma PCR convencional, nas reações GLGI-MPSS somente o iniciador *sense* proporciona especificidade para que a amplificação ocorra. Se o nível de expressão da molécula alvo for baixo, pode ocorrer o anelamento parcial do iniciador *sense* com outras moléculas cujos níveis de expressão sejam maiores, podendo resultar em uma amplificação não específica. Semelhante especificidade de amplificação, em torno de 60%, foi relatada para as *tags* de SAGE com um pequeno número de cópias. Entretanto, a especificidade de amplificação sofreu um aumento, em torno de 85%, para as *tags* com um maior número de cópias (CHEN et al. 2002a). Dados da literatura sugerem que a especificidade de amplificação poderia ser melhorada por meio da seleção de um grande número de colônias de cada produto de GLGI. No caso de GLGI-SAGE uma segunda etapa de seleção de colônias foi adotada como sendo uma maneira de gerar cerca de 10 – 15% a mais de fragmentos 3' de cDNA específicos (CHEN et al. 2002a). Entretanto, resultados semelhantes não foram obtidos para o GLGI-MPSS.

Dez dos 41 fragmentos de GLGI-MPSS foram confirmados como sendo extensões 3' específicas, devido à presença da cauda poli (dA), do sinal de poliadenilação e à ausência de sítios de restrição *Dpn II* dentro das seqüências amplificadas (tabela 10).

Quatro desses dez fragmentos 3' de cDNA, após análise com o programa BLAST, alinharam com seqüências transcritas de transcritos conhecidos e seis fragmentos alinharam com seqüências de ESTs derivadas principalmente de bibliotecas de cDNA normalizadas ou subtraídas, confirmando o baixo nível de expressão desses transcritos. Todas as seqüências transcritas correspondentes a transcritos conhecidos, bem como as seqüências das ESTs, foram submetidas ao GenBank, após a construção do banco de dados utilizado como referência. Esse fato, explica porque as *tags* de MPSS correspondentes a cada um desses 10 fragmentos foram originalmente classificadas como *tags* órfãs. Tais transcritos apesar de já terem sido completa ou parcialmente seqüenciados e disponibilizados em bancos de dados públicos, representam novos transcritos possivelmente regulados por c-erbB2 e, portanto, foram selecionados para validação da expressão diferencial por meio de PCR em Tempo Real.

A análise dos 15 outros fragmentos gerados pelo GLGI-MPSS provou ser útil na identificação de prováveis transcritos *antisense*, na identificação de *tags* alternativas geradas a partir de transcritos polimórficos, na identificação de formas alternativas de poliadenilação de transcritos já conhecidos, além de possibilitar a detecção de *tags* de MPSS artefatuais provenientes do anelamento interno do oligo dT durante a construção da biblioteca de MPSS. Primeiramente, 5 dos 41 fragmentos de GLGI-MPSS apresentaram alinhamento com seqüências de transcritos conhecidos depositadas no GenBank, com um nível de similaridade superior a 95%, mas na orientação oposta (3'- 5'). Todos os 5 fragmentos apresentaram cauda poli A e 3 deles apresentaram sinal de poliadenilação, sugerindo que essas seqüências sejam derivadas de transcritos *antisense* localizados na fita oposta da seqüência conhecida.

A existência desses transcritos *antisense* foi confirmada por meio de RT-PCR fita específica e será descrita em detalhes no item 4.7. Na figura 15, pode-se visualizar um exemplo de alinhamento de um fragmento de GLGI-MPSS correspondente a um transcrito *antisense* e a seqüência do transcrito *sense* já conhecido.

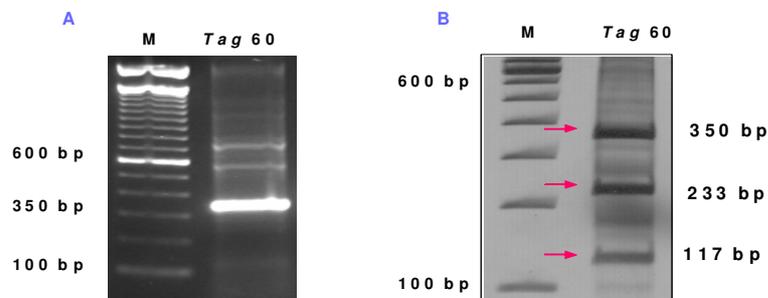


Legenda: (A) Resultado do BLAST N contra o banco de dados do GenBank entre a seqüência do transcrito *antisense* correspondente à tag 12 (Query) e a seqüência do transcrito *sense* (Sbjct) cujo número de acesso no GenBank é BC036460. (B) Representação esquemática do alinhamento entre o transcrito *sense* já conhecido e o transcrito *antisense* representado pela tag 12 (onde o retângulo verde representa o sítio da enzima de restrição e o retângulo azul representa a seqüência da tag).

Figura 15 - Alinhamento entre o transcrito *antisense* correspondente à tag 12 e a seqüência de mRNA do transcrito *sense*.

Para 3 dos 41 fragmentos de GLGI-MPSS, o alinhamento com um transcrito humano conhecido foi quase perfeito na região do iniciador *sense*, exceto pela substituição de uma única base localizada no sítio de restrição *Dpn* II. Após detalhadas análises dessas 3 seqüências, pôde-se concluir que as tags órfãs de MPSS foram derivadas de transcritos polimórficos conhecidos, nos quais a presença de um

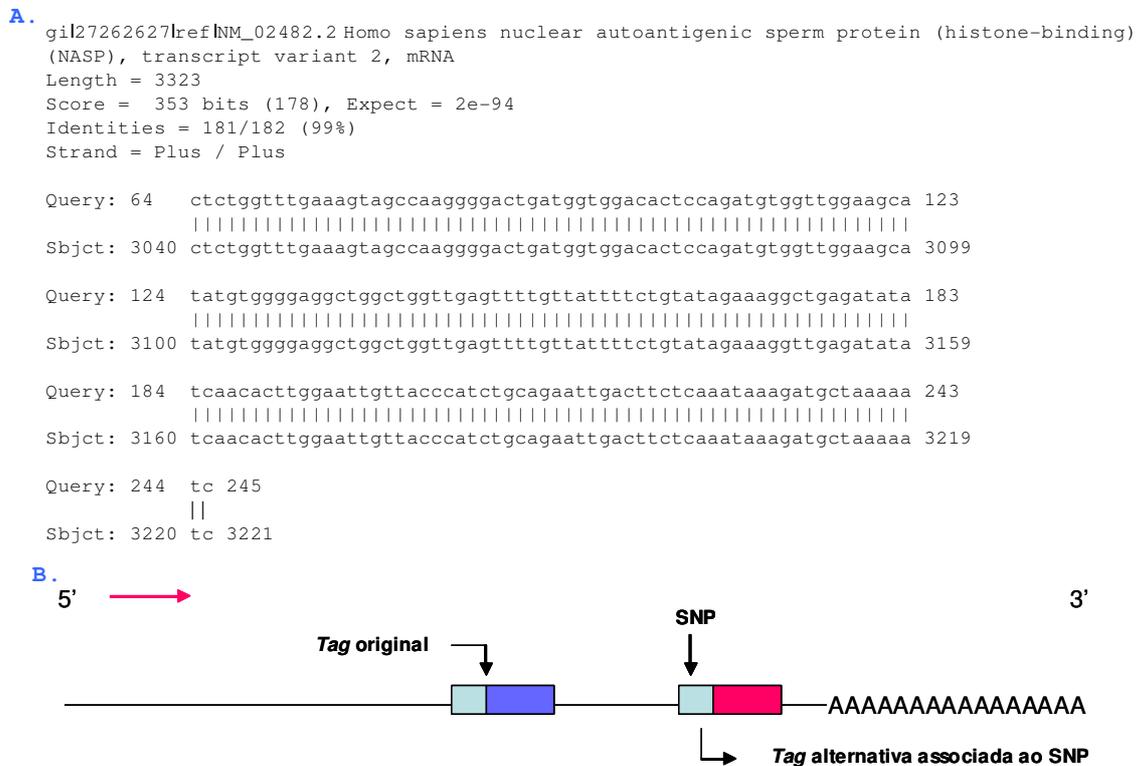
SNP na linhagem celular HB4a criou um sítio de restrição *Dpn* II localizado a 3' da *tag* original de MPSS, e não representado nas seqüências de cDNA utilizadas na associação *tag*-transcrito (tabela 10). A presença desses SNPs nos transcritos expressos na linhagem celular HB4a, gera, então, uma *tag* de MPSS alternativa, que inicialmente foi considerada uma *tag* órfã. Por exemplo, o fragmento de GLGI-MPSS derivado da *tag* órfã de MPSS **GATCTCTGGTTTGAAAG** alinhou com o transcrito NASP (NM_002482) do nucleotídeo 3.040 ao 3.323. O alinhamento observado foi quase perfeito exceto por uma única substituição de base (G/T), a qual cria um novo sítio *Dpn* II *downstream* ao sítio 3' original presente na seqüência depositada no GenBank. Essa substituição de base no transcrito NASP já foi descrita no banco de dados de SNPs do NCBI – versão 108 (SNPid RS1053941), e a presença desse SNP na linhagem celular HB4a foi confirmada por meio da amplificação e digestão com *Dpn* II do DNA genômico (Figura 16).



Legenda: (A) Amplificação GLGI-MPSS da *tag* alternativa de MPSS associada ao SNP correspondente a seqüência de mRNA NM_002482 a partir de cDNA da linhagem HB4a. (B) Tipagem do SNP por meio da amplificação do DNA genômico seguida de digestão com a enzima de restrição *Dpn*II. A região genômica flanqueando o SNP correspondente à *tag* 60 foi amplificada utilizando iniciadores específicos e DNA genômico da linhagem celular HB4a. Os fragmentos foram analisados em gel de agarose 1% e foram corados com brometo de etídeo. O padrão de peso molecular utilizado foi o 100bp (M).

Figura 16 - Validação experimental do transcrito polimórfico correspondente à *tag* 60.

Na Figura 17, pode-se visualizar o alinhamento ocorrido entre a seqüência do transcrito correspondente à *tag* 60 e a seqüência do transcrito polimórfico já conhecido.



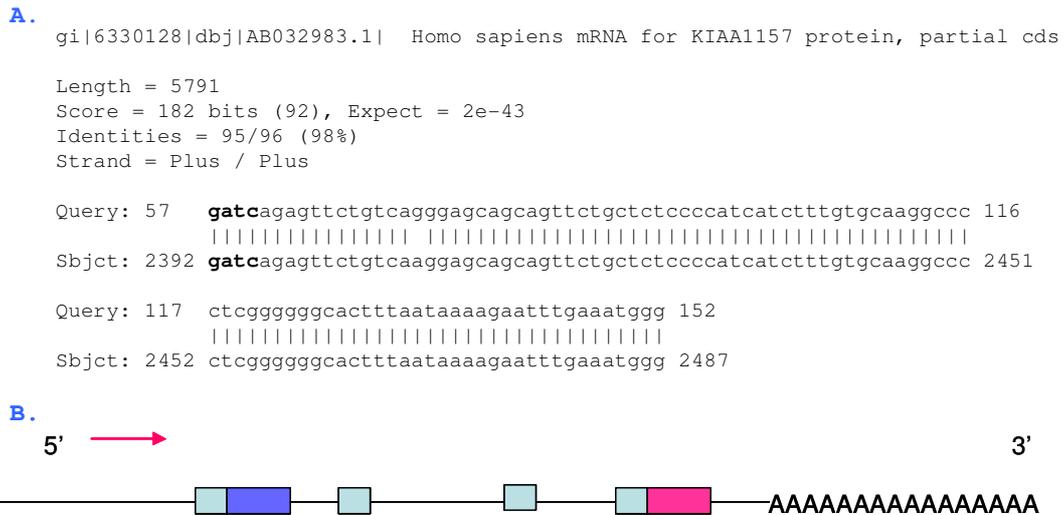
Legenda: (A) Resultado do BLAST N contra o banco de dados do GenBank entre a seqüência do transcrito correspondente à *tag* 60 (Query) e a seqüência do transcrito polimórfico NASP (Sbjct), cujo número de acesso é NM_002482. (B) Representação esquemática do alinhamento entre o fragmento gerado pelo GLGI-MPSS e o transcrito polimórfico em que: os retângulo verde e azul representam, respectivamente, o sítio de restrição e a seqüência da *tag* original encontrada no cDNA do transcrito NASP (2862-2865) e os retângulos verde e rosa representam, respectivamente, *tag* alternativa que foi originada devido ao polimorfismo na linhagem HB4a.

Figura 17 - Alinhamento entre o transcrito correspondente à *tag* 60 e a seqüência de mRNA do transcrito polimórfico NASP.

Esses resultados mostram que o GLGI-MPSS pode ser utilizado na identificação de *tags* alternativas de MPSS derivadas a partir de transcritos polimórficos e deram origem a um estudo paralelo em que o impacto dos SNPs na geração de *tags* alternativas de SAGE e MPSS foi avaliado (SILVA et al. 2004b) (Anexo 2). As *tags* alternativas associadas a SNPs foram identificadas por meio da construção de um banco de dados de referência, no qual as análises das seqüências de mRNA disponibilizadas pelo UniGene foram combinadas com as informações contidas no banco de dados de SNPs disponibilizado pelo NCBI. O resultado das análises permitiu a identificação de 2.020 *tags* alternativas associadas a SNPs (1100 *tags* alternativas de SAGE associadas a SNPs e 920 *tags* alternativas de MPSS associadas a SNPs). Além disso, os resultados indicaram que, até o momento, 8,6% de todos os transcritos humanos apresentam ao menos uma *tag* alternativa associada a um SNP. O banco de dados de *tags* alternativas associadas a SNPs foi integrado ao Sage Genie e pode ser consultado no endereço eletrônico <http://cgap.nci.nih.gov/SAGE>. Esses transcritos, também podem ser regulados por *c-erbB2* e, portanto, foram selecionados para validação da expressão diferencial por meio de PCR em Tempo Real.

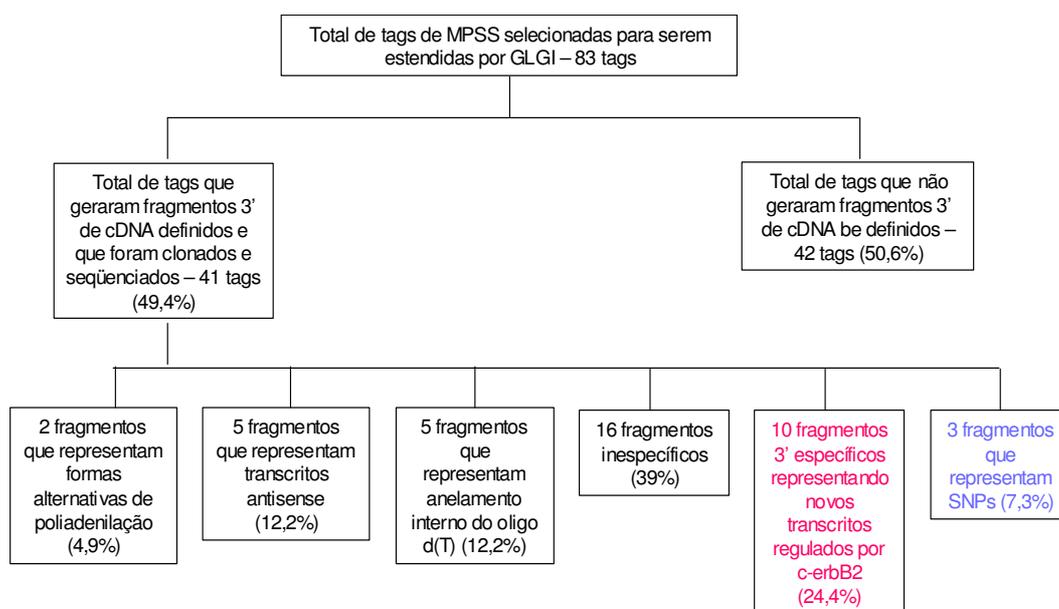
Por fim, 7 dos 41 fragmentos de GLGI-MPSS alinharam na região mediana de transcritos conhecidos na orientação 5'-3'. Cinco dessas extensões de GLGI-MPSS são provavelmente derivadas do anelamento interno do oligo dT durante a síntese de cDNA e, então, essas *tags* órfãs de MPSS podem ser classificadas como artefatuais (Tabela 10). Por exemplo, a seqüência parcial do fragmento de GLGI-MPSS derivado da *tag* órfã GATCCAAAAGTTCACTT alinhou com o transcrito MBP4 (BC034463) do nucleotídeo 373 ao 744. A seqüência completa do transcrito

MBP4 apresenta sítios *Dpn* II adicionais nos nucleotídeos 962, 977, 1363, 1528 e 1898 e segmentos poli A do nucleotídeo 751 a 762 e de 948 a 961 os quais podem ter servido como sítios para o anelamento interno do oligo d(T) (Figura 18). Entretanto, para 2 desses 7 fragmentos de GLGI-MPSS, não foi possível identificar a ocorrência de anelamento interno e eles foram então, classificados como sendo derivados de formas de poliadenilação alternativa ainda não caracterizadas (Figura 19). Por serem consideradas *tags* artefatuais ou correspondentes a formas de poliadenilação alternativa ainda não caracterizadas, os transcritos correspondentes a essas 7 *tags* não foram analisados por PCR em Tempo Real para confirmar a expressão diferencial entre as linhagens HB4a e HB4a-C5.2. Os resultados obtidos a partir das análises dos 41 fragmentos gerados pelo GLGI-MPSS, estão esquematizados de forma resumida na Figura 20.



Legenda: (A) Resultado do BLAST N contra o banco de dados do GenBank entre a sequência do transcrito correspondente à *tag* 65 (Query) e a sequência do transcrito conhecido (Sbjct), cujo número de acesso é AB032983. (B) Representação esquemática do alinhamento entre o fragmento gerado pelo GLGI-MPSS e o transcrito conhecido: os retângulos verde e azul representam, respectivamente, o sítio de restrição e à sequência da *tag* 65. Os retângulos verde e rosa representam, respectivamente, o sítio de restrição e a sequência da *tag* original. Também pode-se observar, na figura, a existência de outros sítios de restrição ao longo da sequência do transcrito porém, não há nenhum segmento de poli A em que possa ter ocorrido anelamento interno do oligo d(T) durante a síntese de cDNA. Por essa razão, acredita-se que o fragmento gerado pelo GLGI-MPSS seja derivado de poliadenilação alternativa.

Figura 19 - Alinhamento entre o transcrito correspondente à *tag* 65 e a sequência de mRNA do transcrito conhecido.



Legenda: Em rosa e em azul estão representados, respectivamente, os 10 fragmentos 3' específicos que representam novos transcritos regulados por c-erbB2 e os 3 fragmentos que representam SNPs. Esses fragmentos terão o padrão de expressão nas linhagens HB4a e HB4a-C5.2 avaliado através de PCR em Tempo Real.

Figura 20 - Representação esquemática dos resultados obtidos a partir da amplificação GLGI para as 83 tags órfãs de MPSS.

4.4 QUANTIFICAÇÃO DA EXPRESSÃO DOS FRAGMENTOS 3' ESPECÍFICOS E DOS TRANSCRITOS CORRESPONDENTES ÀS TAGS ALTERNATIVAS ASSOCIADAS À SNPS ATRAVÉS DE PCR EM TEMPO REAL NAS LINHAGENS CELULARES HB4a E HB4a-C5.2

A técnica de PCR em Tempo Real foi utilizada para validar a expressão diferencial entre as linhagens celulares HB4a e HB4a-C5.2 dos transcritos correspondentes às *tags* órfãs de MPSS que foram estendidas por meio de GLGI-MPSS. Foram selecionados para validação um total de 13 transcritos que correspondiam aos 10 fragmentos de GLGI-MPSS confirmados como sendo extensões 3' específicas (*tags* 7, 10, 13, 16, 28, 39, 44, 52, 76 e 81) e aos 3 fragmentos de GLGI-MPSS derivados de *tags* alternativas de MPSS associadas a SNPs (*tags* 6, 60 e 63).

Os iniciadores utilizados nos experimentos de PCR em Tempo Real foram desenhados com o auxílio do programa Primer Express (Applied Biosystems) e as reações foram realizadas no aparelho 7300 Real Time PCR System (Applied Biosystems). Para avaliar a expressão diferencial dos transcritos selecionados, foi utilizado o sistema de SyberGreen e o método de quantificação relativa, adotando, como amostra de referência, o cDNA da linhagem HB4a e, como gene normalizador, a β -actina. Para a padronização das reações foram variadas as concentrações dos iniciadores (100nM, 200nM, 300nM e 400nM) e as quantidades de cDNA (5ng e

50ng). A especificidade das reações foi avaliada por meio da curva de dissociação do fragmento amplificado seguida de eletroforese em gel de agarose.

As concentrações adequadas de cada um dos iniciadores e a quantidade de cDNA para cada transcrito estão listadas na Tabela 11.

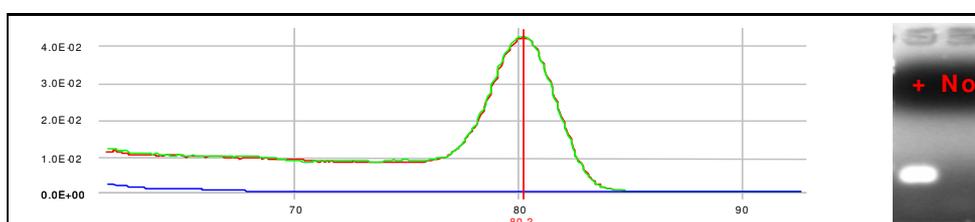
Tabela 11 - Concentração dos iniciadores e quantidade de cDNA utilizadas nas reações de PCR em Tempo Real.

Transcritos representados pelas tags	Concentração dos iniciadores	Quantidade de cDNA
7	400nM	5ng
10	400nM	5ng
13	400nM	50ng
16	400nM	50ng
28	300nM	50ng
39	200nM	5ng
44	400nM	5ng
52	300nM	50ng
76	300nM	5ng
81	300nM	5ng
SNP6	100nM	50ng
SNP60	100nM	50ng
SNP63	100nM	5ng
β -actina	200nM	5 e 50ng

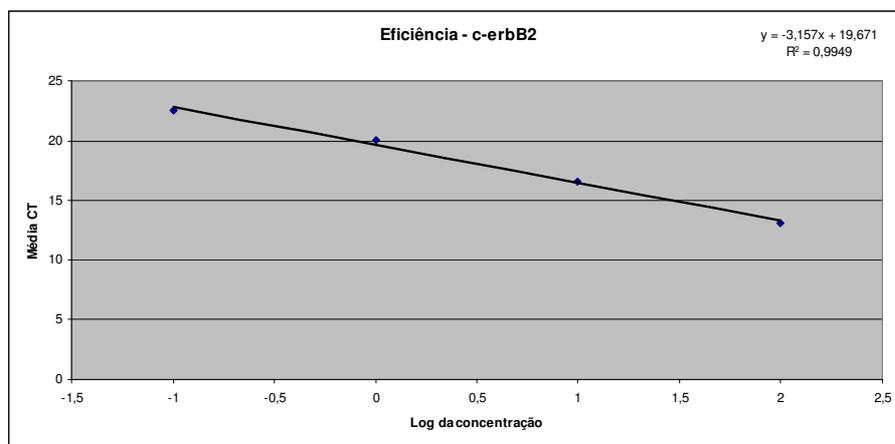
Depois de ajustadas as concentrações dos iniciadores e a quantidade de cDNA a serem utilizadas, foi determinada a eficiência de amplificação dos iniciadores, utilizando, para tanto, a equação $E = 10^{(-1/slope)}$ (PFAFFL 2001), em que o *slope* corresponde à inclinação da reta obtida quando se analisa a variação do C_T em função

do *log* de diferentes quantidades de cDNA. As quantidades de cDNA utilizadas para o cálculo de eficiência foram 100, 10, 1 e 0,1ng de cDNA da linhagem celular HB4a-C5.2. Na figura 21, pode-se visualizar o gráfico referente à curva de dissociação gerada para o fragmento correspondente ao oncogene c-erbB2, o gel de agarose realizado para analisar a especificidade do fragmento amplificado e também o gráfico gerado para o cálculo da eficiência de amplificação.

A.



B.



Legenda: (A) A especificidade da reação de amplificação após a etapa de padronização foi avaliada por meio do perfil da curva de dissociação do fragmento amplificado, realizada após a corrida pelo aparelho de PCR em Tempo Real e por meio de eletroforese em gel de agarose 2%; (B) A eficiência da reação de amplificação foi calculada utilizando diluições seriadas do cDNA como molde. Após a reação foi construída uma reta na qual o valor de CT foi analisado em função do log da diluição. A inclinação da reta foi então utilizada para calcular a eficiência de amplificação segundo a fórmula: $E = 10^{(-1/slope)}$, em que o *slope* corresponde à inclinação da reta.

Figura 21 - Padronização da reação de amplificação por PCR em Tempo Real para o oncogene c-erbB2.

As eficiências obtidas para cada um dos transcritos avaliados e para os genes normalizadores estão listadas na Tabela 12.

Tabela 12 - Eficiência de amplificação dos iniciadores utilizados nas reações de PCR em Tempo Real.

Transcritos	Eficiência	Transcritos	Eficiência
7	100%	76	100%
10	97%	81	99%
13	83%	6	83%
16	100%	60	100%
28	100%	63	100%
39	96%	c-erbB2	100%
44	100%	B-actina	97%
52	100%	GAPDH	91%
BCR	82%	HPRT	97%

Legenda: As eficiências tanto dos transcritos de interesse como dos genes constitutivos foram obtidas por meio da utilização da equação: $E = 10^{(-1/slope)}$.

Como se pode observar na Tabela 12, para alguns transcritos, a eficiência de amplificação foi inferior a 100%, não permitindo a utilização da equação $2^{-\Delta\Delta CT}$ para calcular a expressão relativa dos diferentes transcritos (LIVAK e SCHMITTGEN 2001). Por essa razão, o modelo matemático proposto por PFAFFL (2001) foi utilizado para calcular a expressão relativa dos diferentes transcritos nas linhagens HB4a e HB4a-C5.2. Esse modelo leva em consideração a eficiência de amplificação dos iniciadores no cálculo da expressão relativa, permitindo a quantificação relativa da expressão quando as eficiências de amplificação dos iniciadores são diferentes entre si.

Os dados da expressão relativa dos diferentes transcritos determinados por PCR em Tempo Real estão relacionados na Tabela 13.

Tabela 13 - Dados de expressão relativa dos novos transcritos determinadas por PCR em Tempo Real e a frequência com que cada tag foi encontrada nas bibliotecas HB4a e HB4a-C5.2 de acordo com o MPSS.

Transcritos correspondentes às tags:	Dados de MPSS		Cr amostra	Cr controle endógeno β -actina	$R = (E_{alvo})^{D_{CTalvo}} / (E_{normalizador})^{D_{CTref}} \times f_{frequência}$
	Razão HB4a-C5.2/HB4a	Linhagens			
7	7,8	HB4a	29,03±0,18	19,89±0,43	11,9
		HB4a-C5.2	27,05±0,21	21,52±0,5	
10	6,5	HB4a	33,19±0,18	19,89±0,43	2,4
		HB4a-C5.2	33,51±0,07	21,52±0,5	
13	10,5	HB4a	30,13±0,34	15,08±0,08	2,7
		HB4a-C5.2	29,69±0,37	16,15±0,16	
16	-5	HB4a	25,7±0,064	15,08±0,08	-4
		HB4a-C5.2	28,85±0,15	16,15±0,16	
28	9	HB4a	30,22±0,22	15,08±0,08	20,6
		HB4a-C5.2	26,9±0,05	16,15±0,16	
39	8	HB4a	33,7±0,24	19,89±0,43	2,3
		HB4a-C5.2	34,09±0,07	21,52±0,5	
44	21	HB4a	29,76±0,23	19,89±0,43	8
		HB4a-C5.2	28,35±0,14	21,52±0,5	
52	-6	HB4a	28,04±0,47	15,08±0,08	1,9
		HB4a-C5.2	28,15±0,12	16,15±0,16	
76	-13,4	HB4a	26,19±0,07	19,89±0,43	1,5
		HB4a-C5.2	27,16±0,14	21,52±0,5	
81	-13	HB4a	33,15±0,04	19,89±0,43	2,7
		HB4a-C5.2	33,29±0,00	21,52±0,5	
6	5,3	HB4a	31,34±0,2	15,08±0,08	5,4
		HB4a-C5.2	29,65±0,04	16,15±0,16	
60	-9	HB4a	22,07±0,05	15,08±0,08	3
		HB4a-C5.2	21,43±0,05	16,15±0,16	
63	-6	HB4a	19,79±0,03	19,89±0,43	1
		HB4a-C5.2	18,06±0,02	21,52±0,5	

Legenda: A quantificação relativa da expressão dos diferentes transcritos nas linhagens HB4a e HB4a-C5.2 foi obtida através da utilização do modelo matemático proposto por PFAFFL (2001).

Com base nos dados obtidos, pode-se verificar que dos 13 transcritos avaliados por PCR em Tempo Real, 4 (transcritos correspondentes às tags 7, 28, 44 e 60) apresentaram uma maior expressão na linhagem HB4a-C5.2 em relação à linhagem HB4a e podem ser considerados como novos transcritos putativamente

regulados por c-erbB2. Foram considerados diferencialmente expressos apenas os transcritos que apresentaram um nível de expressão superior a 3. Níveis de expressão abaixo desse *cut off* não foram considerados significativos. Os resultados gerados estão em concordância com os dados inicialmente inferidos pelo MPSS. De acordo com o experimento de MPSS, esses 4 transcritos foram categorizados como sendo superexpressos na linhagem HB4a-C5.2 em relação a HB4a.

O transcrito correspondente à *tag* 16 também se mostrou diferencialmente expresso entre as duas linhagens, porém um maior nível de expressão foi detectado na linhagem HB4a em relação à linhagem HB4a-C5.2. Esse resultado também confirmou o dado inicial de MPSS, no qual a *tag* correspondente a esse transcrito foi encontrada com uma frequência 5 vezes maior na linhagem HB4a em relação à linhagem HB4a-C5.2 (Tabela 9).

4.4.1 Validação da expressão diferencial dos transcritos regulados por c-erbB2 em diferentes linhagens tumorais de mama por meio de PCR em Tempo Real

As análises de expressão diferencial, por meio de PCR em Tempo Real, foram realizadas apenas para os transcritos que apresentaram super-expressão na linhagem HB4a-C5.2.

O nível de expressão dos 4 novos transcritos, putativamente regulados por c-erbB2, foi também analisado em 9 linhagens tumorais de mama que apresentavam ou não super-expressão do oncogene c-erbB2. Para tanto, foram selecionadas 6 linhagens celulares de carcinoma de mama (BT20; MCF7; MDA-MB-468; MDA-MB-453; SKBR3 e ZR75.30), que, segundo dados da literatura, apresentavam ou não super-expressão do oncogene c-erbB2 (KRAUS et al. 1987; SAVINAINEN et al.

2002; RHODES et al. 2004; KAURANIEMI et al. 2004; LACROIX e LECLERCQ 2004). Para essas linhagens, dados relacionados com a super-expressão da proteína c-erbB2 e da amplificação do gene, obtidos por meio de imunohistoquímica e dot blot, respectivamente, estavam disponíveis (Tabela 14) (KRAUS et al. 1987). As linhagens MDA-MB-435, PMC42 e SKBR7 foram selecionadas por estarem disponíveis em nosso laboratório. Entretanto, nenhuma informação a respeito dos níveis de expressão do mRNA, da proteína e da amplificação do gene c-erbB2 estavam disponíveis.

Tabela 14 - Dados de expressão e amplificação do oncogene c-erbB2 em linhagens celulares tumorais de mama.

Linhagen celular	Nível de expressão do mRNA de c-erbB2	Número de cópias do gene c-erbB2
MCF7	1	1
SKBR3	128	4-8
MDA-MB-453	64	2
ZR75.30	4	<1
BT20	1	1
MDA-MB-468	1	-

Fonte: Modificado de KRAUS et al. (1987).

Legenda: Na tabela estão relacionados os resultados referentes a super-expressão do mRNA de c-erbB2 e também dados de amplificação do gene c-erbB2.

Após esse levantamento, os níveis de expressão do oncogene c-erbB2 nessas diferentes linhagens foi avaliado no laboratório por PCR em Tempo Real (Tabela 15). Com base nos resultados obtidos, as linhagens foram classificadas em duas categorias: com baixo nível de expressão de c-erbB2 (MDA-MB-435; PMC42; BT20; SKBR7; MCF7 e MDA-MB-468) e com alto nível de expressão de c-erbB2

(MDA-MB-453; SKBR3 e ZR75.30). É interessante notar que a super-expressão do oncogene c-erbB2 pode ocorrer mesmo na ausência de amplificação gênica e que os níveis de expressão do mRNA, nem sempre, se correlacionam com os níveis de expressão da proteína, sugerindo que diferentes mecanismos moleculares podem resultar na super-expressão de c-erbB2 em células tumorais de mama (KRAUS et al. 1987). Os resultados obtidos nos experimentos estão de acordo com outros trabalhos descritos na literatura, que também avaliaram o nível de expressão do oncogene c-erbB2 utilizando algumas das linhagens acima mencionadas (LACROIX e LECLERQ 2004; KAURANIEMI et al. 2004; RHODES et al. 2004). Os experimentos de PCR em Tempo Real foram realizados como descritos anteriormente. A linhagem utilizada como referência nesses experimentos foi a HB4a, e o gene normalizador utilizado foi a β -actina. Os dados de expressão diferencial foram obtidos por meio do modelo matemático proposto por PFAFFL (2001) (Tabela 15).

Tabela 15 - PCR em Tempo Real para avaliar o nível de expressão diferencial do oncogene c-erbB2 e dos 4 novos transcritos em linhagens tumorais de mama.

Linhagens	c-erbB2	Transcrito representado pela tag 7	Transcrito representado pela tag 28	Transcrito representado pela tag 44	Transcrito representado pela tag 6	C _T controle endógeno: B-actina
	Nível de expressão (C _T amostra)					
HB4a	1 28,7±0,1	1 29,6±0,1	1 31,4±0,1	1 31,2±0,03	1 31,2±0,1	1 18,8±0,0
MDA-MB-435	4 (26,4±0,03)	48,7 (23,7±0,02)	0,9 (31,3±0,3)	24,4 (26,3±0,1)	34,6 (25±0,1)	(18,5±0,1)
PMC42	9,7 (26,1±0,06)	96 (23,7±0,02)	96 (25,5±0,03)	36,4 (26,7±0,06)	17 (27,3±0,06)	(19,5±0,06)
BT20	11,4 (24,7±0,1)	0,4 (30,3±0,31)	28,1 (26,1±0,01)	13,1 (27±0,06)	4,6 (28,1±0,03)	(18,3±0,02)
SKBR7	14,5 (26,7±0,03)	62,2 (25,5±0,1)	142,9 (26,1±0,2)	124,4 (26,1±0,05)	18,5 (28,5±0,1)	(20,7±0,0)
MCF7	15 (24,3±0,1)	0,8 (29,5±0,4)	170,2 (23,5±0,04)	9,9 (27,4±0,1)	32,1 (24,9±0,1)	(18,3±0,06)
MDA-MB-468	26,9 (26,2±0,02)	0,3 (33,6±0,2)	142 (26,5±0,09)	132,5 (26,4±0,02)	103,7 (26,1±0,04)	(21,1±0,03)
MDA-MB-453	279,3 (23,9±0,3)	14,2 (29,1±0,02)	1375,6 (24,3±0,04)	16,3 (30,5±0,06)	112,5 (27,2±0,02)	(22,2±0,03)
SKBR3	1405 (20,8±0,1)	0 -	216,2 (25,6±0,02)	1,4 (32,7±0,3)	164,5 (25±0,04)	(20,8±0,01)
ZR75.30	1930,8 (24,3±0,1)	0,2 (35,6±0,2)	554,5 (26,1±0,03)	91,5 (28,5±0,5)	131,7 (27,5±0,07)	(22,7±0,07)

Legenda: A quantificação relativa da expressão diferencial do oncogene c-erbB2 e dos novos transcritos foi obtida por meio da utilização do modelo matemático proposto por PFAFFL (2001). A linhagem HB4a foi utilizada como amostra de referência e a β -actina foi utilizada como gene normalizador. Em azul, estão representados os valores de C_Ts e desvio padrão obtidos para o c-erbB2 e para cada um dos novos transcritos na linhagem HB4a. Em preto, estão representados os valores de C_Ts e desvio padrão obtidos para o c-erbB2 e para cada um dos novos transcritos nas diferentes linhagens tumorais de mama. Em violeta, estão listados os valores que representam o nível de expressão diferencial da linhagem HB4a, do oncogene c-erbB2 e de cada um dos novos transcritos nas diferentes linhagens tumorais de mama.

O padrão de expressão dos transcritos correspondentes às tags 7 e 44 parecem não estar diretamente relacionados com o padrão de expressão do oncogene c-erbB2, uma vez que os níveis de expressão mais elevados desses dois transcritos foram

detectados respectivamente, nas linhagens PMC42 e MDA-MB-468 que apresentam um baixo nível de expressão de c-erbB2. Para o transcrito correspondente à *tag* 7 a ausência de correlação com a super-expressão de c-erbB2 parece mais evidente, uma vez que esse transcrito apresenta pouca ou nenhuma expressão, respectivamente, nas linhagens celulares ZR75.30 e SKBR3 que apresentam os maiores níveis de expressão de c-erbB2.

Para os transcritos correspondentes às *tags* 28 e 6 foi possível observar uma correlação com a super-expressão de c-erbB2 para a maioria das linhagens analisadas. Os níveis de expressão mais elevados de ambos os transcritos foram observados nas linhagens que apresentam níveis elevados de c-erbB2 (MDA-MB-468; MDA-MB-453; SKBR3 e ZR75.30). Além disso, também foi possível observar que o menor nível de expressão do transcrito correspondente à *tag* 28 foi detectado na linhagem MDA-MB-435, a qual também apresentou o menor nível de expressão de c-erbB2.

Embora não tenha sido possível estabelecer uma correlação perfeita entre a super-expressão de c-erbB2 e a super-expressão dos 4 novos transcritos em algumas das linhagens tumorais utilizadas, os resultados sugerem uma relação funcional entre os novos transcritos e o oncogene c-erbB2.

4.4.2 Quantificação da expressão dos transcritos diferencialmente expressos nas linhagens celulares HB4a e HB4a-C5.2 por meio de PCR em Tempo Real em amostras tumorais de mama

Para dar início aos experimentos de quantificação dos novos transcritos diferencialmente expressos por meio de PCR em Tempo Real, foram selecionadas 49

amostras de carcinomas ductais invasivos de mama. Dentre as 49 amostras tumorais, 27 amostras apresentaram super-expressão da proteína c-erbB2 e 22 amostras não apresentaram super-expressão da proteína c-erbB2. Inicialmente, as amostras foram classificadas com base nos dados de imunohistoquímica disponibilizados pelo Departamento de Anatomia Patológica do Hospital AC Camargo e do Hospital do Câncer Alfredo Abrão. Após a seleção baseada na imunohistoquímica, o nível de expressão de c-erbB2 nessas amostras foi quantificado por PCR em Tempo Real. Nessas análises, foi utilizado um *pool* de 3 amostras de mama normal como referência. No entanto, verificou-se que não havia uma correlação perfeita entre os dados de imunohistoquímica e a quantificação dos níveis de c-erbB2 por PCR em Tempo Real. A exemplo do que havia sido observado para as linhagens, a existência de discrepâncias entre a expressão da proteína, a expressão do mRNA e a amplificação do oncogene c-erbB2 já foi relatada em diversos estudos (MHHALOVÁ et al. 2003). Uma possível explicação para a discrepância encontrada nos dados de expressão da proteína c-erbB2 e super-expressão do mRNA do oncogene seria a subjetividade das análises de imunohistoquímica. As amostras não foram analisadas em um único momento, sendo possível que a detecção de c-erbB2 tenha sido feita utilizando diferentes anticorpos e as amostras analisadas por diferentes patologistas. Dados da literatura (LEBEAU et al. 2001) relatam alguns estudos que avaliaram e correlacionaram a expressão da proteína c-erbB2, por meio de imunohistoquímica, em amostras tumorais e a presença de amplificação gênica, por meio de FISH (Fluorescent In Situ Hybridization) nas mesmas amostras. Esses estudos indicaram que apenas amostras classificadas pela imunohistoquímica como 3+ podem seguramente ser consideradas positivas quanto à super-expressão da

proteína c-erbB2 por amplificação gênica. O mesmo não ocorre com as amostras 2+, que podem ou não apresentar amplificação gênica. Em um outro estudo que também avaliou o nível de expressão da proteína c-erbB2, por meio de imunohistoquímica, foi mostrado que são gerados resultados falso negativos em 28% dos casos detectados como sendo positivos (por FISH), e isso pode ser explicado, por exemplo, pela destruição do epítipo HER2. Por outro lado, a imunohistoquímica também gera resultados falso positivos, em 12% dos casos (ELLIS et al. 2005).

O nível de expressão do oncogene c-erbB2 utilizado para classificar uma amostra como c-erbB2 positiva, varia de forma considerável de um trabalho para outro. Em um trabalho realizado por MHALOVÁ et al. (2003), 39 amostras de carcinoma ductal invasivo tiveram seus níveis de mRNA de c-erbB2 avaliados por meio de PCR em Tempo Real. Esse experimento de quantificação relativa utilizou o gene GAPDH como normalizador, e como amostra referência foi utilizado o RNA total purificado a partir de uma linhagem celular que expressa, de forma constitutiva, o oncogene c-erbB2. Essa amostra referência é parte do kit LightCycler-HER2/neu RNA Quantification (Roche Diagnostics GmbH). Entre as 39 amostras analisadas, 14 amostras que não apresentavam amplificação do oncogene c-erbB2 e nem significativa expressão da proteína foram utilizadas para estabelecer o nível do mRNA de c-erbB2, a partir do qual a expressão do oncogene deve ser considerada aumentada. Os valores atribuídos aos níveis de mRNA considerados não significativos foram >0 e < 19 . A amostra utilizada como controle negativo das reações (sem amplificação do gene e sem expressão da proteína) apresentou um nível de expressão do mRNA de c-erbB2 correspondente a 10.9. Já para os casos em que a

amplificação do oncogene foi detectada e a expressão da proteína mostrou-se acentuada (3+), os níveis de mRNA variaram em torno de 84,5.

Considerando a grande variabilidade existente na literatura com relação aos níveis de expressão do mRNA de c-erbB2 e baseando -se no trabalho realizado por MHHALOVÁ et al. (2003), no presente trabalho, foram consideradas como amostras c-erbB2 negativas, as amostras que apresentaram um nível de expressão de c-erbB2 até 3 vezes maior do que o *pool* de amostras normais. Da mesma forma, foram consideradas amostras c-erbB2 positivas, as amostras que apresentaram um nível de expressão pelo menos 20 vezes maior do que o *pool* de amostras normais utilizadas como referência.

Ao contrário das análises com linhagens celulares, a utilização de amostras tumorais requer cuidados adicionais relacionados, principalmente, com a qualidade da amostra e com a seleção dos genes normalizadores. A qualidade dos RNAs extraídos a partir das amostras tumorais foi analisada em gel de agarose como descrito anteriormente para as linhagens celulares. As amostras nas quais as bandas de 28S e 18S não estavam representadas nas proporções esperadas foram descartadas. No entanto, a análise em gel de agarose não permite avaliar a presença de inibidores que podem influir na síntese de cDNA e na reação de PCR em Tempo Real. A presença desses inibidores interfere diretamente na eficiência de amplificação da reação. Dessa forma, é possível inferir a presença de inibidores em uma dada amostra determinando a eficiência de amplificação com iniciadores já padronizados e de eficiência conhecida.

Com o objetivo de avaliar a qualidade das amostras, foi utilizado o aparelho *Rotor-gene3000 Real Time Thermal Cycler* (Corbett Research), que é capaz de

avaliar a eficiência de amplificação em amostras individualmente por meio de um *software* chamado *Rotor-Gene Real-Time Analysis Software 6.0 (Build 27)*. O resultado desse experimento indicou que a eficiência de amplificação do gene β -actina não variou significativamente entre as amostras, indicando a ausência de substâncias inibidoras e confirmando a boa qualidade do RNA.

A escolha de um normalizador adequado é outro fator de grande importância e influência nos experimentos de PCR em Tempo Real. No método de quantificação relativa, a padronização dos experimentos requer a utilização de um gene normalizador cuja expressão não seja variável nas diferentes amostras analisadas (PFAFFL 2004). Entretanto, sabe-se que o nível de expressão de grande parte dos genes utilizados como normalizadores pode variar consideravelmente entre diferentes indivíduos, diferentes tipos celulares, diferentes estágios de desenvolvimento, em resposta a tratamentos experimentais, ou até mesmo como resultado do crescimento neoplásico. Outro fator importante na escolha do gene normalizador é o seu nível de expressão. Um normalizador torna-se adequado quando seu nível de expressão é comparável ao nível de expressão do transcrito alvo (KOK et al. 2005). Assim, um normalizador universal ainda não foi identificado e, portanto, a forma mais adequada e mais confiável de normalização é a utilização de uma combinação de genes normalizadores cujo nível de expressão seja pouco variável e compatível com o nível de expressão do transcrito alvo (MORSE et al. 2005).

Com base em dados da literatura (VANDESOMPELE et al. 2002), foram selecionados alguns dos genes mais comumente utilizados como normalizadores em experimentos de PCR em Tempo Real: β -actina, GAPDH, HPRT e BCR. A utilização conjunta dos 4 genes normalizadores foi possível através de uma

ferramenta, disponível *on line*, desenvolvida para o *Microsoft Excel* e que foi denominada *geNorm* (VANDESOMPELE et al. 2002). O *geNorm* permitiu determinar os genes normalizadores mais estáveis no conjunto de amostras e calculou um fator de normalização da expressão gênica para cada amostra, com base na média geométrica dos valores obtidos para os diferentes normalizadores. A expressão diferencial de cada transcrito nas amostras tumorais foi calculada a partir desse fator de normalização. As análises do *geNorm* indicaram que o gene da β -actina foi o que apresentou maior variabilidade de expressão nas amostras tumorais sendo, portanto, excluído das análises posteriores.

Assim, foram selecionadas entre as 49 amostras tumorais, 12 amostras c-erbB2 negativas e 11 amostras c-erbB2 positivas. Os níveis de expressão dos 4 novos transcritos foram analisados por meio de PCR em Tempo nas 23 amostras selecionadas. Na tabela 16 estão relacionadas as 23 amostras selecionadas, bem como os níveis de expressão do oncogene c-erbB2, e, também, dos 4 novos transcritos.

Tabela 16 - Dados referentes à expressão diferencial dos novos transcritos e do oncogene c-erbB2 obtidos por meio de PCR em Tempo Real, em amostras de carcinomas de mama.

	Oncogene c-erbB2	Transcrito correspondente a tag 7	Transcrito correspondente a tag 28	Transcrito correspondente a tag 44	Transcrito correspondente a tag 6
12	2,7	0,3	5,1	1,7	7,5
17	0,5	0,5	0,2	0,3	1,4
21	2,1	0,2	1,8	0,5	2,5
51	0,3	0,1	2,2	0,5	1,6
52	0,3	0,2	1,1	0,5	1,0
54	2,9	0,3	2,5	0,5	2,8
55	0,9	0,2	3,3	1,9	6,3
63	3,0	0,9	6,0	0,8	8,5
12T	1,2	0,6	1,4	0,3	2,8
24T	0,9	0,2	2,1	0,1	1,3
38T	1,2	0,4	2,3	0,2	2,0
50T	0,7	0,2	3,8	0,4	3,5
MÉDIA	1,4	0,3	2,6	0,6	3,4
11	22,0	0,3	3,2	0,8	6,7
38	42,0	3,0	12,0	2,0	7,2
45	20,9	0,4	2,4	0,7	1,6
47	48,9	0,2	1,2	0,3	2,3
57	235,9	3,0	12,0	1,1	6,1
58	60,0	1,0	7,5	1,2	4,4
59	47,8	0,6	3,9	0,5	4,9
60	98,0	1,3	14,1	0,8	6,1
66	134,1	0,6	4,2	1,8	7,9
67	122,6	1,9	10,1	5,4	11,3
69	31,1	0,3	1,8	1,0	2,5
MÉDIA	78,5	1,2	6,6	1,4	5,5

Legenda: As amostras tumorais foram selecionadas com base nos níveis do oncogene c-erbB2. Em vermelho, estão representadas as amostras c-erbB2 negativas e, em azul, estão representadas as amostras c-erbB2 positivas. Em destaque, estão relacionados os níveis médios de expressão do oncogene c-erbB2 bem como dos 4 novos transcritos, para o grupo de amostras negativas e para o grupo de amostras positivas.

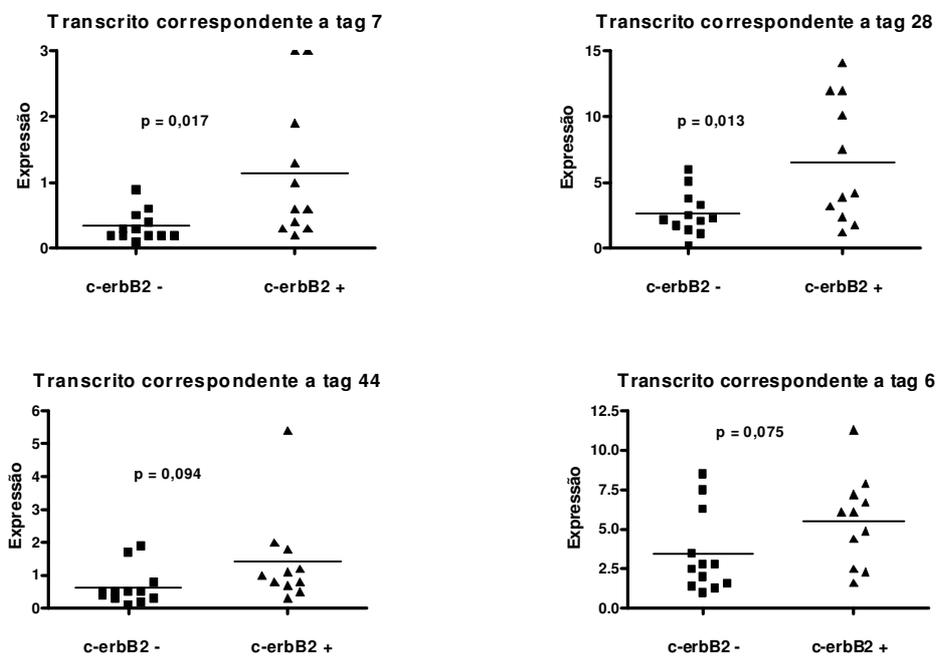
Com base nos dados de expressão apresentados na tabela 16, pode-se observar que não há uma clara correlação entre a super-expressão de c-erbB2 e a expressão aumentada dos novos transcritos, quando as amostras são analisadas individualmente. Nem todas as amostras c-erbB2 positivas apresentam expressão aumentada dos novos transcritos e a mesma pode ser observada em algumas amostras c-erbB2 negativas. Por exemplo, para os transcritos correspondentes às tags 7 e 44,

foi possível observar que, na amostra 57, que apresentou o maior nível de expressão de c-erbB2 (235.9), ambos os transcritos apresentam menores níveis de expressão, 3.0 e 1.1, respectivamente. Um outro exemplo são os transcritos correspondentes às *tags* 28 e 6. Os níveis de expressão de ambos os transcritos foram relativamente elevados, 6.0 e 8.5, respectivamente, na amostra 63, cujo nível de expressão do oncogene c-erbB2 não foi considerado significativo (3.0). No entanto, quando se avalia o nível médio de expressão dos diferentes transcritos nas amostras c-erbB2 positivas e c-erbB2 negativas, é possível notar que existe uma diferença, em alguns casos estatisticamente significativa, entre os dois grupos.

Como se pode observar na Tabela 16, o nível médio de expressão do transcrito correspondente à *tag* 7 foi 4 vezes maior (1,2) no grupo de amostras c-erbB2 positivas em relação ao grupo de amostras c-erbB2 negativas (0,3) sendo esta diferença de expressão estatisticamente significativa ($p=0,017$). Para o transcrito correspondente à *tag* 28, a diferença de expressão entre as amostras c-erbB2 positivas e c-erbB2 negativas também se mostrou estatisticamente significativa ($p=0,013$). O nível médio de expressão desse transcrito, foi 2,5 vezes maior no grupo de amostras c-erbB2 positivas (6,6) em relação ao grupo de amostras c-erbB2 negativas (2,6). No caso dos transcritos correspondentes às *tags* 44 e 6, também foi possível observar uma correlação positiva entre o nível médio de expressão de ambos os transcritos e a super-expressão de c-erbB2. O transcrito correspondente à *tag* 44 apresentou um nível médio de expressão 2,3 vezes maior no grupo de amostras c-erbB2 positivas (1,4) em relação ao grupo de amostras c-erbB2 negativas (0,6). Já o transcrito correspondente à *tag* 6 apresentou um nível médio de expressão 1,6 vezes maior no grupo das amostras c-erbB2 positivas (5,5) em relação ao grupo das amostras c-

erbB2 negativas (3,4). Entretanto, a diferença de expressão entre as amostras c-erbB2 positivas e c-erbB2 negativas não foi estatisticamente significativa para nenhum dos dois transcritos ($p=0,094$ e $p=0,075$ respectivamente).

Os resultados das análises de expressão diferencial dos novos transcritos nas amostras tumorais de mama c-erbB2 positivas e c-erbB2 negativas foram representados sob a forma de gráficos e podem ser visualizados na Figura 22.



Legenda: A expressão relativa foi obtida por PCR em Tempo Real. Nessas análises utilizamos um *pool* de 3 amostras de mama normal como referência e os genes normalizadores utilizados foram: BCR, HPRT e GAPDH.

Figura 22 - Representação gráfica das análises de expressão diferencial dos transcritos correspondentes às *tags* 7, 28, 44 e 6 em amostras tumorais de mama c-erbB2 negativas e c-erbB2 positivas.

A ausência de uma correlação individual entre a super-expressão do oncogene c-erbB2 e a expressão aumentada dos novos transcritos no grupo de amostras pode

ser explicada em um primeiro momento, pela heterogeneidade de cada amostra. Os tecidos tumorais, ao contrário das linhagens celulares, são constituídos por diversos tipos celulares o que, muitas vezes, mascara a avaliação do padrão de expressão de transcritos específicos nas amostras tumorais (SINGER et al. 2003). Embora o presente trabalho não tenha encontrado uma correlação perfeita entre a super-expressão do oncogene c-erbB2 e os 4 novos transcritos, em todas as amostras analisadas, os resultados sugerem a existência de uma associação entre o oncogene c-erbB2 e, principalmente, os transcritos representados pelas *tags* 7 e 28.

4.5 CARACTERIZAÇÃO DOS NOVOS TRANSCRITOS REGULADOS POR C-ERBB2 E VALIDADOS POR PCR EM TEMPO REAL

Entre os quatro transcritos que apresentaram expressão diferencial entre as linhagens celulares HB4a e HB4a-C5.2, os transcritos correspondentes às *tags* 7, 28 e 6 apresentaram similaridade com seqüências de transcritos já conhecidos e o transcrito representado pela *tag* 44 apresentou similaridade apenas com *ESTs*.

O transcrito correspondente à *tag* 7 (AK123675) está localizado no cromossomo 1q21.1 e possui 3.834 pares de bases. A seqüência completa desse transcrito apresenta uma evidente estrutura de *splicing* quando alinhada com a seqüência genômica, sendo formada por 7 éxons. As seqüências correspondentes a esse transcrito foram agrupadas no UniGene *cluster* Hs.515748 e correspondem ao transcrito humano ortólogo do gene *Lix1* de camundongo. Esse transcrito apresenta

uma fase aberta de leitura composta por 337 aminoácidos. O gene Lix1 (Limb Expression 1), que representa um dos membros de uma família gênica, foi isolado em um trabalho realizado para identificar transcritos expressos transiente e localmente durante o desenvolvimento das asas de aves. Em seguida, o transcrito ortólogo de camundongo foi isolado por meio de RT-PCR e as análises das seqüências quanto à similaridade com outros transcritos por meio do programa BLAST N revelaram a existência do gene Lix1 também em drosófila e em humano. Em nível protéico o transcrito Lix1 de ave apresenta 83% de identidade com os ortólogos de camundongo e humano, e 51% de identidade com o ortólogo de drosófila (SWINDELL et al. 2001). Em camundongo, o gene Lix1 apresenta expressão bastante acentuada durante o desenvolvimento do córtex cerebral e compactação da substância nigra (MOELLER et al. 2002). Entretanto, em humanos, ainda não há nenhuma descrição a respeito da função desempenhada por esse transcrito, nem sobre seu padrão de expressão. A função desse novo transcrito em humanos deve ser melhor investigada, uma vez que seu elevado nível de expressão está associado à super-expressão do oncogene c-erbB2 em amostras tumorais de mama. Além disso, em aves, drosófila e camundongo, a expressão desse gene está associada a processos fundamentais de desenvolvimento.

O transcrito correspondente à *tag* 28 (AK093427), localiza-se no cromossomo 11p15.5, e possui 1.867 pares de bases. A seqüência completa desse transcrito também apresenta evidente estrutura de *splicing*, sendo formada por 10 éxons. As seqüências que correspondem a esse transcrito foram agrupadas no UniGene *cluster* Hs.501624 correspondendo ao transcrito SIGIRR (Single Immunoglobulin IL-1R-related). O transcrito apresenta uma fase aberta de leitura,

composta por 411 aminoácidos, que codifica uma proteína transmembrana pertencente à superfamília de receptores de interleucina 1. Todos os membros dessa superfamília possuem um domínio intracelular TIR (Toll-IL-1R) e um domínio extracelular. Com base nesse domínio extracelular, os membros dessa superfamília podem ser divididos em dois sub grupos: os receptores que possuem domínios imunoglobulina, e os receptores que possuem um domínio rico em leucina. O sub grupo de receptores com domínio imunoglobulina inclui o transcrito SIGIRR (WALD et al. 2003; QIN et al. 2005). Entretanto, o transcrito SIGIRR representa um único sub grupo dentro dessa superfamília de receptores devido ao fato de seu domínio extracelular ser formado por apenas um domínio imunoglobulina e devido a substituição de dois aminoácidos no domínio TIR que são essenciais para a sinalização (WALD et al. 2003). Além disso, a localização cromossômica dos demais membros da família de receptores de interleucina 1, ocorre em um *cluster* mapeado no cromossomo 2q12-13. Entretanto, vale destacar, que a região cromossômica em que o SIGIRR foi mapeado também abriga outros genes envolvidos na síndrome de Beckwith-Wiedemann, caracterizada pela ocorrência de diversas anormalidades durante a infância incluindo uma alta incidência de tumores (THOMASSEN et al. 1999).

Embora o SIGIRR tenha um domínio imunoglobulina, o mesmo não é capaz de ligar interleucina 1, ou potencializar a sinalização dependente de interleucina 1. Em um estudo realizado por WALD et al. (2003), a função desempenhada pelo SIGIRR na via de sinalização de interleucina 1 foi avaliada por meio da super-expressão desse transcrito nas células Jurkat e HepG2. Os resultados desse estudo revelaram que a super-expressão desse transcrito reduziu a ativação de NF- κ B

mediada pela interleucina 1 e interleucina 18. Em contraste, a super-expressão de SIGIRR não teve nenhum efeito na ativação de STAT1 mediada por interferon- γ (INF- γ) nas células Jurkat. Esses resultados indicaram que o SIGIRR pode atuar como um regulador negativo da via de sinalização de interleucina 1 e interleucina 18 (WALD et al. 2003).

O papel desempenhado pelo SIGIRR, no mecanismo de resposta imune, foi avaliado por meio da geração de um camundongo em que o transcrito em questão foi alterado por recombinação homóloga. Esses animais receberam, intraperitonealmente, uma injeção de TLR4, um ativador de LPS, e apresentaram uma resposta inflamatória exacerbada quando comparada a reação inflamatória apresentada pelos animais selvagens. Enquanto 70% dos animais selvagens sobreviveram, apenas 10% dos animais portadores de alterações no gene SIGIRR sobreviveram.

Experimentos de co-imunoprecipitação, realizados utilizando células 293, revelaram que o SIGIRR é capaz de formar homodímeros e também interagir com outros membros da superfamília de receptores de interleucina, indicando o possível mecanismo de ação desse transcrito na via de sinalização mediada pela interleucina 1. Os resultados obtidos a partir desses experimentos mostraram que o SIGIRR foi capaz de interagir com o TLR4, TLR5, TLR9 e também com o próprio receptor de interleucina 1 (WALD et al. 2003; QIN et al. 2005). Além disso, em células 293, após tratamento com interleucina 1, foi possível observar a ocorrência de interação entre o SIGIRR endógeno e as proteínas adaptadoras IRAK e TRAF-6, sabidamente envolvidas na via de sinalização mediada por interleucina 1 (WALD et al. 2003).

Com base em diversos estudos, postulou-se um modelo que explica a ativação da via de sinalização mediada pela interleucina 1. Após o receptor de interleucina 1 ser estimulado, ocorre o recrutamento da proteína adaptadora MyD88, seguida de duas quinases, IRAK e TRAF6, favorecendo a formação de um complexo protéico. Em seguida, esse complexo é ativado e então as moléculas de sinalização que estavam acopladas ao mesmo, são liberadas e passam a interagir e ativar outros componentes da via de sinalização *downstream* ao receptor de interleucina 1, culminando com a ativação do fator de transcrição NF κ B. O SIGIRR possivelmente regula negativamente essa via de sinalização interferindo nos processos de recrutamento e ativação dos componentes IRAK ou TRAF6. De forma alternativa, o SIGIRR também poderia atenuar a dissociação entre as moléculas de sinalização ativadas e o receptor de interleucina 1, inibindo a ocorrência de sinalização *downstream* ao receptor (WALD et al 2003; Qin et al 2005).

Dados da literatura relatam que diversos tumores sólidos secretam interleucinas, que atuam como quimioatraentes para linfócitos que desempenham importante papel no processo de resposta inflamatória associada à malignidade. As interleucinas 1- α por exemplo, induzem diversas enzimas proteolíticas em fibroblastos presentes no estroma celular, sugerindo que as mesmas desempenhem uma função no processo de degradação da matrix extracelular e invasão tumoral (SINGER et al. 2003). Em um estudo que avaliou a expressão das interleucinas 1- α e 1- β em um painel de 9 linhagens celulares de mama (MCF-7, T47D, ZR75-1, SKBR3, BT474, BT20, BT549, HS578T e MDA-MB-231), os resultados revelaram que a expressão dessas interleucinas está associada a um fenótipo mais maligno e também com à negatividade para o receptor de estrógeno α . As linhagens BT20,

BT549, HS578T e MDA-MB-231, fenotipicamente mais malignas, foram as que apresentaram expressão dessas interleucinas (SINGER et al. 2003).

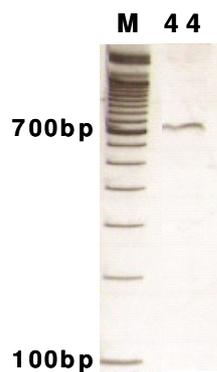
Em um outro estudo, que avaliou a expressão da interleucina 1- β em amostras de carcinomas invasivos de mama e em amostras de carcinomas *in situ*, os resultados revelaram a presença de expressão dessa interleucina em aproximadamente 90% dos carcinomas invasivos. Além disso, nas amostras dos carcinomas invasivos, a positividade para o oncogene c-erbB2 foi avaliada por meio de imunohistoquímica, revelando que 14% dos tumores eram positivos para esse oncogene. Nesses tumores, o nível de expressão de interleucina 1- β foi mais elevado em relação aos tumores c-erbB2 negativos, embora a diferença não tenha sido estatisticamente significativa (JIN et al. 1997).

Com base nos resultados dos trabalhos, acima mencionados, é possível associar a expressão das interleucinas aos tumores mais agressivos e com pior prognóstico. Nesse contexto, a super-expressão do SIGIRR nas amostras de carcinoma ductal invasivo que super-expressam c-erbB2, utilizadas nesse estudo, em um primeiro momento, parece um dado discrepante. Considerando a atividade inibitória desse transcrito, o esperado seria que o mesmo fosse *down*-regulado nas amostras tumorais c-erbB2 positivas. Entretanto, uma possível explicação para o resultado obtido, refere-se ao mecanismo de escape do sistema imune. A super-expressão do SIGIRR, pode representar um mecanismo desenvolvido pelas células tumorais para tentar conter a resposta inflamatória exacerbada, possivelmente desencadeada pela interleucina 1, e assim limitar a resposta imune do hospedeiro. Essa hipótese pode ser sustentada, com base nos achados de um estudo, que foi realizado para avaliar os efeitos causados pela super-expressão do antagonista

endógeno de interleucina 1 em camundongos infectados pela bactéria *Listeria monocytogenes*. Os resultados desse estudo revelaram que o antagonista de interleucina 1, *in vivo*, têm a função de limitar a resposta inflamatória e a resposta imune inata durante a infecção. A super-expressão do antagonista de interleucina 1, reduz a atividade biológica da interleucina 1, conseqüentemente reduz a sinalização mediada por essa citocina, levando a uma redução da resposta imune (IRIKURA et al. 1999). Assim, elucidar os mecanismos moleculares envolvidos na interação desse novo transcrito com a super-expressão do oncogene c-erbB2 torna-se essencial para que o potencial terapêutico desse transcrito possa ser avaliado.

O transcrito correspondente à *tag* 6 (BX640636), localiza-se no cromossomo 8q21.3, e possui 2.334 pares de bases. A seqüência completa desse transcrito também apresenta evidente estrutura de *splicing*, sendo formado por 4 éxons. As seqüências que correspondem a esse gene foram agrupadas no UniGene cluster Hs.145386. O transcrito apresenta uma fase aberta de leitura composta por 112 aminoácidos, que codificam uma proteína hipotética sem domínios protéicos característicos.

O transcrito correspondente à *tag* 44, representada pela *EST* BE385820, teve sua seqüência completa gerada por meio da técnica de RACE. O kit utilizado favorece a síntese de moléculas de cDNA completas. Com o objetivo de aumentar a eficiência de amplificação da reação, a estratégia de PCR *Nested* foi adotada. O fragmento obtido foi analisado em gel de poliacrilamida 8% e pôde-se observar a presença de uma banda única de forte intensidade, de aproximadamente 700 bp, para o transcrito correspondente à *tag* 44. O fragmento obtido pode ser visualizado na Figura 23.



Legenda: Fragmento obtido utilizando os iniciadores específicos *nested* para o transcrito correspondente à *tag* 44. O fragmento foi analisado em gel de agarose 1% corado com brometo de etídeo. O marcador utilizado foi o 100bp.

Figura 23 - Geração da seqüência completa do transcrito representado pela *tag* 44 através da técnica de RACE.

Em seguida, o fragmento obtido para o transcrito foi purificado a partir de um gel de agarose 1%. Esse fragmento foi clonado e 12 colônias foram seqüenciadas nos dois sentidos.

Para garantir que a seqüência completa do transcrito havia sido gerada, uma segunda reação de RACE foi realizada, utilizando novos iniciadores que foram baseados na seqüência anteriormente obtida. A reação ocorreu como descrita anteriormente e o fragmento obtido foi clonado e 12 colônias foram seqüenciadas nos dois sentidos. O resultado do seqüenciamento revelou que a seqüência completa já havia sido gerada, uma vez que não foi possível estender mais o fragmento obtido. Uma outra forma de determinar se a seqüência completa do transcrito foi gerada seria por meio da técnica de Northern blot. Entretanto, essa estratégia não foi adotada devido ao baixo nível de expressão do transcrito.

A seqüência completa do transcrito correspondente à *tag* 44 foi obtida reunindo o fragmento gerado por RACE ao fragmento gerado por MPSS-GLGI. Na figura 24 a seqüência completa do transcrito, que é composta por 662 pares de bases distribuídos ao longo de 3 éxons, pode ser visualizada. Esse transcrito se localiza no cromossomo 12q14. A análise da seqüência desse transcrito revelou a ausência de uma fase aberta de leitura, podendo tratar-se de um RNA não codificante.

```
ACGCGGGGGACGCACGGTGTCTGGGAGGAGTTACGGAGGCAGGGTTGTGGTTGCTGGCTGTTACCAGGACAACCGGAGGCGATT
GACCGTTATCTGCGGTTGGAGCCGTTAGCGGGAGAGGCAGAGATATTCAGAGGTCTTTTAGGATGCGCTAAAGGGTCGTGAGGG
CTCTCTTAAAAATTTTCTTACAAAGCGGCTATCCAGTCGTGCCCGCGGCCCTGCTGCTGGCCCCGGGGATCTGAGTCGTACCCTC
TTGTTTTTCTCTGAGTCAGTCTTAAGGTGAAATGAAGTGTGGCCAGTGGCTCCTCACTGTCGCTTCTCTAGTTTTCTGCCTCCT
TTTAGAAAATGAATTGAAAAGACAGGATGAAGTGGACACAGCATGTGAAGACAATTTCTGTTCAAGAAGTTTGGCTGTCAAGGAA
ACAGAGAATGTGCTAAAGAACATACAGACACAGAGCAGACAGGCCACCTTTGCAACCACATGGAGGTTTGTCTGATATGAAGCT
AAGGAAGCTAAGCTGGAAGACAGAGAGACCAAGTCCTGATGACATTGTTTGAACCCAGAGATCCAGACATGCCTGAAAACTAGTT
TTACCCTGGACTTATCCGTTGAATGAGCCAATAAACTCTCTTTTATACTTAAAAAAAAAAAAAAAAAAAA
```

Legenda: A seqüência completa do transcrito correspondente à *tag* 44 foi gerada a partir da montagem entre as seqüências geradas por RACE e as seqüências obtidas por meio de MPSS-GLGI. Em vermelho, está representado o sítio da enzima *DpnII* e, em azul, a seqüência da *tag* de MPSS.

Figura 24 - Seqüência consenso do transcrito correspondente à *tag* 44.

Durante muito tempo, os RNAs foram considerados moléculas acessórias envolvidas apenas nos processos de transcrição e tradução. Entretanto, nas últimas décadas, diversas evidências têm mostrado que as moléculas de RNA são bastante versáteis e desempenham outras funções nas células eucarióticas. Essa versatilidade se deve, principalmente, às propriedades químicas do RNA, as quais possibilitam a formação de estruturas terciárias capazes de desempenhar funções até então conferidas apenas às proteínas (SZYMANSKI et al. 2003).

O término do seqüenciamento do genoma humano revelou a existência de uma pequena fração de seqüências codificadoras. Estima-se que mais de 98% da fração transcrita do genoma dos eucariotos, seja constituído por moléculas de RNA que não codificam proteínas. As moléculas de RNA não codificantes têm sido

envolvidas em diferentes mecanismos moleculares e celulares que ocorrem nas células eucarióticas tais como inativação do cromossomo X, manutenção da estrutura da cromatina, *imprinting* genômico, etc. Alterações nos níveis de expressão dos RNAs não codificantes têm sido detectadas em doenças complexas como o câncer. O transcrito não codificante BC200 (200 pares de bases), por exemplo, localiza-se no braço curto do cromossomo 2 e sua super-expressão está associada, entre outros tipos de câncer, ao carcinoma invasivo de mama. Esse transcrito apresenta expressão restrita ao tecido tumoral e, por essa razão, tem sido considerado um novo marcador molecular para identificar tumores com pior prognóstico (SZYMANSKI et al. 2003; COSTA 2005). Com base nessas informações, pode-se perceber que os RNAs não codificantes estão emergindo como uma classe de transcritos funcionais nos organismos eucariotos. Cada vez mais, esses transcritos estão sendo associados a processos moleculares e celulares que ocorrem em eucariotos, e seus mecanismos de desregulação (deleção, super-expressão ou *down*-regulação) precisam ser esclarecidos. Por essa razão, a identificação e a caracterização de novos transcritos não codificantes representam um grande avanço no sentido de compreender o papel desempenhado por essas moléculas de RNA, não apenas nos processos fisiológicos normais, como também em processos complexos que ocorrem em doenças como o câncer.

4.6 CARACTERIZAÇÃO DE TRANSCRITOS *ANTISENSE* REGULADOS POR C-ERBB2

Dentre os 41 fragmentos 3' de cDNA obtidos por meio da técnica de MPSS-GLGI, 5 fragmentos representaram prováveis transcritos *antisense*. Cada uma das seqüências geradas, a partir dos fragmentos acima mencionados, foi analisada por meio do programa BLAST, e apresentou alinhamento com seqüências de transcritos conhecidos depositadas no GenBank, com alto nível de similaridade, porém na orientação oposta (3' - 5'). Para quatro dos cinco fragmentos, houve identificação de cauda poli A e um deles apresentou sinal de poliadenilação, sugerindo que essas seqüências eram derivadas de transcritos *antisense* localizados na fita oposta da seqüência conhecida (Tabela 17). A figura 25 mostra os alinhamentos obtidos para os transcritos *antisense*.

Segundo os dados iniciais de MPSS, os transcritos correspondentes às *tags* 12 e 46 apresentam maior nível de expressão na linhagem HB4a-C5.2 em relação à linhagem HB4a enquanto os transcritos correspondentes às *tags* 57, 69 e 70 apresentam maior nível de expressão na linhagem HB4a em relação à linhagem HB4a-C5.2.

A - Fragmento de GLGI-MPSS - tag 57

GATCCTAGAAATCGGTGACAACTGGTCTTTGCCTTATTCTCAGCAAATTCAGCTACTGATATCTTTGAGACGAGCCTTGTATT
 TATTCATTTGTCCACAACAATCAGTTCACACAAGTGGTAGTCTTACTGATGTCCTGTACACTTTGTAGTGGGTCAGTGGTGTGTC
 AGGGCAGCGGAGAATGCAGGGCGCAAACAAATGGCCAAAGCATTAGCAGACATTGATTAGTGTCTTCTGCAGAGCAATCCTA
 AGAATTAACAAAAA

B - [gi|5902011|ref|NM_006901.1](#) Homo sapiens myosin IXA (MYO9A), mRNA

Length = 8473

Score = 482 bits (243), Expect = e-133
 Identities = 246/247 (99%)
 Strand = Plus / Minus

Query: 84 gaatcggtgacaacctggctctttgccttattctcagcaaattccaagctactgatatctt 143
 ||||| |||||||
 Sbjct: 7068 gaatcagtgacaacctggctctttgccttattctcagcaaattccaagctactgatatctt 7009

Query: 144 tgagacgagccttgtatatttattcatttggttccacaacaatcagttccacacaagtggtag 203
 |||||||
 Sbjct: 7008 tgagacgagccttgtatatttattcatttggttccacaacaatcagttccacacaagtggtag 6949

Query: 204 tcttactgatgtcctgtacactttgtagtggtcagtggtgtcagggcagcgagaaatgc 263
 |||||||
 Sbjct: 6948 tcttactgatgtcctgtacactttgtagtggtcagtggtgtcagggcagcgagaaatgc 6889

Query: 264 agggcgcaaacacaatggccaaagcatttagcagacattcgattagtgcttctcctgcagag 323
 |||||||
 Sbjct: 6888 agggcgcaaacacaatggccaaagcatttagcagacattcgattagtgcttctcctgcagag 6829

Query: 324 caatcct 330
 |||||
 Sbjct: 6828 caatcct 6822

A - Fragmento de GLGI-MPSS - tag 69

GATCAGTCTTTCAGAGTCTGGCGGCATGAACCTCTTCAATTTCTCAGGCACAGTGATGCCCTTCTCTGTCTGGTAGTTCTCCAGG
 ATGGCGCAGATGGTACGGGTAGTGGCGCACATGGTAGCATTGAGCATATGGACAACTCCACCTTGCCATCATCTTCTGGTTT
 GCCCATATCGGATTCGAAGCGCCGCTCTAGATAGTAATCGAATTCGCCGCGCCGATGGCGCCGGAAGCATGCGACGTCGGG
 CCCAATTCGCCCTATAGTGAGCCGTNATAAANNNNNNN

B - [gi|16306547|ref|NM_006513.2](#) Homo sapiens seryl-tRNA synthetase (SARS), mRNA Length = 1942

Score = 377 bits (190), Expect = e-101
 Identities = 190/190 (100%)
 Strand = Plus / Minus

Query: 80 gatcagttcttgagtcctggcgcatgaactccttcaatttctcagggcacagtgatgcc 139
 |||||||
 Sbjct: 1479 gatcagttcttgagtcctggcgcatgaactccttcaatttctcagggcacagtgatgcc 1420

Query: 140 cttctctgtctggtagttctccaggatggcgagatggtagggtagggcgacatggt 199
 |||||||
 Sbjct: 1419 cttctctgtctggtagttctccaggatggcgagatggtagggtagggcgacatggt 1360

Query: 200 agcattgagcatatggacaaactccaccttgtccatcatcttcttggtttgccatcatcg 259
 |||||||
 Sbjct: 1359 agcattgagcatatggacaaactccaccttgtccatcatcttcttggtttgccatcatcg 1300

Query: 260 gattcgaagc 269
 |||||
 Sbjct: 1299 gattcgaagc 1290

A - Fragmento de GLGI-MPSS - *tag* 70

GATCTGACAGTGGTCACATCCTGTGCAAAAACTTGATACAAAAATGATAGCACATGGTATCTGAGCTGCTTACATTACAAGAAA
 AAGGAAATACAGTAGCTGAAATATGGCACTCCTGGGAATCAACTTCTAAACCAAATAGAATGCCTTTGAAATGATTAATTTATT
 TGTGTATTAGTAAGAAAGCCCCACCACCATAAATAGTACAATATTTAAAAAAAAAAAAAAAAAAAAA

B - [gi|7656909|ref|NM_014382.1](#) Homo sapiens ATPase, Ca++ transporting, type 2C, member 1 (ATP2C1), mRNA

Length = 3637

Score = 428 bits (216), Expect = e-117
 Identities = 216/216 (100%)
 Strand = Plus / Minus

```

Query: 80   gatctgacagtggtcacatcctgtgcaaaaaacttgatacaaaaatgatagcacatggta 139
          |||
Sbjct: 3567 gatctgacagtggtcacatcctgtgcaaaaaacttgatacaaaaatgatagcacatggta 3508

Query: 140  tctgagctgcttacattacaagaaaaggaaatacagtagctgaaatatggcactcctgg 199
          |||
Sbjct: 3507 tctgagctgcttacattacaagaaaaggaaatacagtagctgaaatatggcactcctgg 3448

Query: 200  gaatcaacttctaaaccaaatagaatgcctttgaaatgattaaatttatttgtgtattag 259
          |||
Sbjct: 3447 gaatcaacttctaaaccaaatagaatgcctttgaaatgattaaatttatttgtgtattag 3388

Query: 260  taagaaagccccaccaccataaatagtacaatatatt 295
          |||
Sbjct: 3387 taagaaagccccaccaccataaatagtacaatatatt 3352

```

Legenda: Em (A) estão, representadas as seqüências dos 5 fragmentos gerados por MPSS-GLGI. Em vermelho, está representado o sítio da enzima *DpnII* e, em azul, a seqüência da *tag* de MPSS. Em (B), estão representados os alinhamentos obtidos por meio da análise da seqüência de cada transcrito por meio da utilização do programa BLAST.

Figura 25 - Análise dos fragmentos gerados por MPSS-GLGI que correspondem a transcritos *antisense*.

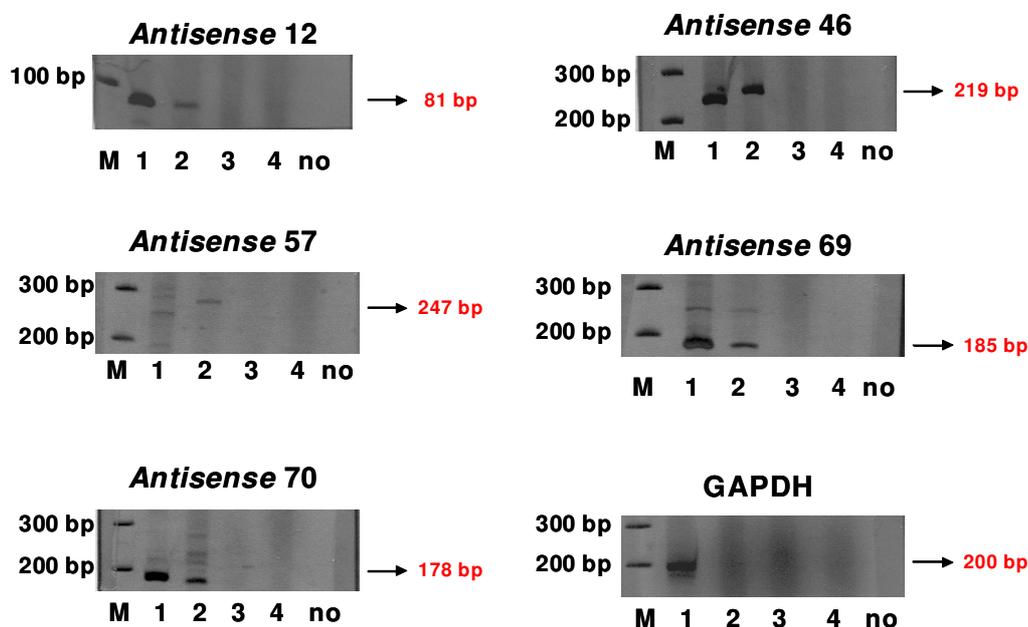
Com o objetivo de comprovar a existência dos transcritos *antisense*, os mesmos foram validados experimentalmente. Para tanto foi utilizada a técnica de RT-PCR fita específica. Essa metodologia baseou-se na utilização apenas do iniciador complementar ao transcrito que se quer estudar durante a síntese da primeira fita de cDNA.

Foram realizadas quatro reações de síntese de cDNA fita específica para a validação de cada um dos transcritos *antisense*. Na primeira reação, foi utilizado apenas o iniciador complementar ao transcrito *sense*. Na segunda reação, foi utilizado

apenas o iniciador complementar ao transcrito *antisense*. A terceira reação ocorreu em presença apenas do iniciador complementar ao transcrito *antisense*, porém na ausência da transcriptase reversa. A quarta reação, ao contrário da terceira, ocorreu na ausência do iniciador complementar ao transcrito *antisense*, porém na presença da enzima transcriptase reversa. A primeira reação realizada, representou o controle positivo, e nos permitiu avaliar a expressão do transcrito *sense* e as duas últimas reações descritas foram realizadas como controles negativos do experimento, de modo a garantir a ausência de contaminação genômica e a ausência de formação de estruturas secundárias durante a síntese de cDNA. O gene GAPDH foi utilizado como controle da PCR fita específica, pois esse gene não apresenta um gene *antisense*.

Em seguida, foi realizada uma PCR e, de acordo com o esperado, na primeira reação, foi possível visualizar a amplificação do transcrito *sense*. Na segunda reação, que ocorreu na presença apenas do iniciador complementar ao transcrito *antisense*, foi possível visualizar a amplificação de um fragmento, confirmando a existência do *antisense*. Na terceira reação, que ocorreu na ausência da transcriptase reversa, nenhum fragmento foi amplificado, garantindo a ausência de contaminação genômica. Finalmente, a quarta reação que ocorreu na presença da transcriptase reversa, também não apresentou amplificação de nenhum fragmento. Esse resultado confirmou a ausência de formação de estruturas secundárias durante a síntese de cDNA, as quais poderiam servir como iniciadores para a transcriptase reversa. Após a realização da RT-PCR fita específica, foi possível confirmar a existência de quatro entre os cinco transcritos *antisense*. Apenas o transcrito correspondente à *tag 57* não

foi validado, uma vez que o tamanho do fragmento obtido não correspondeu ao tamanho esperado. Os dados descritos acima podem ser visualizados na Figura 26.



Legenda: Validação experimental dos transcritos *antisense* e GAPDH. (1) presença de transcriptase reversa e iniciador complementar ao transcrito *sense*; (2) presença de transcriptase reversa e iniciador complementar ao transcrito *antisense*; (3) ausência de transcriptase reversa e presença de iniciador complementar ao transcrito *antisense*; (4) presença de transcriptase reversa e ausência de iniciador complementar ao transcrito *antisense*. Em vermelho, estão representados os tamanhos esperados para os fragmentos amplificados. Os fragmentos foram analisados em gel de agarose 1% e foram corados com brometo de etídeo. O padrão de peso molecular utilizado é o 100bp.

Figura 26 - RT-PCR fita específica.

Entre os quatro transcritos *antisense* que tiveram sua existência confirmada, os transcritos correspondentes às *tags* 12 e 46, ambos superexpressos na linhagem celular HB4a-C5.2 em relação à linhagem HB4a, apresentaram similaridade com dois transcritos *sense* envolvidos com expressão gênica, transdução de sinal e proliferação celular.

O transcrito correspondente à *tag* 12 apresentou similaridade com o transcrito *sense* que contém um domínio tireodoxina (TXNDC – NM_030755). Esse transcrito, que representa um dos membros de uma família gênica que contém domínio tireodoxina, está localizado no cromossomo 14q22.1 e possui 2.473 pares de bases. A seqüência completa desse transcrito apresenta uma evidente estrutura de *splicing* quando alinhada com a seqüência genômica, sendo formada por 8 éxons. As seqüências correspondentes a esse transcrito foram agrupadas no UniGene cluster Hs.125221. O transcrito TXNDC apresenta uma fase aberta de leitura composta por 280 aminoácidos que codifica uma proteína transmembrana. A proteína tireodoxina foi identificada originalmente em *Escherichia coli* e está evolucionariamente conservada desde procariotos até os eucariotos superiores. Essa proteína está envolvida em diversos processos celulares incluindo expressão gênica, transdução de sinais e proliferação. Além disso, a tireodoxina também modula a atividade de ligação de fatores de transcrição, tais como NFκB e proteína de ativação 1, ao DNA (MATSUO et al. 2001).

O transcrito correspondente à *tag* 46, apresentou similaridade com o transcrito *sense* trombospondina 1 que localiza-se no cromossomo 15q15 e possui 5.819 pares de bases. A seqüência completa desse transcrito também apresenta evidente estrutura de *splicing*, sendo formada por 22 éxons. As seqüências que correspondem a esse transcrito foram agrupadas no UniGene *cluster* Hs.164226. A trombospondina 1 é uma glicoproteína secretada que está associada à matrix extracelular e possui diversas funções biológicas, incluindo uma potente atividade angiogênica. Essa proteína é sintetizada e secretada não apenas pelas plaquetas como também pelas células endoteliais, fibroblastos, células musculares lisas e pneumócitos. Em 2001,

Rodriguez-Manzaneque e colaboradores, realizaram um estudo para avaliar a participação do inibidor endógeno de trombospondina 1 na progressão tumoral. Para tanto, foi gerado um camundongo com ausência do gene da trombospondina ou com super-expressão desse gene na glândula mamária. Os resultados desse estudo revelaram que a ausência da trombospondina, resulta em aumentada associação do fator de crescimento endotelial vascular com seu receptor e com altos níveis de matrix metaloproteinase 9 ativa (molécula que favorece tanto angiogênese como invasão tumoral). Por outro lado, a super-expressão de trombospondina mostrou retardar o desenvolvimento do tumor. Estudos clínicos revelaram que a expressão de trombospondina está inversamente correlacionada com a progressão maligna em câncer de mama, melanoma e carcinoma de pulmão. Esses achados dão suporte à hipótese de que a trombospondina inibe o crescimento tumoral por meio de mecanismos anti-angiogênicos. (RODRIGUEZ-MANANEQUE et al. 2001).

Dados da literatura baseados em estudos experimentais atribuem algumas prováveis funções aos transcritos *antisense*. Esses transcritos podem, basicamente, contribuir para a regulação da expressão gênica por meio de três principais mecanismos: (1) degradação de um RNA alvo, pela formação de fragmentos de RNA dupla fita que servirão de substrato para a degradação do mRNA homólogo; (2) regulação do mecanismo de *splicing* do pré mRNA; ou, ainda, (3) pela alteração do controle do mecanismo de tradução (LEHNER et al. 2002; SHENDURE e CHURCH 2002, LAVORGNA et al. 2004). Devido aos mecanismos acima citados, por meio dos quais os transcritos *antisense* podem regular a expressão gênica, torna-se possível associar alterações no mecanismo de transcrição desses transcritos e o desenvolvimento de doenças. Até o momento existem poucos trabalhos que

comprovam a associação entre transcritos *antisense* e câncer. Entretanto, os dados gerados por SHENDURE e CHURCH (2002) sugerem que transcritos *antisense* que apresentam níveis de expressão elevados em tecidos tumorais podem desempenhar um importante papel com relação à supressão do potencial gene supressor de tumor *sense*. Em um outro trabalho, uma plataforma de cDNA *microarray* foi enriquecida para transcritos *antisense* intrônicos com o objetivo de avaliar o papel biológico desses transcritos em câncer de próstata. Os resultados desse trabalho revelaram que uma grande fração dos transcritos *antisense* intrônicos foram correlacionados com o grau de diferenciação dos tumores de próstata quando comparados aos transcritos oriundos de regiões genômicas ainda não anotadas ou transcritos provenientes de éxons de transcritos conhecidos (REIS et al. 2004).

Considerações Finais

5 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo isolar e caracterizar transcritos diferencialmente expressos entre as linhagens celulares de mama HB4a e HB4a-C5.2, que expressam diferentes níveis do oncogene c-erbB2. Para tanto foi utilizada uma nova metodologia para análise global da expressão gênica denominada MPSS (*Massively Parallel Signature Sequencing*). O MPSS é uma técnica bastante robusta que além de gerar um grande volume de dados, favorece a realização de análises estatísticas, a caracterização de transcritos diferencialmente expressos e a identificação de transcritos pouco abundantes e geralmente ainda não caracterizados. Por essas razões, essa metodologia vem sendo cada vez mais utilizada não apenas para a identificação de novos transcritos humanos como também de outros organismos, contribuindo para a caracterização de novos genomas.

Uma limitação do MPSS refere-se ao pequeno tamanho das *tags* geradas, e em nosso trabalho, essa limitação foi solucionada por meio da adaptação da técnica GLGI (*Generation of Longer cDNA fragments from SAGE tags for Gene Identification*) para a extensão de tags de MPSS. A técnica GLGI foi desenvolvida recentemente (2002), com o objetivo de estender as *tags* de SAGE a fragmentos 3' de cDNA. O GLGI representa uma abordagem extremamente útil para que as informações geradas por SAGE ou MPSS, possam ser utilizadas na validação experimental e caracterização de novos transcritos. A metodologia do GLGI-MPSS, adaptada nesse trabalho, pode ser utilizada para qualquer organismo, sendo de especial importância para aqueles que possuem um número reduzido de sequências

transcritas caracterizadas e depositadas em banco de dados públicos e consequentemente um número significativo de tags órfas nas análises de MPSS.

A combinação entre as técnicas de MPSS e GLGI, utilizadas neste trabalho, nos permitiu identificar 10 extensões 3' específicas, 5 transcritos *antisense*, 3 transcritos polimórficos, nos quais a presença de *SNPs* gerou *tags* de MPSS alternativas, 2 formas de poliadenilação alternativa e *tags* de MPSS artefatuais.

A expressão diferencial das 10 extensões 3' específicas e dos 3 transcritos correspondentes à *tags* geradas por *SNPs*, foi avaliada por meio de PCR em Tempo Real e dentre os 13 transcritos analisados, 5 transcritos apresentaram expressão diferencial entre as linhagens, sendo 4 transcritos super-expressos na linhagem HB4a-C5.2 em relação a HB4a. Esses resultados foram confirmados, por meio de PCR em Tempo Real, em amostras tumorais c-erbB2 positivas e negativas, sugerindo a existência de uma correlação entre a super-expressão do oncogene c-erbB2 e a expressão aumentada dos novos transcritos.

A super-expressão do oncogene c-erbB2 ocorre em aproximadamente 30% dos carcinomas de mama e está associada a um pior prognóstico. No entanto, os mecanismos moleculares de atuação desse oncogene no câncer de mama permanecem pouco compreendidos. Além disso, diversas pesquisas têm demonstrado que apenas parte das pacientes com câncer de mama, que apresentam super-expressão de c-erbB2, respondem aos tratamentos convencionais disponíveis. Por essa razão a identificação de transcritos regulados por c-erbB2 bem como o desenvolvimento de novas abordagens terapêuticas têm atraído a atenção de diversos grupos de pesquisa. Os transcritos identificados no presente trabalho, podem representar novos marcadores moleculares para o diagnóstico ou prognóstico do

câncer de mama ou ainda podem estar envolvidos em mecanismos de resistência a drogas, podendo tornarem-se alvos terapêuticos.

Um grande obstáculo para o sucesso do tratamento do câncer, é a resistência intrínseca ou adquirida pelos tumores aos agentes citotóxicos utilizados no tratamento. Atualmente existem dois tipos principais de inibidores de c-erbB2 em uso clínico: os anticorpos humanizados como por exemplo a Herceptina, que agem diretamente sobre o domínio extracelular do c-erbB2, e as moléculas inibidoras de atividade tirosina quinase. Sabe-se que apenas a super-expressão do oncogene c-erbB2 não é suficiente para induzir resistência à drogas em células de mama normal, sugerindo que a alteração na quimiosensibilidade das células depende não apenas do nível de expressão de c-erbB2, como também da interação desse oncogene com outros transcritos presentes na célula tumoral.

Nesse contexto, um dos novos transcritos identificado, neste trabalho, merece atenção especial. Trata-se do SIGIRR, uma proteína transmembrana que pertence à super-família de receptores de interleucina 1 e que regula negativamente a via de sinalização mediada por essa citocina. Neste trabalho, a super-expressão do SIGIRR foi associada à super-expressão do oncogene c-erbB2 e pode representar um mecanismo de escape do sistema imune uma vez que a super-expressão de um antagonista da interleucina 1, reduz a atividade biológica da interleucina 1, conseqüentemente reduz a sinalização mediada por essa citocina, levando a uma redução da resposta imune. A exemplo de outros receptores transmembrana, a proteína codificada por esse transcrito pode representar um novo alvo terapêutico, podendo dar origem, por exemplo, ao desenvolvimento de um novo anticorpo.

Uma vez que não foi observada a ocorrência de uma correlação perfeita entre a super-expressão de c-erbB2 e a expressão aumentada dos 4 novos transcritos, a realização de estudos funcionais se mostra muito importante para que o papel desempenhado por esses transcritos, durante o desenvolvimento do tumor de mama, possa ser elucidado. Nesse sentido, dar-se-á continuidade a esse estudo, em um primeiro momento confirmando o envolvimento desses novos transcritos nas vias de sinalização associadas ao oncogene c-erbB2. Para tanto, avaliar-se-á o impacto causado pelo uso de inibidores de c-erbB2, anticorpo monoclonal Herceptina, e inibidores de vias clássicas de sinalização reguladas por esse oncogene (PI3K, MAPK e PKC) no nível de expressão dos novos transcritos diferencialmente regulados por c-erbB2.

Posteriormente, será analisada a ocorrência de alterações fenotípicas relacionadas ao processo de formação tumoral, na linhagem celular de mama HB4a transfectada com os novos transcritos diferencialmente regulados por c-erbB2. O fenótipo resultante será analisado experimentalmente *in vitro*, por meio de ensaios funcionais que determinarão a capacidade desses transcritos quanto a promover proliferação, migração e morte celular. Além disso, serão realizados ensaios *in vivo*, para verificar a capacidade tumorigênica da linhagem HB4a transfectada com os novos transcritos diferencialmente regulados por c-erbB2, utilizando camundongos atímicos que serão injetados com células que super-expressam os transcritos de interesse.

Conclusões

6 CONCLUSÕES

1. A partir da padronização da técnica de GLGI-MPSS foi possível estender 83 *tags* órfãs de MPSS e para 41 dessas *tags* selecionadas, pôde-se obter um fragmento bem definido;
2. As seqüências desses 41 fragmentos foram analisadas e nos permitiram identificar 10 extensões 3' específicas, 5 transcritos *antisense*, 3 *tags* de MPSS alternativas que correspondem a transcritos polimórficos, 2 formas de poliadenilação alternativa e 5 *tags* de MPSS artefatuais;
3. As análises das 10 extensões 3' específicas e das 3 *tags* correspondentes à *SNPs*, por meio de PCR em Tempo Real, indicaram que 5 transcritos são diferencialmente expressos entre as linhagens HB4a e HB4a-C5.2, sendo 4 transcritos mais expressos na linhagem HB4a-C5.2;
4. As análises realizadas, por meio de PCR em Tempo Real, para avaliar o nível de expressão dos 4 transcritos em amostras tumorais de mama c-erbB2 positivas e negativas revelaram uma associação entre a super-expressão do oncogene c-erbB2 e a expressão aumentada dos novos transcritos. Porém, a diferença no nível de expressão dos novos transcritos nas amostras tumorais foi estatisticamente significativa apenas para 2 transcritos.
5. A seqüência completa de um dos transcritos, com expressão aumentada na linhagem HB4a-C5.2, foi gerada por meio da técnica de RACE, porém esse transcrito não apresentou nenhuma fase aberta de leitura sugerindo tratar-se de um RNA não codificante;

6. Dentre os 5 transcritos *antisense* identificados por meio do MPSS-GLGI, foi possível confirmar a existência de 4 transcritos por meio de RT-PCR fita específica, sendo 2 transcritos mais expressos na linhagem HB4a-C5.2 em relação à linhagem HB4a.

Referências Bibliográficas

7 REFERÊNCIAS BIBLIOGRÁFICAS

Adams GP, Weiner LM. Monoclonal antibody therapy of cancer. **Nat Biotechnol** 2005; 23:1147-57.

Arciero C, Somiari SB, Shriver CD, et al. Functional relationship and gene ontology classification of breast cancer biomarkers. **Int J Biol Markers** 2003; 18:241-7.

Arteaga CL. Trastuzumab, an appropriate first-line single-agent therapy for HER2-overexpressing metastatic breast cancer. **Breast Cancer Res** 2003; 5:96-100.

Badache A, Hynes NE. A new therapeutic antibody masks erbB2 to its partners. **Cancer Cell** 2004; 5:299-300.

Bargmann CI, Hung MC, Weinberg RA. Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. **Cell** 1986; 45:649-57.

Baylin SB, Ohm JE. Epigenetic gene silencing in cancer: a mechanism for early oncogenic pathway addiction? **Nat Rev Cancer** 2006 6:107-16.

Benusiglio PR, Lesueur F, Luccarini C, et al. Common ERBB2 polymorphisms and risk of breast cancer in a white British population: a case-control study. **Breast Cancer Res** 2005; 7/2/R204-R9.

Bièche I, Lidereau R. Genetic alterations in breast cancer. **Genes Chromosomes Cancer** 1995; 14:227-51.

Brenner S, Williams SR, Vermaas EH, et al. In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. **Proc Natl Acad Sci U S A** 2000a; 97:1665-70.

Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. **Nat Biotechnol** 2000b; 18:630-4 Erratum in: **Nat Biotechnol** 2000b; 18:1021.

Burstein HJ. The distinctive nature of HER2- positive breast cancers. **N Engl J Med** 2005; 353:1652-4.

Camargo AA, Costa FF. Oncogenes e genes supressores de tumor. In: Brentani MM, Coelho FRG, Kowalski LP, editores. **Bases da oncologia**. 2rd ed. São Paulo: Lemar; 2003; p.71-86.

Camargo AA, Silva APM, Carraro DM, Chen J, Wang SM. **SAGE: current technologies and applications**. Chicago: Horizon Bioscience; 2004. Conversion of MPSS orphan tags into 3' ESTs. p.129-38.

Casalini P, Iorio MV, Galmozzi E, Ménard S. Role of HER receptors family in development and differentiation. **J Cell Physiol** 2004; 200:343-50.

Chen JJ, Rowley JD, Wang SM. Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. **Proc Natl Acad Sci U S A**. 2000; 97:349-53.

Chen J, Lee S, Zhou G, Wang SM. High-throughput GLGI procedure for converting a large number of serial analysis of gene expression tag sequences into 3' complementary DNAs. **Genes Chromosomes Cancer** 2002a; 33:252-61.

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

Costa FF. Non-coding RNAs: new players in eukaryotic biology. **Gene** 2005; 357:83-94.

Dowsett M. Improved prognosis for biomarkers in breast cancer [Commentary] **Lancet** 1998; 351:1753-4.

Dumitrescu RG, Cotarla I. Understanding breast cancer risk where do we stand in 2005? **J Cell Mol Med** 2005; 9:208-21.

Ellis CM, Dyson MJ, Stephenson TJ, Maltby EL. HER2 amplification status in breast cancer: a comparison between immunohistochemical staining and fluorescence in situ hybridisation using manual and automated quantitative image analysis scoring techniques. **J Clin Pathol** 2005; 58:710-714.

Esteva FJ, Sahin AA, Cristofanilli M, Arun B, Hortobagyi GN. Molecular prognostic factors for breast cancer metastasis and survival. **Semin Radiat Oncol** 2002; 12:319-28.

Esteva FJ, Hortobagyi GN. Prognostic molecular markers in early breast cancer. **Breast Cancer Res** 2004; 6:109-18.

Folgueira MAK, Brentani MM. **Oncologia molecular**. São Paulo: Atheneu; 2004. Câncer de mama. p.135-44.

Gross ME, Shazer RL, Agus DB. Targeting the HER-kinase axis in cancer. **Semin Oncol** 2004; vol 31(1 suppl 3):9-20.

Harris RA, Eichholtz TJ, Hiles ID, Page MJ, O'Hare MJ. New model of ErbB-2 over-expression in human mammary luminal epithelial cells. **Int J Cancer** 1999; 80:477-84.

Hortobagyi GN. Introduction. **Semin Oncol** 2001; 28(6 suppl 18):1-2.

Hortobagyi GN. Trastuzumab in the treatment of breast cancer. **N Engl J Med** 2005; 1734-1736.

Huggett J, Dheda K, Bustin S, Zumla A. Real-time RT-PCR normalization; strategies and considerations. **Genes Immun** 2005; 6: 279-284.

Hynes NE, Lane HA. ERBB receptors and cancer: the complexity of targeted inhibitors. **Nat Rev Cancer** 2005; 5:341-54.

Irikura VM, Hirsch E, Hirsh D. Effects of interleukin-1 receptor antagonist overexpression on infection by *Listeria monocytogenes*. **Infection and Immunity** 1999; 4:1901-1909.

Ishii S, Iamamoto F, Yamanashi Y, Toyoshima K, Yamamoto T. Characterization of the promoter region of the human c-erbB2 protooncogene. **Proc Natl Acad Sci USA** 1987; 84:4374-8.

Jardines L, Berger D. **The breast: comprehensive management of benign and malignant diseases**. 2nd ed. Philadelphia: W. B. Saunders; 1998. Integration of molecular biology into prognostic parameters for breast carcinoma. p.419-25.

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

Jin L, Yuan RQ, Fuchs A, et al. Expression of interleukin-1 β in human breast carcinoma. **Cancer** 1997; 80:421-34.

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

Karunagaran D, Tzahar E, Beerli RR, et al. ErbB2 is a common auxiliary subunit of NDF and EGF receptors: implications for breast cancer. **Embo J** 1996; 15:254-64.

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.

Keen J, Davidson NE. The biology of breast carcinoma. **Cancer** 2003; 97(3 Suppl):825-33.

King R. **Cancer biology**. 2nd ed. London: Prentice Hall; 2000. Natural history: the life of a cancer. p.8-30

Kouraklis G, Theocharis S. Histone deacetylase inhibitors: a novel target of anticancer therapy. **Oncol Rep** 2006; 2:489-94.

Kok JB, Roelofs RW, Giesendorf BA, et al. Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. **Lab Invest** 2005; 85:154-9.

Kraus MH, Popescu NC, Amsbaugh SC, King CR. Overexpression of the EGF receptor-related proto-oncogene erbB2 in human mammary tumor cell lines by different molecular mechanisms. **Embo J** 1987; 6:605-10.

Kute T, Lack CM, Willingham M, et al. Development of herceptin resistance in breast cancer cells. **Cytometry Part A** 2004; 57A:86-93.

Lacroix M, Leclercq G. Relevance of breast cancer cell lines as models for breast tumors: an update. **Breast Cancer Res Treat** 2004; 83:249-89.

Lane HA, Beuvink I, Motoyama AB, Daly JM, Neve RM, Hynes NE. ErbB2 potentiates breast tumor proliferation through modulation of p27^{Kip1}- Cdk2 complex formation: receptor overexpression does not determine growth dependency. **Mol Cell Biol** 2000; 20:3210-23.

Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. In search of antisense. **Trends Biochem Sci** 2004; 29:88-94.

Lakhani SR, Chaggar R, Davies S, et al. Genetic alterations in 'normal' luminal and myoepithelial cells of the breast. **J Pathol** 1999; 189:496-503.

LeBeau A, Deimling D, Kaltz C, et al. HER2/neu analysis in archival tissue samples of human breast cancer: a comparison of immunohistochemistry and fluorescent in situ hybridization. **J Clin Oncol** 2001; 19:354-63.

Lee S, Zhou G, Clark T, Chen J, Rowley JD, Wang SM. The pattern of gene expression in human CD15+ myeloid progenitor cells. **Proc Natl Acad Sci U S A** 2001; 98:3340-5.

Lee S, Clark T, Chen J, et al. Correct identification of genes from serial analysis of gene expression tag sequences. **Genomics** 2002; 79:586-602.

Lehner B; Williams RDC; Sanderson CM. Antisense transcripts in the human genome. **Trends Genet** 2002; 18:63-5.

Lenferink AEG, Busse D, Flanagan WM, Yakes FM, Arteaga CL. ErbB2/neu kinase modulates cellular p27^{Kip1} and cyclin D1 through multiple signaling pathways. **Cancer Research** 2001; 61:6583-6591.

Li YM, Pan Y, Wel Y, et al. Upregulation of CXCR4 is essential for HER2-mediated tumor metastasis. **Cancer Cell** 2004; 6:459-69.

Liu E, Thor A, He M, Barcos M, Ljung BM, Benz C. The HER2 (c-erbB-2) oncogene is frequently amplified in in situ carcinomas of the breast. **Oncogene** 1992; 7:1027-32.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. **Methods** 2001; 25:402-8.

Lohrisch C, Piccart M. An overview of HER2. **Semin Oncol** 2001; (6 suppl 18):3-11.

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.

Mai A, Massa S, Rotili D, Cerbara I, Valente S, Pezzi R et al. Histone deacetylation in epigenetics: an attractive target for anticancer therapy. **Med Res Rev** 2005; 3:261-309.

Marmor MD, Skaria KB, Yarden Y. Signal transduction and oncogenesis by erbB/HER receptors. **Int J Radiat Oncol Biol Phys** 2004; 58:903-13.

Matsuo Y, Akiyama N, Nakamura H, Yodoi J, Noda M, Kizaka-Kondoh S. Identification of a novel thioredoxin-related transmembrane protein. **J. Biol. Chem** 2001; 276:10032-10038.

Mei S, Ho AD, Mahlknecht U. Role of histone inhibitors in the treatment of cancer. **Int J Oncol** 2004; 6:1509-19.

Mhhalová M, Kodet R, Kalinová M, Hilská I. Relative quantification of ERBB2 mRNA in invasive duct carcinoma of the breast: correlation with ERBB2 protein expression and ERBB2 gene copy number. **Pathol Res Pract** 2003; 199:453-61.

Ménard S, Casalini P, Campiglio M, Pupa SM, Tagliabue E. Role of HER2/neu in tumor progression and therapy. **Cell Mol Life Sci** 2004; 61:2965-78.

Mendelsohn J, Howley PM, Israel MA, Liotta LA. **The molecular basis of cancer**. Philadelphia: WB Saunders; 1995.

Ministério da Saúde. **Câncer de mama: epidemiologia**. Disponível em: <URL:<http://www.inca.gov.br/cancer/mama>>. [2006 fev 10].

Moeller C, Yaylaoglu MB, Alvarez-Bolado G, Ihallu C, Eichele G. Murine Lix 1, a novel marker for substantia nigra, cortical layer 5, and hindbrain structures. **Brain Res Gene Expr Patterns** 2002; 1:199-203.

Montoyama AB, Hynes NE, Lane HA. The efficacy of erbB receptor-target anticancer therapeutics is influenced by the availability of epidermal growth factor-related peptides. **Cancer Res** 2002; 62:3151-8.

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

Morse DL, Carroll D, Weberg L, Borgstrom MC, Ranger-Moore J, Gillies RJ. Determining suitable internal standards for mRNA quantification of increasing cancer progression in human breast cells by real-time reverse transcriptase polymerase chain reaction. **Anal Biochem** 2005; 342:69-77.

[NCBI] National Center for Biotechnology Information. **Nucleotide – Nucleotide BLAST**. Available from: <URL:<http://www.ncbi.nlm.nih.gov/BLAST/>> [2005 Feb 20].

Negm RS, Verma M, Srivastava S. The promise of biomarkers in cancer screening and detection. **Trends Mol Med** 2002; 8:288-93.

Olayioye MA, Badache A, Daly JM, Hynes NE. An essential role for Src kinase in ErbB receptor signaling through the MAPK pathway. **Exp Cell Res** 2001; 267:81-7.

Orr MS, O'Connor PM, Kohn KW. Effects of c-erbB2 overexpression on the drug sensitivities of normal human mammary epithelial cells. **J Natl Cancer Inst** 2000; 92:987-94.

Page MJ, Amess B, Townsend RR, Parekh R, Herath A, Brusten L. Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties. **Proc Natl Acad Sci U S A** 1999; 96:12589-94.

Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. **Nucleic Acids Res** 2001; 29:2002-7.

Placido S, Carlomagno C, Laurentiis M, Bianco AR. C-erbB2 expression predicts tamoxifen efficacy in breast cancer patients. **Breast Cancer Res Treat** 1998; 52:55-64.

Qin J, Qian Y, Yao J, Grace C, Li X. SIGIRR inhibits interleukin-1 receptor and toll-like receptor 4 mediated signaling through different mechanisms. **J Biol Chem** 2005; 280:25233-41.

Rabindran SK. Antitumor activity of HER-2 inhibitors. **Cancer Lett** 2005; 227:9-23.

Radice D, Redaelli A. Breast cancer management: quality-of-life and cost considerations. **Pharmacoeconomics** 2003; 21:383-96.

Rahman N, Stratton MR. The genetics of breast cancer susceptibility. **Annu Rev Genet** 1998; 32:95-121.

Reis EM, Nakaya HI, Louro R, et al. Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. **Oncogene** 2004; 23:6684-92

Révillion F, Bonnetterre J, Peyrat JP. ERBB2 oncogene in human breast cancer and its clinical significance. **Eur J Cancer** 1998; 34:791-808.

Rhodes A, Borthwick D, Sykes R, Al-Sam S, Path FRC, Paradiso A. The use of cell line standards to reduce HER2/neu assay variation in multiple cancer centers and the potential of automated image analysis to provide for more accurate cut points for predicting clinical response to trastuzumab. **Am J Clin Pathol** 2004; 122:51-60.

Rodrigues-Manzaneque JC, Lane TF, Ortega MF, Hynes RO, Lawler J, Iruela-Arispe ML. Thrombospondin-1 suppresses spontaneous tumor growth and inhibits activation of matrix metalloproteinase-9 and mobilization of vascular endothelial growth factor. **Proc. Nat. Acad. Sci.** 2001; 98:12485-12490.

Roskoski R Jr. The ErbB/HER receptor protein-tyrosine kinases and cancer. **Biochem Biophys Res Commun** 2004; 319:1-11.

Savinainen KJ, Saramäki OR, Linja MJ, et al. Expression and gene copy number analysis of ERBB2 oncogene in prostate cancer. **Am J Pathol** 2002; 160:339-45.

Shendure J, Church GM. Computational discovery of sense-antisense transcription in the human and mouse genomes. **Genome Biol** 2002; 3:Research0044.

Sherman CD, Hossfeld DK. Câncer de mama. In: Hossfeld DK, Sherman CD, Love RR, Bosh FF, editor. **Manual de oncologia clínica**. Trad da UICC. 5 ed. São Paulo: Fundação Oncocentro de São Paulo; 1993. p.259-78

Silva AP, Chen J, Carraro DM, Wang SM, Camargo AA. Generation of longer 3' cDNA fragments from massively parallel signature sequencing tags. **Nucleic Acids Res** 2004a; 32:e94.

Silva AP, De Souza JE, Galante PA, Riggins GJ, De Souza SJ, Camargo AA. The impact of SNPs on the interpretation of SAGE and MPSS experimental data. **Nucleic Acids Res** 2004b; 32 32:6104-10.

Singer CF, Kronsteiner N, Hudelist G, et al. Interleukin system and sex steroid receptor expression in human breast cancer: interleukin 1 α protein secretion is correlated with malignant phenotype. **Clin Cancer Res** 2003; 9:4877-83.

Srinivas PR, Kramer BS, Srivastava S. Trends in biomarker research for cancer detection. **Lancet Oncol** 2001; 2:698-704.

Srivastava S, Gopal-Srivastava R. Biomarkers in cancer screening: a public health perspective. **J. Nutri** 2002; 132 (8 suppl): 2471S-5S.

Stamps AC, Davies SC, Burman J, O'Hare MJ. Analysis of proviral integration in human mammary epithelial cell lines immortalized by retroviral infection with a temperature-sensitive SV40 T-antigen construct. **Int J Cancer** 1994; 57:865-74.

Stern DF. Tyrosine Quinase signalling in breast cancer erbB family receptor tyrosine kinases. **Breast Cancer Res** 2000; 2:176-83.

Styblo MT, Wood WC. **The breast: comprehensive management of benign and malignant diseases**. 2nd ed. Philadelphia: W. B. Saunders; 1998. p.419-25: Prognostic factors for breast cancer.

Sunpaweravonh S, Sunpaweravonh P. Recent developments in critical genes in the molecular biology of breast cancer. **Asian J Surg** 2005; 28:71-5.

Swindell EC, Moeller C, Thaller C, Eichele G. Cloning and expression analysis of chicken *Lix1*, a founding member of a novel gene family. **Mech Dev** 2001; 109:405-8.

Szymanski M, Barciszewska MZ, Zywicki M, Barciszewski J. Noncoding RNA transcripts. **J Appl Genet** 2003; 44:1-19.

Tanner M, Kapanen AI, Junttila T, et al. Characterization of a novel cell line established from a patient with Herceptin-resistant breast cancer. **Mol Cancer Ther** 2004; 3:1585-92.

Thomassen E, Renshaw BR, Sims JE. Identification and characterization of SIGIRR, a molecule representing a novel subtype of the IL-1R superfamily. **Cytokine** 1999; 6:389-399

Thomssen C, Janicke F, Harbeck N. Clinical relevance of prognostic factors in axillary node-negative breast cancer. **Onkologie** 2003; 26:438-45.

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

Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome Biol** 2002; 3:Research 0034.1.

Van de Vijver MJ, He YD, Van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. **N Engl J Med** 2002; 347:1999-2009.

Van Poznak C, Seidman AD. Breast cancer. In: Bertino JR, editor. **Encyclopedia of cancer**. 2nd ed. Amsterdam: Elsevier Science; 2002. p.287-99.

Van't Veer LJ, Dai H, Van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. **Nature** 2002; 415: 530-536.

Wakeling AE. Inhibitors of growth factor signalling. **Endocrine-Related Cancer**. 2005; 12:S183-S187.

Wald D, Qin J, Zhao Z, et al. SIGIRR, a negative regulator of Toll-like receptor-interleukin 1 receptor signalling. **Nature Immunol** 2003; 4:920-7.

Weinberg RA. How cancer arises. **Sci Am** 1996; 275:62-70.

Witton CJ. Structure of HER receptors and intracellular localisation downstream effector elements gives insight into mechanism of tumor growth promotion. **Breast Cancer Res** 2003; 5:206-7.

Yamauchi H, Stearns V, Hayes DF. When is a tumor marker ready for prime time? A case study of c-erbB-2 as a predictive factor in breast cancer. **J Clin Oncol** 2001; 19:2334-56.

Yu Qunyan, Yan G, Sicinski P. Specific protection against breast cancer by cyclin D1 ablation. **Nature** 2001; 411:1017-1021.

Xie D, Shu XO, Deng Z, et al. Population-based, case-control study of HER2 genetic polymorphism and breast cancer risk. **J Natl Cancer Inst** 2000; 92:412-7.

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

Zhang X, Ding L, Sandford AJ. Selection of reference genes for gene expression studies in human neutrophils by real-time PCR. **BMC Mol Biol** 2005; 6:1-7.

Zhou BP, Hung MC. Dysregulation of cellular signaling by HER2/neu in breast cancer. **Semin Oncol** 2003; 30(5 Suppl 16):38-48.

Anexos

Anexo 1

Anexo 2

Anexo 3

Anexo 4