

**IDENTIFICAÇÃO DE MARCADORES MOLECULARES
EM CÂNCER DE MAMA ATRAVÉS DA TÉCNICA
DE *MICROARRAY* UTILIZANDO UMA PLATAFORMA
DE EXONS TUMOR-ASSOCIADOS**

MARIA CRISTINA RODRIGUES RANGEL

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

Área de Concentração: Oncologia

Orientadora: Dra. Dirce Maria Carraro

Co-Orientadora: Dra. Helena Paula Brentani

São Paulo

2008

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Rangel, Maria Cristina Rodrigues

Identificação de marcadores moleculares em câncer de mama através da técnica de microarray utilizando uma plataforma de exons tumor-associados / Maria Cristina Rodrigues Rangel – São Paulo, 2008.

166p.

Tese(Doutorado)-Fundação Antônio Prudente.

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

Orientadora: Dra. Dirce Maria Carraro

Descritores: 1. CÂNCER DA MAMA. 2. PROCESSAMENTO ALTERNATIVO. 3. EXONS. 4. ANALISES MICROARRAY.

*“Mas a ciência e a vida de todo dia não podem e não devem ser separadas.
A ciência, para mim, (...) é baseada em fato, conhecimento e
experimentação.”*

Rosalind Franklin

*“Plante seu jardim e decore sua alma, ao invés de esperar que alguém lhe
traga flores...E você aprende que realmente pode suportar...que realmente é
forte, e que pode ir mais longe, depois de pensar que não pode mais...”*

Willian Shakespeare

DEDICATÓRIA

A Deus, meu criador

Ao meu marido Eros, minha fortaleza

*À minha família, que sempre acreditou em mim, rezou e torceu por cada
conquista*

Aos meus amigos, companheiros incondicionais

Às pacientes que contribuíram para este trabalho

*Ao querido tio João que lutou até o fim contra o câncer, mas teve sua missão
encerrada aos 03 de maio deste ano,*

Dedico.

AGRADECIMENTOS

A Deus pela vida, por me permitir chegar até aqui e por me dar forças e sabedoria para continuar nesta caminhada.

Em especial à minha orientadora Dra. Dirce Maria Carraro pela amizade, e por todo o apoio e dedicação. Obrigada pela paciência e confiança ao longo desses anos, e pela oportunidade de realizar este trabalho da melhor forma.

À minha co-orientadora Dra. Helena P. Brentani pela fundamental contribuição e pelas importantes sugestões ao longo deste trabalho. Obrigada por estar sempre disponível a ajudar.

Ao Hospital A.C. Camargo e Fundação Antônio Prudente, São Paulo, na pessoa do ilustre diretor Dr. Ricardo Renzo Brentani.

Ao Instituto Ludwig de Pesquisa Sobre o Câncer, São Paulo, na pessoa da ilustre diretora Dra. Luisa Lina Villa.

À Dra. Luisa Villa agradeço em especial pelo grande exemplo, e por todo o carinho e amizade. Muito obrigada pelo incentivo e disponibilidade em ajudar.

Ao Dr. Sandro J. de Souza pela colaboração e importantes sugestões.

Ao Departamento de Anatomia Patológica do Hospital A.C. Camargo em nome do Dr. Fernando Augusto Soares.

A todos os integrantes do Laboratório de Biologia Molecular e Genômica do Hospital A.C. Camargo pela convivência, amizade, apoio e ensinamentos. Aderbal, Alexandre, Elen, Elisa, Eloísa, Fábio, Felipe, Gustavo, Louise, Luana, Mariana, Mev, Nádia, Sabrina, Thiago e Vera. Obrigada pelos bons momentos vividos dentro e fora do laboratório.

Aos grandes amigos: Thiago ('Kakóviski'), por todo ensinamento sobre petulância, pela amizade incondicional e muitas, muitas risadas; Nádia ('Friendoca') pela força e incentivo durante todos esses anos, por toda dedicação e tempo dispensados, pelos bons conselhos e por sempre me compreender; Sabrina ('Salssitcha'), pelo exemplo de organização e resolução de problemas, pela imensa ajuda com tudo, pelas risadas e pelos inesquecíveis abraços de urso! Elen ('Tchuca'), pelo exemplo de doação e perseverança, obrigada por toda a ajuda, pela amizade sincera e por sempre me ouvir e

incentivar; Mev ('Mevita') pelos momentos de reflexão sobre a vida no cafézinho ou Rei do Mate, obrigada por tudo 'amiguinha', principalmente pelo exemplo de otimismo; Aderbal ('Dedé') pela amizade repleta de risadas, bons momentos e 'lindas' canções. Enfim, meus amigos obrigada por acreditarem em mim, pelos ensinamentos e por sempre lerem em meu olhar o que estava sentindo, me apoiarem incondicionalmente e serem responsáveis por momentos inesquecíveis que me lembrarei com muita saudade! Amizades sinceras que sempre cultivarei!

A todos os amigos mais que especiais que colaboraram de alguma forma para a execução deste trabalho: Elisa, Elen, Jane, Laura, Louise e Reimar. Sem vocês, cada um do seu modo, certamente tudo teria sido mais difícil. Muito obrigada por tudo e pelo carinho! À amiga Mariana ('Mari') quero agradecer pela disponibilidade em me ajudar muitas vezes também e pela amizade sincera!

À amiga Natanja Kirschbaum-Slager pela importante contribuição neste trabalho, pela amizade e as inesquecíveis conversas sobre o futuro.

A todos os inesquecíveis ex-integrantes do Laboratório de Biologia Molecular que fizeram parte da minha jornada aqui, mas seguiram antes de mim: Adriana ('Heydrian'), Jamila, Jane ('Jeyne'), Laura ('Laurita'), Leandra, Paulo ('Meso'), Rafael, Reimar ('Rei do Mar'), Waleska ('Wal') e Wilson ('Wirso'). Obrigada pelo imenso carinho e amizade!

Aos queridos 'agregados' do Laboratório de Biologia Molecular e Genômica: David ('Davidson'), Fernanda, L. Paulo, Rodolfo ('Rodis') e Rodrigo, com quem pude compartilhar muitos bons momentos ao longo desses anos!

À Dra. Maria do Socorro Maciel e à Dra. Adriane Pimentel pela grande ajuda na revisão dos prontuários, em especial à Dra. Socorro pela amizade e todo aprendizado.

Ao Dr. João Gonçalves pelas análises estatísticas da parte clínica e pelo aprendizado.

À Dra. Sabrina Silva, pela imensa ajuda com a parte clínica deste trabalho.

Às meninas da Patologia, Fernanda, Alexandra, Yukie, Luciane e Edaíse pela amizade e pelos bons momentos passados juntas (até os com o prof 'De!').

Às amigas Walleka Martins, Elza Durham, Nair Muto e Bárbara Melo pelo carinho, pelas inesquecíveis conversas e bons momentos vividos juntas.

Ao André Silva pela boa vontade em ajudar na localização das amostras.

Aos colegas do Laboratório de Biologia Computacional do Hospital A. C. Camargo. Ao Luiz Paulo pela análise dos dados de microarray, ao Daniel pelo auxílio com o mapeamento das membranas, e ao Dr. Eduardo Abrantes, Diogo, César e Arthur pela atenção de sempre e disponibilidade em ajudar.

Aos colegas do Laboratório de Biologia Molecular e Genômica do Instituto Ludwig de Pesquisa sobre o Câncer em nome da Dra. Anamaria A. Camargo. Em especial à Anna Chris que me ajudou muito com as primeiras hibridizações, ao Murilo Vieira (ex-integrante) pela ajuda com o GeNorm, à Valéria Paixão, sempre pronta a ajudar em tudo, ao Ricardo Moura pela atenção de sempre, à MaCris pela ajuda com o seqüenciamento, à Ana Paula Medeiros pela disponibilidade em ajudar com o Real Time, enfim à todos os integrantes e companheiros pela atenção de sempre e disponibilidade em ajudar.

Aos funcionários Carlinhos, Severino e Miuky pela confecção das lâminas histopatológicas utilizadas neste estudo.

Ao Dr. Hugo Campos e Dra. Cíntia Ozório pela revisão das lâminas histopatológicas. À Dra. Cíntia em especial pela atenção e boa vontade em esclarecer minhas dúvidas no corredor.

Ao Dr. Humberto Torloni e a Sra. Hirde Contesini, responsáveis pelo SAME (Serviço de Arquivo Médico) e a todos os funcionários pelo auxílio na obtenção dos prontuários.

À pós-graduação da Fundação Antônio Prudente, pela grande contribuição à pesquisa científica. Em especial ao diretor Dr. Luiz Fernando L. Reis, à coordenadora Ana Maria A. Kuninari e à secretária Luciana Pitombeira, pela excelente administração, auxílio constante e grande competência.

Às queridas funcionárias da biblioteca da Fundação Antônio Prudente em nome da Sra. Suely Francisco. Agradeço por sempre dispensarem sua atenção e tempo comigo. Agradeço também pela amizade e imensa colaboração na obtenção do material bibliográfico para o desenvolvimento deste manuscrito e sua revisão criteriosa. Em especial agradeço à 'Su', 'Fran' e Rosi' que acompanham minha jornada desde o começo e sempre foram muito amáveis.

À Banca de qualificação: Dra. Anamaria Aranha Camargo, Dr. Eduardo Moraes Rego Reis, Dra. Maria Aparecida Koike Folgueira e à Dra. Ana Tereza Ribeiro de Vasconcelos, pela importante contribuição durante esses anos,

através de críticas e sugestões que foram fundamentais no decorrer do desenvolvimento deste trabalho.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro e ao parecerista pelas críticas e sugestões durante a análise dos relatórios científicos.

Ao Dr. Alex Fiorini de Carvalho e Chamberlein Mendonça, do Microarray Facility, pela confecção da membrana de microarray e por todo o aprendizado.

Às pacientes que cooperaram com esta pesquisa. Sem vocês simplesmente não seria possível.

Aos queridos amigos Erika Taniguti ('Gueixita'), Carolina Cappi ('Carol'), Andrea Glatt ('Dêa'), Jeanne Andrade ('Jê'), Marcelo Zerillo ('Ma'), Marcos Ternero ('Suba'), Mônica Serra ('Monikinha'), (UNESP), Anastásia Staack ('Stase'), Ana Márcia Cardoso, Laila Salman ('Lailita'), Liane Leobons ('Li'), Renata Cobianchi ('Rê'), Leonardo Moraes ('Leo'), Maria Tereza Aquino ('Teresoca'), Thamy Quintanilha, (LORENA), Daniela Truffi ('Dani'), Thayne Munhoz ('Tha'), Viviane Rezende ('Vivi'), (USP-ESALQ), que mesmo longe, estarão sempre presentes. Obrigada pelo carinho e amizade!

Ao meu amigo inseparável 'Ma' agradeço em especial pelos momentos inesquecíveis desde o tempo da Faculdade, pela paciência, carinho e toda a confiança durante esses anos. Obrigada também por me fazer dar tanta risada e me tornar uma pessoa mais feliz!

À querida D. Irene, que me acolheu com muito amor em minha chegada a Rio Claro, e aos amigos Andréa, Baço e Pixuco pelo carinho e acolhida.

Aos meus grandes amigos e padrinhos Nádía Castro e Marcelo Zerillo, que além de me ouvirem, apoiarem incondicionalmente e sempre se importarem comigo, mudaram o rumo da minha vida. Obrigada por sempre me incentivarem e por me ajudarem a realizar este trabalho da melhor forma possível.

À Dra. Cláudia B. Monteiro-Vitorello, grande amiga e exemplo de pesquisadora que levarei por toda a minha vida. Obrigada pela grande ajuda, pelos ensinamentos e por sempre ter acreditado em mim!

Ao Dr. Luis Eduardo Aranha Camargo, pela oportunidade, carinho e amizade. Devo a você muito do que sou hoje!

Ao Dr. Carlos F. M. Menck, exemplo admirável de inteligência e amor à ciência, obrigada pela oportunidade e por todo o aprendizado.

Ao Dr. Januário B. Cabral Neto, que me abriu as portas para a Ciência e me inspirou a ser quem sou hoje. Obrigada pela oportunidade de ter te conhecido como pessoa e cientista.

Em especial aos meus pais queridos, Sérgio e Telma, pelo exemplo de caráter, determinação e união familiar. Por terem anulado seus sonhos muitas vezes em função dos meus e por seu esforço constante em me oferecer as melhores condições para que eu chegasse até aqui. Obrigada por todo o amor e por sempre acreditarem em mim!

Aos meus amados irmãos Sérgio, Renato e Thaís (Tati), e minha cunhada Kariza por todo apoio, ajuda, paciência e cuidado ao longo desses anos. Obrigada principalmente pela amizade verdadeira!

Tati, em especial quero agradecer a você por todo companheirismo e compreensão nesta caminhada. Obrigada por me livrar de alguns apuros também! Ao Serginho, em especial, muito obrigada pela ajuda fundamental com o inglês! Renato e Kariza obrigada principalmente pelo maior presente que pude receber durante esses anos: Gigi e Gabi!

Às minhas lindas princesas Giovanna e Gabriela. Vocês são a alegria da 'tia-dinda' e meu principal incentivo!

À minha família no Rio em nome da minha avó Alidéa e da tia-dinda Vany, e à família do Eros em Lorena em nome da minha sogra Dona 'Marilyn' (todos os meus amados sobrinhos e cunhados), que sempre estiveram em meu coração e me apoiaram e incentivaram durante esta longa caminhada. Obrigada a cada um pelo amor, carinho e pela compreensão da minha ausência.

Aos meus queridos padrinhos, Vany e Moacyr, grandes exemplos de sabedoria, dedicação e oração.

Ao meu inesquecível e amado primo Robson. Obrigada pelo amor e lição de vida. Você será sempre um exemplo alegria, bondade e altruísmo pra mim!

Ao meu amado marido Eros. Obrigada pela compreensão e companheirismo ao longo desses anos e pelo amor incondicional. Pelo olhar carinhoso, quando eu mais precisava e pelas palavras de consolo quando eu mais me desesperava. Por seu apoio constante, imensa ajuda, paciência e presença em todos os momentos, principalmente nos mais difíceis. Obrigada meu amor por ser quem você é! Sem você ao meu lado eu simplesmente não teria conseguido!

RESUMO

Rangel, MCR. **Identificação de marcadores moleculares em câncer de mama através da técnica de *microarray* utilizando uma plataforma de exons tumor-associados.** São Paulo; 2008. [Tese de Doutorado-Fundação Antônio Prudente]

Análises recentes têm mostrado a ocorrência de *splicing* alternativo (AS) do mRNA em pelo menos 60% dos genes humanos, sendo que 80% desses eventos ocorrem dentro da região codificadora, aumentando a diversidade proteômica. Sabe-se que algumas variantes geradas por AS são preferencialmente expressas em tumores humanos e são potenciais marcadores moleculares, podendo contribuir para o desenvolvimento de fatores diagnósticos e prognósticos mais precisos, assim como novos alvos terapêuticos. Para identificar variantes de *splicing* diferencialmente reguladas em câncer de mama, 270 exons expressos em tecidos tumorais foram selecionados através de uma análise computacional, dos quais 75 foram associados a mama, por apresentarem expressão em bibliotecas originadas de tecido tumoral de mama e não no correspondente normal. Esses exons foram imobilizados em membranas de *nylon* juntamente com controles positivos e negativos, e hibridizados contra amostras tumorais e normais de mama. Para identificar os exons mais expressos em tecidos tumorais, três comparações foram realizadas: 4 amostras de linhagem tumoral contra 2 de linhagem normal de mama (LTxLN), 27 amostras de tecido tumoral contra 5 de tecido não neoplásico de mama (TxN), e comparação entre 4 amostras de tecido pareadas tumor-normal (PTxPN). Foi aplicado o teste *TStudent* para determinar os exons diferencialmente expressos ($p < 0,05$) em cada comparação (LTxLN; TxN; PTxPN), sendo identificados como superexpressos 24 exons na comparação LTxLN, 79 exons na comparação TxN e 195 exons na comparação PTxPN. Para validação técnica, foram selecionados aqueles exons que apresentaram *fold*

change ≥ 3 em amostras tumorais e que foram identificados em pelo menos 2 comparações. Quatorze exons foram identificados pelos experimentos de *microarray*, os quais foram avaliados através de RT-PCR quantitativo (qRT-PCR), usando o mesmo grupo de amostras já utilizadas anteriormente. Para testar se os exons selecionados por *microarray* pertenciam, de fato, à variantes de *splicing* superexpressas e não ao gene como um todo, foram adotados 2 critérios: (1) validação por qRT-PCR da variante que contém o exon superexpresso (VCE) (*fold change* ≥ 3); (2) para os candidatos confirmados pelo critério (1), avaliação da expressão do gene como um todo (através da análise da expressão constitutiva do gene). Três VCE foram confirmadas através de qRT-PCR como superexpressas em tumor de mama, sendo elas TRIM37-VCE, MK-STYX-VCE e BRRN1-VCE. A expressão constitutiva do gene (EC) foi analisada por qRT-PCR, através do desenho de iniciadores em exons constitutivos, isto é, presente em todas as variantes do gene. Os resultados mostraram que a razão do nível de expressão entre as 3 VCE e a expressão constitutiva do gene (VCE/EC) foi significativamente maior em amostras tumorais de mama quando comparadas às amostras normais ($p < 0,05$), sugerindo que TRIM37-VCE, MK-STYX-VCE e BRRN1-VCE são, de fato, variantes de *splicing* superexpressas em câncer de mama. Essas variantes foram também avaliadas em um grupo independente de 40 amostras tumorais de mama para validar biologicamente a superexpressão da VCE em amostras tumorais de mama quando comparadas às amostras normais. Todos os dados foram correlacionados com características clínicas e histopatológicas das amostras, e algumas associações significativas foram encontradas, tais como a expressão de receptores de estrógeno (ER+) e progesterona (PgR+) com TRIM37-VCE. Embora um maior grupo de dados experimentais seja requerido para a completa exploração do estudo desenvolvido, os resultados sugerem 3 variantes de *splicing* superexpressas em carcinoma ductal de mama, candidatos a marcadores moleculares.

SUMMARY

Rangel, MCR. **[Identification of breast cancer molecular markers through microarray technology using a tumor-associated exons platform]**. São Paulo; 2008. [Tese de Doutorado-Fundação Antônio Prudente]

Current analyses have shown that alternative mRNA splicing (AS) appears in at least 60% of human genes and 80% of these events occurs within the coding region, increasing the proteomic diversity. Some AS variants have been preferentially expressed in human tumors and are potential molecular markers, contributing to the development of more accurate diagnostic and prognostic factors as well as therapeutic targets. To identify differentially regulated splicing variants in breast cancer, 270 exons expressed in tumor tissues were selected by a computational analysis, of which 75 were associated with breast, because they are found to be expressed in libraries that originated from breast tumors and they were not found in the corresponding normal libraries. These exons were immobilized on nylon membranes together with positive and negative controls, and hybridized against tumor and normal breast samples. To identify the most highly expressed exons in tumor tissues, 3 comparisons were performed: 4 tumor against 2 normal breast cell lines (LTxLN), 27 tumor against 5 non-neoplastic breast tissues (TxN) and 4 matched tumor-normal samples (PTxPN). A T-student test was used to select for differentially expressed exons ($p < 0.05$) in each comparison (LTxLN; TxN; PTxPN). Here, 24 were selected as over expressed exons in the LTxLN comparison, 79 exons in the TxN comparison and 195 exons in the PTxPN comparison. For technical validation, those exons having a fold change of ≥ 3 in tumor samples and being present in at least 2 comparisons were selected. Fourteen exons were identified by microarray experiments and evaluated through quantitative RT-PCR (qRT-PCR), using the same sample set utilized previously. In order to test whether

the exons selected by microarray experiments belonged to a prone over expressed splicing variant, 2 criteria were adopted: (1) Validation by qRT-PCR of the variant that comprises the selected over expressed exon (VCE) (fold change ≥ 3); (2) For those exons confirmed in criterion (1), evaluation of whole gene expression (through the analysis of constitutive gene expression) is adopted as a secondary criterium. Three of the VCE's were confirmed through qRT-PCR as being over expressed in breast tumor, such as TRIM37-VCE, MK-STYX-VCE and BRRN1-VCE. The constitutive expression of the gene (EC) was analyzed by qRT-PCR, through the primer design within constitutive exons, that is, present in all variants of the gene. The results showed that the expression level ratio between the 3 VCE's and the constitutive expression of genes (VCE/EC) was significantly higher in tumor samples when compared to normal samples ($p < 0.05$), suggesting that TRIM37-VCE, MK-STYX-VCE and BRRN1-VCE are indeed breast tumor associated variants. These variants were also evaluated in an independent set of 40 breast tumor samples to biologically validate the VCE over expression in tumor samples as compared to normal samples. All data were correlated with clinical and histopathological samples features, and some significative associations were found, such as the expression of estrogen (ER+) and progesterone (PgR+) receptors with TRIM37-VCE. Although an increase of the experimental data is required for the complete exploration of this study, the results suggest 3 splicing variants that are over expressed in ductal carcinoma of the breast, and are candidates for molecular markers.

LISTA DE FIGURAS

Figura 1	Tipos de <i>splicing</i> alternativo	18
Figura 2	Seqüenciamento do exon 115538_6	39
Figura 3	Representação das bandas 28S e 18S de RNA ribossomal de 3 amostras de tecido tumoral extraídas por CsCl	45
Figura 4	Representação de RNAa de 3 linhagens celulares de mama-1: MDA-MB-435; 2: MDA-MB-231 e 3: MCF7.	47
Figura 5	Representação de RNAa de 4 amostras de tecido tumoral de mama- 1: 19T, 2: 26T; 3: 27T; 4: 28T.	47
Figura 6	Representação esquemática do pareamento dos iniciadores na VCE (a) e na VSE (b).	54
Figura 7	Esquema 1 = Representação hipotética do pareamento inespecífico do iniciador F-SE na variante com o exon (VCE).	57
Figura 8	Modelo matemático para quantificação relativa em qRT-PCR.	63
Figura 9	Representação de alguns produtos amplificados.	66
Figura 10	Representação de membrana hibridizada com amostra de tecido tumoral de mama.	68
Figura 11	Gráfico de Correlação de <i>Pearson</i> .	70
Figura 12	Gráfico de dispersão - amostra MN04.	71
Figura 13	Gráfico representativo da intensidade média dos exons em relação à extremidade 3'.	72
Figura 14	Diagrama de <i>Venn</i> , representando os exons que apareceram superexpressos nas comparações.	76
Figura 15	<i>Bay Boots</i> . a: exon 4363_4; b: exon 89442_12 e c: exon 115216_14.	78
Figura 16	Representação esquemática do desenho dos iniciadores nas seqüências da VCE do gene TRIM37.	83

Figura 17	Representação dos produtos amplificados por RT-PCR para o gene GAPDH ₁₁₀₁ , confirmando o tamanho esperado de 102pb.	84
Figura 18	Representação das VCE amplificadas com a enzima <i>Taq Gold (Applied Biosystems)</i> na temperatura de 54°C.	84
Figura 19	Análise pelo programa geNorm para a identificação dos genes mais estáveis para o conjunto de amostras utilizadas no presente estudo.	86
Figura 20	Gráfico da média dos valores de estabilidade de expressão dos genes GAPDH, ACTB, HPRT, RPLP0 e BCR.	87
Figura 21	Análise pelo programa geNorm dos normalizadores para a identificação dos genes mais estáveis no conjunto de amostras utilizadas nesse estudo.	88
Figura 22	Gráfico da média dos valores de estabilidade de expressão dos genes GAPDH, ACTB e HPRT.	89
Figura 23	Gráficos representando à curva de dissociação, os CTs das triplicatas e a curva padrão para o gene GAPDH e VCE do gene TRIM37 respectivamente.	92
Figura 24	Gráficos de amplificação das amostras e a curva de dissociação para os genes GAPDH e da VCE do gene TRIM37 respectivamente.	94
Figura 25	Representação esquemática do pareamento dos iniciadores.	97
Figura 26	Representação dos produtos amplificados utilizando iniciadores no exon 1 e 3. Em todos os genes houve amplificação da variante VCE, mais forte, e VSE, mais fraca.	98

Figura 27	Alinhamento das seqüências das 2 variantes (VCE e VSE) seqüenciadas do gene TRIM37 contra o genoma humano e seqüências de mRNA disponíveis nos bancos de dados, com auxílio da ferramenta BLAT (UCSC).	100
Figura 28	Representação esquemática do desenho dos iniciadores nas seqüências da variante VSE do gene TRIM37.	101
Figura 29	Esquema 2	103
Figura 30	Representação do teste de especificidade do iniciador F-VSE1.3 para os genes TRIM37, MK-STYX e BRRN1.	104
Figura 31	Análise entre VCE e VSE pertencentes aos genes TRIM37 e MK-STYX.	106
Figura 32	Análise entre VCE e o fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1.	107
Figura 33	Análise entre VCE e o fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1.	109
Figura 34	Análise entre VCE e o fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1.	110
Figura 35	Mapeamento do exon associado a tumor pertencente à VCE do gene TRIM37 na seqüência protéica.	111
Figura 36	Mapeamento do exon associado a tumor pertencente à VCE do gene MK-STYX na seqüência protéica.	112
Figura 37	Mapeamento do exon associado a tumor pertencente à VCE do gene BRRN1 na seqüência protéica.	112

LISTA DE TABELAS

Tabela 1	Número de exons associados a tumores para cada tecido analisado através de análise computacional.	28
Tabela 2	Exons superexpressos em tumor em 3 comparações.	75
Tabela 3	Exons selecionados para marcador molecular entre as comparações TxN, PTxPN e LTxLN.	80
Tabela 4	Iniciadores <i>forward</i> e <i>reverse</i> para cada gene. F-CE: <i>forward</i> , desenhado inteiro no exon 2; R: <i>reverse</i> , desenhado no exon 3.	82
Tabela 5	Genes e seus respectivos valores de slope, eficiência e coeficiente de regressão (R ²).	93
Tabela 6	Resumo dos resultados de expressão relativa das amostras tumorais em relação às normais das variantes com o exon (VCE) para os 13 genes avaliados.	95
Tabela 7	Análise da VSE dos 3 genes através de seqüenciamento e evidências em bancos de dados.	99
Tabela 8	Iniciadores <i>forward</i> e <i>reverse</i> para cada gene. F-VSE1.3: <i>forward</i> da variante sem o exon, desenhado na junção entre o exon 1 e o exon 3; R: <i>reverse</i> , desenhado no exon 3.	101
Tabela 9	VSE dos genes TRIM37 e MK-STYX e seus respectivos valores de slope, eficiência e coeficiente de regressão (R ²).	105
Tabela 10	Fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1, e seus respectivos valores de slope, eficiência e coeficiente de regressão (R ²)	107
Tabela 11	Dados clinicopatológicos, demográficos e sobre o tratamento dos pacientes portadores de carcinoma ductal invasivo de mama incluídos neste estudo.	116

Tabela 12	Associação entre a VCE do gene TRIM37 e as variáveis clinicopatológicas.	118
Tabela 13	Associação entre a expressão da VCE do gene TRIM37 nas amostras ER+ e a variável menarca.	119
Tabela 14	Associação entre a expressão da VCE do gene TRIM37 nas amostras PgR+ e a variável faixa etária.	119
Tabela 15	Associação entre a relação VCE/EC do gene TRIM37 e as variáveis clinicopatológicas.	120
Tabela 16	Associação entre a expressão da VCE do gene MK-STYX e as variáveis clinicopatológicas.	121
Tabela 17	Associação entre a expressão da razão VCE/EC do gene MK-STYX e as variáveis clinicopatológicas.	122
Tabela 18	Associação entre a relação VCE/EC do gene MK-STYX nas amostras não menopausa e os estádios clínicos I e II.	123
Tabela 19	Associação entre a expressão da VCE do gene BRRN1 e as variáveis clinicopatológicas.	123
Tabela 20	Associação entre a relação VCE/EC do gene BRRN1 e as variáveis clinicopatológicas.	124

LISTA DE ABREVIACOES

AJCC	do ingls <i>The American Joint Committee on Cancer</i>
AS	do ingls <i>Alternative Splicing</i>
BLAST	do ingls <i>Basic Local Alignment Search Tool</i>
BRCA 1	do ingls <i>Breast Cancer Gene 1</i>
BRCA 2	do ingls <i>Breast Cancer Gene 2</i>
CDS	do ingls <i>Coding Sequence</i>
CsCl	Cloreto de Csio
cDNA	do ingls <i>complementary DNA</i>
cm	centmetro
C_T	do ingls <i>Cycle Treshold</i>
DCIS	do ingls <i>Ductal Carcinoma in situ</i>
DNA	do ingls <i>deoxyribonucleic acid</i>
dNTP	do ingls <i>deoxynucleoside 5' triphosphato</i>
DEPC	do ingls <i>Di-etil pirocarbonato</i>
DTT	do ingls <i>1,4-dithiothreitol</i>
EC	Expresso constitutiva do gene
EDTA	do ingls <i>ethylenediamine tetraacetate</i>
ER	do ingls <i>Estrogen Receptor</i>
ERBB2	Membro da famlia de receptores de fator de crescimento epidermal
ESE	do ingls <i>Exonic Splicing Enhancers</i>
ESS	do ingls <i>Exonic Splicing Silencer</i>
ESTs	do ingls <i>Expressed Sequence Tags</i>
FAPESP	Fundao de Amparo  Pesquisa do Estado de So Paulo
HB4a	Linhagem de clulas epiteliais mamrias humanas
HB4a C5.2	Linhagem de clulas epiteliais mamrias humanas transfectada com o oncogene <i>ERBB2</i>
hnRNP A1	do ingls <i>heterogeneous nuclear ribonueoprotein A1</i>

IDC	do inglês <i>Invasive Ductal Carcinoma</i>
IPTG	do inglês <i>isopropyl-β-D-thiogalactoside</i>
ISE	do inglês <i>Intronic Splicing Enhancers</i>)
ISS	do inglês <i>Intronic Splicing Silencer</i>
M	Molar
MgCl₂	Cloreto de Magnésio
mL	Mililitro
mM	Milimolar
MM	Microdissecção Manual
mm	milímetro
mRNA	do inglês <i>messenger RNA</i>
NaCl	do inglês <i>sodium chloride</i>
NCBI	do inglês <i>National Center for Biotechnology Information</i>
ng	nanograma
nm	nanômetro
ORF	do inglês <i>Open Reading Frame</i>
PCR	do inglês <i>polimerase chain reaction</i>
pb	pares de base
PgR	do inglês <i>Progesterone Receptor</i>
pH	do inglês <i>hydrogen ionic potential</i>
Poli-A	do inglês <i>polyadenylate</i>
qRT-PCR	do inglês <i>Quantitative Reverse transcriptase – polymerase chain reaction</i>
RNA	do inglês <i>acid ribonucleic</i>
rpm	do inglês <i>rotation per minute</i>
RT-PCR	do inglês <i>Reverse transcriptase – polymerase chain reaction</i>
RNAa	do inglês <i>amplified RNA</i>
rpm	rotações por minuto
SAGE	do inglês <i>Serial Analysis of Gene Expression</i>
siRNA	do inglês <i>small interfering RNA</i>
snRNPs	do inglês <i>small nuclear ribonucleoproteins</i>

Taq	<i>Thermus aquaticus</i>
Tris	do inglês <i>hydroxymethyl aminomethane</i>
UICC	do francês <i>Union Internationale Contre le Cancer</i>
U	Unidade
VCE	Variante com o exon associado a tumor
VSE	Variante sem o exon associado a tumor
µg	microgramas
µg/µL	microgramas por microlitro
µL	microlitro
µm	micrometro
µM	micromolar
°C	graus <i>Celsius</i>
>	Maior
<	Menor
X-GAL	do inglês <i>5-bromo-4-chloro-3-indolyl-β-D-galactoside</i>

ÍNDICE

1	INTRODUÇÃO	1
1.1	Câncer de mama	1
1.2	A genética do câncer de mama	5
1.3	Marcadores moleculares e câncer	9
1.4	<i>Splicing</i> Alternativo	15
1.5	O estudo do <i>splicing</i> alternativo na busca por novos marcadores moleculares	21
2	OBJETIVOS	30
2.1	Objetivo Geral	30
2.2	Objetivos específicos	30
3	MATERIAIS E MÉTODOS	32
3.1	Comitê de Ética em Pesquisa	32
3.2	Construção da Plataforma de <i>microarray</i>	32
3.2.1	Escolha dos iniciadores para extensão dos fragmentos	32
3.2.2	DNA molde para extensão dos fragmentos	33
3.2.3	Extensão dos fragmentos	34
3.2.4	Desenho dos iniciadores para extensão de fragmentos dos genes <i>GAPDH</i> e <i>ACTB</i>	35
3.2.5	Purificação dos fragmentos	37
3.2.6	Clonagem, seqüenciamento e análise das seqüências	37
3.2.7	Construção das membranas	39
3.3	Amostras	41
3.3.1	Amostras utilizadas nos experimentos de <i>microarray</i> e validação através de RT-PCR quantitativo (qRT-PCR)	41
3.3.2	Amostras utilizadas para validação em um grupo independente através de qRT-PCR	43
3.4	Extração de RNA	44

3.4.1	Avaliação da qualidade do RNA extraído	44
3.5	Amplificação de RNA (RNAa)	45
3.5.1	Avaliação da qualidade do RNA amplificado (RNAa)	46
3.6	Marcação e Hibridização	47
3.7	Captura do dados	49
3.8	Correção do valor de sinal obtido	49
3.9	Remoção dos <i>outliers</i>	50
3.10	Normalização do dados	50
3.11	Controle de qualidade experimental das membranas	50
3.12	Análise estatística	51
3.13	Análise dos genes em bancos de SAGE	52
3.14	Validação através da análise do nível de expressão por RT-PCR em tempo real (qRT-PCR)	53
3.14.1	Desenho dos iniciadores	53
3.14.2	Síntese de cDNA	55
3.14.3	Estabelecimento das condições da reação de RT-PCR	56
3.14.4	Teste de especificidade do iniciador <i>forward</i> da variante sem o exon (VSE) desenhado na junção exon1-exon3	57
3.14.5	RT-PCR em tempo real (qRT-PCR)	58
3.14.6	Padronização das concentrações de cDNA e de iniciadores	59
3.14.7	Seleção dos genes normalizadores	61
3.14.8	Análise dos dados de expressão gênica	62
3.14.9	Análises Estatísticas	63
3.15	Análises clínicas e histopatológicas	64
4	RESULTADOS	66
4.1	Construção da Plataforma	66
4.1.1	Extensão dos fragmentos	66
4.1.2	Clonagem, seqüenciamento e análise das seqüências	67
4.2	Marcação e Hibridização	67
4.3	Controle de qualidade experimental das membranas	69
4.4	Análise dos dados de <i>microarray</i>	73

4.4.1	Resultados das comparações	74
4.4.2	Representatividade dos exons previamente selecionados pela análise computacional como associados a tumor de mama	75
4.4.3	Seleção dos exons superexpressos para confirmação através de RT-PCR quantitativo (qRT-PCR)	76
4.4.4	Análise de expressão dos genes correspondentes a cada exon em bancos de SAGE	79
4.5	Confirmação da superexpressão do exon através de qRT-PCR	81
4.5.1	Avaliação da VCE através de qRT-PCR	81
4.6	Confirmação da superexpressão da VCE e não do gene como um todo	95
4.6.1	Identificação da seqüência correspondente à junção exon1-exon3	98
4.6.2	Teste de especificidade dos iniciadores desenhados nas junções exon1-exon3	102
4.6.3	Avaliação por qRT-PCR da variantes sem o exon (VSE)	105
4.6.4	Avaliação da expressão por qRT-PCR do gene como um todo através de iniciadores desenhados em 2 exons constitutivos	106
4.7	Confirmação da superexpressão da VCE em relação ao seu respectivo gene em um grupo independente de amostras	108
4.8	Identificação dos domínios funcionais através das seqüências protéicas do gene e da VCE	110
4.9	Características clínicas e histopatológicas das amostras	113
4.10	Correlação entre as variáveis clínicas e histopatológicas e os dados de expressão obtidos	117
5	DISCUSSÃO	126
6	CONCLUSÕES	144
7	REFERÊNCIAS BIBLIOGRÁFICAS	145

ANEXOS

- Anexo 1** cTNM
- Anexo 2** pTNM
- Anexo 3** Parecer do Comitê de Ética em Pesquisa
- Anexo 4** Consentimento Informado
- Anexo 5** Gel de agarose 1%
- Anexo 6** Amostras
- Anexo 7** Linhagens Celulares
- Anexo 8** Extração de RNA por CsCl
- Anexo 9** Extração de RNA por TRIzol® Reagent
- Anexo 10** Amplificação de RNA (RNAa)
- Anexo 11** Gráficos de amplificação das amostras por qRT-PCR e curva de dissociação obtida para todas as VCE estudadas
- Anexo 12** Amostras Independentes

1 INTRODUÇÃO

1.1 CÂNCER DE MAMA

O câncer de mama está entre as neoplasias de maior incidência no mundo todo e é uma das mais importantes causas de mortalidade por câncer em mulheres, sendo responsável por um quinto dos óbitos em mulheres na faixa etária entre 40 e 50 anos (RADICE e REDAELLI 2003). No Brasil, de acordo com as estimativas de incidência de neoplasias para 2008, será o segundo câncer mais incidente, com 49.400 casos e um risco médio estimado de 51 casos a cada 100 mil mulheres. Na região Sudeste, o câncer de mama é o mais incidente entre as mulheres com um risco estimado de 68 novos casos por 100 mil. Depois dos tumores de pele não melanoma, o câncer de mama é também o mais freqüente nas mulheres das regiões Sul (67/100.000), Centro-Oeste (38/100.000) e Nordeste (28/100.000). Na região Norte é o segundo câncer mais incidente (16/100.000) (Ministério da Saúde 2008).

Histologicamente o câncer de mama pode se apresentar sob a forma de carcinoma *in situ* e invasivo, e 90% dos casos pertence aos carcinomas que se originam das células luminais do epitélio dos lóbulos e ductos da glândula mamária, sendo o carcinoma ductal mais frequente, representando cerca de 80% dos casos. Os carcinomas lobular *in situ* (cuja sigla do inglês é LCIS) e ductal *in situ* (cuja sigla do inglês é DCIS) são o resultado de uma

proliferação anormal de células neoplásicas confinada aos lóbulos ou ductos da glândula mamária, respectivamente. Já no carcinoma lobular invasivo (cuja sigla do inglês é ILC) e no carcinoma ductal invasivo (cuja sigla do inglês é IDC) as células rompem a membrana basal do epitélio lobular e ductal, respectivamente, invadindo tecidos adjacentes e penetrando no endotélio de vasos sangüíneos e/ou linfáticos.

A etiologia do câncer de mama é bastante complexa, sendo influenciada tanto por fatores endógenos (história familiar e níveis hormonais), quanto exógenos (dieta e uso de contraceptivos). Outros fatores, como menarca precoce, idade avançada (a cada 10 anos as taxas de incidência dobram), menopausa tardia (após os 50 anos), e a nuliparidade ou a primeira gestação após os 30 anos de idade podem ocasionar uma maior susceptibilidade a essa doença (Ministério da Saúde 2008). Os sintomas relacionados ao aparecimento da doença incluem a presença de nódulos palpáveis, que nem sempre provocam dor, alterações na pele que recobre a mama e o aparecimento de nódulos na axila. A detecção precoce da doença em conjunto com tratamentos como a intervenção cirúrgica, terapia hormonal, quimio e radioterapia podem levar à cura em muitos casos. Apesar de no Brasil o índice de detecção precoce ter aumentado nos últimos anos, ainda são diagnosticados casos de câncer de mama em estágios bastante avançados, o que dificulta muito o tratamento e diminui as chances de cura da doença. Assim, há diversas campanhas de âmbito nacional que estimulam o rastreamento e detecção precoce da doença, incluindo o auto-exame das mamas, o exame clínico palpatório e a

mamografia. Há também exames auxiliares, como a ultra-sonografia, exames citológicos (PAAF- punção aspirativa com agulha fina e citologia de descarga papilar) e histopatológicos (biópsia), que combinados aumentam ainda mais a acurácia no diagnóstico da doença.

O estadiamento do câncer de mama é baseado no sistema de classificação TNM, que é o método mais utilizado para a classificação de tumores malignos e a descrição de sua extensão anatômica. Esse sistema foi criado pela *Union Internationale contre le Cancer* (UICC) e pelo *The American Joint Committee on Cancer* (AJCC) e está baseado tanto em evidências anteriores ao tratamento, obtidas normalmente pelo exame físico e diagnóstico por imagem, caracterizando o cTNM (Anexo 1), quanto em evidências após o tratamento cirúrgico - pTNM (Anexo 2), sendo evidenciado pelo exame histopatológico do espécime cirúrgico e dos linfonodos, para detectar a presença ou ausência de metástases em linfonodos regionais. O sistema TNM de classificação considera a extensão do tumor primário (T), presença e extensão de metástases em linfonodos regionais (N) e presença de metástases à distância (M). Os estadios são numerados de 0 à IV, em ordem decrescente em relação à sobrevida, sendo estadio 0: carcinoma *in situ*, estadio I: carcinoma invasivo localizado, estadio II: carcinoma invasivo localmente limitado ou espalhado regionalmente, estadio III: carcinoma invasivo localmente extensivo ou espalhado regionalmente, e estadio IV: carcinoma invasivo extensivo com presença de metástases à distância (STYBLO e WOOD 1998). O tamanho do tumor e a presença de metástases linfonodais são os mais significantes fatores de risco para a recorrência do

câncer de mama e a piora da sobrevida global (SMART et al. 1997). No entanto, algumas pacientes sem acometimento de linfonodos apresentam recorrência da doença (VAN'T VEER et al. 2002; CLARKE et al. 2005).

O grau histológico tem como base a combinação de parâmetros como a razão de índice mitótico, grau nuclear e arranjo do tumor. De acordo com a intensidade de cada parâmetro (discreta, moderada ou elevada) são atribuídas pontuações (1 a 3). Assim, a soma dos 3 parâmetros fornece um determinado valor (score) que corresponde ao grau histológico, sendo grau 1: score 3-5, grau 2: score 6 ou 7 e grau 3: score 8 ou 9, de acordo com Scarff-Bloom-Richardson (SBR) (DOUSSAL et al. 1989; HOWEL et al. 1994) (ANEXO 2).

Desse modo, a classificação atual é baseada principalmente em análise clínica e histológica aliada à utilização de alguns marcadores moleculares disponíveis, agrupando muitas vezes tumores com comportamento clínico distinto em classes únicas (SOTIRIOU et al. 2003). Isso se deve ao comportamento altamente heterogêneo da tumorigênese da mama, desencadeado por alterações em genes envolvidos em vários processos celulares que devem ser altamente controlados em uma célula saudável, tais como proliferação e adesão celular, apoptose, reparo de DNA e outros (SOTIRIOU et al. 2006; LONNING et al. 2007). A perda de controle desses processos podem por sua vez desencadear uma diversidade de alterações em outros processos celulares, conferindo a alta heterogeneidade no câncer de mama que resulta em comportamentos clínicos bastante

particulares em relação à resposta a quimio e hormônioterapia, influenciando de forma distinta na evolução da doença em cada paciente.

1.2 A GENÉTICA DO CÂNCER DE MAMA

Diversas alterações genéticas e epigenéticas (que alteram o padrão de expressão gênica sem alterar a seqüência de DNA) têm sido detectadas em tumores de mama envolvendo proto-oncogenes e genes supressores de tumor (FOLGUEIRA e BRENTANI 2004; COSTA et al. 2004; CARRAWAY et al. 2008), sendo as alterações genéticas mais comuns a amplificação gênica e a perda de heterozigose (LOH). Além das alterações envolvendo grandes regiões cromossômicas (duplicações, translocações, inserções e deleções), mutações pontuais (substituições, deleções e inserções) têm sido detectadas em tumores esporádicos e hereditários da mama (CORVELLO 1999). A forma esporádica, quando a mutação ocorre em células somáticas, constitui mais de 90% de todos os casos, e cerca de 5-10% corresponde à forma hereditária, a qual está associada com mutações na linhagem germinativa.

No câncer de mama, a seqüência de eventos genéticos que conduzem à formação de um tumor ainda não está claramente compreendida. Entretanto, algumas alterações comumente encontradas nesses tumores já foram descritas. Nesse processo de malignização, ocorre o envolvimento de diversos proto-oncogenes (c-MYC, ERBB2, ERBB1 e

ciclina D1) e genes supressores de tumor (BRCA1, BRCA2, TP53, p16, E-caderina entre outros) (FOLGUEIRA e BRENTANI 2004).

Os genes *BRCA1* (*BReast CAncer suceptibility gene 1*) e *BRCA2* (*BReast CAncer suceptibility gene 2*) são os dois maiores genes que causam susceptibilidade genética ao câncer de mama e ovário (WOOSTER e WEBER 2003). São genes supressores de tumor, desempenhando função de proteção, uma vez que têm um papel importante na sinalização e reparo de DNA, regulação transcricional, regulação do ciclo celular, apoptose e duplicação do centrôssomo (PHILLIPS 2001; VENKITARAMAN 2002; YOSHIDA e MIKI 2004). O gene *BRCA1* está localizado no cromossomo 17q21 e sua proteína compreende 1.863 aminoácidos. *BRCA2* está no cromossomo 13q12 e é ainda maior, codificando uma proteína de 3.418 aminoácidos. Dado o tamanho desses genes, não é surpreendente que por volta de 1.500 mutações distintas e variantes de *splicing* têm sido documentadas em cada um deles. Mutações nesses genes explicam a maioria dos cânceres de mama hereditários (ROSEN et al. 2005).

O gene *TP53* é o supressor de tumor que apresenta a maior frequência de mutações pontuais somáticas no câncer de mama, e aproximadamente 50% dos tumores primários de mama apresentam alguma alteração nesse gene que codifica uma fosfoproteína nuclear, cuja principal função é a regulação do ciclo celular (GREENBLATT et al. 1994). Esse gene é conhecido como o “guardião do genoma”, pois a proteína codificada por ele impede a progressão do ciclo celular quando há algum dano na molécula de DNA, permitindo que ela seja reparada, ou pode alternativamente induzir a

célula à apoptose (morte celular programada) caso o DNA não seja reparado. Portanto, quando *Tp53* sofre mutações, as células com danos no DNA podem desencadear a transformação maligna, pois escapam do mecanismo de reparo e de apoptose, podendo dar início a um clone maligno (WEINERT 1998; NAKAMURA et al. 2002). A proteína mutante p53 está associada à negatividade para os receptores hormonais em câncer de mama e à maior proliferação celular, conferindo maior resistência das células neoplásicas a estímulos apoptóticos (MEGHA et al. 2002; ROLLAND et al. 2007).

A busca por alterações genéticas presentes nos tumores de mama e a associação destas alterações com os dados clínicos e histopatológicos têm permitido identificar alguns fatores genéticos associados à progressão da doença. Assim, além dos fatores envolvidos no prognóstico do câncer de mama como tamanho do tumor, presença de metástases em linfonodos regionais e metástases à distância, subtipo histológico e grau histológico, alguns marcadores moleculares já são bastante utilizados na classificação desta neoplasia, o que determina algumas vezes mudança na conduta clínica. Esses marcadores incluem a expressão de receptores de estrógeno (ER) e progesterona (PgR), e a amplificação e ativação de proto-oncogenes, como o *ERBB2*, os quais tem grande importância na predição da resposta ao tratamento.

Estrógeno desempenha um papel importante no crescimento e diferenciação da glândula mamária normal, assim como no desenvolvimento e progressão do carcinoma de mama. Ele age via seu receptor (ER),

estando envolvido na regulação de processos transcricionais relacionados ao crescimento e diferenciação celular da glândula mamária. Os efeitos biológicos do estrógeno são mediados através dos receptores de estrógeno α e β ($ER\alpha$ e $Er\beta$), os quais são membros de uma grande família de receptores nucleares (MATTHEWS e GUSTAFSSON 2003).

O receptor tirosina quinase *ERBB2*, também conhecido como *HER2* ou *HER-2/neu*, é um membro da família de genes que codificam receptores transmembrana para fatores de crescimento, incluindo o receptor de fator de crescimento epidermal *EGFR* (ou *ERBB1/HER1*), *ERBB3* (ou *HER3*) e *ERBB4* (ou *HER4*). O domínio intracelular de *HER2* tem atividade tirosina quinase que regula etapas importantes da fisiologia, crescimento e diferenciação de células (YARDEN e SLIWKOWSKI 2001; CHO et al. 2003). A amplificação do oncogene *ERBB2* (localizado no cromossomo 17q21.1) e de elementos genéticos relacionados à região amplificada no cromossomo 17 causam um aumento na sua expressão na superfície de células de tumor de mama de até 100 vezes o nível usual (BURSTEIN 2005). A superexpressão desse gene pode transformar células em cultura a um fenótipo maligno e tumorigênese acelerada. Foi estabelecido um sistema modelo para investigar o papel do gene *ERBB2* no processo tumorigênico do câncer de mama, através da transfecção da linhagem HB4a de mama normal com o cDNA do gene *ERBB2*, originando as linhagens HB4a C5.2 e HB4a C3.6 (HARRIS et al. 1995). A linhagem HB4a foi originada a partir de células epiteliais de lúmen de mama imortalizadas pelo antígeno T do vírus SV40 (STAMPS et al. 1994), as quais são cubóides, bem organizadas e

apresentam inibição de crescimento por contato (HARRIS et al. 1995). As células da linhagem HB4a C5.2 são finas e alongadas com perda de inibição de crescimento por contato, e cada célula expressa aproximadamente 106 receptores ERBB2, que é uma quantidade similar à de linhagens tumorais de mama como SKBR3, ou ainda superior a outras linhagens tumorais de mama, como BT474. Este modelo tem sido utilizado por muitos trabalhos que procuram investigar a influência deste gene no processo de tumorigênese bem como caracterizar as proteínas envolvidas nas vias de ativação de *ERBB2* (TIMMS et al. 2002; MACKAY et al. 2003; SILVA 2006).

1.3 MARCADORES MOLECULARES E CÂNCER

Inúmeros avanços têm sido realizados nos últimos anos na melhor caracterização molecular dos diferentes tipos de câncer, que aliados a um aprimoramento contínuo dos métodos de estudo têm fornecido uma grande variedade de candidatos a marcadores tumorais de importância clínica. Esses marcadores podem ser proteínas ou frações delas, detectáveis no tecido por imunohistoquímica ou no soro por Elisa ou radioimunoensaio, além de marcadores genéticos variados, detectáveis pela avaliação de mutações gênicas ou alterações cromossômicas, e através do nível de expressão de RNA mensageiro (COSTA e DEL GIGLIO 2003).

Na maioria dos tipos de câncer, o diagnóstico é realizado através do exame microscópico dos tecidos, auxiliado pela avaliação de marcadores moleculares, resultando em uma classificação patológica usada para

diferenciar classes de tumores que podem apresentar características distintas tanto no prognóstico quanto na resposta à terapia.

Alguns exemplos de marcadores moleculares introduzidos rotineiramente na prática clínica são o antígeno prostático superficial (PSA), utilizado no rastreamento, diagnóstico e pesquisa de doença residual após tratamento do câncer de próstata (FENELEY e PARTIN 2000); a enzima desidrogenase láctica no soro de pacientes com linfoma não-*Hodgkin*, que exerce uma correlação entre o marcador e a resposta do câncer ao tratamento (SHIPP et al. 1993); o antígeno cárcino-embriônico (CEA) utilizado no diagnóstico, prognóstico e predição de tratamento em tumores colorretais (POSNER e MAYER 1994; COMPTON et al. 2000); a superexpressão da proteína codificada pelo oncogene *ERBB2* que é um fator prognóstico adverso e está associado a tumores de mama de alto grau, pouco diferenciados, com altas taxas de proliferação e envolvimento linfonodal, conferindo uma relativa resistência a certos tipos de quimioterapia (BURSTEIN 2005).

Marcadores prognósticos são utilizados para fornecer informações sobre o provável comportamento de um tumor no futuro, fornecendo informações sobre a evolução do tumor, independente de uma terapia adjuvante sistêmica, enquanto fatores preditivos são utilizados, por exemplo, para selecionar prospectivamente a resposta ou resistência a um tratamento específico (DUFFY 2005).

O receptor ER α é expresso em aproximadamente 70-80% de todos os carcinomas de mama (PARL 2003). A presença deste receptor não apenas

prediz resposta à terapia endócrina, como se correlaciona com um melhor prognóstico, sendo que pacientes com o fenótipo ER α negativo apresentam menor sobrevida livre de doença e menor sobrevida global quando comparados a pacientes com fenótipo positivo para ER α (MOKBEL 2003).

Em pacientes ER positivo podem ser utilizados tratamentos com hormonioterápicos bem conhecidos, como o tamoxifeno, que tem alta afinidade em ligar-se a ER, bloqueando a ação do hormônio. Os benefícios do tamoxifeno e sua relativa baixa toxicidade têm garantido seu uso num período de 5 anos como profilaxia em tumores ER positivos (FISHER et al. 2001; BUZDAR et al. 2003).

Quando medido acuradamente, o *status* do receptor de progesterona (PgR) é um fator preditivo independente para o benefício da terapia endócrina adjuvante e poderia ser utilizado para estimar redução de risco relativo esperado do tratamento endócrino com pacientes individuais. O *status* PgR melhora significativamente a predição de resultado à terapia endócrina sobre o *status* ER isoladamente (BARDOU et al. 2003). Aproximadamente 70% dos tumores ER/PgR positivos apresentam resposta ao tamoxifeno, enquanto apenas 34% dos tumores ER positivos/PgR negativos respondem. Além disso, são observados 45% de respostas em pacientes ER negativos/PgR positivos (CLARKE et al. 2001).

ERBB2 é considerado tanto um indicador de mau prognóstico para câncer de mama, quanto um marcador preditivo de resposta a agentes terapêuticos utilizados no tratamento dessa doença (PETIT et al. 2001; DUFFY 2005). Tem sido demonstrada a superexpressão deste gene em 20

a 30% de todos os tumores de mama (ALBANELL e BASELGA 2001; SLAMON et al. 2001) e alguns estudos indicam uma correlação significativa entre a sua superexpressão e redução da sobrevida de pacientes com câncer de mama (CAMP et al. 2003). Alguns estudos mostram que a presença concomitante de alterações nas proteínas codificadas por *ERBB2* e *Tp53* se correlaciona com um prognóstico ainda pior e maiores taxas de recorrência da doença (TAKIKAWA et al. 1994; BARBATI et al. 1997; RUDOLPH et al. 2001). Porém a superexpressão de ERBB2 é também um marcador preditivo de resposta a trastuzumab (Herceptina[®]), que é um anticorpo monoclonal humanizado (mAb) contra ERBB2, dirigido contra o domínio extracelular da proteína. Esse mAb é utilizado para o tratamento do câncer de mama e tem uma excelente atividade anti-tumoral, inibindo o crescimento das células que têm expressão aumentada de ERBB2, particularmente quando usado em combinação com doxorubicina e paclitaxol (KUMAR e MADISON 2001; SLAMON e PEGRAM 2001).

Desse modo, marcadores moleculares, tanto preditivos como de prognóstico, têm sido extensivamente investigados em câncer de mama com o objetivo de fornecer informações cada vez mais acuradas para tratamentos mais individualizados, conferindo melhores resultados aos pacientes acometidos por essa doença. No entanto, apesar do emprego desses marcadores, muitas pacientes com mesmo diagnóstico e prognóstico clínico podem apresentar resultados muito diferentes tanto em nível de resposta a tratamento, como progressão, recidiva e surgimento de metástase (SORLIE

et al. 2001; SOTIRIOU et al. 2003), por isso é muito importante a busca por novos marcadores moleculares tumorais.

Poderosas tecnologias, desenvolvidas através de avanços da biologia molecular e da bioinformática têm proporcionado o amplo conhecimento dos processos celulares através da análise da expressão gênica diferencial. A análise do perfil de expressão gênica tem sido uma ferramenta poderosa na classificação prognóstica do câncer de mama. Além disso, tem se mostrado útil na subclassificação de tumores que apresentam a mesma classificação morfológica e características distintas quanto a resposta à quimioterapia e à progressão tumoral, permitindo estabelecer correlações com finalidades preditivas do comportamento tumoral (FOLGUEIRA et al. 2005).

A tecnologia de cDNA *microarray* ou *biochip* de cDNA está sendo empregada com muita eficiência na área de câncer, permitindo que pesquisadores analisem a expressão de milhares de genes simultaneamente e relacionem seus dados com parâmetros clínicos. O método consiste na hibridização de genes imobilizados em uma lâmina de vidro ou membrana de *nylon*, por amostras fluorescentes ou radioativas de cDNA, respectivamente. As amostras são preparadas a partir de RNA de células, tecidos ou outras fontes biológicas. Moléculas da amostra fluorescente ou radioativa reagem com suas seqüências complementares ordenadas na lâmina de vidro ou membrana, que emitem fluorescência ou radiação em cada *spot* (várias moléculas de cDNA de mesma seqüência arranjadas na lâmina). A intensidade da fluorescência ou radiação emitida é proporcional

ao nível de expressão do gene que corresponde àquela determinada seqüência (SCHENA 2003).

Vários estudos têm sido desenvolvidos, comprovando a importância da utilização desta técnica para o melhor entendimento do processo de tumorigênese (NISHIDATE et al. 2004) e progressão do câncer de mama (MA et al. 2003), assim como para identificar assinatura molecular de predição de resposta ou resistência à quimioterapia (WOLFF e DAVIDSON 2000; CHANG et al. 2003; SHIMIZU et al. 2004; FOLGUEIRA et al. 2005). Além disso essa metodologia pode auxiliar na busca por grupos de genes com valor preditivo de potencial metastático e associação do perfil de expressão com possibilidade de recidiva para pacientes linfonodo negativo (VAN'T VEER et al. 2002; SOTIRIOU et al. 2003). Neste sentido foi desenvolvido o teste prognóstico multigene MammaPrint[®] (VAN DE VIJVER ET AL. 2002), ensaio composto por 70 genes, com capacidade de prever recorrência da doença para pacientes linfonodo negativo.

Esses estudos mostram que a utilização dos dados de expressão gênica gerados pela técnica de *microarray* pode nos fornecer melhor embasamento biológico na determinação de prognósticos mais precisos para os tumores de mama e na predição de resposta a tratamento, quando o perfil de expressão gênica está vinculado a observações clínicas e histopatológicas.

1.4 *SPLICING* ALTERNATIVO

A arquitetura da maioria dos genes em eucariotos superiores consiste em exons codificantes intercalados por íntrons não-codificantes. A remoção dos íntrons e junção dos exons é realizada através de eventos de *splicing* do RNA imaturo (pré-mRNA), que é um passo essencial no processamento de mRNA funcional. Em cada evento de *splicing*, um íntron é removido através de duas reações de trans-esterificação, unindo dois exons. Essa reação é catalizada pelo spliceossomo, um complexo macromolecular que contém cinco moléculas de RNA ricas em uracila U1, U2, U4, U5 e U6, conhecidas como snRNAs (*small nuclear RNA*), as quais se associam a 6 a 10 proteínas formando partículas ribonucleoprotéicas, snRNPs (*small nuclear ribonucleoprotein particles*), que por sua vez estão associadas a um grande número de proteínas adicionais, formando um grande complexo. Há estudos que relatam a existência de até 300 proteínas associadas ao spliceossomo (JURICA e MOORE 2003; NILSEN 2003). Durante a montagem do spliceossomo diversos complexos são formados em uma ordem definida para o reconhecimento dos sítios de *splicing*, excisão dos íntrons e junção dos exons. Sítios de *splicing* são sinais seqüenciais específicos de reconhecimento nas junções exon/íntron, sendo geralmente os dinucleotídeos conservados GU na porção 5' do íntron e os dinucleotídeos AG bem na porção 3' do íntron. Outras seqüências conservadas são encontradas nessa região, como o ponto de quebra ou ramificação e o trato de polipirimidinas (ALBERTS et al. 2002; NILSEN

2003). Entretanto, essas seqüências não são suficientes para definir a ocorrência do *splicing*, e seqüências adicionais ao redor desses sítios e até distante deles são requeridas, os elementos reguladores *cis*: *splicing enhancers* e *splicing silencers*. *Splicing enhancers* estão localizados tanto nos exons, quanto nos íntrons (ESE - *Exonic Splicing Enhancers*; ISE- *Intronic Splicing Enhancers*) e auxiliam a maquinaria do spliceossomo no reconhecimento dos sítios de *splicing*, enquanto *splicing silencers* podem atuar inibindo o reconhecimento desses sítios (ESS - *Exonic Splicing Silencer*; ISS - *Intronic Splicing Silencer*) (BLACK 2003). Há ainda proteínas conhecidas como fatores de *splicing* (reguladores *trans*) que se ligam a essas seqüências também regulando a definição dos sítios de *splicing*, e, de acordo com o grupo funcional ao qual pertencem, atuam ativando ou inibindo o processamento de um determinado exon, ocasionando na sua inclusão ou exclusão do transcrito. Os dois grupos mais conhecidos são a família de proteínas SR (*serine/arginine rich*) que são fatores de *splicing* essenciais, que recrutam direta ou indiretamente componentes do spliceossomo e se ligam geralmente a seqüências ESE, aumentando a taxa de *splicing*; e as proteínas hnRNP, em particular hnRNP A1 (*heterogeneous nuclear ribonucleoprotein A1*), que se ligam em sua maioria às seqüências ESS, inibindo o processamento (MATLIN et al. 2005). Desse modo, a composição, localização e concentração desses fatores reguladores *trans* e sua influência nos elementos reguladores *cis* podem alterar o padrão de *splicing* do pré-mRNA, e em muitos casos mais de um mRNA maduro pode ser produzido pela escolha alternativa de sítios 5' e 3' pela maquinaria de *splicing*,

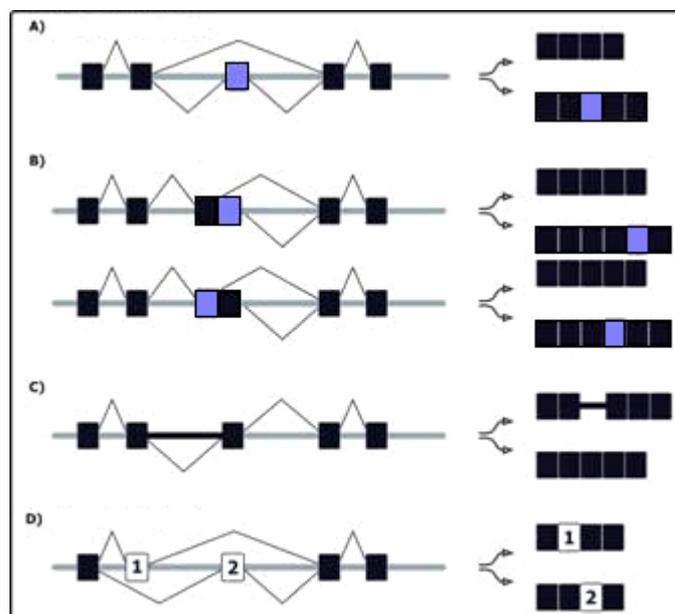
fenômeno conhecido como *splicing* alternativo (AS). Mudanças na expressão relativa de vários fatores de *splicing* podem modular o tipo e a concentração relativa das variantes correspondentes a vários genes (CACERES e KORNBLIHTT 2002), sendo observadas em diferentes tipos celulares e em algumas doenças, como, por exemplo, o câncer (HANAMURA et al. 1998; STICKELER et al. 1999).

Splicing alternativo (AS) é o processo pelo qual um único transcrito primário gera diferentes RNAs maduros, podendo levar à produção de isoformas de proteínas com funções diversas e até mesmo antagônicas, sendo este processo importante no desenvolvimento normal das células como um meio de criar diversidade protéica em organismos complexos (MANIATIS e TASIC 2002; BLACK 2003).

O fato da maioria dos genes humanos serem constituídos por dois ou mais exons permite que, através de diferentes combinações, múltiplos RNA mensageiros maduros distintos sejam formados a partir de uma única molécula de mRNA imaturo pelo processo de *splicing* alternativo. Estimativas sugerem que mais de 70% dos genes humanos sofrem *splicing* alternativo (JOHNSON et al. 2003), havendo casos em que um único gene pode gerar inúmeros transcritos diferentes (BLACK 2000; CARNINCI et al. 2005) aumentando a diversidade transcricional dos organismos. Aparentemente a maior parte dos eventos de *splicing* alternativo (74%) ocorre na região codificante do gene (*coding sequence*), também gerando uma grande contribuição para a diversidade proteômica (MODREK e LEE 2002). Assim, este fenômeno explica em parte a complexidade de

organismos superiores frente ao pequeno número de genes (GRAVELEY 2001; WIEBEN 2003; BRINKMAN 2004; KALNINA et al. 2005).

Há diversos tipos de *splicing* alternativo (Figura 1): exons cassetes (ou regulados) podem ser incluídos ou excluídos (exon *skipping*) de um determinado transcrito; sítios doadores e/ou aceptores crípticos podem ser usados ao invés dos sítios de *splicing* originais, o que gera exons mais curtos ou longos; seqüências intrônicas podem ser usadas como exons em alguns transcritos (*intron retention*) e exons podem sofrer *splicing* mutuamente exclusivo (os dois exons nunca serão encontrados simultaneamente no mesmo transcrito). A exclusão de exons parece ser o tipo de *splicing* alternativo mais freqüente (BRETT et al. 2002) e a retenção de íntrons o mais raro (KIM et al. 2007).



Legenda: (A) uso alternativo de exons, (B) uso de sítios crípticos 5' ou 3', (C) retenção de íntrons e (D) uso mutuamente exclusivo de exons. Os retângulos azuis representam os exons que são alternativamente processados. As linhas que ligam retângulos (exons) mostram quais exons serão combinados depois dos diferentes tipos de *splicing* alternativo.

Figura 1 - Tipos de *splicing* alternativo.

O seqüenciamento do genoma humano e a introdução de análises em larga escala de etiquetas de seqüências expressas (do inglês *expressed sequence tags*- ESTs) tem tornado possível avaliar variantes de *splicing* em nível genômico. Os primeiros estudos sobre o genoma humano mostraram haver um número relativamente pequeno de genes (30.000-40.000) (LANDER et al. 2001; VENTER et al. 2001) em relação à complexidade observada através de 100.000-150.000 genes preditos, no nível de mRNA. Diante destas observações, foi sugerida a grande importância do AS na geração da diversidade transcricional, sendo o maior responsável pela complexidade funcional do genoma humano (BLACK 2000; GRAVELEY 2001; MODREK e LEE 2002). Esse mecanismo pode causar impactos significativos na atividade funcional das próprias proteínas além de alterar suas interações com outras proteínas (MINNEMAN 2001; THAI e KEARNEY 2004; SCHEPER et al. 2004).

A bioinformática tem sido fundamental na caracterização de novas variantes de *splicing* através do desenvolvimento de estratégias e ferramentas computacionais (FERREIRA et al. 2007). Muitos esforços têm sido realizados neste sentido, onde, por exemplo, análises computacionais através do alinhamento de ESTs contra a seqüência genômica humana, em conjunto com o uso de *softwares* de predição gênica, têm fornecido ferramentas para a identificação de novos eventos de *splicing* alternativo (BLACK 2000; GALANTE et al. 2004; GUPTA et al. 2004; HSU et al. 2005).

Uma abordagem experimental interessante para a identificação de AS é a construção de bibliotecas de cDNA enriquecidas para eventos de

splicing, através da captura de heteroduplexes formados entre duas diferentes variantes de *splicing* do mesmo gene (WATAHIKI et al. 2004). Essa metodologia é muito promissora, uma vez que qualquer tipo de evento de *splicing* pode ser identificado, sem conhecimento prévio dos transcritos.

A tecnologia de *microarray* é uma poderosa ferramenta para o estudo de expressão gênica em larga escala (SCHENA et al. 2003), utilizando uma grande quantidade de amostras. Considerando-se que essa tecnologia pode ser muito útil na caracterização do perfil de expressão de diferentes transcritos de um gene, diversos estudos em larga escala têm sido desenvolvidos através da utilização da tecnologia de *microarray* para analisar eventos de *splicing* alternativo. A utilização de plataformas que cobrem grandes regiões genômicas (*tiling microarrays*) tem se mostrado eficiente na identificação de novas isoformas de *splicing* (HU et al. 2001; FAN et al. 2006). Além disso, a construção de plataformas de sondas de oligonucleotídeos que representam junções exon-exon conhecidas permite a avaliação do perfil de expressão de diversas variantes além de obter uma medida quantitativa de sua abundância (JOHNSON et al. 2003; CUPERLOVIC-CULF et al. 2006; GARDINA et al. 2006).

Uma variante de *splicing* pode ser definida como sendo funcional se for requerida durante o ciclo de vida normal do organismo e se for ativada de maneira regulada (SOREK et al. 2004). Entretanto, não está claro quantas das variantes de *splicing* preditas a partir de ESTs são funcionais e quantas representam *splicing* aberrantes ou artefatos da metodologia utilizada para sua detecção (GRAVELEY 2001; MANIATIS e TASIC 2002; MODREK e LEE

2002). Tanto mutações somáticas em sítios de *splicing* ou nas seqüências *cis enhancers* e *silencers*, além de SNPs (*single nucleotide polymorfism*) nessas regiões podem influenciar na seleção dos sítios de *splicing* (FAUSTINO e COOPER 2003) ou ainda resultar em *splicing* aberrante, gerando proteínas truncadas (SOREK et al. 2004). Além disso, erros na maquinaria do spliceossomo também têm sido propostos como um mecanismo que pode resultar em transcritos não funcionais que produzem geralmente proteínas truncadas (GRAVELEY 2001).

Um dos maiores objetivos de estudo atualmente é identificar eventos de *splicing* alternativo fisio e patologicamente relevantes, para determinar onde e quando eles ocorrem, quais são seus papéis específicos e como estão regulados (PAN et al. 2005). Essas questões fundamentais requerem avanços de tecnologias em larga escala para o monitoramento do AS.

1.5 O ESTUDO DO *SPLICING* ALTERNATIVO NA BUSCA POR NOVOS MARCADORES MOLECULARES

Tem sido mostrado que o *splicing* alternativo está envolvido em diversas doenças humanas (FAUSTINO e COOPER 2003; GARCIA-BLANCO et al. 2004; PAGANI e BARALLE 2004) e o seu controle pode estar desregulado nessas doenças, e estudos mostram ainda que está notavelmente alterado em câncer (STIMPFL et al. 2002; BRINKMAN 2004; VENABLES 2004; KIM et al. 2008), sendo evidenciada uma correlação direta entre a expressão diferencial de determinadas variantes de *splicing* e o

processo de tumorigênese em diferentes tipos de tumor (GE et al. 1999; BAUDRY et al. 2000; CRAGG et al. 2002; WANG Z et al. 2003).

Têm sido documentadas numerosas isoformas de *splicing* alternativo associadas a câncer que podem afetar a função de proteínas, incluindo fatores de transcrição, componentes de sinalização celular, reguladores da apoptose e componentes da matriz-extracelular (VENABLES 2004). Em muitos exemplos de AS associado a tumores, a função da isoforma protéica parece ser consistente com um possível papel no câncer, embora muitas vezes esteja relacionada a processos não exclusivos de células tumorais, onde as funções exercidas pelas isoformas alternativas, em conjunto com um possível desbalanço entre elas, poderiam ter um efeito na transformação maligna através de, por exemplo, indução positiva do ciclo celular pela perda da função de um gene supressor de tumor ou ganho de função de um oncogene. Os exemplos mais conhecidos são o receptor multifuncional *CD44* (SNEATH e MANGHAM 1998; NAOR et al. 2002) e o fator de transcrição *WT1- Wilms' tumor gene* (BAUDRY et al. 2000; WAGNER et al. 2003). O gene *CD44* é um receptor glicoprotéico de membrana constituído por 20 exons, dos quais 9 são variáveis (do 6 ao 15 ocorre *splicing* alternativo), levando à formação de variantes preferencialmente expressas em tumores. Uma variante particular de *CD44* (v6), é encontrada em câncer de mama (HERRERA-GAYOL e JOTHY 1999), bexiga (COOPER 1995) e cólon (HERRLICH et al. 2000), estando relacionada com pior prognóstico, e a expressão elevada de variantes de *CD44* contendo o exon 6 (*CD44v6*) está positivamente relacionada com estágios avançados em carcinoma

gástrico, indicando progressão tumoral (XIN et al. 2001). O gene *WT1* possui quatro domínios de dedos de zinco, o que lhe confere a capacidade de se ligar ao DNA, agindo como fator de transcrição. *Splicing* alternativo de *WT1* produz várias isoformas de proteínas com distintos domínios de ligação ao DNA, proporcionando atividades transcricionais altamente reguladas durante, por exemplo, a diferenciação do rim. Foi mostrado por BAUDRY et al. (2000) que mutações em *WT1* são observadas em 10% dos casos de tumor de Wilms esporádicos, e 63% dos casos apresentam alterações de *splicing*, sendo sugerido que o desbalanço entre as variantes (uma contendo o exon 5 e outra não) afeta vários aspectos da fisiologia celular deste tumor.

Um exemplo de AS de um gene supressor de tumor é a proteína neurofibromatose tipo1 (NF1). NF1 é um supressor de tumor neuronal, que funciona em parte inativando a sinalização do oncogene RAS. Uma variante alternativa de NF1 com a inserção de 63 nucleotídeos em seu quadro aberto de leitura (*ORF*), perde parte dessa função supressora de tumor e está superexpressa em meduloblastomas e tumores ectodérmicos primitivos (SCHEURLIN e SENF 1995). Outro exemplo é o gene *spleen tyrosine kinase* (Syk), um supressor de metástase tumoral que está altamente expresso em células epiteliais mamárias normais e sua perda de expressão está associada com o aumento da capacidade invasiva em tumores de mama. Uma variante de Syk, conhecida como Syk(S) está superexpressa nesse tipo de tumor e não exerce o papel supressor de metástase. A única diferença entre as duas formas está na perda de 23 resíduos de aminoácidos dentro do interdomínio B (IDB) da variante Syk(S). A ausência desses

resíduos leva à alteração de localização da proteína e perda da função (WANG L et al. 2003).

Outros exemplos de genes que apresentam variantes associadas a tumores são *EGFR* (WIKSTRAND et al. 1995), *BIN1* (GE et al. 1999), *MUC-1* (BARUCH et al. 1999) e *CD79b* (CRAGG et al. 2002). Também os genes envolvidos no processo de apoptose *BCL-X*, *CASPASE-9*, *CED-4*, *CASPASE-2/ICH-1* e *HTID-1* codificam diferentes variantes pró e anti-apoptose (BRINKMAN 2004).

Recentemente têm sido descobertas novas variantes de *splicing* que ocorrem no gene *PTEN*, um supressor de tumor que leva à parada do ciclo celular na fase G1, o qual aparece inativado na síndrome de Cowden, caracterizada por alto risco de câncer de mama e tireóide. Tem sido sugerido que a expressão diferencial de *PTEN* e suas variantes de *splicing* poderiam desempenhar um papel na patogênese de câncer de mama esporádico podendo servir como alvo terapêutico (AGRAWAL e ENG 2006). Outra variante de *splicing* associada ao câncer de mama é a variante do gene *BRCA1* - *BRCA1-IRIS*, cuja proteína não interage com *BARD1*, não formando o complexo *BRCA1: BRAD1*, o qual medeia a atividade ubiquitina ligase, que pode ser importante para a supressão de tumores (HASHIZUME et al. 2001). *BRCA1-IRIS* está associada a proteínas envolvidas na replicação do DNA, e tem sido postulado que influencia positivamente a replicação de DNA (EL SHAMY e LIVINGSTON 2004).

Esses exemplos mostram que o *splicing* alternativo está notavelmente alterado em câncer, gerando muitas vezes variantes aberrantes que podem

desempenhar um papel importante na tumorigênese e progressão tumoral (CABALLERO et al. 2001; VENABLES 2004; PETTIGREW e BROWN 2008), e poderiam ser utilizadas como marcadores tumorais, ou ainda como estratégia terapêutica, através do bloqueio da função dessas variantes. Entretanto não se sabe ao certo se o padrão de *splicing* em câncer é um fator causal para células normais tornarem-se tumorais ou se o processo tumorigênico por si só altera a taxa de replicação celular e, ainda, o modo de síntese de mRNA, de forma que a maquinaria de *splicing* passe a gerar isoformas espúrias (Xu & Lee, 2003).

A exploração da variabilidade transcricional gerada por *splicing* alternativo pode ser promissora, especialmente na identificação de novos marcadores para alvo terapêutico (AGRAWAL e ENG 2006; LU et al. 2007; SAMPATH e PELUS 2007), uma vez que a droga seria mais específica, atingindo a variante e não o gene como um todo. Por exemplo, Ding et al. (2002) mostraram que o aumento da expressão de uma isoforma alternativa do receptor de secretina impede a ligação do receptor constitutivo ao seu ligante (Ding et al., 2002). Assim, foi demonstrado que a redução da ativação do receptor original devido à competição com o receptor variante permitiu o crescimento tumoral. Neste caso, o bloqueio da isoforma alternativa do receptor, tanto através do uso de anticorpos específicos ou eliminando seu respectivo mRNA por siRNA (RNA de interferência), poderia apresentar um valor terapêutico.

Neste sentido, recentes estudos em larga escala têm sido realizados para avaliação da associação entre *splicing* alternativo e câncer (XIE et al. 2002;

XU et al. 2002; WANG Z et al. 2003; HUI et al. 2004; KIRSCHBAUM-SLAGER et al. 2005), sendo que alguns estudos têm aplicado a tecnologia de *microarray* com esse objetivo (RELOGIO et al. 2005; LI et al. 2006; MILANI et al. 2006; THORSEN et al. 2008).

Nesse contexto, um estudo computacional foi realizado no Laboratório de Bioinformática do Instituto Ludwig, sob coordenação do Dr. Sandro J. de Souza, com objetivo de identificar variantes de *splicing* associadas a tumores, cujos exons apresentavam expressão predominante em bibliotecas de tecido tumoral em comparação com seu respectivo normal. Foi utilizado um banco de dados relacional (MySQL) contendo seqüências de cDNA alinhadas com a seqüência genômica humana, com informações do mapeamento no genoma, expressão tecidual e anotação funcional dos genes. O padrão de expressão de isoformas de 318.272 *clusters* (grupos de seqüências que compartilham regiões de similaridade, supostos genes únicos), foi analisado, dos quais 21.306 possuíam ao menos 1 seqüência completa de mRNA. Desses 21.306 *clusters*, foram selecionados aqueles que continham pelo menos uma variante com um evento de inserção de exon (variantes que contêm o exon alternativo) expressa em bibliotecas de tecido tumoral e não em bibliotecas do correspondente normal. Após análise estatística foram identificados 1.295 genes, contendo 2.878 exons com expressão exclusiva em tecidos tumorais em relação ao seu respectivo normal, sendo denominados exons 'associados a tumor'.

Como em nenhum trabalho da literatura foi demonstrado uma variante de *splicing* que possua expressão 'tumor-específica', ou seja, que apresente

expressão apenas em tecido tumoral, e não seja expressa em nenhum outro tecido normal, decidiu-se utilizar o termo 'associado a tumor' para variantes que mostraram uma expressão mais abundante em tecido tumoral em relação a tecido normal, e cujo 'protótipo' (onde não foi observada a presença do exon inserido) não apareceu superexpresso em tecido tumoral. Os exons foram alinhados contra o banco de dados Unigene para excluir aqueles que alinham com mais de um *cluster* de *ESTs*, sendo identificados, então, 932 exons 'associados a tumor' (Tabela 1). Uma validação experimental desses candidatos foi realizada através de experimentos de RT-PCR, onde alguns deles foram randomicamente selecionados, sendo que 4 de 10 (40%) e 5 de 6 (83,3%) puderam ser confirmados como superexpressos em linhagem e tecido tumoral, respectivamente (KIRSCHBAUM-SLAGER et al. 2005).

Tabela 1 - Número de exons associados a tumores para cada tecido analisado através de análise computacional.

Tecido	Exons associados a tumores
Glândula adrenal	63
Bexiga	118
Ossos	226
Cérebro	54
Mama	80
Cartilagem	127
Cólon	156
<i>Denis-drash</i>	51
Ouvido	9
Embrião	30
Epidídimo	32
Olho	16
Cabeça e pescoço	339
Rim	135
Fígado	30
Pulmão	62
Linha	130
Boca	206
Músculo	145
Sistema Nervoso	30
Ovário	351
Pancreas	156
Próstata	43
Pele	219
Estômago	35
Testículo	1
Útero	54

Assim, foram delineados os objetivos desse trabalho, aliando o potencial da metodologia de *microarray* em analisar o nível de expressão de um grande número de seqüências a uma análise computacional (KIRSCHBAUM-SLAGER et al. 2005), com o objetivo de identificar variantes de *splicing* superexpressas em tumores de mama. Para isso, uma plataforma de *microarray* com seqüências correspondentes a exons associados a tumores foi construída e hibridizada contra amostras tumorais e normais de mama. Os exons mais expressos em câncer de mama identificados através de hibridizações contra amostras de tecido tumoral e normal de mama foram selecionados. Neste trabalho, as variantes que contêm os exons analisados serão chamadas VCE (variante com o exon). Os exons mais expressos em tumores selecionados pela técnica de *microarray* foram validados por RT-PCR quantitativo (qRT-PCR) através de iniciadores desenhados no próprio exon e no exon flanqueador a 3'. Essa medida de expressão se refere à VCE. Para a confirmação da superexpressão dessa variante e não do gene como um todo, foi realizada avaliação da expressão constitutiva do gene (EC) através do desenho de iniciadores em 2 exons constitutivos, pertencentes a todas as variantes. O balanço entre a expressão da VCE e a EC foi avaliado em amostras tumorais e normais de mama. Aqueles candidatos que foram confirmados como variantes de *splicing* superexpressas em câncer de mama no grupo de amostras iniciais e posteriormente em um grupo independente de amostras foram correlacionados com características clínicas e histopatológicas.

2 OBJETIVO

2.1 OBJETIVO GERAL

Identificar variantes de *splicing* superexpressas em câncer de mama e correlacionar com características clínicas e histopatológicas.

2.2 OBJETIVOS ESPECÍFICOS

1. Construir plataforma de *microarray* com seqüências de DNA correspondentes a exons superexpressos em tumores selecionados por análise computacional;
2. Hibridizar contra amostras de tecido normal e tumoral de mama;
3. Identificar exons superexpressos em amostras de tecidos tumorais de mama;
4. Validar por RT-PCR quantitativo (qRT-PCR) os exons identificados como superexpressos em câncer de mama no mesmo grupo inicial de amostras através do desenho de iniciadores no exon superexpresso e no flaqueador a 3' (essa expressão se refere à VCE) ;
5. Verificar se as variantes de *splicing* que contêm o exon (VCE) se tratam, de fato, de variantes superexpressas em câncer de mama, através da análise da expressão do gene como um todo (através do desenho de iniciadores em exons constitutivos);

6. Validação dos achados em um grupo independente de amostras tumorais;
7. Correlacionar com características clínicas e histopatológicas.

3 MATERIAIS E MÉTODOS

3.1 COMITÊ DE ÉTICA EM PESQUISA

O desenvolvimento deste projeto, bem como a utilização das amostras, foram autorizados pelo Comitê de Ética em Pesquisa do Hospital A.C. Camargo, sob o processo de número 586/04 (Anexo 3). Para todas as amostras há um termo de consentimento livre e esclarecido pós-informado assinado (Anexo 4).

3.2 CONSTRUÇÃO DA PLATAFORMA DE *MICROARRAY*

3.2.1 Escolha dos Iniciadores para Extensão dos Fragmentos (Exons)

Os iniciadores *forward* (F) e *reverse* (R) foram desenhados utilizando-se o programa *PRIMER 3* (<http://www.basic.neu.edu/biotools/Primer3.html>). De acordo com os critérios estabelecidos, os fragmentos estendidos por esses iniciadores deveriam ter no mínimo 100 pb e apresentar um conteúdo GC de 50 a 60%. O par (F+R) deveria ser desenhado no mesmo exon associado a tumor, começar e terminar com G ou C, ter 16 a 21 pares de bases, temperatura de pareamento variando entre 58 e 68°C, sendo que a diferença em graus de temperatura de pareamento permitida entre os pares de iniciadores deveria ser de no máximo 4°C. Em seguida eles passaram por uma análise manual, com o auxílio do programa *OLIGOTECH* versão 1.00

(Copyright[©] 1995), para avaliação das estruturas secundárias (*stem loop* e *homodimer*), permitindo temperaturas de até 30°C. Foram selecionados 270 pares de iniciadores (F+R), sendo que 75 deles correspondem a exons associados a tumor de mama, 54 a tumor de cérebro e 141 a outros tecidos. Todos os 540 iniciadores foram diluídos a uma concentração estoque de 100uM, de onde foram feitas diluições trabalho de 0,2uM para serem utilizadas nas reações de amplificação do DNA molde.

3.2.2 DNA Molde para Extensão dos Fragmentos

Foi extraído DNA de leucócitos normais humanos, que serviram como molde na extensão dos fragmentos de exons associados a tumor. A extração seguiu o método fenol:clorofórmio. Foi acrescentado ao sangue 2ml de TRIS-EDTA-SDS, agregando 10ul de proteinase K (20mg/ml) na concentração final de 100ug/ml. Incubou-se a 55°C *overnight*. Após digestão com proteinase K, o DNA foi extraído das amostras com igual volume de fenol saturado com 1 M TrisHCl, pH 8,0. Esse material foi agitado por 5 minutos e centrifugado por 10 minutos a 3000 rpm. O sobrenadante (fase aquosa), foi então transferido para outro tubo e acrescentou-se o mesmo volume de clorofórmio e álcool isoamílico (24:1). Agitou-se por 5 minutos e centrifugou-se por 10 minutos a 3.000 rpm. O sobrenadante foi separado (fase aquosa), sendo transferido para outro tubo. Acrescentou-se o mesmo volume de clorofórmio e álcool isoamílico (24:1), e agitou-se por 5 minutos, centrifugando-se em seguida por 10 minutos a 3.000 rpm. O sobrenadante foi separado e transferido para um tubo Corex. O DNA contido na fase

aquosa foi então precipitado com 2,5 volumes de etanol absoluto, em presença de 10% de 3M acetato de sódio e mantido por 14-16 horas a -20°C ou 1 hora a -80°C. Depois centrifugou-se por 20 minutos a 12.000 rpm (ultra centrífuga), desprezando-se o sobrenadante. Adicionou-se então 5ml de etanol 70% e centrifugou-se por 10 minutos a 12.000 rpm, desprezando o sobrenadante, e o precipitado foi submetido à temperatura ambiente por aproximadamente 20 minutos e ressuspendido em um volume de tampão TE (10 mM Tris-HCl, 1 mM EDTA, pH 8,0) de cerca de 200ul. A concentração do DNA foi determinada por leitura espectrofotométrica a 260/280nm e realizou-se eletroforese em gel de agarose 1% corado com Brometo de Etídeo para avaliação da qualidade do DNA (Anexo 5).

3.2.3 Extensão dos Fragmentos

Os fragmentos foram amplificados através de uma reação de PCR, utilizando-se a enzima *Taq Pht* (PHONEUTRIA Biotec. e Serviços Ltda) a partir de aproximadamente 100ng de DNA genômico. Para as reações foram necessárias as seguintes concentrações finais de cada reagente: 1,5mM de MgCl₂; 0,05μM de cada iniciador; 0,2 mM de cada dNTP; 0,05 unidade de *Taq Pht*, 1X tampão adequado e 1M do *enhancer* Betaína. As reações foram realizadas em um volume final de 100μL. O aparelho utilizado para incubação foi o termociclador PTC 100™ (MJ RESEARCH, INC).

As condições de termociclagem seguiram os passos de *touchdown* descritos a seguir: **1 ciclo:** 96°C por 3 min. **6 ciclos:** 96°C por 1 min; 60°C por 1 min; 72°C por 1 min. **6 ciclos:** 96°C por 1 min; 58°C por 1 min; 72°C

por 1 min. **12 ciclos:** 96°C por 30 seg; 55°C por 30 seg; 72°C por 1 min. **16 ciclos:** 96°C por 30 seg; 53°C por 30 seg; 72°C por 1 min. **1 ciclo:** 72°C por 7 min.

Alíquotas de 5,0µL das reações de PCR foram, em seguida, aplicadas em gel de agarose 1,0% corado com Brometo de Etídio em uma corrida com duração de 60 minutos na voltagem de 60 mV, a fim de se confirmar os tamanhos dos fragmentos esperados.

3.2.4 Desenho dos iniciadores para extensão de fragmentos dos genes *GAPDH* e *ACTB*

As seqüências dos genes utilizados como normalizadores foram obtidas através do banco de dados *Ref Seq*, sendo que 8 iniciadores foram desenhados, 2 pares para cada gene, resultando em 1 fragmento referente à região exônica e 1 à região intrônica, que se referem a controles positivos e negativos, respectivamente.

***GAPDH* (NM_002046.2)**

Iniciadores para a região EXÔNICA: Desenhados no exon 8 do gene a 333 pb da região 3' (tamanho do fragmento: 199 pb)

GAPDH-EXON-Foward (GGAAGCTCACTGGCATGG)

GAPDH-EXON-Reverse (GAGTGGGTGTCGCTGTTG)

Iniciadores para a região INTRÔNICA: Desenhados entre os exons 7 e 8 (tamanho do fragmento: 147 pb)

GAPDH-INTRON-Foward (GCTCCCACCTTTCTCATCC)

GAPDH-INTRON-Reverse (CTGCAGCGTACTCCCCAC)

ACTB (NM_001101.2)

Iniciadores para a região EXÔNICA: Desenhados no exon 5 do gene a 742 pb da região 3' (tamanho do fragmento: 156 pb)

ACTB-EXON-Foward (CACGAAACTACCTTCAACTC)

ACTB-EXON-Reverse (CTTCATTGTGCTGGGTGC)

Iniciadores para a região INTRÔNICA: Desenhados entre os exons 3 e 4 (tamanho do fragmento de 211 pb)

ACTB-INTRON-Foward (GTGTTGTGGGTGTAGGTAC)

ACTB-INTRON-Reverse (CATGTCACACTGGGGAAG).

Os iniciadores foram diluídos a uma concentração estoque de 100uM, de onde foram feitas diluições trabalho de 10µM para serem utilizadas nas reações de amplificação do DNA molde. Para as reações foram necessárias as seguintes concentrações finais de cada reagente: 1,5mM de MgCl₂; 0,05µM de cada iniciador; 0,2 mM de cada dNTP; 0,05 unidade de *Taq Pht*; 1X tampão adequado e 1M do *enhancer* Betaína. As reações foram realizadas em um volume final de 100µL. O aparelho utilizado para incubação foi o termociclador PTC 100™ (MJ RESEARCH, INC).

As condições de termociclagem seguiram os seguintes passos: **1 ciclo:** 95°C por 2 min. **40 ciclos:** 95°C por 45 seg; 57°C por 45 seg; 72°C por 1 min. **1 ciclo:** 72°C por 7 min.

Assim, 5 μ L de cada um dos 4 fragmentos foram corridos em gel de agarose 1% corado com Brometo de Etídio em uma corrida com duração de 60 minutos na voltagem de 60 mV, onde foram confirmados os tamanhos mencionados acima.

3.2.5 Purificação dos Fragmentos

Os fragmentos a serem imobilizados, depois de confirmados os seus tamanhos em gel de agarose 1% corado com Brometo de Etídeo, foram concentrados a um volume de aproximadamente 25 μ L e aplicados em gel de agarose 1,0%, para serem purificados de acordo com o tamanho esperado com o auxílio do *Kit GFX PCR DNA and Gel Band Purification (GE Healthcare Life Sciences)*, seguindo-se as recomendações do fabricante. Os fragmentos foram então quantificados em espectrofotômetro (*Gene Quant pro*) e corridos em gel de acrilamida 8%. Esse procedimento foi adotado com a finalidade de descartar a presença de possíveis bandas inespecíficas que pudessem favorecer hibridizações cruzadas.

3.2.6 Clonagem, Seqüenciamento e Análise das Seqüências

Após o procedimento de purificação dos fragmentos, parte do volume deles foi utilizada para a construção da membrana e o restante preparado para clonagem e seqüenciamento. Todos os 270 fragmentos, mais os fragmentos correspondentes a íntrons e exons dos genes *GAPDH* e *ACTB* foram clonados no plasmídeo pTZ57R/T (*InsT/A clone PCR Product Cloning Kit* - Fermentas) e seqüenciados em seguida. Foi utilizado o kit de

seqüenciamento *BigDye Terminator v3.1 cycle sequencing kit* (Applied Biosystems), seguindo-se as especificações do fabricante. O aparelho seqüenciador utilizado foi o ABI Prism 3100 (*Applied Biosystems*).

O procedimento de clonagem dos fragmentos foi adotado com a finalidade de incluí-los em outra plataforma de cDNA em vidro que foi construída pela equipe do *Microarray facility* do Hospital AC Camargo.

A identidade das seqüências foi analisada com o auxílio da ferramenta disponível no site do NCBI (*National Center for Biotechnology Information*), *Blast 2 Sequences* (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>) (Figura 2).

A) Seqüência ABI Prism 3100:

>C04 seqüência exportada do cromatograma

CCCAAGCCCCGCTTTTACCCAGCATATCTGAGTCACCTGCAGACAGGATGAGAACCCA
GACCCCTTCCCGTGAGTCCACATGCCCGGCTGCCCTGGGACCAGCTGCCTGGCTCC
TACCTAGGTGGCTCTCCTCTCCTTTTCTCTGCCTTGACCTCTCTGCAGCTCTGATTCTG
GAAGATTCTGTCCCCTTAAAATCCATGAAGCAGCACTGGCCTGGGTCTCCCCTCACCCCT
TCTCATAGGCTTGGCTATGACTTC

B) Seqüência do exon 115538 6:

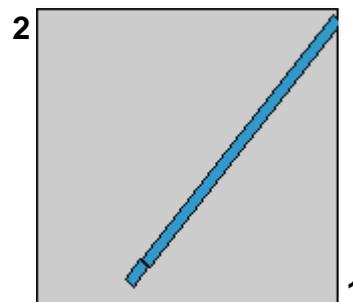
>115538_6

CTCACCACAGACCCCGCCGAGAAAGGCACAGCCTGAGGCACTCAGCACAGTCCAGGGC
CCCAAGCTGAGACACACTCCAAGAACAACACTCACAACCCCAAGCAATATCTGAGTCAC
CTGCAAGACAGGATGAGAAACCAGACCCCTTCCCGTGAGTCCACATGCCCGGCTGCC
CTGGGACCAGCTGCCTGGCTCCTACCTAGGTGGCTCTCCTCTCCTTTTCTCTGCCTTG
ACCTCTCTGCAGCTCTGATTCTGGAAGATTCTGTCCCCTTAAAATCCATGAAGCAGCAC
TGGCCTGGGTCTCCCCTCACCCCTTCTCATAGGCTTGGCT

C) Blast 2 Sequences:

Sequence 1 lcl|115538_6 **Length** 331 (1..331)

Sequence 2 lcl|C04 **Length** 258 (1..258)



Legenda: **A)** Seqüência obtida com o seqüenciamento do exon 115538_6; **B)** seqüência FASTA do exon; **C)** *Blast 2 Sequences* - Gráfico representativo da comparação da seqüência do exon 115538_6 com a seqüência obtida através de seqüenciamento.

Figura 2 - Seqüenciamento do exon 115538_6.

3.2.7 Construção das Membranas

Aproximadamente 150 nanolitros de cada fragmento (em média 5-10 ng) foram arranjados em membrana de *nylon* (Hybond N+ Amersham) em quadruplicata com o auxílio do robô Flexys (Genome Solutions, UK). Todas as etapas de confecção das membranas foram coordenadas pelo Dr. Alex Fiorini de Carvalho do grupo do Dr. Luiz Fernando Lima Reis, da equipe de *Microarray Facility* do Instituto Ludwig e Hospital AC Camargo. Os exons foram imobilizados em grupos de 96 *subarrays*, cada um contendo 16 *spots*. Em cada *subarray* foram imobilizados 2 controles negativos (água milliQ) e 2 controles positivos de marcação e hibridização (cDNA correspondente a um fragmento do lambda fago Q gene). Foram imobilizados também os dois fragmentos, um de 199 pb, do exon 8 e outro de 156 pb do exon 5, correspondentes ao genes normalizadores *GAPDH* e *ACTB*,

respectivamente. Para garantir que os sinais captados correspondem à região de cDNA, e monitorar a contaminação genômica das amostras, um fragmento intrônico de 147 pb (intron entre os exons 7 e 8) e outro de 211 pb (intron entre os exons 3 e 4) dos mesmos genes foram também imobilizados. As membranas foram, em seguida, desnaturadas por 2 minutos em solução de 0,5N NaOH; 1,5M NaCl e neutralizadas por 2 minutos em solução de 1,5M NaCl; 0,5M Tris-Hcl (pH 7,4). Em seguida os fragmentos de DNA foram fixados à membrana pela exposição à luz ultravioleta ($120\mu\text{J}/\text{cm}^2$) do aparelho *Crosslinker* UV (Fisher Biotech, USA) e à temperatura de 80°C por 2 horas.

Em resumo, a plataforma contém um total de 1.536 *spots*, distribuídos em 96 *subarrays*, contendo cada um 16 *spots*: 12 correspondentes a 6 exons associados a tumor em duplicata, 02 controles positivos de marcação (λ Qgene) e 02 controles negativos (água). Cada um dos 270 exons associados a tumor foi imobilizado 4 vezes na mesma membrana. Foram imobilizadas também 24 réplicas de exons dos genes normalizadores *GAPDH* e *ACTB* e 12 réplicas de íntrons dos mesmos genes para controle negativo de expressão.

3.3 AMOSTRAS

3.3.1 Amostras utilizadas nos experimentos de *microarray* e validação através de RT-PCR quantitativo (qRT-PCR)

Para o grupo inicial, foram selecionadas 27 amostras de tecido tumoral de mama de pacientes operadas no Hospital AC Camargo arquivadas no Banco de Tumores desta Instituição (Biorrepositório Hospital AC Camargo) e 09 amostras de tecido normal de mama, sendo 04 bordas de fibroadenoma (MN04, MN12, MN13, MN16), 01 amostra de mamoplastia (MN46) e 04 bordas de carcinoma invasor (MN06, MN08, MNP12 e MN32) que eram bordas livres de doença de amostras tumorais (6T, 8T, 12T e 32T), respectivamente.

As amostras foram obtidas cirurgicamente e congeladas em nitrogênio líquido (-196°C). Antes de serem processadas, todas as amostras foram submetidas à confirmação histopatológica. Esta validação visa não somente a confirmação do diagnóstico, mas também a avaliação da amostra quanto à presença de necrose, infiltrado inflamatório, estroma, epiderme e tecido adiposo. As amostras foram incluídas em *Tissue-Tek*® e, com auxílio de um criostato (a -30°C), foram cortadas a 5 μm de espessura e fixadas em lâminas histológicas, as quais foram coradas com hematoxilina e eosina (H&E). As lâminas foram avaliadas por um patologista e o critério mínimo da presença do tecido de interesse (tumoral ou normal) em todas as amostras foi de 70%. Quando necessário as amostras foram dissecadas manualmente pelo patologista, guiado pelas lâminas de H&E.

Todas as amostras foram selecionadas de pacientes do sexo feminino com diagnóstico de carcinoma ductal invasivo, e não submetidas a tratamento prévio à cirurgia (quimioterapia, hormônio terapia ou radioterapia), sendo que 04 pertenciam ao estadio I; 14 ao estadio IIA; 07 ao estadio IIB; e 02 ao estadio IIIA.

A partir das informações coletadas, 11 casos foram classificados como linfonodo-positivo e 16 casos linfonodo-negativo. A idade das pacientes variou entre 32 e 87 anos, apresentando uma média de 54,7 anos e mediana de 55 anos. Dados referentes aos marcadores moleculares receptor de estrógeno (ER), ERBB2, receptor de Progesterona (PgR) também foram coletados. Em relação ao marcador ER, 22 casos eram positivos, 05 negativos; em relação ao marcador ERBB2, 21 casos eram positivos, 06 negativos; em relação ao marcador PgR, 15 casos eram positivos, 12 negativos.

Todas as informações detalhadas a respeito dessas amostras estão apresentadas no Anexo 6.

Foram também utilizadas no estudo 04 amostras correspondentes a linhagens celulares tumorais de mama da “*American Type Culture Collection*” (ATCC): MCF7 (ATCC: HTB-22 - célula epitelial de adenocarcinoma), MDA-MB-231 (ATCC: HTB-26 - célula epitelial de adenocarcinoma), MDA-MB-435S (ATCC: HTB-129 - célula epitelial de carcinoma ductal invasivo), e HB4a C5.2 (célula epitelial de lúmen de mama normal transfectada com o oncogene *ERBB2*) e 02 correspondentes a linhagens de mama normal: HB4a (célula epitelial de lúmen de mama

normal) e MCF10A (ATCC: CRL-10317- célula epitelial de doença fibrocística). No Anexo 7 estão relacionadas todas as linhagens utilizadas no estudo e as informações detalhadas a respeito do *status* dos marcadores de prognóstico ER e PgR.

3.3.2 Amostras utilizadas para validação em grupo independente através de qRT-PCR

Foram utilizadas 40 amostras tumorais de mama obtidas de pacientes operadas no Hospital AC Camargo arquivadas no Banco de Tumores de nossa Instituição (Biorrepositório Hospital AC Camargo). Para essa validação foram incluídas duas amostras normais bordas de fibroadenoma (MN02 e MN23) e duas amostras de glândula mamária normal (*Clontech*, cat#636576, sendo uma do lote#7020050 e outra de lote desconhecido), uma vez que não restou material do grupo inicial. O critério para seleção das amostras foi o mesmo utilizado anteriormente, sendo todas as amostras de pacientes do sexo feminino com diagnóstico de carcinoma ductal invasivo, e não submetidas a tratamento prévio à cirurgia (quimioterapia, hormônio terapia ou radioterapia). Todas as informações detalhadas a respeito dessas amostras estão apresentadas no Anexo 12.

Uma tabela caracterizando clinicamente toda a casuística utilizada neste trabalho está apresentada em resultados (página 116).

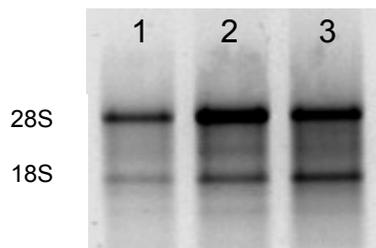
3.4 EXTRAÇÃO DE RNA

A metodologia de extração de RNA total utilizada para o grupo inicial e para todas as amostras normais foi através de gradiente isopícnico de Cloreto de Césio - CsCl (GLISIN et al. 1974) (Anexo 8), uma vez que esta técnica possibilita maior recuperação do RNA total com melhor qualidade das amostras de mama normal quando comparada à extração por TRIZOL[®] Reagent. Foi também extraído RNA das 06 linhagens celulares de mama (HB4a, HB4a C5.2, MCF7, MDA-MB-231, MDA-MB-435S e MCF10A), as quais seguiram o método TRIZOL[®] (*Invitrogen Life Technologies™*), seguindo-se as especificações do fabricante (Anexo 9). Esse procedimento também foi utilizado para amostras tumorais do grupo independente.

3.4.1 Avaliação da qualidade do RNA extraído

A qualidade do RNA extraído foi analisada através de eletroforese em gel de agarose 1,0% utilizando-se 5,0µL de tampão de corrida desnaturante (*3X Loading Buffer bromofenol blue* e 7M uréia) antes de ser aplicada no gel. Em seguida adicionou-se 2,0µL de brometo de etídio 0,5 µg/µL à amostra. A corrida foi realizada a 60 mV por 50 minutos. A análise da integridade foi feita avaliando-se a relação entre as bandas correspondentes aos RNA ribossomais 28S e 18S, cuja relação deve ser \geq a 1 (Figura 3). A quantificação do RNA total foi feita pela leitura da absorbância em espectrofotômetro (*Gene Quant pro*), com comprimento de onda equivalente a 260 nm, considerando-se que 1 OD a 260 nm equivale a 40 ng/µl de RNA.

A razão das leituras 260/280 foi utilizada como parâmetro na estimativa do grau de contaminação do RNA por fenol ou proteínas, sendo utilizado RNA com razão $\geq 1,8$.



Legenda: 1 - amostra 18T; 2 - amostra 19T; 3 - amostra 21T; 0,5ug RNA total - Gel de agarose 1,0%. Relação entre as bandas 28S/18S: 2:1.

Figura 3 - Representação das bandas 28S e 18S de RNA ribossomal de 3 amostras de tecido tumoral extraídas por CsCl.

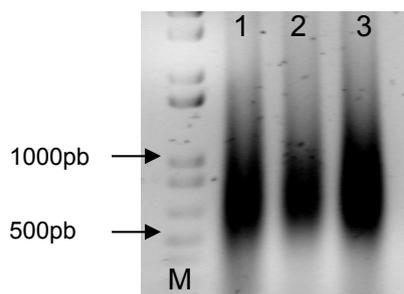
3.5 AMPLIFICAÇÃO DE RNA (RNAa)

Os RNA totais foram submetidos à metodologia de amplificação de mRNA, baseado em GOMES et al. (2003), para obtenção de quantidade suficiente para a realização dos experimentos de hibridização. Foram utilizados 4ug de RNA total na presença do iniciador: Oligo dT (18)-T7 iniciador. O iniciador possui uma cauda poli T que ancora na cauda poli-A dos RNA mensageiros e possuem região correspondente ao promotor da RNA polimerase do fago T7, para a reação de transcrição *in vitro*. Esse método utiliza dois *rounds* de amplificação para que a quantidade de RNA

seja suficiente para os experimentos que se seguirão, permitindo a amplificação de mRNA na ordem de 3×10^2 a 3×10^3 (Anexo 10).

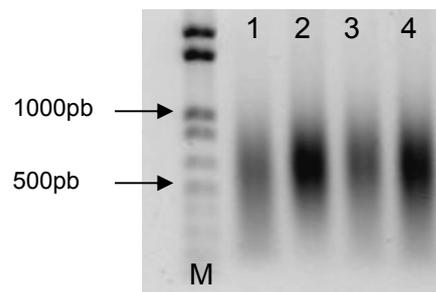
3.5.1 Avaliação da qualidade do RNA amplificado (RNAa)

A avaliação das amostras de RNA amplificado foi realizada em gel de agarose 1,0%, em uma corrida com duração de 50 minutos na voltagem de 60 mV. 1µg de RNA amplificado foi adicionada a 5,0µL de tampão TAE com uréia 7M, 3X Loading Buffer azul de bromofenol e 2,0µL de brometo de etídio 0,5 µg/µL antes de ser aplicada no gel. Somente as amostras que apresentaram um perfil de amplificação onde a maior concentração de RNA amplificado (RNAa) estava pelo menos entre 350-750 pb foram utilizadas, pois correspondem a uma qualidade de hibridização satisfatória. A quantificação do RNA amplificado foi feita pela leitura da absorbância em espectrofotômetro (*Gene Quant pro*), com comprimento de onda equivalente a 260 nm, considerando-se que 1 OD a 260 nm equivale a 40 ng/µl de RNAa. A razão das leituras 260/280 foi utilizada como parâmetro na estimativa do grau de contaminação do RNAa por fenol ou proteínas, sendo utilizado RNA com razão $\geq 1,8$. As figuras 4 e 5 representam alguns exemplos de RNAs amplificados.



Legenda: 1 ug RNAa. Gel de agarose 1,0%. M: marcador de peso molecular 1kb Plus (Invitrogen).

Figura 4 - Representação de RNAa de 3 linhagens celulares de mama-1: MDA-MB-435; 2: MDA-MB-231 e 3: MCF7.



Legenda: 1 ug RNAa. Gel de agarose 1,0%. M: marcador de peso molecular 1kb Plus (Invitrogen).

Figura 5 - Representação de RNAa de 4 amostras de tecido tumoral de mama- 1: 19T, 2: 26T; 3: 27T; 4: 28T.

3.6 MARCAÇÃO E HIBRIDIZAÇÃO

A hibridização das membranas foi realizada em duplicata de acordo com o método de CHURCH e GILBERT (1984). Para tanto, 5µg de RNA amplificado foram marcados com $\alpha^{32}\text{P}$ -dCTP em uma reação de transcrição reversa segundo o procedimento descrito por HAUSER et al. (1998). A cada tubo contendo RNAa foram adicionados 3µL de iniciadores *random primer* dN6 (3µg/µl), 0,3µl de RNA de fago λ Qgene. O volume da reação foi completado com água DEPC para 15,9µl e foi realizada uma incubação à 70°C durante 10 min. Em seguida foram adicionados 3,0µL da enzima *Transcriptase Reversa* (200u/µL) (*SuperScript II*, *Invitrogen Life Technologies*TM), 0,5µL de RNase IN (40u/µL), 60µCi de $\alpha^{32}\text{P}$ -dCTP, 1,66µM de dCTP, 0,25 mM de cada um dos demais dNTPs (dATP, dTTP e dGTP),

3,0 μ L de DDT (0,1M), 10,0 μ l de tampão 5X e água DEPC q.s.p. o volume final de 50 μ L de reação para a síntese da primeira fita de cDNA. A reação foi incubada durante 2 horas a 43°C e em seguida foi feita a hidrólise do RNA através da adição de 1,5 μ L de 1% SDS, 1,5 μ L de 0,5M EDTA, 3,0 μ L de 3M NaOH, durante 30 minutos a 67°C e 15 minutos à temperatura ambiente. A seguir a reação foi neutralizada adicionando-se 1,5 μ L de 1M Tris-HCL pH8,0 e 4,5 μ L de 2N HCl. Os nucleotídeos radioativos não incorporados foram removidos por centrifugação em coluna Sephadex G50 previamente equilibrada com água MilliQ.

A membrana foi pré-hibridizada em solução de 0,25M Na₂HPO₄ pH 7,2, 7% SDS, 1% BSA e 1mM EDTA durante 1 hora a 65°C em forno giratório. A seguir foram adicionados os cDNAs marcados radioativamente para que ocorresse a hibridização sob a temperatura de 65°C durante um período de 16 horas. A membrana foi, então, lavada duas vezes por 30 min. a 65°C em solução contendo 0,5M Na₂HPO₄ pH 7,2, 1% SDS, e 1mM EDTA e uma vez em solução contendo 0,25M Na₂HPO₄ pH 7,2, 1% SDS, e 1mM EDTA. Em seguida a membrana foi exposta a *Phosphorimager* (*Molecular Dynamics*, USA) durante um período de 96 horas e escaneada em aparelho *STORM 840* (*Molecular Dynamics*, USA).

3.7 CAPTURA DOS DADOS

Os sinais gerados nas membranas foram capturados em *Phosphosimager* (*Molecular Dynamics*, USA) e a imagem gerada foi analisada pelo programa *Array Vision*[™] (*Imaging Research Inc*, USA). Este programa localiza automaticamente os sinais gerados pela hibridização, devendo ser feito posteriormente um ajuste manual ponto a ponto. Em seguida o programa quantifica a intensidade deste sinal através da média de distribuição dos *pixels*. Para cada elemento, o valor fornecido pelo *software* é chamado de volume. O volume total é correspondente ao valor da média de todos os *pixels* contidos no *spot* multiplicado pela área do *spot*.

3.8 CORREÇÃO DO VALOR DE SINAL OBTIDO

As análises matemáticas e estatísticas foram realizadas pelo Laboratório de Bioinformática do Hospital AC Camargo (LBHC) sob a coordenação da Dra. Helena P. Brentani. O valor de intensidade do *background* local foi subtraído do valor de intensidade de expressão do *foreground* obtido para cada elemento, sendo considerados apenas os pontos que apresentaram valor positivo. Após esta etapa todos os valores foram trabalhados na escala logarítmica com base 2.

3.9 REMOÇÃO DOS OUTLIERS

Como há oito medidas correspondentes a cada exon (duplicata de membrana com 4 *spots* do mesmo exon em cada uma), foi feita uma média dos 8 valores de intensidade para cada exon e os valores que ultrapassaram $\pm 2\sigma$ da média foram excluídos da análise.

3.10 NORMALIZAÇÃO DOS DADOS

A normalização foi realizada através da média de intensidade das 24 réplicas de exons dos genes normalizadores *GAPDH* e *ACTB* imobilizados na membrana. Assim, foi criada uma constante através da média dos 24 pontos (K) em cada membrana e a normalização foi feita membrana a membrana e ponto a ponto, isto é, cada exon foi normalizado individualmente através da subtração do valor de K.

3.11 CONTROLE DE QUALIDADE EXPERIMENTAL DAS MEMBRANAS

Após a subtração do valor de intensidade do *background* e normalização, as membranas passaram por um controle de qualidade experimental.

Para verificar a reprodutibilidade experimental entre a membrana principal e sua réplica foi calculada a Correlação de *Pearson* (R^2), onde somente foram aceitas amostras com correlação acima de 0,8.

Foi analisada a variabilidade entre cada exon em cada experimento. Para isso, foi feita uma distribuição dos valores de intensidade após normalização para cada um dos 270 exons.

3.12 ANÁLISE ESTATÍSTICA

Após a normalização dos dados, foi aplicado o teste estatístico *TStudent*, sendo considerados exons com expressão diferencial significativa aqueles que apresentaram $p < 0,05$ entre os diferentes grupos de amostras.

Para verificar a representatividade dos exons diferencialmente expressos que estavam entre os 75 previamente selecionados pela análise computacional como associados a tumor de mama, levou-se em consideração todos os exons diferencialmente expressos da análise e os 270 imobilizados na membrana, aplicando-se o teste de Qui-quadrado (X^2) com intervalo de confiança 1 e considerando-se um $p < 0,05$ (5%).

Para avaliação da variabilidade dos valores de intensidade entre os exons pertencentes a cada classe analisada foi utilizada uma ferramenta da *web* que implementa a ferramenta estatística *Bay Boots* (*bayesian bootstrap*) (<http://blasto.iq.usp.br/~rvencio/bayboots/index.php>), baseada no BER (*Bayes error rate*), onde são considerados diferencialmente expressos os valores de intensidade cuja comparação não apresente sobreposição de

curvas maior que 30% ($BER < 0,3$). Essa ferramenta estatística foi aplicada em todos os exons diferencialmente expressos em todos os grupos e entre todas as comparações.

3.13 ANÁLISE DOS GENES EM BANCOS DE SAGE

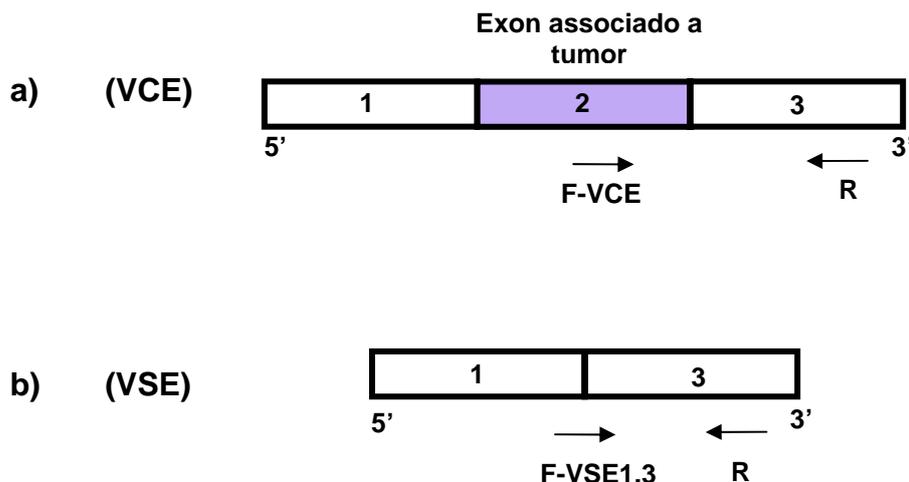
Com o intuito de aumentar a chance de selecionar variantes de *splicing* mais expressas em tumores e não genes mais expressos, uma análise prévia em bancos de SAGE foi realizada pelo Laboratório de Bioinformática do Hospital AC Camargo (LBHC). Para selecionar os exons superexpressos que tinham probabilidade de pertencer somente à variante de *splicing* que contém o exon (VCE) e não ao gene como um todo, foi investigada a expressão de cada cDNA *full length* em bancos de dados de SAGE. Foram geradas *tags* virtuais das seqüências completas que continham uma cauda poliA. Para isso foram inicialmente mapeadas as *tags* virtuais com a predição de uma seqüência de 10 pb determinada pelo sítio de restrição da enzima *NlaIII*, a partir da extremidade 3' da seqüência. Foram excluídas *tags* localizadas em regiões repetitivas, uma vez que as regiões genômicas repetitivas (*repetitive regions*) aparecem em diferentes localizações do genoma, e *tags* inespecíficas, isto é, que representam mais de um gene. A freqüência da *tag* correspondente ao cDNA *full length* foi comparada nos dois tipos de bibliotecas de mama (tumoral e normal), utilizando-se uma ferramenta da *web* denominada SAGEbetaBin (VENCIO et al. 2004), um modelo bayesiano beta-binomial que analisa a variabilidade

entre classes e está baseado no BER (*Bayes error rate*), o qual considera diferencialmente expressos os valores de intensidade cuja comparação não apresente sobreposição de curvas maior que 30% (BER<0,3).

3.14 VALIDAÇÃO ATRAVÉS DA ANÁLISE DO NÍVEL DE EXPRESSÃO POR RT-PCR QUANTITATIVO (qRT-PCR)

3.14.1 Desenho dos iniciadores

Para avaliar o nível de expressão através de RT-PCR quantitativo foram desenhados iniciadores para as 2 variantes: VCE (com o exon associado a tumor), VSE (sem o exon associado a tumor) e para o gene como um todo. Para a variante contendo exon, o primer *forward* foi desenhado dentro do exon 2 associado a tumor (F-VCE, Figura 6a) e para a variante sem o exon, o primer *forward* foi desenhado na junção exon-exon (exon1-exon3) do transcrito que não contém o exon associado a tumor (F-VSE1.3, Figura 6b). Foi utilizado um iniciador *reverse* comum (R) para as duas formas desenhado no exon 3, que flanqueia à 3' o exon associado a tumor (Figura 6a e b). Para a avaliação do gene como um todo os iniciadores foram desenhados em exons constitutivos (página 29).



Legenda: (VCE) - Variante com o exon; (VSE) - variante sem o exon; 1- exon flanqueador *upstream*; 2- exon associado a tumor; 3- exon flanqueador *downstream*; F-VCE- iniciador *forward* desenhado no exon 2, associado a tumor; F-VSE1.3- iniciador *forward* desenhado na junção exon1-exon3 da variante sem o exon; R- iniciador *reverse* desenhado no exon 3, comum às duas variantes (VCE e VSE).

Figura 6 - Representação esquemática do pareamento dos iniciadores na VCE (a) e na VSE (b).

Para evitar extensão inespecífica, no caso da variante sem o exon (VSE), para os iniciadores desenhados nas junções exon1-exon3, foi estabelecido um critério de sobreposição da porção 3' do iniciador *forward* ser de no máximo 4 bases pareando no exon 3, segundo recomendação de SHULZHENKO et al. (2003). Esse critério foi estabelecido, uma vez que os fragmentos gerados a partir das duas combinações de iniciadores correspondentes aos dois tipos de transcritos (VCE e VSE) compartilham seqüências comuns (exon 3), e o que os diferencia é apenas a trecho da seqüência correspondente ao exon 1 da VSE e do exon 2 (associado a tumor) da VCE. Os iniciadores apresentaram de 11 a 19 bases, temperatura

de pareamento variando entre 52 e 62°C, sendo que a diferença em graus de temperatura permitida entre os pares de iniciadores foi de no máximo 3°C. Para verificar a presença de estruturas secundárias (*stem loop* e *homodimer*) foi utilizado o programa *OLIGOTECH* versão 1.00 (Copyright® 1995), permitindo temperaturas de no máximo 20°C. Todos os iniciadores foram diluídos a uma concentração estoque de 100uM, de onde foram feitas diluições trabalho de 10uM, as quais foram utilizadas nas reações de qRT-PCR.

3.14.2 Síntese de cDNA

O cDNA de todas as amostras foi sintetizado a partir de 1ug de RNA total, adicionando-se 1ul de iniciadores 0,5 ug/ul oligo dT, e em seguida incubou-se a 70°C por 10 minutos. Uma mistura composta por 4ul de 5x *First strand buffer*; 2ul 0,1 M DTT; 1ul de *Rnase IN*; 2ul de 10mM dNTP; 1ul de *Superscript II (Invitrogen Life Technologies™)*; água DEPC q.s.p. 20ul foi adicionado, e em seguida incubou-se a reação 42°C em termociclador PTC 100™ (MJ Research, INC). Após 45 minutos foi adicionado mais 1ul de *Superscript II* e incubou-se por mais 45 minutos a 42°C. Para a verificação da integridade do cDNA foram feitas reações de RT-PCR para o gene normalizador GAPDH₁₁₀₁, utilizando-se iniciadores localizados a 1.101pb da porção 3' do gene (F: GAAGGTGAAGGTCGGA; R: GGGTCATTGATGGCAAC), em um volume final de 20µl, sendo 2,0 µl de tampão (10X); 1,5 mM de MgCl₂; 0,4 mM de iniciador *forward*; 0,4 mM de iniciador *reverse*; 1 U de *Platinum Taq DNA Polymerase (Invitrogen Life*

*Technologies*TM) e 10,0ng de cDNA. O programa apresentou as seguintes etapas: 4 minutos a 94°C, seguidos de 23 ciclos de 45 segundos a 94°C, 45 segundos a 60°C para o pareamento dos iniciadores e 1 minuto a 72°C. Por fim, estendeu-se 6 minutos a 72°C, terminando-se em 16°C infinito. Alíquotas de 5,0µL das reações de RT-PCR foram, em seguida, aplicadas em gel de agarose 1,0% corado com Brometo de Etídio em uma corrida com duração aproximada de 60 minutos na voltagem de 80 mV, a fim de se confirmar a amplificação de um fragmento de 102pb correspondente ao gene normalizador GAPDH₁₁₀₁.

3.14.3 Estabelecimento das condições da reação de RT-PCR

As reações de RT-PCR foram feitas em um volume final de 20µl, sendo 2,0 µl de tampão (10X); 1,5 mM de MgCl₂; 0,4µl de iniciador *forward*; 0,4µl de iniciador *reverse*; 1 U de *Platinum* Taq DNA Polymerase (*Invitrogen Life Technologies*TM) e quantidade de cDNA correspondente a 20 ng de RNA total. O programa apresentou a seguinte ciclagem de temperatura: 4 minutos a 94°C, seguidos de 35 ciclos de 45 segundos a 94°C, 45 segundos na temperatura específica de pareamento de cada iniciador e 1 minuto a 72°C. Por fim, estendeu-se por 6 minutos a 72°C, terminando-se em 16°C infinito.

Alíquotas de 5,0µL das reações de RT-PCR foram, em seguida, aplicadas em gel de poliacrilamida 8,0% em uma corrida com duração aproximada de 60 minutos na voltagem de 80 mV, a fim de se confirmar os tamanhos dos fragmentos esperados.

3.14.4 Teste de especificidade do iniciador *forward* da variante sem o exon (VSE) desenhado na junção exon1-exon3

O iniciador *forward* da VSE (F-VSE1.3) ancora com especificidade total na junção exon1-exon3 desta variante, e pode ancorar de forma parcial na junção exon2-exon3 da VCE, pareando de forma específica somente na região correspondente ao exon 3 (Figura 7-Esquema 1). Isso pode ocorrer mesmo que o pareamento desses iniciadores seja de apenas 4pb conforme já descrito anteriormente. Dessa forma, para teste de especificidade foi utilizado como molde os *amplicons* constituídos pelos exons 2+3 da VCE, que representavam a junção exon2-exon3, e os iniciadores F-VSE1.3 e R, complementares à junção exon1-exon3 e ao exon 3, respectivamente. A presença ou ausência de amplificação do fragmento de tamanho esperado sinaliza ocorrência de amplificação inespecífica ou específica, respectivamente.

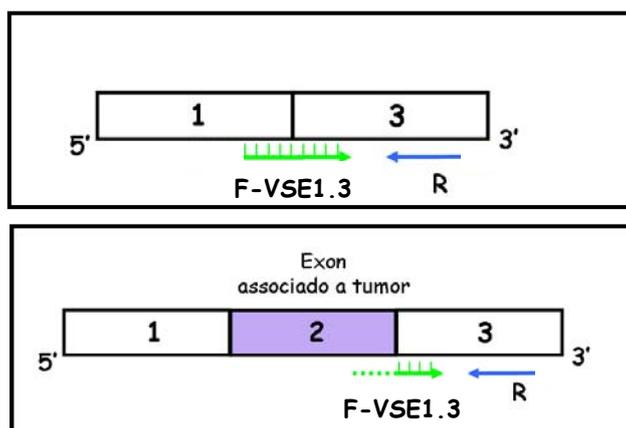


Figura 7 - Esquema 1 = Representação hipotética do pareamento inespecífico do iniciador F-SE na variante com o exon (VCE).

3.14.5 RT-PCR quantitativo (qRT-PCR)

qRT-PCR ou RT-PCR em tempo real avalia o acúmulo do produto na fase logarítmica da reação de amplificação e está diretamente relacionado à quantidade de transcritos existentes no início da reação. O sistema de detecção utilizado neste trabalho foi o *SYBR® Green*, que se baseia na utilização de uma molécula fluorescente denominada *SYBR® Green I* que, quando intercalada à dupla fita de DNA, passa a emitir fluorescência. Durante os ciclos iniciais da reação de PCR o sinal de fluorescência emitido pelo *SYBR® Green I* é fraco para ser detectado (não ultrapassando o sinal de fluorescência de fundo). Entretanto, no decorrer dos ciclos da reação de PCR, há um aumento do produto amplificado e conseqüentemente do sinal fluorescente, que passa então a ser detectável, sendo que quanto menor o ciclo em que a fluorescência atinge o limite determinado pelo aparelho, maior é o nível de expressão do transcrito analisado (MORRISON et al. 1998). Desta forma, valores quantitativos são obtidos do ponto em que a amplificação do produto foi detectada, diferente da reação de PCR convencional, na qual a detecção acontece depois de um número fixo de ciclos, o que representa a quantidade final de cada produto acumulado. Nesta metodologia, a quantificação da expressão gênica é baseada no parâmetro de *CTs* (*Threshold Cycle* ou ciclo limite) ou *CPs* (*Crossing Point* ou ponto de cruzamento), definido como o número do ciclo em que a fluorescência gerada atinge um limite acima da fluorescência de fundo (Livak et al., 2001), sendo os dados de quantificação analisados pelo programa de análise do aparelho (MORRISON et al. 1998).

As reações foram realizadas no aparelho *ABI Prism® 7500 Sequence Detection System (Applied Biosystems)* em um volume final de 20µl em placas de 96 wells - *MicroAmp® Optical (Applied Biosystems)*. Cada reação foi realizada em duplicata na presença de 10µl de *SYBR® Green PCR Master Mix (Applied Biosystems)* e as reações que apresentaram valores de desvio padrão acima de 0,6 foram desconsideradas.

3.14.6 Padronização das concentrações de cDNA e de iniciadores

Para cada gene estudado foi realizada uma cinética de amplificação numa reação de qRT-PCR usando diferentes concentrações de iniciadores e diferentes diluições de cDNA convertido de RNA de uma mesma amostra, e as quantidades ideais foram padronizadas individualmente. Para seu estabelecimento foram utilizadas as seguintes quantidades finais de cada iniciador: 200, 500 ou 800nM, e cDNA a partir de 10 e 30ng de RNA total.

Para a padronização da quantidade de cDNA foi selecionada, preferencialmente, a mínima concentração que mostrasse um C_T não muito variável em relação a uma maior concentração, a fim de economizar amostra.

As concentrações dos iniciadores foram padronizadas considerando-se dois fatores: a utilização, sempre que possível, das menores concentrações de iniciadores a fim de evitar a formação de dímeros, e a obtenção dos menores valores de C_T possíveis.

As condições de amplificação consistiram em 2 minutos a 50°C para a ativação da enzima *AmpErase*, 10 minutos a 95°C para a ativação da

enzima *AmpliTaq Gold DNA Polymerase*, seguidos de 40 ciclos de desnaturação a 95°C durante 15 segundos e um minuto na temperatura ideal de cada iniciador para o seu pareamento e extensão. Ao final da amplificação foi adicionada 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), a partir da qual se obtém uma curva de dissociação, utilizada para confirmação da especificidade da amplificação. A especificidade de cada fragmento foi também analisada em gel de poliacrilamida a 8%.

Para determinar a eficiência de amplificação dos iniciadores específicos para os transcritos de interesse (transcritos alvo) ou normalizadores, foi construída uma curva padrão utilizando 5 diferentes concentrações de cDNA em triplicata (100ng, 20ng, 4ng, 0,8ng e 0,16ng) obtidas por meio da diluição seriada de cDNA calculado a partir da quantidade de RNA total de linhagem celular de mama normal HB4a. O cálculo da eficiência de amplificação foi feito para cada gene através da fórmula $[Ef = 10^{(-1/slope)}]$, recomendada pelo manual da *Applied Biosystems*, em que *slope* corresponde à inclinação da reta obtida quando se analisa a variação do C_T dos transcritos alvo e do gene normalizador em função do *log* das diferentes quantidades de cDNA.

Foram aceitos valores mínimos de eficiência de 70% e do coeficiente de regressão da reta (R^2) de 0,97.

3.14.7 Seleção dos genes normalizadores

Para a análise dos dados de expressão das amostras foram testados 5 genes normalizadores, GAPDH (NM_002046.2), ACTB (NM_001101.2) - ambos utilizados como genes normalizadores nos experimentos de *microarray* -, HPRT (NM_000194), BCR (NM_004327) e RPLP0 (NM_001002) - selecionados com base nos dados da literatura como apresentando expressão constitutiva (VANDESOMPELE et al. 2002; DE KOK et al. 2004; HUGGETT et al. 2005). Para seleção do gene normalizador, os valores foram tabulados em planilha EXCEL 9.0/2000 (Microsoft, EUA) e foi utilizada a ferramenta geNorm (Visual Basic Application - VBA), disponível *online* (<http://medgen.urgent.be/~jvdesomp/genorm/>) desenvolvida para o *Microsoft Excel* (VANDESOMPELE et al. 2002). Esse programa calculou automaticamente quais dos genes testados foram os mais estáveis no grupo de cDNA de amostras para serem utilizados como normalizadores.

A estabilidade dos genes normalizadores é determinada partindo-se do princípio de que dois genes normalizadores ideais possuem razões de expressão idênticas em todas as amostras de cDNA analisadas, independente de quaisquer condições experimentais ou tipos celulares. A razão de expressão de cada gene normalizador em relação aos demais é analisada sempre aos pares, o que permite a exclusão dos genes menos estáveis. A cada gene normalizador excluído, o programa automaticamente atribui um novo valor referente à estabilidade dos demais genes 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 genes normalizadores a ser utilizada nas análises de RT-PCR quantitativo.

3.14.8 Análise dos dados de expressão gênica

A expressão diferencial dos transcritos alvo foi determinada pelo método de quantificação relativa a um gene normalizador. Para cálculo da medida relativa de expressão foi utilizado o modelo matemático proposto por PFAFFL (2001) que analisa individualmente as eficiências de amplificação de cada gene (E) e leva em consideração a eficiência de amplificação de cada par de iniciadores (Figura 8). Nesta equação, **E alvo** corresponde à eficiência de amplificação dos iniciadores específicos para o transcrito alvo, **E normalizador** corresponde à eficiência de amplificação dos iniciadores específicos para o gene normalizador, **ΔC_T alvo** corresponde à 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 calibradora, **ΔC_T normalizador** corresponde à diferença entre o C_T do gene normalizador obtido para a amostra em análise e o C_T do gene normalizador obtido para a amostra calibradora.

A linhagem celular de mama normal HB4a foi utilizada como amostra calibradora em todos os experimentos de RT-PCR quantitativo. Assim, o cDNA desta linhagem foi utilizado em duplicata em cada uma das placas de reação de qRT-PCR.

$$\text{Expressão relativa} = \frac{(E_{\text{alvo}} + 1) \Delta C_{\text{t alvo}}}{(E_{\text{endógeno}} + 1) \Delta C_{\text{t endógeno}}}$$

Fonte: PFAFFL (2001)

Figura 8 - Modelo matemático para quantificação relativa em qRT-PCR.

3.14.9 Análises Estatísticas

A expressão diferencial dos transcritos de interesse (transcritos alvo) foi determinada a partir dos dados normalizados pelo método de quantificação relativa considerando um gene normalizador (PFAFFL 2001), apresentado no item anterior. Para cada gene alvo foram obtidas 3 medidas relativas: 1) da variante que contém o exon (VCE), 2) da variante que não contém o exon (VSE) e 3) da expressão constitutiva do gene (EC).

Para avaliação da significância estatística das diferenças de expressão foi utilizado o programa computacional *GraphPad Prism version 4.0* (GraphPad Software Inc., San Diego, CA) aplicando o teste T *Student* não-pareado, e foram considerados como significativos valores de $p < 0,05$. Foram consideradas como variantes superexpressas em amostras tumorais de mama aquelas que: 1) apresentaram um *fold-change* ≥ 3 no cálculo da média do nível de expressão da VCE nas amostras tumorais em relação à média do nível de expressão da VCE nas amostras normais e, 2) um valor relativo de expressão T/N da VCE maior que a VSE [(VCE tumor / VCE normal) $>$ (VSE tumor / VSE normal)], ou ainda um valor relativo de expressão T/N da VCE maior que a EC [(VCE tumor / VCE normal) $>$ (EC

tumor / EC normal)] avaliados pelo teste estatístico *TStudent*, sendo considerados valores estatisticamente significativos aqueles com $p < 0,05$. Para esses cálculos foram considerados os valores de expressão normalizados pelo gene normalizador de cada amostra tumoral dividido pela média dos respectivos valores das amostras normais. Os detalhes para essa análise serão apresentados nos resultados.

3.15 ANÁLISES CLÍNICAS E HISTOPATOLÓGICAS

Os dados coletados dos prontuários dos pacientes foram transcritos em ficha padronizada e digitados em banco de dados específico confeccionado no programa “ Excel 2003 – Microsoft Office”. As informações histopatológicas e dos marcadores moleculares também foram passadas para o mesmo banco de dados onde estavam registradas as características clínicas dos pacientes. Após conclusão do banco de dados com todas as informações, a análise estatística dos dados foi realizada utilizando o software estatístico SPSS 15.0 for Windows (*Statistical Package for Social Science*, SPSS 15.0).

A caracterização da população de estudo e do tratamento efetuado foi feita por meio de estatística descritiva, calculando-se, inicialmente, as frequências de todas as variáveis estudadas. Para associação dos valores de expressão obtidos através dos experimentos de RT-PCR quantitativo (qRT-PCR) com as variáveis clínicas e histopatológicas, os valores foram divididos em quartis, e a associação foi realizada através do teste estatístico

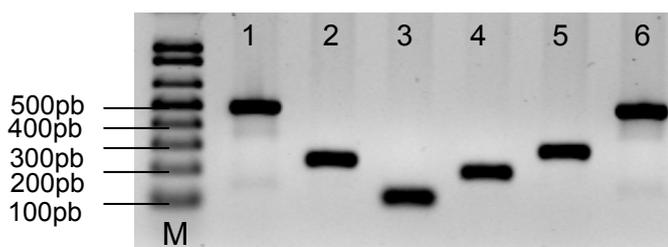
qui-quadrado de *Pearson* ou pelo teste exato de *Fisher* quando aplicável. As diferenças foram consideradas estatisticamente significativas quando o valor de $p < 0,05$.

4 RESULTADOS

4.1 CONSTRUÇÃO DA PLATAFORMA DE *MICROARRAY*

4.1.1 Extensão dos Fragmentos

Todos os 270 fragmentos puderam ser amplificados com tamanho esperado, representando uma aproveitamento de 100%. Inicialmente, obteve-se aproveitamento de 90% nas amplificações de fragmentos seguindo-se o protocolo de *touchdown* estabelecido. Os fragmentos que não estenderam nessas condições, foram submetidos a reações de PCR com pequenas variações de temperatura e concentração de cloreto de magnésio. Na Figura 9 estão representado alguns dos fragmentos amplificados.



Legenda: 1: exon 113282_9 (486pb), 2: exon 114644_12 (228pb), 3: exon 1152216_20 (100pb), 4: exon 115216_23 (190pb), 5: exon 115216_25 (274pb), 6: exon 115216_26 (485pb). Gel de agarose 1,0%. Foram aplicados 5,0 μ L de cada amostra. M: marcador de peso molecular 1Kb Plus.

Figura 9 - Representação de alguns produtos amplificados.

4.1.2 Clonagem, Seqüenciamento e Análise das Seqüências

Todos os fragmentos amplificados foram clonados e seqüenciados para confirmação de sua identidade. Os resultados de seqüenciamento mostraram que 95,2% das seqüências (257 de 270) imobilizadas apresentaram identidade confirmada.

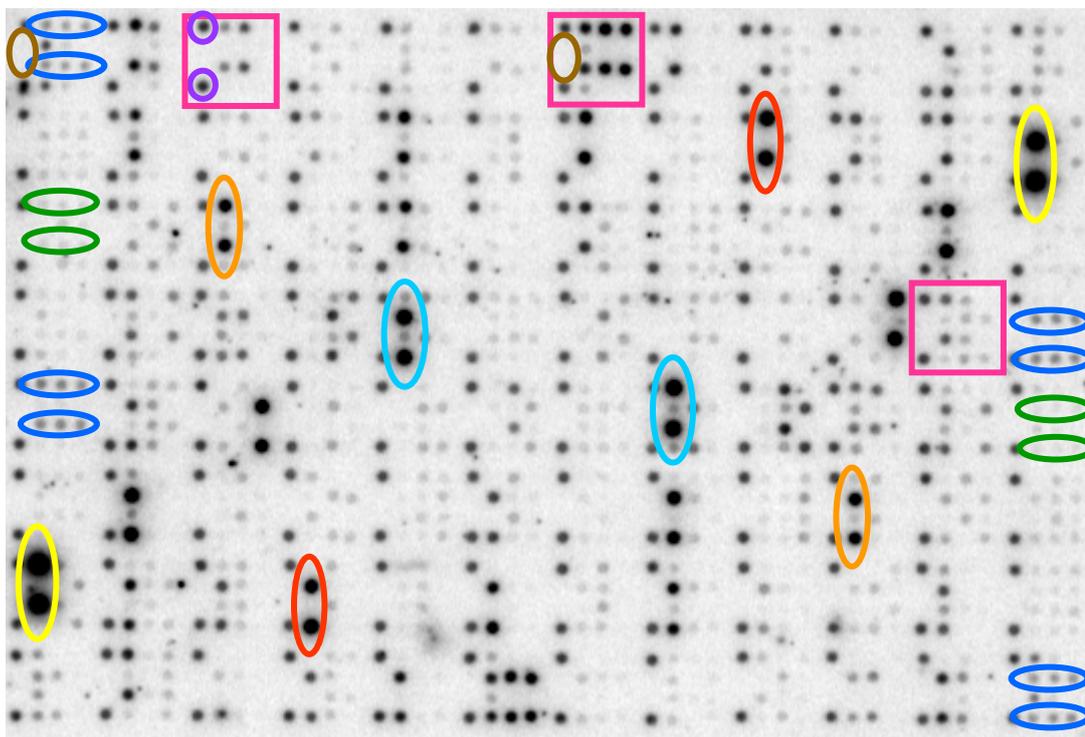
Foi confirmada também através de seqüenciamento a identidade de todos os 4 fragmentos amplificados correspondentes aos exons e íntrons do genes normalizadores *GAPDH* e *ACTB*.

Os exons que não tiveram confirmação de sua identidade estão listados a seguir: 118648_12, 120490_2, 154681_10, 4435_11, 4845_2, 232465_9, 282067_7, 9779_10, 50245_6, 156333_8, 89010_10, 230795_9 e 84250_12.

4.2 MARCAÇÃO E HIBRIDIZAÇÃO

Foram hibridizadas 27 amostras de tecido tumoral de mama, 09 amostras de tecido normal de mama e 06 amostras de linhagem celular de mama, sendo 04 tumorais e 02 normais.

Na Figura 10 está representada uma foto da membrana hibridizada correspondente à amostra tumoral BS-IDC-ECIIA-9T-46, para apresentar a localização dos controles positivos, negativos e as réplicas dos exons imobilizados.



Legenda: rosa: representação de um *subarray*; marrom: controles negativos - água milliQ; roxo: controles positivos - λ phago Q gene; azul: exons dos genes normalizadores GAPDH e ACTB; verde: íntrons dos genes normalizadores GAPDH e ACTB; laranja, amarelo, vermelho e azul turquesa: ilustração da distribuição das 4 replica tas de alguns exons na membrana AS01-2103, amostra BS-IDC-ECIIA-9T-46 (carcinoma ductal invasivo, estadio IIA, ERBB2 positivo e ER positivo, com 46 horas de exposição no *Phosphosimager*).

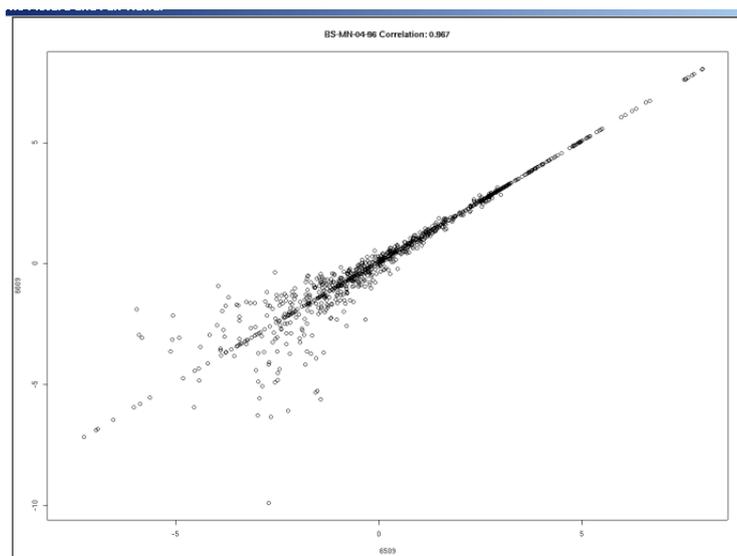
Figura 10 - Representação de membrana hibridizada com amostra de tecido tumoral de mama.

Como citado no material e métodos (item 3.2.7) cada membrana apresentou um total de 1.536 *spots*, que estão distribuídos em 96 *subarrays*, contendo cada um 16 *spots*: 12 correspondentes a 6 exons associados a tumor em duplicata, 02 controles positivos de marcação (λ Qgene) e 02 controles negativos (água). Cada um dos 270 exons associados a tumor foi imobilizado 4 vezes na mesma membrana, além de 24 fragmentos que

correspondem a genes normalizadores para controle de expressão positiva e 12 que correspondem a regiões intrônicas dos genes normalizadores para controle negativo de expressão. Assim, considerando que cada amostra foi hibridizada em replicata e cada membrana apresenta 4 replicas de cada exon, foram gerados 8 medidas de expressão para cada um dos exons.

4.3 CONTROLE DE QUALIDADE EXPERIMENTAL DAS MEMBRANAS

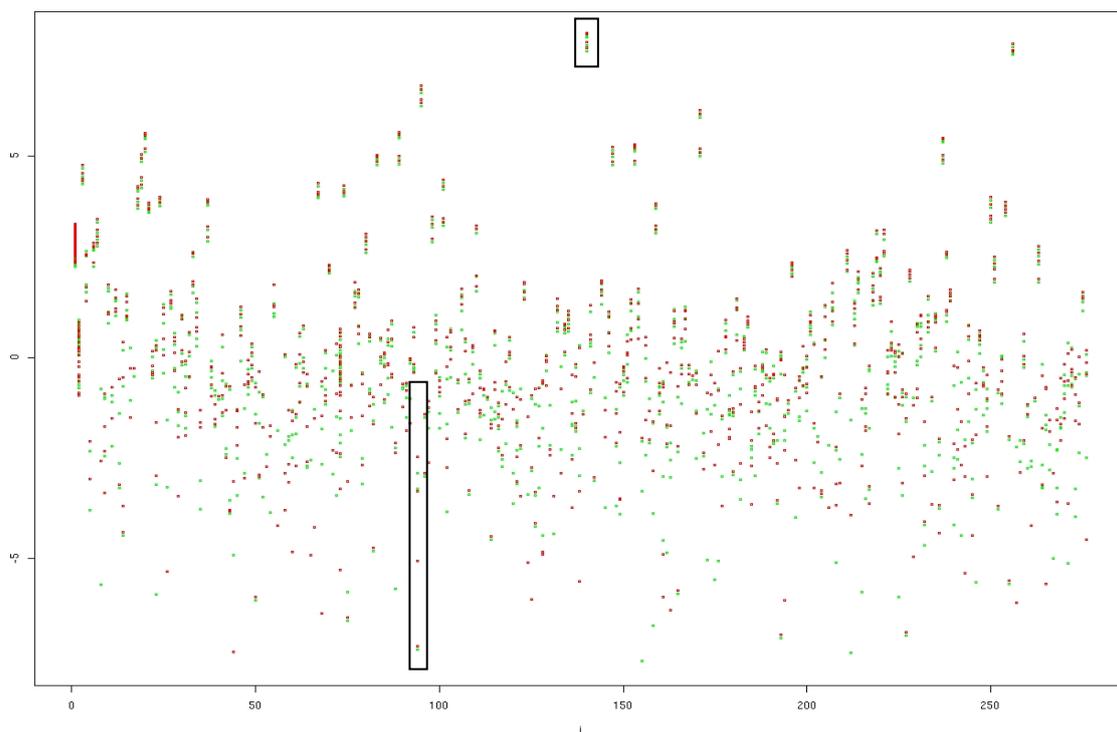
Para todas as membranas foi obtido o valor da correlação entre os valores de intensidade da membrana principal e sua réplica, os quais variaram entre 0,80 e 0,97, apresentando uma correlação média de 0,90 entre todos os experimentos. Pode ser observado na Figura 10 uma maior dispersão dos valores de intensidade baixos, o que é esperado, uma vez que o tipo de normalização utilizada não leva em conta as faixas de variação de intensidade de sinal. Isso foi verificado em todos os gráficos correspondentes de cada amostra.



Legenda: $R^2=0,967$ - amostra MN04. Na abscissa está representada a membrana principal e na ordenada a sua réplica.

Figura 11 - Gráfico de Correlação de *Pearson*.

Para se verificar a variabilidade entre as 8 réplicas de cada exon (4 exons da membrana principal e 4 exons da membrana controle) em cada experimento foi construído um gráfico de dispersão através da distribuição dos valores de intensidade após normalização para cada um dos 270 exons, os quais estão representados na abscissa da Figura 12. Como esperado, verificou-se uma maior variabilidade nos pontos de baixa intensidade, o que foi observado em todos os gráficos correspondentes de cada amostra.

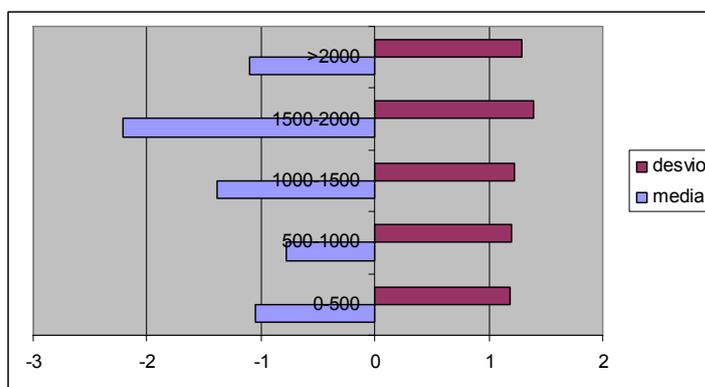


Legenda: pontos verdes: exons da membrana principal; vermelhos: exons da membrana réplica). Na abscissa estão representados os 270 exons e na ordenada o valor de intensidade normalizada de cada uma das 8 réplicas de cada exon.

Figura 12 - Gráfico de dispersão - amostra MN04.

Neste trabalho como a quantidade de RNA total obtido das amostras foi, na maioria das vezes, insuficiente para a realização dos experimentos, decidiu-se utilizar amostras de RNA amplificado. Para isso foi avaliada a correlação dos dados das intensidades médias de cada exon (8 medidas) obtidos de membranas hibridizadas em duplicatas com RNA total de HB4aC5.2 e RNA antisense amplificado (RNAa) do mesmo RNA total (dados não mostrados). O valor de correlação entre elas foi aceitável (correlação de *Pearson* 0,71). Dessa forma decidiu-se utilizar RNAa amplificado em 2 ciclos para todas as amostras utilizadas nesse trabalho.

Relatos da literatura mostram que a amplificação de RNA mensageiro não gera viéses na quantidade relativa dos transcritos (WANG et al. 2000, GOMES et al. 2003). No entanto, a metodologia resulta em menor representação da porção 5' dos transcritos, sendo que a recomendação no desenho das plataformas é que a sonda representativa de cada transcritos seja desenhada no máximo a 1.500 pb da região 3' (BRENTANI et al. 2005). Nesse trabalho, como os exons são distribuídos ao longo de todo o transcrito, apresentando diferentes distâncias da extremidade 3', variando de 110 a 3.338pb, nós resolvemos avaliar a intensidade média dos diferentes exons em relação a sua distância 3'. Os dados obtidos mostraram que o sinal de intensidade média não apresentou decréscimo inversamente proporcional à distância dos exons a extremidade 3' dos transcritos (Figura 13).



Legenda:

Distância da extremidade 3' (pb)	Número de exons	Média	Desvio
0-500	46	-1,04	1,19
500-1000	50	-0,78	1,2
1000-1500	61	-1,38	1,23
1500-2000	35	-2,21	1,40
>2000	36	-1,10	1,28

Figura 13 - Gráfico representativo da intensidade média dos exons em relação à extremidade 3'. Valores logados na base 2.

4.4 ANÁLISE DOS DADOS DE *MICROARRAY*

Para identificação dos exons superexpressos tumores de mama foram realizadas várias comparações no sentido de evitar resultados artefatuais e conseguir dados mais robustos. Essas comparações se apresentam da seguinte maneira:

- 1) **amostras tumorais x normais (TxN):** para essa comparação foram utilizadas 27 amostras de carcinoma ductal de mama como amostras tumorais e 5 amostras de tecido normal (sendo que foram consideradas como normais aquelas amostras que eram bordas de fibroadenoma);
- 2) **amostras tumorais x normais pareadas (PTxPN):** para essa comparação foram utilizadas 4 amostras pareadas, sendo considerado como morfologicamente normal o tecido adjacente (6N, MN08, MNP12, MN32) ao tecido tumoral (6T, 8T, 12T, 32T);
- 3) **linhagens tumorais x normais (LTxLN):** para essa comparação foram consideradas as linhagens MCF7, MDA-MB-231, MDA-MB-435S e HB4a C5.2 no grupo de amostras tumorais e as linhagens HB4a e MCF10A no grupo de amostras normais.

4.4.1 Resultados das Comparações

Foram selecionados os exons superexpressos em tumor de mama apenas por diferença de expressão baseadas em significância estatística (*T student*; $p < 0,05$). Foi analisada a representatividade dos exons associados a tumor a mama nos grupos de exons superexpressos de cada comparação nos experimentos de *microarray*. Os dados estão apresentados a seguir.

- 1) **Amostras tumorais x normais (27Tx5N):** 79 exons apresentaram-se superexpressos em amostras tumorais em relação à amostras normais, sendo que a variação no número de vezes (*fold*) no nível de intensidade entre as classes variou de 1,1 a 2,8. Dezoito dos 79 estavam entre os 75 exons associados a tumor de mama selecionados nas análises computacionais.
- 2) **amostras pareadas (04 amostras):** 195 exons apresentaram-se superexpressos em amostras tumorais em relação à amostras normais, sendo que a variação no número de vezes (*fold*) no nível de intensidade entre as classes variou de 1,3 a 5,9. Cinquenta e nove dos 195 estavam entre os 75 exons associados a tumor de mama selecionados nas análises computacionais.
- 3) **linhagens tumorais x linhagens normais (4LTx2LN):** 24 exons apresentarem-se superexpressos em linhagens tumorais em relação à linhagens normais, sendo que a variação no número de vezes (*fold*)

no nível de intensidade entre as classes variou de 1,4 a 4,1. Dezoito dos 24 estavam entre os 75 exons associados a tumor de mama selecionados nas análises computacionais.

4.4.2 Representatividade dos exons previamente selecionados pela análise computacional como associados a tumor de mama

Para verificar a significância estatística da representatividade dos exons previamente selecionados pela análise computacional como associados a tumor de mama em relação aos superexpressos em tumor em cada comparação, aplicou-se o teste estatístico Qui-quadrado (X^2), levando em consideração os 270 exons imobilizados na membrana. Essa análise mostrou que apenas na comparação linhagem tumoral contra linhagem normal (LTxLN) houve uma super representatividade de exons associados a mama ($p < 0,001$).

Os resultados dessas análises estão apresentados de forma resumida na Tabela 2.

Tabela 2 - Exons superexpressos em tumor em 3 comparações.

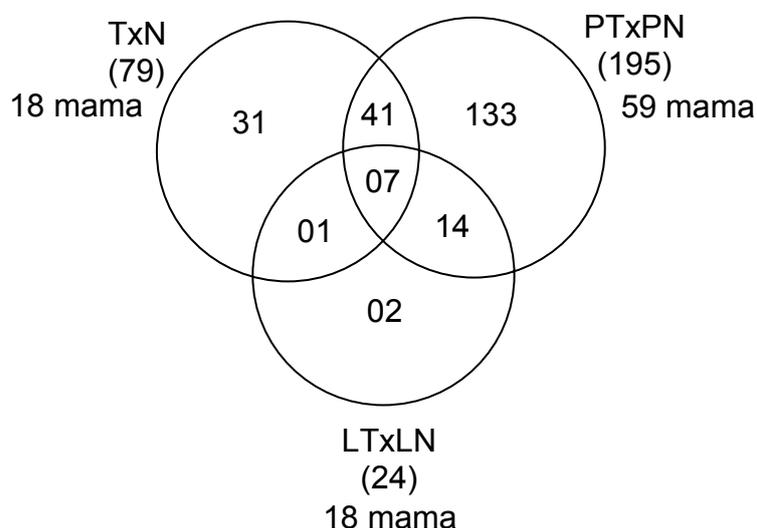
Comparação	Nº total de exons superexpressos*	Fold change	Nº exons de mama superexpressos*	X^2
TxN	79	1,1 a 2,8	18	$p=0,2$
PTxPN	195	1,3 a 5,9	59	$p=0,5$
LTxLN	24	1,4 a 4,1	18	$p < 0,001$

Legenda: Tumor x Normal (TxN), Pareadas (PTxPN) e Linhagem Tumor x Linhagem Normal (LTxLN), *considerando-se $p < 0,05$ e X^2 com grau de liberdade 1.

4.4.3 Seleção dos exons superexpressos para confirmação através de RT-PCR quantitativo (qRT-PCR)

Para seleção de exons para validação através de qRT-PCR, decidiu-se utilizar 2 critérios de análise, que estão descritos a seguir:

- 1) Exons superexpressos que foram comuns em mais de 1 comparação e apresentaram $fold > 3,0$. Para isso, foi construído inicialmente um diagrama de *Venn* (Figura 14).



Legenda: TxN: amostras tumorais x normais (79 exons superexpressos, sendo 18 identificados pela análise computacional como associados a tumor de mama), PTxPN: amostras pareadas tumorais x normais (195 exons superexpressos, sendo 59 associados a mama) e LTxLN: linhagens tumorais x normais (24 exons superexpressos, sendo 18 associados a mama). As intersecções mostram os exons comuns entre duas comparações (01 exon entre TxN e LTxLN; 14 exons entre LTxLN e PTxPN; 41 exons entre TxN e PTxPN), e entre as três comparações (07 exons).

Figura 14 - Diagrama de *Venn*, representando os exons que apareceram superexpressos nas comparações.

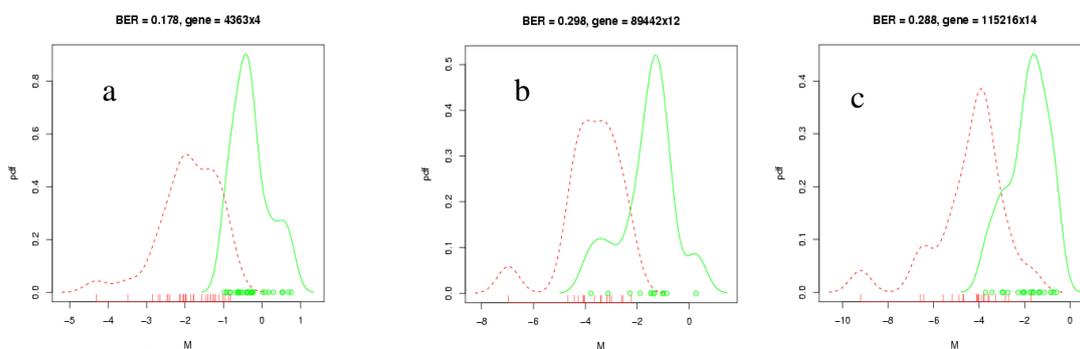
Essa análise mostrou que 56 exons diferencialmente expressos foram comuns entre 2 comparações (41 entre amostras e amostras pareadas; 14 entre amostras pareadas e linhagem; 01 entre amostras e linhagem), sendo que desses, 24 foram previamente selecionados pela análise computacional como associados a tumor de mama, constatando-se uma representatividade significativa de exons associados a mama nesse grupo ($p < 0,001$). Entre os 7 exons que foram comuns nas 3 comparações, 04 foram previamente selecionados pela análise computacional como associados a tumor de mama, o que também mostra, de acordo com o teste Qui-quadrado, uma representatividade significativa desse grupo ($p < 0,001$).

Considerando-se o critério de aparecer em pelo menos 2 comparações e apresentar *fold* maior que 3, 12 candidatos foram selecionados, sendo que 5 deles foram diferencialmente expressos nas 3 comparações (TxN, PTxPN e LTxLN). Os 7 exons restantes estavam diferencialmente expressos entre 2 comparações. Dos 12 exons selecionados, foi verificado que 8 foram previamente identificados pela análise computacional como associados a tumor de mama, mostrando uma representatividade significativa ($p < 0,001$).

2) O segundo critério considerou exons superexpressos que apresentaram $fold > 3$ e $BER < 0,3$. A ferramenta estatística *Bay Boots* (*bayesian bootstrap*) é baseada no BER (*Bayes error rate*), e considera diferencialmente expressos os valores de intensidade cuja comparação não apresente sobreposição de curvas maior que 30%

(BER<0,3). Essa ferramenta estatística foi aplicada em todos os exons diferencialmente expressos entre todas as comparações (TxN, PTxPN e LTxLN) com a finalidade de identificar os exons diferencialmente expressos com menor variabilidade de intensidade dentro das classes avaliadas.

Considerando-se esse segundo critério, somente na comparação PTxPN foram identificados 3 exons que apresentaram BER<0,3 (sobreposição de curvas de no máximo 30%). Sendo eles: 4363_4 (BER=0,178; *fold* 3,0), 89442_12 (BER=0,298; *fold* 4,3) e 115216_14 (BER=0,288; *fold* 5,9) (Figuras 14 a, b e c). Os 3 exons foram selecionados, uma vez que apresentaram valor de BER<0,3 em relação aos dados de intensidade do *array* e apresentaram *fold* maior que 3.



Legenda: Na abscissa estão distribuídos os pontos referentes a todos os valores de intensidade dos respectivos exons nas duas classes estudadas: tumoral (verde) e normal (vermelho) e o valor de sobreposição entre as classes (BER). Na ordenada estão representados os valores de densidade de probabilidade da população.

Figura 15 - Bay Boots. a: exon 4363_4; b: exon 89442_12 e c: exon 115216_14.

Como o exon 89442_12 foi identificado nos 2 critérios, o número total de exons selecionados como candidatos pertencentes à variantes de *splicing* associadas a câncer de mama foi 14.

4.4.4 Análise de expressão dos genes correspondentes a cada exon em bancos de SAGE

Considerando-se que a superexpressão dos 14 exons pode ser devido à superexpressão do gene como um todo, e não da variante de *splicing*, foi investigada a expressão dos cDNAs *full lengths* correspondentes aos genes em bancos de dados de SAGE. Essa análise identificou 171 *tags* únicas dos 270 analisados, que tiveram informação de expressão nas bibliotecas de SAGE normais e tumorais. Dessa forma foi comparada a frequência da *tag* correspondente ao gene candidato nos dois tipos de bibliotecas (tecido normal e tumoral). Foram considerados como diferencialmente expressos, aqueles genes cujas *tags* apresentaram $BER < 0,3$. Essa análise mostrou que, dos 171, apenas 14 deles eram superexpressos em bibliotecas tumorais em relação às bibliotecas normais, o que sugere que os 157 restantes são candidatos à superexpressão da variante.

Dos 14 candidatos previamente selecionados, para 4 deles não foi possível obter informação de SAGE. Para os 10 candidatos com informação, nenhum apresentou $BER < 0,3$ nas análises das bibliotecas em bancos de SAGE, o que significa que a superexpressão do exon provavelmente é

devido à superexpressão da variante e não do gene, tornando esses candidatos mais promissores para validação por qRT-PCR.

Os resultados referentes aos 14 exons selecionados estão apresentados de forma resumida na Tabela 3.

Tabela 3 - Exons selecionados para marcador molecular entre as comparações TxN, PTxPN e LTxLN.

Exon	Gene Symbol	mRNA Accession Number	Nome	Fold em cada comparação	SAGE-BER
50321_8	BAP1	NM_004656	BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase)	TxN (1,5), <u>PTxPN (4,8)</u> e LTxLN (2,2)	0,9
234267_14	LIP8	NM_053051	LYST-interacting protein LIP8	TxN (2,3), PTxPN (2,7) e <u>LTxLN (4,0)</u>	0,4
116016_5	MK-STYX	NM_016086	map kinase phosphatase-like protein MK-STYX	TxN (1,8), <u>PTxPN (3,9)</u> e LTxLN (2,0)	0,3
15801_14	BRRN1	NM_015341	barren homolog (Drosophila)	TxN (1,8), <u>PTxPN (3,6)</u> e LTxLN (1,7)	Sem informação
230795_7	MBTPS1	NM_003791	membrane-bound transcription factor protease, site 1	TxN (2,2), <u>PTxPN (3,0)</u> e LTxLN (2,7)	0,5
89442_12	KIAA0676	NM_198868	KIAA0676 protein	TxN (1,8) e <u>PTxPN (4,3)</u>	0,9
274389_10	PRKD2	AK095884	protein kinase D2	TxN (2,3) e <u>PTxPN (4,5)</u>	0,4
393_5	FLJ10315	AK091861	hypothetical protein FLJ10315	TxN (2,8) e <u>PTxPN (4,4)</u>	0,9
281313_8	ASC1p100	AY013289	ASC-1 complex subunit P100	TxN (2,3) e <u>PTxPN (3,0)</u>	0,6
9779_9	FLJ12528	BC007824	threonyl-tRNA synthetase	PTxPN (3,1) e <u>LTxLN (4,1)</u>	Sem informação
310552_20	GRIA 3	NM_181894	glutamate receptor, ionotropic, AMPA3	PTxPN (2,6) e <u>LTxLN (3,2)</u>	Sem informação
267445_3	TRIM37	NM_001005207	tripartite motif-containing 37	PTxPN (2,1) e <u>LTxLN (3,0)</u>	0,6
4363_4	PPP1R8	NM_138558	protein phosphatase 1, regulatory (inhibitor) subunit 8	<u>PTxPN (3,0)</u>	0,6
115216_14	SND1	NM_014390	staphylococcal nuclease domain containing 1	<u>PTxPN (5,9)</u>	Sem informação

Legenda: Negrito: exons previamente selecionados pela análise computacional como associados a tumor de mama. Sublinhado: comparação com fold \geq 3,0.

4.5 CONFIRMAÇÃO DA SUPEREXPRESSÃO DO EXON ATRAVÉS DE qRT-PCR

Para confirmação da superexpressão do exon através de qRT-PCR foram desenhados iniciadores em 2 exons distintos. Um iniciador no exon superexpresso analisado no *microarray* e outro no exon flangeador a 3'. A expressão obtida se refere à variante que contém o exon (VCE). Para essa avaliação foram utilizadas as mesmas amostras do grupo inicial.

4.5.1 Avaliação da VCE através de qRT-PCR

A Desenho dos iniciadores

Como citado anteriormente, foram desenhados iniciadores *forward* e *reverse* para a VCE de cada um dos 14 genes selecionados, sendo os iniciadores *forward* (F-CE) desenhados no exon 2, ou superexpresso, e os iniciadores *reverse* (R) desenhados no exon 3, flangeados a 3' (Tabela 4). Esses iniciadores variaram de 12 a 18 mers em tamanho e apresentaram T_M de 54,8°C a 62,2°C de acordo com o programa *OLIGOTECH* versão 1.00 (Copyright© 1995).

Tabela 4 - Iniciadores *forward* e *reverse* para cada gene. F-CE: *forward*, desenhado inteiro no exon 2; R: *reverse*, desenhado no exon 3.

NOME	SEQÜÊNCIA	AMPLICON (F+R) pb
3BAP1-F-VCE	GTACTAGAGGCTCTG	100
3BAP1-R	CTGACTTGGACTCC	
3LIP8-F-VCE	CACAAGGCAGTGG	104
3LIP8-R	CTGGTAGAGTCGC	
3MK-STYX-F-VCE	GATCATCTGGATGC	106
3MK-STYX-R	CTTGGGGTCACAG	
3TRIM37-F-VCE	GGGTTTCACGACTC	122
3TRIM37-R	CATCTGTATTGAAGCTG	
3FLJ103-15-F-VCE	CTTCAGGACCTACTC	121
3FLJ103-15-R	GGAGGCCATGTGG	
3BRRN1-F-VCE	GGCCATGTTGCTG	121
3BRRN1-R	CTTCGGAGACATTG	
3MBTPS1-F-VCE	GATGGCCTGTATG	118
3MBTPS1-R	GGTCCTTGAAAGTC	
3KIAA0696-F-VCE	GACACGGCCAAG	66
3KIAA0696-R	CAGCTCTTCAATGG	
3PRKD2-F-VCE	CAGTTTGGAGTGGTC	87
3PRKD2-R	GGTAGGGAAGCGC	
3FLJ125-28-F-VCE	CAAGGGTGTATAATGCAC	88
3FLJ125-28-R	GAAAACAGTGTGGGAG	
3ASC1p100-F-VCE	CGGAGGAAGAGG	79
3ASC1p100-R	GAGATGAGAGAGTC	
3GRIA3-F-VCE	GTTCCACCAGGTGG	111
3GRIA3-R	CTCAAGAAGGGAGG	
3PPPIR8-F-VCE	GGCTTCCAGAGG	79
3PPPIR8-R	GTAAGGGTAGAAATC	
3SDN1-F-VCE	GAGACAGTGCCTG	114
3SDN1-R	CTGCCGGTATCTG	

Na seqüência representada na Figura 16 está ilustrado o desenho dos 2 iniciadores para a VCE do gene TRIM37.

Variante com o exon (VCE):

```

GGCTAATGCTAAAGGAGGTCATCTGGAAGGACTGCAGATGACTGATTTGGAAAATAATTC
TGAAACTGGAGAGTTACAGCCTGTACTACCTGAAGGAGCTTCAGCTGCCCTGAAGAAG
GAATGAGTAGCGACAGTGACATTGAATGTGACACTGAGAATGAGGAGCAGGAAGAGCAT
ACCAGTGTGGGCGGGTTTCACGACTCCTTCATGGTCATGACACAGCCCCGGATGAAGA
TACACATTCCAGTTTTCTGATGGTGAACAAATAGGCCCTGAAGATCTCAGCTTCAATAC
AGATGAAAATAGTGGAAG
  
```

Legenda: 1) VCE - variante com a presença do exon 2 (associado a tumor); seqüência em vermelho - exon 2; seqüências em azul: à esquerda do exon 2 = exon 1, e à direita do exon 2 = exon 3.

GGGTTTCACGACTC: 3TRIM37-F-VCE; CAGCTTCAATACAGATG:3TRIM37-R.

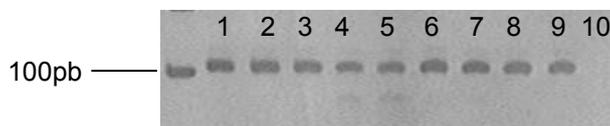
Figura 16 - Representação esquemática do desenho dos iniciadores nas seqüências da VCE do gene TRIM37.

B Síntese de cDNA das amostras e padronização das reações de RT-PCR para cada gene

Para confirmação da superexpressão da VCE através de qRT-PCR foram utilizadas nas reações as mesmas amostras utilizadas na primeira etapa do projeto (experimentos de *microarray*). Das 27 amostras de tecido tumoral de mama usadas nos experimentos de *microarray*, foram utilizadas 19 amostras para os experimentos de qRT-PCR, uma vez que para 8 amostras não havia RNA total ou cDNA disponível. Para essa validação foram incluídas 2 amostras normais bordas de fibroadenoma e duas amostras de glândula mamária normal (*clontech*), uma vez que para as amostras normais iniciais não havia RNA total ou cDNA disponível.

A qualidade da síntese do cDNA das 19 amostras tumorais e 2 normais (borda de fibroadenoma), assim como das 2 amostras normais da *clontech* foi verificada através de RT-PCR para o gene normalizador

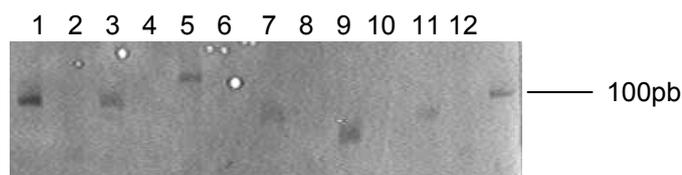
GAPDH₁₁₀₁ (102pb), conforme representado na Figura 17. Todas as amostras apresentaram qualidade satisfatória.



Legenda: Gel de poliacrilamida 8%. (M) Marcador 100pb (*Invitrogen Life Technologies*TM); 1 e 2: amostras normais; 3 a 9: amostras tumorais; 10: NO.

Figura 17 - Representação dos produtos amplificados por RT-PCR para o gene GAPDH₁₁₀₁, confirmando o tamanho esperado de 102pb.

Para 13 dos 14 genes conseguiu-se o estabelecimento das condições da reação de RT-PCR para amplificação da VCE. A temperatura de pareamento ideal para todos os pares de iniciadores variou de 54°C a 58°C. Foi utilizado cDNA de linhagem de mama normal HB4a como molde. Nessas condições os fragmentos foram amplificados com sucesso, e se mostraram específicos (Figura 18).



Legenda: Gel de poliacrilamida 8%. (M) Marcador 100pb (*Invitrogen Life Technologies*TM); 1- LIP8ce; 2- NO; 3- BAP1ce; 4- NO; 5- BRRN1ce; 6- NO; 7- ASC1p100ce; 8- NO; 9- KIAA0696ce; 10- NO ; 11- PPP1R8ce ; 12- NO.

Figura 18 - Representação das VCE amplificadas com a enzima *Taq Gold* (*Applied Biosystems*) na temperatura de 54°C.

Desse modo 13 variantes com o exon dos genes selecionados seguiram para os experimentos de qRT-PCR, exceto o gene FLJ103-15, para o qual não se obteve amplificação desta variante.

C RT-PCR quantitativo (qRT-PCR)

✓ Escolha dos genes normalizadores utilizando o programa *GENORM*

Antes de iniciar os experimentos de qRT-PCR com as VCE de 13 dos 14 genes selecionados, foi escolhido através do programa geNorm (VANDESOMPELE et al. 2002) o melhor gene normalizador. O programa geNorm calcula automaticamente uma medida de estabilidade de expressão, denominada M , para todos os genes normalizadores no grupo de amostras utilizadas. O programa permite a eliminação do gene normalizador com a pior pontuação, ou seja, aquele com o mais alto valor M , que representa a maior variabilidade de expressão no grupo de amostras avaliado, e recalcula um novo valor M para os genes restantes. O valor M sugerido pelo programa é de 1,5, abaixo do qual se tem os genes normalizadores mais estáveis.

Na primeira análise do geNorm foram incluídos os cinco genes normalizadores GAPDH, ACTB, HPRT1, BCR e RPLP0, conforme resultado mostrado na Figura 19.



Change Data	GAPDH	BACT	BCR	HPRT1	huPO
1T	0,12777012	0,687266339	0,619864995	0,198470183	0,767258248
2T	0,312563528	0,824747748	0,648220202	0,622001422	1,00E+00
6T	0,179005885	0,48719019	0,561000716	0,666310543	1,08E+09
7T	0,184001529	0,821914937	0,282881538	0,289777367	0,7968535
9T	0,193081715	0,626296389	0,706446368	0,282881538	0,713776084
11T	0,29786439	0,777890786	0,704019893	0,485516809	0,936718198
14T	0,282881538	0,641563674	0,48053107	0,637164007	0,807896164
19T	0,13546647	0,462684078	0,341813492	0,143626416	0,50772637
20T	0,161450483	0,392247167	0,447035385	0,592750136	0,410190187
21T	0,140691783	0,54016507	0,780571863	0,301992145	0,889600415
22T	0,317987192	0,748999811	0,639360056	0,609292434	0,696790362
27T	0,319083166	0,483849174	0,582640051	0,52549958	0,47396297
29T	0,138292111	0,428953994	1,66E+09	0,477235721	0,654945795
30T	0,248211855	0,401809013	1,64E+09	0,718704763	0,516536537
32T	0,174145873	0,908156236	0,497352296	0,652696213	0,576656961
37T	0,118048697	0,493941591	0,247359307	0,205417729	0,377679052
MN02	0,179622846	0,959552664	0,129095795	0,153329375	1,09E+09
M < 1,5	5,501	5,599	11,882	5,561	11,626

Legenda: (verde) gene normalizador mais estável, (vermelho) gene normalizador menos estável, M: medida de estabilidade gênica. Os valores apresentados para cada amostra em cada gene correspondem à eficiência de cada gene normalizador elevada ao ΔC_T de cada amostra para o gene normalizador avaliado.

Figura 19 - Análise pelo programa geNorm para a identificação dos genes mais estáveis para o conjunto de amostras utilizadas no presente estudo.

Os genes BCR e RPLP0 apresentaram valores muito elevados de *M*, ou seja, mostraram-se os menos estáveis nas amostras testadas, como pode ser comprovado pelo gráfico da Figura 20, que é disponibilizado pelo geNorm e demonstra a maior instabilidade de um dado gene em relação aos demais.



Legenda: Os genes que estão localizados à esquerda do gráfico apresentam menor estabilidade comparado aos genes que se localizam à direita do gráfico. Assim, os genes mais estáveis determinados pelo geNorm são: GAPDH e ACTB e o menos estável é BCR.

Figura 20 - Gráfico da média dos valores de estabilidade de expressão dos genes GAPDH, ACTB, HPRT, RPLP0 e BCR.

Essa análise mostrou que os genes que receberam os menores valores de M , sendo, portanto, os que menos variam a expressão no grupo de amostras desse estudo, foram: GAPDH, ACTB e HPRT1. Sendo assim, foi realizada uma nova análise na qual os genes BCR e RPLP0 foram excluídos. O resultado dessa nova análise pode ser visualizado na Figura 21, e mostra valores de M menores que 1,5 para os 3 genes.

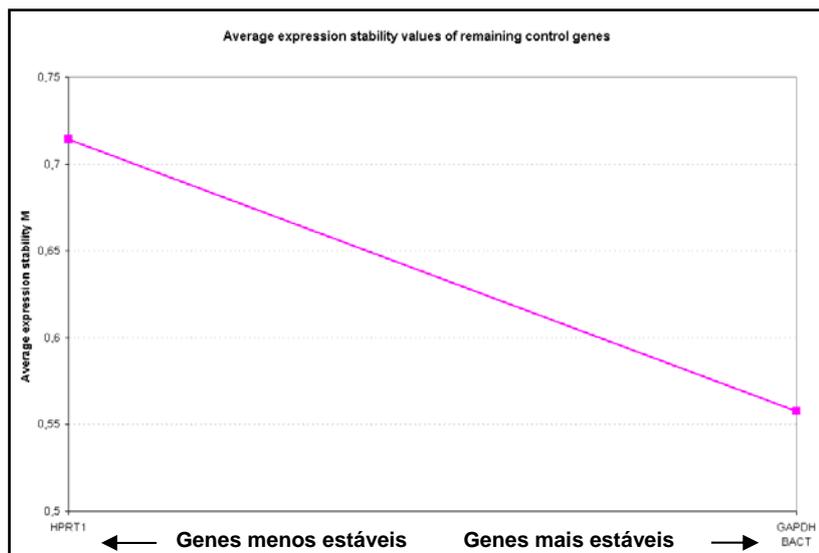


Change Data	GAPDH	BACT	HPRT1
1T	0,12777012	0,687266339	0,198470183
2T	0,312563528	0,824747748	0,622001422
6T	0,179005885	0,48719019	0,666310543
7T	0,184001529	0,821914937	0,289777367
9T	0,193081715	0,626296389	0,282881538
11T	0,29786439	0,777890786	0,485516809
14T	0,282881538	0,641563674	0,637164007
19T	0,13546647	0,462684078	0,143626416
20T	0,161450483	0,392247167	0,592750136
21T	0,140691783	0,54016507	0,301992145
22T	0,317987192	0,748999811	0,609292434
27T	0,319083166	0,483849174	0,52549958
29T	0,138292111	0,428953994	0,477235721
30T	0,248211855	0,401809013	0,718704763
32T	0,174145873	0,908156236	0,652696213
37T	0,118048697	0,493941591	0,205417729
MN02	0,179622846	0,959552664	0,153329375
M < 1,5	0,594	0,756	0,792

Legenda: (verde) gene normalizador mais estável, (vermelho) gene normalizador menos estável, M: medida de estabilidade gênica. Os valores apresentados para cada amostra em cada gene correspondem à eficiência de cada gene normalizador elevada ao ΔCT de cada amostra.

Figura 21 - Análise pelo programa geNorm dos normalizadores para a identificação dos genes mais estáveis no conjunto de amostras utilizadas nesse estudo.

Essa segunda análise apontou o gene GAPDH como o mais estável no conjunto de amostras testadas, e o gene HPRT1 como o menos estável com base nos valores de *M* calculados pelo programa (Figura 22).



Legenda: Os genes que estão localizados à esquerda do gráfico apresentam menor estabilidade comparado aos genes que se localizam à direita do gráfico. Assim, os genes mais estáveis determinados pelo geNorm são: GAPDH e ACTB.

Figura 22 - Gráfico da média dos valores de estabilidade de expressão dos genes GAPDH, ACTB e HPRT.

Dentre os quatro genes normalizadores testados nas amostras deste trabalho, os genes de menor variação, segundo os dados gerados pela ferramenta geNorm (Visual Basic Application - VBA) foram GAPDH e a ACTB. O gene BCRP apresentou maior variabilidade de expressão entre as amostras. Além de determinar o gene mais estável, o geNorm calculou a variação sistemática através da análise de variação pareada (V) em que valores de V abaixo de 0,15 indicam que a inclusão de um gene normalizador adicional não é necessário. Neste cálculo o valor de $V_{2/3}$ correspondeu a 0,001 ($V_{2/3} < 0,15$), indicando que não haveria diferença real na utilização de dois ou três genes normalizadores. Desta forma, os dados de expressão diferencial de qRT-PCR para cada transcrito de

interesse foi realizada individualmente relativa ao gene apontado pelo programa como o mais estável: GAPDH.

✓ **Padronização dos experimentos de qRT-PCR**

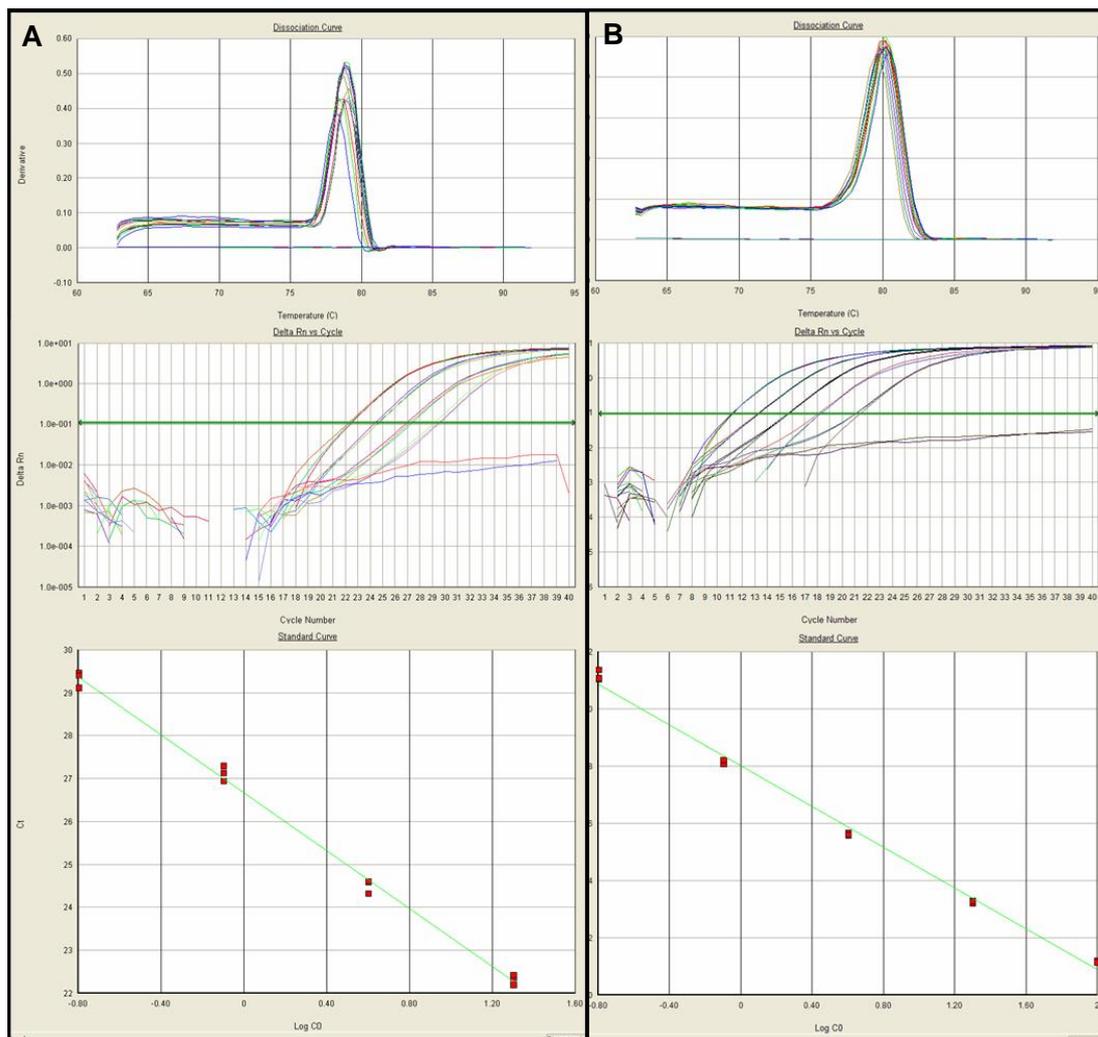
Através de um ensaio de padronização, foram estabelecidas a partir da análise dos *C_t*s, as quantidades de cDNA e iniciadores ideais para cada gene. Em relação à concentração de cDNA foi selecionada preferencialmente a menor concentração na qual houve extensão do fragmento a fim de economizar amostra. A concentração de iniciadores escolhida foi a menor concentração a partir do momento em que um aumento na quantidade de iniciador não acarretou em uma diferença significativa do ciclo onde a fluorescência é captada (*C_t*). As quantidades de iniciadores e cDNA estabelecidas para todos os genes foi de 800nM e 40ng respectivamente. A temperatura de pareamento dos iniciadores variou de 54 °C a 60°C.

✓ **Eficiência de Amplificação dos Iniciadores**

Os valores das eficiências foram calculados utilizando o coeficiente angular da equação da reta da curva padrão (*slope*). O cálculo da eficiência foi realizado para os genes de interesse, além do gene normalizador GAPDH. As reações foram feitas em triplicatas, sendo que triplicatas com valores de desvio padrão acima de 0,6 foram desconsideradas. Portanto, nos casos onde as triplicatas apresentavam valores de desvio padrão maiores que o estabelecido, o ponto correspondente a essas triplicatas foi

desconsiderado e a curva de diluição foi calculada a partir dos 3 ou 4 pontos restantes. A eficiência variou de 79 a 100% e o valor mínimo do coeficiente de regressão da reta (R^2) obtido foi de 0,97.

Na Figura 23A pode-se visualizar 3 gráficos: o primeiro (superior) referente à curva de dissociação gerada para o fragmento correspondente ao gene GAPDH₁₁₀₁, o segundo (meio) representando os CTs das triplicatas de diferentes diluições do cDNA molde de HB4a e o terceiro (inferior) gerado para o cálculo da eficiência de amplificação do gene GAPDH₁₁₀₁. Na Figura 23B pode-se visualizar também 3 gráficos: o primeiro (superior) referente à curva de dissociação gerada para o fragmento correspondente à VCE do gene TRIM37, o segundo (meio) representando os CTs das triplicatas de diferentes diluições do cDNA molde de HB4a e o terceiro (inferior) gerado para o cálculo da eficiência de amplificação da VCE do gene TRIM37.



Legenda: (A) gene GAPDH1101; (B) VCE do gene TRIM37. Superior: especificidade da reação de amplificação avaliada por meio do perfil da curva de dissociação do fragmento amplificado; meio: relação entre os CTs das triplicatas de diferentes diluições do cDNA molde de HB4a e *threshold* estabelecido (verde); inferior: eficiência da reação de amplificação calculada utilizando diluições seriadas do cDNA molde de HB4a. Após a reação, o programa do aparelho constrói uma reta na qual o valor do CT foi analisado em função do logaritmo da diluição. A inclinação da reta foi utilizada para calcular a eficiência de amplificação segundo a fórmula $E = 10^{(-1/slope)}$, em que o *slope* corresponde à inclinação da reta.

Figura 23 - Gráficos representando à curva de dissociação, os CTs das triplicatas e a curva padrão para o gene GAPDH e VCE do gene TRIM37 respectivamente.

Os iniciadores utilizados não tiveram o mesmo desempenho durante as reações de qRT-PCR, logo, as reações não apresentaram a mesma eficiência de amplificação, porém, estavam dentro do intervalo de confiança acima de 97% ($R^2 > 0,7$). A eficiência de amplificação foi calculada usando o *slope* da curva padrão de cada gene. A eficiência média obtida foi de 95%. Os valores de *slope* e da eficiência de cada gene estão representados na Tabela 5.

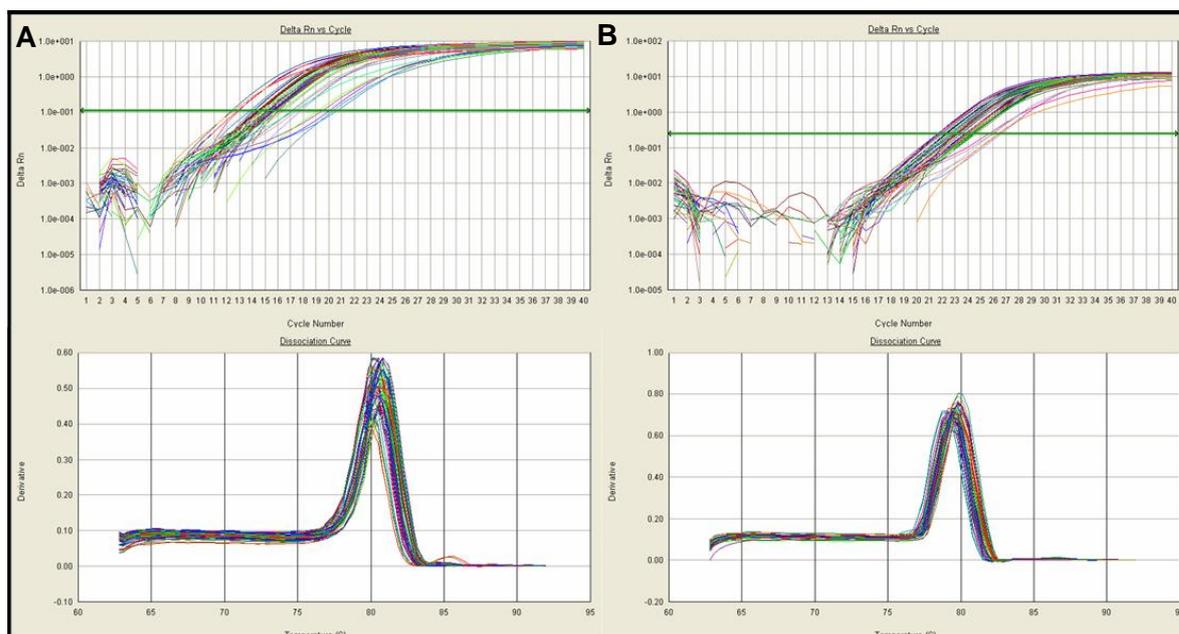
Tabela 5 - Genes e seus respectivos valores de *slope*, eficiência e coeficiente de regressão (R^2).

GENE	Slope	Eficiência	Eficiência (%)	R^2
PRKD2ce	-3,86	0,815797	81	0,9929
LIP8ce	-3,5	0,930698	93	0,9966
SND1ce	-3,95	0,791284	79	0,9821
MBTPS1ce	-3,66	0,875962	87	0,9975
TRIM37ce	-3,37	0,980322	98	0,9961
MK-STYXce	-3,33	0,996642	100	0,9701
BRRN1ce	-3,48	0,938011	94	0,9937
PPP1R8ce	-3,27	1,022136	100	0,9985
BAP1ce	-3,44	0,952979	95	0,9971
KIAA0696ce	-3,24	1,035363	100	0,9964
ASC1p100ce	-3,46	0,945438	94	0,9814
FLJ125-28ce	-3,34	0,992513	99	0,9788
GAPDH1101	-3,36	0,984354	98	0,9899

✓ Validação das VCE Superexpressas através de qRT-PCR

Depois do estabelecimento dos valores de eficiência de amplificação de 12 VCE, esses foram submetidos às reações de qRT-PCR usando as amostras de pacientes do grupo inicial. Para o gene GRIA3 os resultados dos experimentos, tanto de eficiência, quanto de amplificação das amostras,

não foram satisfatórios, uma vez que o gráfico da curva de dissociação apresentou diversos picos, resultantes de provável não especificidade da reação. Os gráficos de amplificação das amostras e da curva de dissociação para os genes GAPDH e da VCE do gene TRIM37 estão representados na Figura 24, os demais estão apresentados no Anexo11.



Legenda: qRT-PCR utilizando amostras tumorais e normais. (A) gene GAPDH₁₁₀₁; (B) VCE do gene TRIM37. Superior: CTs das duplicatas de diferentes amostras e *threshold* estabelecido (verde); inferior: especificidade da reação de amplificação avaliada por meio do perfil da curva de dissociação do fragmento amplificado.

Figura 24 - Gráficos de amplificação das amostras e a curva de dissociação para os genes GAPDH e da VCE do gene TRIM37 respectivamente.

Assim, após a realização dos experimentos de qRT-PCR foram feitas análises da expressão através da razão (*fold change*) entre as médias de expressão das amostras tumorais e das amostras normais (VCE tumor / VCE normal), sendo considerados como critério de validação valores de *fold*

≥ 3 . Seguindo-se esse critério, 3 genes (25%) mostraram superexpressão da VCE nas amostras tumorais em relação às normais, sendo eles: TRIM37 (5,69), MK-STYX (3,29) e BRRN1 (10,17). Os resumos dos dados de expressão relativa de todos os genes estão representados de forma resumida na Tabela 6.

Tabela 6 - Resumo dos resultados de expressão relativa (*fold change*) da média das 19 amostras tumorais em relação à média das amostras normais das variantes com o exon (VCE) para os 12 genes avaliados.

GENE	<i>Fold change</i> VCE tumor / VCE normal
BAP1	1,56
MK-STYX	3,29
BRRN1	10,17
MBTPS1	-10,68
KIAA0676	1,61
PRKD2	2,53
FLJ12528	1,35
ASC1P100	-1,20
TRIM37	5,69
PPP1R8	-1,46
SND1	1,59
LIP8	2,92

Legenda: Negrito - genes cuja relação VCE tumor / VCE normal apresentou valor ≥ 3 .

4.6 CONFIRMAÇÃO DA SUPEREXPRESSÃO DA VCE E NÃO DO GENE COMO UM TODO

Para testar se a VCE é realmente superexpressa nas amostras tumorais de mama e não o gene como um todo, foram adotados 2 critérios:

- 1) *Fold-change* ≥ 3 , obtido através da média do nível de expressão da variante que contém o exon (VCE) nas amostras tumorais em relação

à média do nível de expressão da VCE nas amostras normais [(VCE tumor / VCE normal) \geq 3];

- 2) Valor relativo de expressão da variante que contém o exon (VCE) no tumor em relação ao normal seja significativamente maior ($p < 0,05$) que o valor relativo de expressão do gene como um todo (EC), definido aqui como expressão de variantes que possuem exons constitutivos, na comparação (T/N) [(VCE tumor / VCE normal) $>$ (EC tumor / EC normal)].

As 3 VCE (genes TRIM37, MK-STYX e BRRN1) preencheram o critério 1, sendo validadas por qRT-PCR como superexpressas em câncer de mama. Assim, inicialmente pretendia-se confirmar a superexpressão da VCE e não do gene como um todo através da avaliação da expressão da variante que não continha o exon superexpresso (associado a tumor), a qual foi denominada VSE. Várias análises foram realizadas, e alguns problemas encontrados, que serão apresentados a seguir. Além disso, algumas atividades foram realizadas em paralelo às reações de qRT-PCR da VCE (critério 1), sem esperar a conclusão de cada etapa para passar para a seguinte que definiria quais variantes seriam superexpressas. Desse modo, algumas atividades, como o estudo das 2 variantes (VCE e VSE) em bancos de dados e o seu seqüenciamento, assim como a avaliação da especificidade da VSE, foram realizados simultaneamente. Dessa forma, para os 3 genes validados, a expressão da variante que não contém o exon (VSE) foi avaliada no mesmo grupo de amostras.

Para amplificação das duas variantes (VCE e VSE), foram desenhados iniciadores nos exons flangeadores (exon1 e exon3). Assim, amplificou-se as duas variantes numa única reação de PCR a partir de iniciadores *forward* desenhados no exon 1, e *reverse* desenhados no exon 3 (ambos flanqueiam o exon associado a tumor) (Figura 25), utilizando cDNA da linhagem de mama normal HB4a (STAMPS et al. 1994) como molde. Assim, *amplicons* de cada variante, sendo um maior (da variante que contém o exon 2 associado a tumor) e outro menor (da variante sem o exon 2), foram isolados a partir de gel de agarose 1%, purificados e seqüenciados. Esse procedimento foi utilizado com a finalidade de mapear o sítio de junção exon1-exon3 para cada gene, confirmando a presença do transcrito sem o exon 2 (VSE).

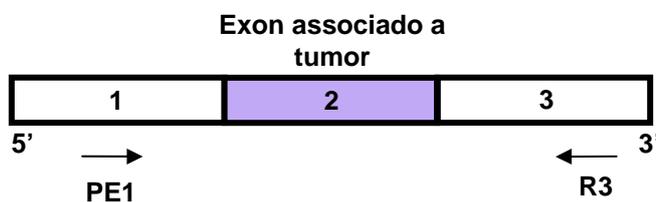
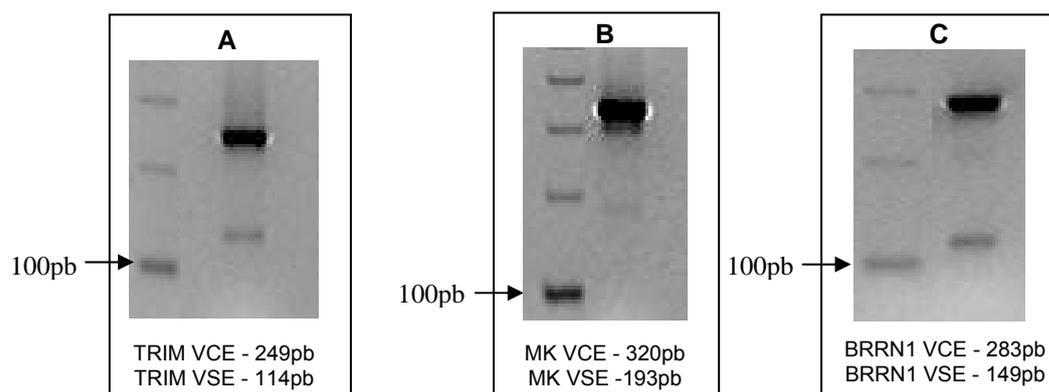


Figura 25 - Representação esquemática do pareamento dos iniciadores.

Para os genes TRIM37, MK-STYX e BRRN1 conseguiu-se amplificar as duas variantes utilizando o par de iniciadores PE1+R3, as quais apresentaram tamanho esperado (Figura 26), confirmando a existência da VSE.



Legenda: A: TRIM37; B: MK-STYX; C: BRRN1

Figura 26 - Representação dos produtos amplificados utilizando iniciadores no exon 1 e 3. Em todos os genes houve amplificação da variante VCE, mais forte, e VSE, mais fraca.

4.6.1 Identificação da seqüência correspondente à junção exon1-exon3

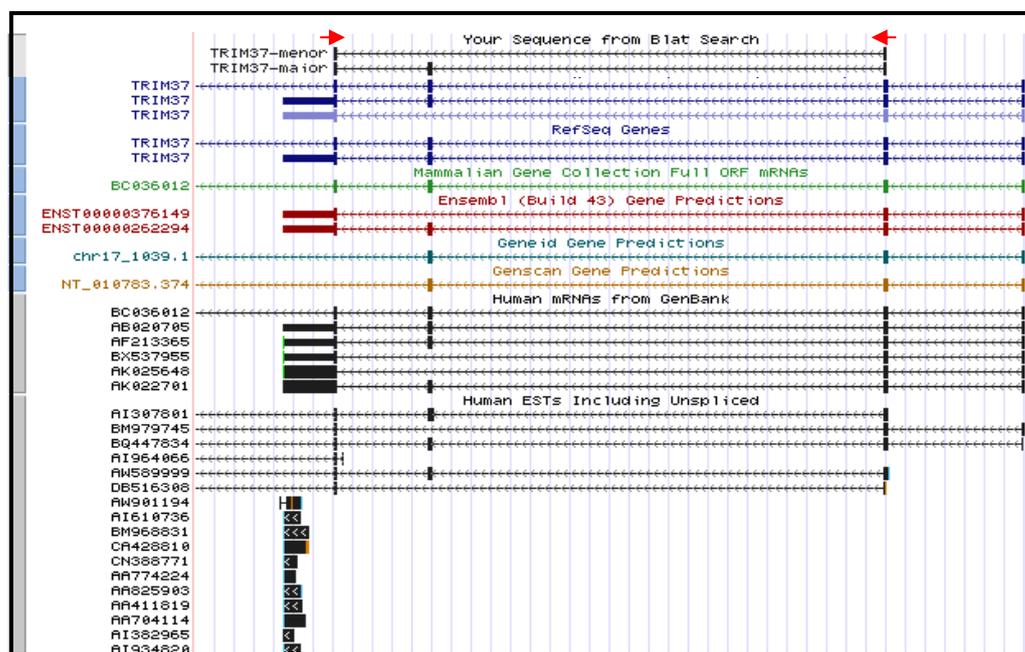
O seqüenciamento da VSE foi muito importante para confirmação da junção exon1-exon3, como citado anteriormente, na qual seriam desenhados os iniciadores para avaliação da VSE. Esse procedimento permitiu a confirmação de variantes sem o exon (VSE) de 2 dos 3 genes (TRIM37 e MK-STYX), identificando a seqüência da junção exon1-exon3 dessa variante e a presença de sítios de *splicing* conservados. Para BRRN1 não foi possível o seqüenciamento da VSE, uma vez que o isolamento do fragmento correspondente à VSE apresentou quantidade insuficiente para seu seqüenciamento. Desse modo foi realizado o alinhamento desse gene no genoma humano e seqüências de mRNA disponíveis nos bancos de dados (RefSeq, MGC e GenBank), predições gênicas (Ensembl, Gene id e Genscan) e seqüências de ESTs (*Expressed Sequence Tags*) com auxílio da ferramenta BLAT (UCSC, University of Califórnia, Santa Cruz - *Assembly*:

mar 2006) para identificar uma seqüência correspondente a essa variante, e assim confirmar indiretamente a seqüência da junção exon1-exon3. Os resultados estão apresentados na Tabela 7.

Tabela 7 - Análise da VSE dos 3 genes através de seqüenciamento e evidências em bancos de dados.

GENE	Confirmação através de seqüenciamento da VSE	Presença da VSE nos bancos de dados	Avaliação manual da presença de SS conservados
MK-STYX	OK	Sim, variante sem o exon 2 (associado a tumor)	OK
BRRN1	—	Não	OK
TRIM37	OK	Sim, variante sem o exon 2 (associado a tumor)	OK

Na Figura 27 pode ser visualizada uma representação do alinhamento das seqüências obtidas por seqüenciamento das duas variantes em estudo do gene TRIM37.



Legenda: As flechas em vermelho determinam a localização dos iniciadores nos exons 1 e 3. TRIM37-menor representa a variante sem o exon (VSE); TRIM37-maior representa a variante com o exon (VCE).

Figura 27 - Alinhamento das seqüências das 2 variantes (VCE e VSE) seqüenciadas do gene TRIM37 contra o genoma humano e seqüências de mRNA disponíveis nos bancos de dados, com auxílio da ferramenta BLAT (UCSC).

Dessa forma, os iniciadores *forward* foram desenhados na junção entre o exon 1 e 3, e foi utilizado o mesmo iniciador *reverse* empregado na avaliação da VCE (desenhado no exon flangeador 3), o qual é compartilhado pelas 2 variantes (Tabela 8).

Esses iniciadores variaram de 14 a 17 *mers* em tamanho e apresentaram TM de 53,8°C a 55,6°C de acordo com o programa OLIGOTECH versão 1.00 (Copyright© 1995).

Tabela 8 - Iniciadores *forward* e *reverse* para cada gene. F-VSE1.3: *forward* da variante sem o exon, desenhado na junção entre o exon 1 e o exon 3; R: *reverse*, desenhado no exon 3.

NOME	SEQÜÊNCIA	AMPLICON (F+R) pb
3MK-STYX-F-VSE1.3	GATGGCAAAGGAAC	97
3MK-STYX-R	CTTGGGGTCCACAG	
3TRIM37-F-VSE1.3	CCCTGAAGAAGATAC	77
3TRIM37-R	CATCTGTATTGAAGCTG	
3BRRN1-F-VSE1.3	CAACCAACTTTAAAATG	122
3BRRN1-R	CTTCGGAGACATTG	

Na seqüência representada na Figura 28 estão ilustrados os 2 iniciadores utilizados para avaliar a variante sem o exon (VSE) do gene TRIM37.

Variante sem o exon (VSE):

GGCTAATGCTAAAGGAGGTCATCTGGAAGGACTGCAGATGACTGATTTGGAAA
 ATAATTCTGAAACTGGAGAGTTACAGCCTGTACTACCTGAAGGAGCTTCAGCTG
 CCCTGAAGAAG*ATACACATTCCAGTTTTCTGATGGTGAACAAATAGGCCCT
 GAAGATCTCAGCTTCAATACAGATGAAAATAGTGAAG

Legenda: 2) VSE - variante sem o exon 2, apresenta apenas o exon 1 e 3. Asterisco vermelho: junção exon1-exon3.

CCCTGAAGAAG*ATAC:3TRIM37-F-VSE1.3; CAGCTTCAATACAGATG:3TRIM37-R.

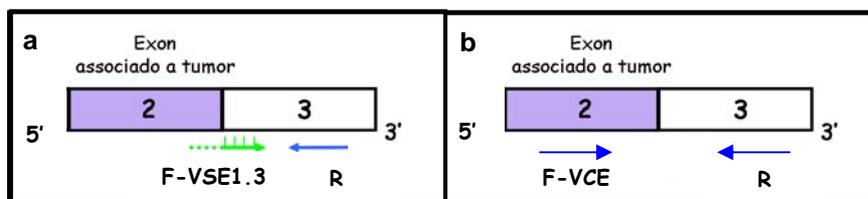
Figura 28 - Representação esquemática do desenho dos iniciadores nas seqüências da variante VSE do gene TRIM37.

Para os genes TRIM37 e MK-STYX conseguiu-se o estabelecimento das condições da reação de RT-PCR para amplificação da VSE, sendo que a temperatura ideal de pareamento dos iniciadores estabelecida foi de 57°C e 58°C, respectivamente, e foi utilizada como molde a linhagem de mama normal HB4a. Já para o gene BRRN1 não foi possível a amplificação desta

variante. É importante lembrar que para TRIM37 e MK-STYX foi confirmada a presença da VSE tanto através de seqüenciamento, quanto nos bancos de dados analisados. Já para o gene BRRN1 não foi verificada a presença desta variante nem nos bancos de dados analisados, o que, de certa forma, pode explicar a dificuldade em amplificar essa variante utilizando os iniciadores F-VSE1.3, provavelmente pelo baixo nível de sua expressão em amostras tumorais. No entanto, a existência da VSE do gene BRRN1 foi confirmada pela extensão dos 2 fragmentos numa única reação utilizando os iniciadores flanqueadores do exon associado a tumor (PE1 e R3) na linhagem HB4a (Figura 26C, página 98), o que confirma a sua existência.

4.6.2 Teste de especificidade dos iniciadores desenhados nas junções exon1-exon3

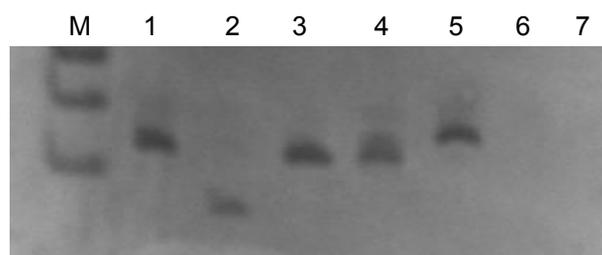
Testes foram realizados para avaliação de especificidade de amplificação do iniciador F-VSE1.3 (*forward* da variante sem o exon - pareamento nos limites exon1-exon3), juntamente com o iniciador R. Foi contemplada a possibilidade do pareamento parcial do iniciador da junção somente nas 4 bases do exon flanqueador 3 da VCE, como ilustrado a seguir (Figura 29, Esquema 2a). Para isso foram utilizados como molde *amplicons* constituídos pelos exons 2+3 (presentes na VCE), que trazem a junção do exon2-exon3, e os iniciadores das junções exon1-exon3 (F-VSE1.3). Como controle positivo (C+) foram utilizados os iniciadores que pareiam no exon 2 (F-VCE) (Figura 29, Esquema 2b).



Legenda: a) Representação do pareamento inespecífico do iniciador F-VSE1.3 no molde 2+3 da VCE; b) Representação do controle positivo utilizando o iniciador F-VCE e o molde 2+3 da VCE.

Figura 29 - Esquema 2

Os testes foram realizados para os 3 genes validados, dos quais apenas BRRN1 mostrou especificidade para o iniciador F-VSE1.3, não amplificando nenhum fragmento na presença do molde 2+3 e dos iniciadores F-VSE1.3+R, e amplificando o controle positivo representado pelos iniciadores F-VCE+R na presença do mesmo molde 2+3 (Figura 30). Os genes TRIM37 e MK-STYX apresentaram expressão na presença dos iniciadores F-VSE1.3 quando foi utilizado como molde o *amplicon* 2+3, o que representa amplificação inespecífica, resultante do pareamento parcial das 4 bases do exon 3 flanqueador a 3. É importante ressaltar que para os genes TRIM37 e MK-STYX a expressão inespecífica do molde 2+3 na presença dos iniciadores F-VSE1.3, apesar de menor do que a específica, pode prejudicar a validação da superexpressão da VCE em relação à VSE, uma vez que a expressão da VSE será contabilizada de forma artefactual, podendo diminuir a razão do nível de expressão da VCE em relação à VSE.



Legenda: Gel de poliacrilamida 8%. (M) Marcador 100pb (*Invitrogen Life Technologies*TM); 1- C+ TRIM37 (molde 2+3 e iniciadores F-VCE+R); 2- TRIM37 inespecífico (houve amplificação utilizando iniciadores F-VSE1.3+R na presença do molde 2+3); 3- C+ MK-STYX (molde 2+3 e iniciadores F-VCE+R); 4- MK-STYX inespecífico (houve amplificação utilizando iniciadores F-VSE1.3+R na presença do molde 2+3); 5- C+ BRRN1 (molde 2+3 e iniciadores F-VCE+R); 6- BRRN1 específico (não houve amplificação utilizando iniciadores F-VSE1.3+R na presença do molde 2+3); 7- NO.

Figura 30 - Representação do teste de especificidade do iniciador F-VSE1.3 para os genes TRIM37, MK-STYX e BRRN1.

Como citado anteriormente, o teste de especificidade dos iniciadores *Forward* para as VSE foram concluídos após os experimentos de qRT-PCR dessa variante para os 3 genes selecionados. Como foi detectada a amplificação inespecífica da VSE, o cálculo relativo da VCE foi realizado também em relação à exons constitutivos, isto é, presentes nas 2 variantes (VCE e VSE). Assim, para os genes TRIM37 e MK-STYX foi avaliada a expressão do gene através de desenho de iniciadores em exons constitutivos. A idéia seria que o desbalanço entre as variantes fosse salientado nessa segunda situação, uma vez que a expressão artefactual da VSE não estaria sendo contabilizada. Dessa forma para os genes TRIM37 e MK-STYX serão apresentados 2 cálculos relativos, com a variante sem o exon (VSE) e com o fragmento correspondente à expressão constitutiva.

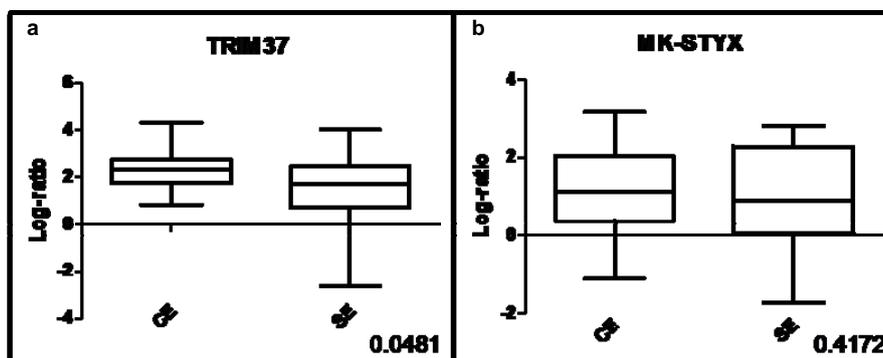
4.6.3 Avaliação por qRT-PCR da variante sem o exon (VSE)

A eficiência de amplificação foi avaliada para as VSE dos genes TRIM37 e MK-STYX estão representadas na Tabela 9.

Tabela 9 - VSE dos genes TRIM37 e MK-STYX e seus respectivos valores de *slope*, eficiência e coeficiente de regressão (R^2).

GENE	Slope	Eficiência	Eficiência (%)	R^2
TRIMse	-3,39	0,97	97	0,99
MK-STYXse	-3,70	0,86	86	0,98

A expressão da VSE dos genes TRIM37 e MK-STYX foi avaliada por qRT-PCR no mesmo grupo de amostras utilizadas previamente nos experimentos de *microarray* e em seguida comparada com os resultados obtidos para a VCE no mesmo grupo de amostras. Na comparação [(VCE tumor / VCE normal) / (VSE tumor / VSE normal)] observou-se que o balanço entre as 2 variantes foi significativamente diferente entre amostras tumorais e normais para o gene TRIM37 ($p=0,0481$; Figura 31a), confirmando que a VCE parece ser, de fato, uma variante associada a tumor. Para o gene MK-STYX o balanço entre as duas variantes não foi significativamente diferente no grupo de amostras em questão ($p=0,4172$; Figura 31b), provavelmente pela amplificação inespecífica da VCE pelos iniciadores F-VSE1.3, o que pode ter diminuído indiretamente a razão do nível de expressão da VCE em relação à VSE.



Legenda: TRIM37 (a) e MK-STYX (b). CE: VCE tumor / VCE normal. SE: VSE tumor / VSE normal. Programa *Graph Pad Prism* versão 5.0 (pvalor < 0,05).

Figura 31 - Análise entre VCE e VSE pertencentes aos genes TRIM37 e MK-STYX, considerando as amostras iniciais.

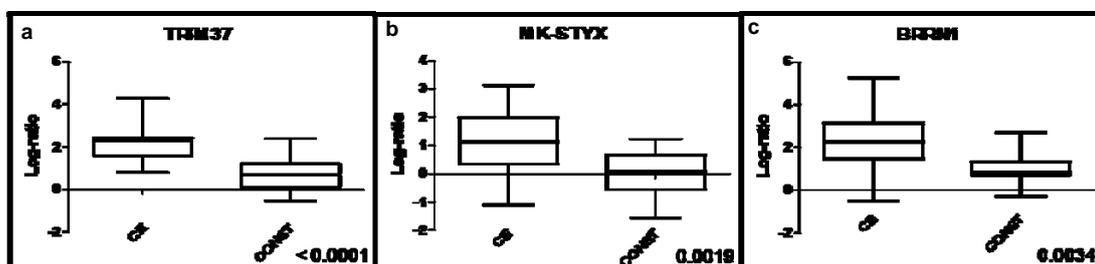
4.6.4 Avaliação da expressão por qRT-PCR do gene como um todo através de iniciadores desenhados em 2 exons constitutivos

Como a VSE apresentou resultados artefatuais de expressão, a confirmação da superexpressão da VCE (e não do gene como um todo) em tumor foi avaliada através da relação da VCE e dos exons constitutivos (EC). Os iniciadores foram desenhados nos exons 1 e -1 (*downstream* ao exon 1) para o gene TRIM37, e 4 e 5 para o gene MK-STYX. Para o gene BRRN1 o desenho de iniciadores foi realizado nos exons constitutivos -1 (*downstream* ao exon 1) e -2 (*downstream* ao exon -1). A eficiência de amplificação dos iniciadores está representada Tabela 10.

Tabela 10 - Fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1, e seus respectivos valores de *slope*, eficiência e coeficiente de regressão (R^2).

GENE	Slope	Eficiência	Eficiência (%)	R^2
TRIM37const	-4,33	0,70195	70	0,9964
MK-STYXconst	-3,95	0,791284	79	0,9993
BRRN1const	-4,33	0,70195	70	0,9904

Para os 3 genes foi observado um balanço significativamente diferente entre a VCE e o fragmento correspondente à expressão constitutiva (EC) em amostras tumorais em relação às normais (TRIM37, $p < 0,0001$; MK-STYX, $p = 0,0019$; BRRN1, $p = 0,0034$). Isso sugere que a VCE é realmente uma variante superexpressa em tumor de mama. Os resultados estão demonstrados nas Figuras 32a, 32b e 32c.



Legenda: TRIM37 (a), MK-STYX (b) e BRRN1 (c). CE: VCE tumor / VCE normal. CONST: EC tumor / EC normal. Programa *Graph Pad Prism* versão 5.0 (p valor $< 0,05$).

Figura 32 - Análise entre VCE e o fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1, considerando as amostras iniciais.

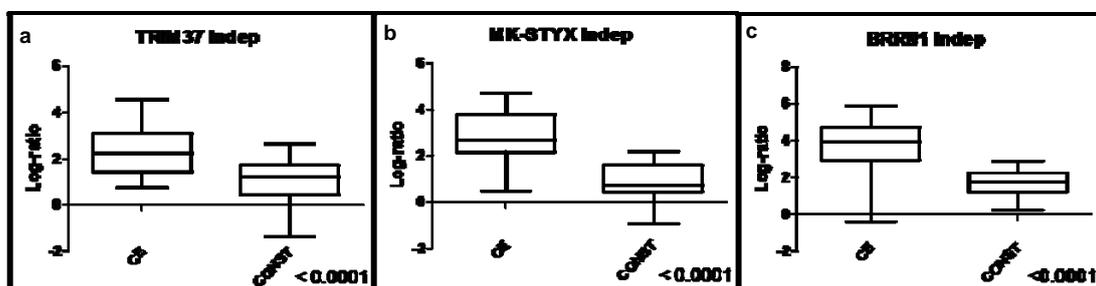
Pode-se observar um aumento significativo na diferença da média da razão de expressão entre a VCE e o fragmento correspondente à expressão constitutiva do gene TRIM37 (Figura 32a, $p < 0,0001$) em relação à média da razão da expressão entre a VCE e VSE (Figura 31a, $p = 0,0481$) do mesmo gene, reforçando que a VCE é uma variante superexpressa em tumor de mama. Do mesmo modo, a relação VCE/VSE para o gene MK-STYX não foi significativa, provavelmente pela quantificação inespecífica de transcritos da VCE na presença dos iniciadores F-VSE1.3.

4.7 CONFIRMAÇÃO DA SUPEREXPRESSION DA VCE EM RELAÇÃO AO SEU RESPECTIVO GENE EM UM GRUPO INDEPENDENTE DE AMOSTRAS

Para os genes TRIM37, MK-STYX e BRRN1 a VCE e o fragmento correspondente à expressão constitutiva foram avaliados em um grupo independente de 40 amostras tumorais de mama (Anexo 12) para validar biologicamente a superexpressão da VCE em amostras tumorais de mama quando comparadas às amostras normais. Os mesmos critérios para testar se, de fato, os exons selecionados pertencem a uma variante superexpressa e não ao gene como um todo, foram adotados (página 95).

Os experimentos foram realizados de acordo com o item 3.14.5 de material e métodos e os iniciadores foram os mesmos utilizados nos experimentos anteriores. Os 3 genes apresentaram *fold* ≥ 3 , respondendo ao critério 1, TRIM37 (6,20), MK-STYX (9,57) e BRRN1 (17,37). Os resultados

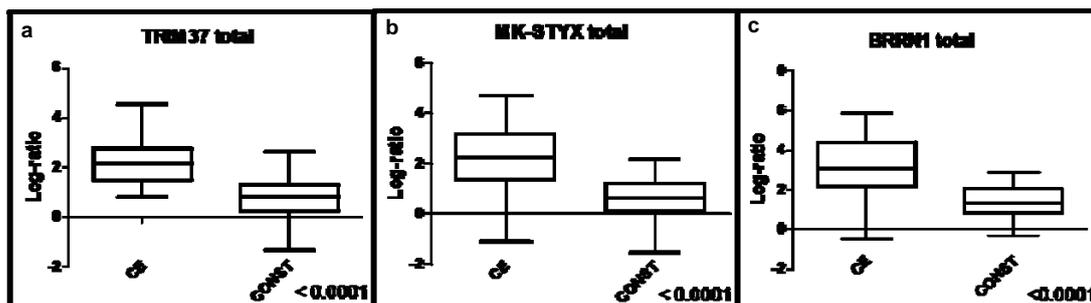
de análise da expressão constitutiva do gene revelaram um aumento significativo na diferença de expressão entre a VCE e o fragmento correspondente à expressão constitutiva para 2 genes (MK-STYX, $p < 0,0001$; BRRN1, $p < 0,0001$), e o gene TRIM37 manteve a mesma diferença ($p < 0,0001$), conforme mostra a Figura 33, comprovando que são, de fato, variantes superexpressas em tumor de mama.



Legenda: TRIM37 (a), MK-STYX (b) e BRRN1 (c). CE: VCE tumor / VCE normal. CONST: EC tumor / EC normal. Programa *Graph Pad Prism* versão 5.0 (pvalor < 0,05).

Figura 33 - Análise entre VCE e o fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1, considerando as amostras independentes.

Em seguida, a razão entre a VCE e o fragmento correspondente à expressão constitutiva foi avaliada para cada um dos 3 genes utilizando os 2 grupos de amostras (19+40) ao mesmo tempo. Esse resultado pode ser observado na Figura 34.



Legenda: TRIM37 (a), MK-STYX (b) e BRRN1 (c). CE: VCE tumor / VCE normal. CONST: EC tumor / EC normal. Programa *Graph Pad Prism* versão 5.0 (pvalor < 0,05).

Figura 34 - Análise entre VCE e o fragmento correspondente à expressão constitutiva dos genes TRIM37, MK-STYX e BRRN1, considerando todas as amostras (iniciais + independentes).

4.8 IDENTIFICAÇÃO DOS DOMÍNIOS FUNCIONAIS ATRAVÉS DAS SEQÜÊNCIAS PROTEICAS DO GENE E DA VCE

As diferentes isoformas protéicas geradas por variantes de *splicing* podem acarretar mudanças na estrutura da proteína que podem esconder ou expor domínios funcionais, ou alterar a atividade de interação com outras proteínas ou complexos protéicos. Para conhecer a provável influência da inserção do exon associado a tumor na função da proteína, primeiramente foi verificado se todos os 3 exons pertencentes às variantes identificadas como superexpressas neste trabalho (TRIM37, MK-STYX e BRRN1) estavam na fase aberta de leitura, isto é, se estavam contidos na região codificadora do gene (CDS). Foi confirmado que todos os exons estavam contidos na CDS. Para identificar eventual presença de algum domínio funcional presente nas variantes identificadas como superexpressas em

câncer de mama, os domínios funcionais foram mapeados na seqüência de aminoácidos, utilizando a ferramenta Pfam (<http://www.sanger.ac.uk/Software/Pfam>). Para a VCE do gene TRIM37 pode-se observar que o exon associado a tumor estava a aproximadamente 500aa (1.500pb) do domínio funcional mais próximo: *MATH_TRIM37 domain*, seguido por um *RING-finger (Really Interesting New Gene) domain* (Figura 35). Já o exon associado a tumor do gene MK-STYX aparece inserido dentro de um domínio funcional denominado *RHOD, Rhodanese homology domain* e está próximo de outro domínio denominado *DSPc, Dual specificity phosphatase domain* (Figura 36). O exon associado a tumor do gene BRRN1, por sua vez, aparece inserido no domínio funcional denominado *Barren* (Figura 37).

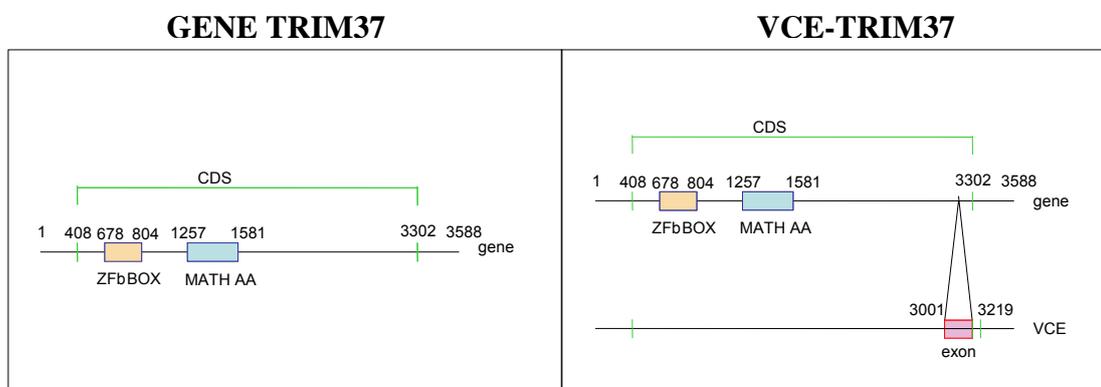


Figura 35 - Mapeamento do exon associado a tumor pertencente à VCE do gene TRIM37 na seqüência protéica.

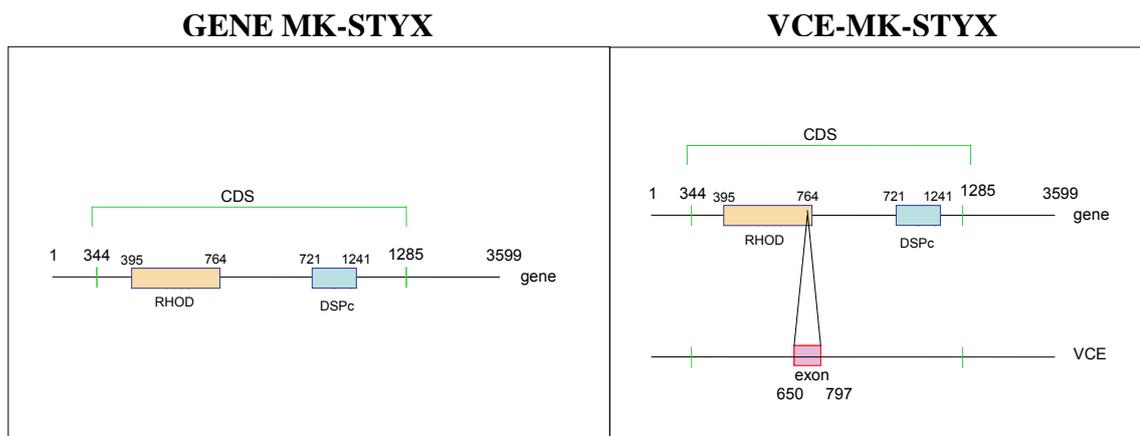


Figura 36 - Mapeamento do exon associado a tumor pertencente à VCE do gene MK-STYX na seqüência protéica.

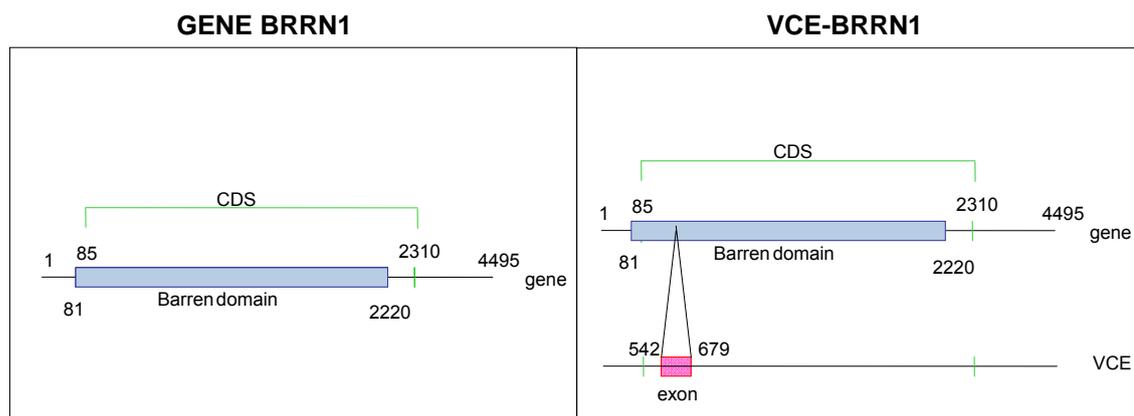


Figura 37 - Mapeamento do exon associado a tumor pertencente à VCE do gene BRRN1 na seqüência protéica.

4.9 CARACTERÍSTICAS CLÍNICAS E HISTOPATOLÓGICAS DAS AMOSTRAS

Foram estudados 59 casos (19+40) de carcinoma ductal invasivo de mama, que foram acompanhados por um período mínimo de 10 meses e máximo de 93 meses, sendo todos os pacientes do sexo feminino, e tumores primários sem nenhum tratamento prévio.

A idade das pacientes no momento do diagnóstico variou de 26 a 88 anos, sendo a média de 54 e mediana de 56 anos. A raça branca foi preponderante, sendo 54 casos (91,5%) e apenas 5 (8,5%) não-brancas. A média e mediana de idade da menarca dessas pacientes foi de 13 anos. Das 59 pacientes, para 5 não havia informação sobre *status* menopausal. Para as 54 com informação, 24 (44,4%) não eram menopausadas e 30 (55,6%) sim, sendo a média de idade da menopausa de 47 anos e mediana de 49 anos. Dezoito pacientes (30,5%) nunca tiveram filhos e 41 (69,5%) tiveram pelo menos 1 parto. Trinta e dois pacientes (54,2%) nunca fizeram uso de anticoncepcionais, 25 (42,4%) fizeram e para 2 (3,4%) não havia informação. Das 57 pacientes para as quais havia informação sobre reposição hormonal (TRH), 41 (69,5%) não fizeram e apenas 16 (27,1%) fizeram TRH, para 2 não havia informação (3,4%). Das 59 pacientes, 20 (33,9%) tinham história familiar de câncer de mama e 36 (61,0%) não, e para 3 (5,1%) não havia informação.

Para analisar os casos que compuseram esta amostra foram realizados cortes histológicos de 5µm corados com hematoxilina e eosina

(H&E) que revelaram presença de células epiteliais malignas, com variados graus de pleomorfismo celular e nuclear. Segundo a gradação histopatológica das amostras (de acordo com Scarff-Bloom-Richardson, SBR), 7 casos (11,9%) foram classificados como grau I, ou seja, o tumor estava bem diferenciado exibindo morfologia semelhante a do tecido epitelial normal. A maioria dos espécimes (30 casos ou 50,8%) foi classificada como grau II, ou moderadamente diferenciado e 22 casos (37,3%) foram enquadrados como grau III ou pouco diferenciados (Tabela 11). Microscopicamente, apenas 1 caso (1,7%) apresentou invasão vascular, 19 (32,2%) invadiram vasos linfáticos, e em 4 casos (6,8%) foi notada presença de invasão perineural. Os pacientes incluídos neste estudo foram tratados de acordo com o protocolo vigente no Departamento de Mastologia do Hospital A.C. Camargo, São Paulo. Dos casos estudados, 4 pacientes (6,8%) foram tratados somente com cirurgia, 8 (13,5%) casos com cirurgia associada à quimioterapia, 10 casos (16,9%) com cirurgia associada à radioterapia e quimioterapia, 6 casos (10,2%) com cirurgia associada à radioterapia e hormônio terapia, 4 casos (6,8%) com cirurgia associada à quimioterapia e hormônio terapia, 26 casos (44,1%) com cirurgia associada à radioterapia, quimioterapia e hormônio terapia e 1 caso (1,7%) com cirurgia associada somente à hormônio terapia. Dos pacientes que tiveram linfonodos removidos, 27 casos (45,8%) confirmaram o comprometimento microscópico, ou seja, havia infiltração de células malignas nos linfonodos. O tempo médio de detecção das recorrências do tumor foi de 22 meses após o

início do tratamento, sendo o tempo mínimo de 3 meses e máximo de 59 meses.

De acordo com a última informação obtida (23/01/2008), 8 pacientes (13,7%) evoluíram para óbito, sendo 4 (6,8%) devido à doença e 4 (6,8%) de outras causas. Dos pacientes vivos, 43 (72,9%) estão livres da doença e 7 (11,9%) apresentam doença. Apenas 1 (1,7%) caso foi perdido de vista e não se conseguiu nenhuma informação do paciente após o tratamento.

Na Tabela 11 estão resumidas as principais informações sobre os aspectos clínicos e histopatológicos estudados neste trabalho.

Tabela 11 - Dados clinicopatológicos, demográficos e sobre o tratamento dos 59 pacientes portadores de carcinoma ductal de mama incluídos neste estudo.

Variáveis		n (%)
Raça	Branca	54 (91,5)
	Não Branca	5 (8,5)
Idade	≤ 40 anos	16 (27,1)
	> 40 anos	43 (72,9)
Menarca	≤ 12 anos	24 (40,7)
	> 12 anos	35 (59,3)
Parto	Não	18 (30,5)
	Sim	41 (69,5)
Status Menopausal	Pré-menopausa	26 (44,1)
	Pós-menopausa	33 (55,9)
Estadiamento T	T1+T2	47 (79,6)
	T3+T4	12 (20,4)
Estadiamento N	N0	31 (52,5)
	N+	28 (47,5)
Estadio Clínico	EC I e II (doença localizada)	46 (77,9)
	EC III e IV (doença avançada)	13 (22,1)
Grau SBR	1 (bem diferenciado)	7 (11,9)
	2 (moderadamente)	30 (50,8)
	3 (pouco diferenciado)	22 (37,3)
Status Nodal (presença de metástase linfonodal)	Não	31 (52,5)
	Sim	27 (45,8)
	Ignorado	1 (1,7)
Receptor de Estrógeno	Negativo	17 (28,8)
	Positivo	42 (71,2)
Receptor de Progesterona	Negativo	26 (44,1)
	Positivo	33 (55,9)
ERBB2	0+1 (Negativo)	25 (42,4)
	2+3 (Positivo)	32 (54,2)
	Ignorado	2 (3,4)
p53	Negativo	36 (61)
	Positivo	11 (18,6)
	Ignorado	12 (20,3)
Radioterapia adjuvante	Não	17 (28,8)
	Sim	42 (71,2)
Quimioterapia adjuvante	Não	11 (18,6)
	Sim	48 (81,4)
Hormônio terapia adjuvante	Não	22 (37,3)
	Sim	37 (62,7)
Recorrência	Não	46 (78)
	Sim	13 (22)
Morte	Não	51 (86,4)
	Sim	8 (13,6)

4.10 ASSOCIAÇÃO ENTRE AS VARIÁVEIS CLÍNICAS E HISTOPATOLÓGICAS E OS DADOS DE EXPRESSÃO OBTIDOS

Para cada gene realizou-se comparação dos dados de expressão da variante que contém o exon associado a tumor (VCE), assim como da relação entre a VCE e o fragmento correspondente à expressão constitutiva do gene (VCE/EC) com as variáveis clínicas e histopatológicas de cada paciente.

- ✓ Os resultados referentes à associação entre a VCE do gene TRIM37 e as variáveis clínicopatológicas estão representados na Tabela 12.

Tabela 12 - Associação entre a VCE do gene TRIM37 e as variáveis clinicopatológicas.

Variáveis clinicopatológicas		TRIM37ce				pvalor TRIM37ce
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Faixa etária	≤ 40 anos	6 (37,5)	3 (18,8)	1 (6,3)	6 (37,5)	0,087
	> 40 anos	9 (20,9)	12 (27,9)	14 (32,6)	8 (18,6)	
Menarca	≤ 12 anos	8 (33,3)	8 (33,3)	3 (12,5)	5 (20,8)	0,188
	> 12 anos	7 (20)	7 (20)	12 (34,3)	9 (25,7)	
Status menopausal	Não	8 (30,8)	7 (26,9)	4 (15,4)	7 (26,9)	0,456
	Sim	7 (21,2)	8 (24,2)	11 (33,3)	7 (21,2)	
Estádio Clínico	I+II	11 (23,9)	11 (23,9)	13 (28,3)	11 (23,9)	0,792
	III+IV	4 (30,8)	4 (30,8)	2 (15,4)	3 (23,1)	
Grau SBR	1	1 (14,3)	1 (14,3)	4 (57,1)	1 (14,3)	0,571
	2	8 (26,7)	7 (23,3)	7 (23,3)	8 (26,7)	
	3	6 (27,3)	7 (31,8)	4 (18,2)	5 (22,7)	
Status nodal	Negativo	10 (32,3)	5 (16,1)	10 (32,3)	6 (19,4)	0,161
	Positivo	5 (17,9)	10 (35,7)	5 (17,9)	8 (28,6)	
Receptor de Estrógeno	Negativo	8 (47,1)	6 (35,3)	3 (17,6)	0 (0)	0,009*
	Positivo	7 (16,7)	9 (21,4)	12 (28,6)	14 (33,3)	
Receptor de Progesterona	Negativo	9 (34,6)	10 (38,5)	4 (15,4)	3 (11,5)	0,024*
	Positivo	6 (18,2)	5 (15,2)	11 (33,3)	11 (33,3)	
ERBB2	Negativo	7 (28)	4 (16)	7 (28)	7 (28)	0,486
	Positivo	7 (21,9)	11 (34,4)	7 (21,9)	7 (21,9)	
p53	Negativo	6 (16,7)	9 (25)	10 (27,8)	11 (30,6)	0,021*
	Positivo	6 (54,5)	4 (36,4)	1 (9,1)	0 (0)	
Recorrência	Não	10 (21,7)	10 (21,7)	14 (30,4)	12 (26,1)	0,189
	Sim	5 (38,5)	5 (38,5)	1 (7,7)	2 (15,4)	

* p<0,05

A expressão de receptores de estrógeno (ER+) apresentou uma associação significativa (0,009) com a VCE do gene TRIM37. Assim todos os 42 casos ER+ foram comparados às demais variáveis, sendo observada uma associação estatisticamente significativa (p=0,033) dos casos ER+ com a variável idade da menarca superior a 12 anos (Tabela 13).

Tabela 13 - Associação entre a expressão da VCE do gene TRIM37 nas amostras ER+ e a variável menarca.

Variáveis clinicopatológicas		TRIM37ce ER+				pvalor TRIM37ce ER+
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Menarca	≤ 12 anos	4 (22,2)	7 (38,9)	2 (11,1)	5 (27,8)	0,033*
	> 12 anos	3 (12,5)	2 (8,3)	10 (41,7)	9 (37,5)	

* p<0,05

A expressão de receptores de progesterona (PgR+) também apresentou uma associação significativa (0,024) com a VCE do gene TRIM37. Assim, todos os 33 casos PgR+ foram comparados às demais variáveis, sendo observada uma associação estatisticamente significativa (p=0,037) dos casos PgR+ com a variável idade das pacientes superior a 40 anos (Tabela 14).

Tabela 14 - Associação entre a expressão da VCE do gene TRIM37 nas amostras PgR+ e a variável faixa etária.

Variáveis clinicopatológicas		TRIM37ce PgR+				pvalor TRIM37ce PgR+
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Faixa etária	≤ 40 anos	2 (22,2)	1 (11,1)	0 (0)	6 (66,7)	0,037*
	> 40 anos	4 (16,7)	4 (16,7)	11 (45,8)	5 (20,8)	

* p<0,05

- ✓ Os resultados referentes à associação entre a relação VCE/EC do gene TRIM37 e as variáveis clinicopatológicas estão representados na Tabela 15.

Tabela 15 - Associação entre a relação VCE/EC do gene TRIM37 e as variáveis clinicopatológicas.

Variáveis clinicopatológicas		TRIM37cec				pvalor TRIM37cec
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Faixa etária	≤ 40 anos	3 (18,8)	2 (12,5)	4 (25)	7 (43,8)	0,233
	> 40 anos	10 (27)	11 (29,7)	9 (24,3)	7 (18,9)	
Menarca	≤ 12 anos	7 (30,4)	5 (21,7)	5 (21,7)	6 (26,1)	0,841
	> 12 anos	6 (20)	8 (26,7)	8 (26,7)	8 (26,7)	
Status menopausal	Não	4 (16,7)	6 (25)	5 (20,8)	9 (37,5)	0,335
	Sim	9 (31)	7 (24,1)	8 (27,6)	5 (17,2)	
Estadio Clínico	I+II	10 (23,8)	9 (21,4)	12 (28,6)	11 (26,2)	0,534
	III+IV	3 (27,3)	4 (36,4)	1 (9,1)	3 (27,3)	
Grau SBR	1	3 (42,9)	1 (14,3)	2 (28,6)	1 (14,3)	0,291
	2	3 (11,5)	7 (26,9)	9 (34,6)	7 (26,9)	
	3	7 (35)	5 (25)	2 (10)	6 (30)	
Status nodal	Negativo	8 (25,8)	7 (22,6)	9 (29)	7 (22,6)	0,753
	Positivo	5 (22,7)	6 (27,3)	4 (18,2)	7 (31,8)	
Receptor de Estrógeno	Negativo	2 (11,8)	7 (41,2)	4 (23,5)	4 (23,5)	0,205
	Positivo	11 (30,6)	6 (16,7)	9 (25)	10 (27,8)	
Receptor de Progesterona	Negativo	5 (20)	9 (36)	5 (20)	6 (24)	0,329
	Positivo	8 (28,6)	4 (14,3)	8 (28,6)	8 (28,6)	
ERBB2	Negativo	7 (30,4)	4 (17,4)	3 (13)	9 (39,1)	0,065
	Positivo	4 (14,3)	9 (32,1)	10 (35,7)	5 (17,9)	
p53	Negativo	4 (12,9)	6 (19,4)	11 (35,5)	10 (32,3)	0,040*
	Positivo	2 (20)	6 (60)	2 (20)	0 (0)	
Recorrência	Não	9 (22,5)	10 (25)	11 (27,5)	10 (25)	0,801
	Sim	4 (30,8)	3 (23,1)	2 (15,4)	4 (30,8)	

* p<0,05

A variável p53 negativo apresentou uma associação significativa (0,040) com a relação VCE/EC do gene TRIM37. Assim os 31 casos p53- foram comparados às demais variáveis, porém não apresentaram associação estatisticamente significativa com nenhuma delas.

Os resultados referentes à associação entre a VCE do gene MK-STYX e as variáveis clinicopatológicas estão representados na Tabela 16.

Tabela 16 - Associação entre a expressão da VCE do gene MK-STYX e as variáveis clinicopatológicas.

Variáveis clinicopatológicas		MK-STYXce				pvalor MK-STYXce
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Faixa etária	≤ 40 anos	4 (25)	3 (18,8)	4 (25)	5 (31,3)	0,812
	> 40 anos	10 (25,6)	11 (28,2)	10 (25,6)	8 (20,5)	
Menarca	≤ 12 anos	5 (20,8)	7 (29,2)	8 (33,3)	4 (16,7)	0,475
	> 12 anos	9 (29)	7 (22,6)	6 (19,4)	9 (29)	
Status menopausal	Não	4 (15,4)	6 (23,1)	8 (30,8)	8 (30,8)	0,298
	Sim	10 (34,5)	8 (27,6)	6 (20,7)	5 (17,2)	
Estadio Clínico	I+II	10 (23,3)	13 (30,2)	9 (20,9)	11 (25,6)	0,257
	III+IV	4 (33,3)	1 (8,3)	5 (41,7)	2 (16,7)	
Grau SBR	1	1 (16,7)	3 (50)	2 (33,3)	0 (0)	0,447
	2	8 (29,6)	4 (14,8)	8 (29,6)	7 (25,9)	
	3	5 (22,7)	7 (31,8)	4 (18,2)	6 (27,3)	
Status nodal	Negativo	9 (31)	9 (31)	6 (20,7)	5 (17,2)	0,375
	Positivo	5 (19,2)	5 (19,2)	8 (30,8)	8 (30,8)	
Receptor de Estrógeno	Negativo	6 (35,3)	5 (29,4)	4 (23,5)	2 (11,8)	0,459
	Positivo	8 (21,1)	9 (23,7)	10 (26,3)	11 (28,9)	
Receptor de Progesterona	Negativo	7 (26,9)	10 (38,5)	6 (23,1)	3 (11,5)	0,09
	Positivo	7 (24,1)	4 (13,8)	8 (27,6)	10 (34,5)	
ERBB2	Negativo	7 (30,4)	5 (21,7)	3 (13)	8 (34,8)	0,211
	Positivo	7 (22,6)	9 (29)	10 (32,3)	5 (16,1)	
p53	Negativo	10 (29,4)	9 (26,5)	6 (17,6)	9 (26,5)	0,211
	Positivo	3 (7,3)	4 (36,4)	4 (36,4)	0 (0)	
Recorrência	Não	11 (26,2)	11 (26,2)	10 (23,8)	10 (23,8)	0,966
	Sim	3 (23,1)	4 (30,8)	3 (23,1)	3 (23,1)	

* p<0,05

A expressão da VCE do gene MK-STYX não apresentou associação estatisticamente significativa com nenhuma das variáveis analisadas.

- ✓ Os resultados referentes à associação entre a relação VCE/EC do gene MK-STYX e as variáveis clinicopatológicas estão representados na Tabela 17.

Tabela 17 - Associação entre a relação VCE/EC do gene MK-STYX e as variáveis clinicopatológicas.

Variáveis clinicopatológicas		MK-STYXcec				pvalor MK-STYXcec
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Faixa etária	≤ 40 anos	2 (12,5)	6 (37,5)	2 (12,5)	6 (37,5)	0,123
	> 40 anos	11 (30,6)	7 (19,4)	11 (30,6)	7 (19,4)	
Menarca	≤ 12 anos	6 (26,1)	6 (26,1)	5 (21,7)	6 (26,1)	0,972
	> 12 anos	7 (24,1)	7 (24,1)	8 (27,6)	7 (24,1)	
Status menopausal	Não	2 (8)	7 (28)	6 (24)	10 (40)	0,018*
	Sim	11 (40,7)	6 (22,2)	7 (25,9)	3 (11,1)	
Estadio Clínico	I+II	11 (27,5)	11 (27,5)	7 (17,5)	11 (27,5)	0,158
	III+IV	2 (16,7)	2 (16,7)	6 (50)	2 (16,7)	
Grau SBR	1	3 (50)	1 (16,7)	1 (16,7)	1 (16,7)	0,254
	2	9 (34,6)	6 (23,1)	5 (19,2)	6 (23,1)	
	3	1 (5)	6 (30)	7 (35)	6 (30)	
Status nodal	Negativo	8 (27,6)	10 (34,5)	6 (20,7)	5 (17,2)	0,204
	Positivo	5 (21,7)	3 (13)	7 (30,4)	8 (34,8)	
Receptor de Estrógeno	Negativo	4 (23,5)	4 (23,5)	5 (29,4)	4 (23,5)	0,967
	Positivo	9 (25,7)	9 (25,7)	8 (22,9)	9 (25,7)	
Receptor de Progesterona	Negativo	6 (24)	6 (24)	7 (28)	6 (24)	0,972
	Positivo	7 (25,9)	7 (25,9)	6 (22)	7 (25,9)	
ERBB2	Negativo	4 (18,2)	8 (36,4)	5 (22,7)	5 (22,7)	0,435
	Positivo	9 (31)	5 (17,2)	7 (24,1)	8 (27,6)	
p53	Negativo	9 (28,1)	9 (28,1)	5 (15,6)	9 (28,1)	0,339
	Positivo	3 (30)	1 (10)	4 (40)	2 (20)	
Recorrência	Não	10 (25,6)	11 (28,2)	8 (20,5)	10 (25,6)	0,583
	Sim	3 (23,1)	2 (15,4)	5 (38,5)	3 (23,1)	

* p<0,05

A variável status menopausal negativo apresentou uma associação significativa (0,018) com a relação VCE/EC do gene MK-STYX. Assim, todos os 25 casos menopausa negativos foram isolados e comparados às demais variáveis. Foi observada uma associação estatisticamente significante (p=0,012) dos casos menopausa negativos com os estadios clínicos I e II (Tabela 18).

Tabela 18 - Associação entre a da relação VCE/EC do gene MK-STYX nas amostras não menopausa e os estádios clínicos I e II.

Variáveis clinicopatológicas		MK-STYXcec não menopausa				pvalor MK- STYXce c
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Estadio Clínico	I+II	2 (10)	7 (35)	2 (10)	9 (45)	0,012*
	III+IV	0 (0)	0 (0)	4 (80)	1 (20)	

* p<0,05

- ✓ Os resultados referentes à associação entre a VCE do gene BRRN1 e as variáveis clinicopatológicas estão representados na Tabela 19.

Tabela 19 - Associação entre a expressão da VCE do gene BRRN1 e as variáveis clinicopatológicas.

Variáveis clinicopatológicas		BRRN1ce				pvalor BRRN1ce
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Faixa etária	≤ 40 anos	1 (9,1)	2 (18,2)	3 (27,3)	5 (45,5)	0,222
	> 40 anos	9 (31)	8 (27,6)	7 (24,1)	5 (17,2)	
Menarca	≤ 12 anos	4 (22,2)	4 (22,2)	5 (27,8)	5 (27,8)	0,939
	> 12 anos	6 (27,3)	6 (27,3)	5 (22,7)	5 (22,7)	
Status menopausal	Não	4 (22,2)	3 (16,7)	6 (33,3)	5 (27,8)	0,568
	Sim	6 (27,3)	7 (31,8)	4 (18,2)	5 (22,7)	
Estadio Clínico	I+II	6 (13)	8 (17,4)	7 (15,2)	25 (54,3)	0,844
	III+IV	3 (23,1)	2 (15,4)	2 (15,4)	6 (46,2)	
Grau SBR	1	3 (60)	1 (20)	1 (20)	0 (0)	0,084
	2	6 (33)	6 (33)	3 (16,7)	3 (16,7)	
	3	1 (5,9)	3 (17,6)	6 (35,3)	7 (41,2)	
Status nodal	Negativo	6 (19,4)	5 (16,1)	6 (19,4)	14 (45,2)	0,543
	Positivo	3 (10,7)	5 (17,9)	3 (10,7)	17 (60,7)	
Receptor de Estrógeno	Negativo	1 (8,3)	4 (33,3)	3 (25)	4 (33,3)	0,414
	Positivo	9 (32,1)	6 (21,4)	7 (25)	6 (21,4)	
Receptor de Progesterona	Negativo	1 (5,3)	6 (31,6)	5 (25,3)	7 (36,8)	0,04*
	Positivo	9 (42,9)	4 (19)	5 (23,8)	3 (14,3)	
ERBB2	Negativo	4 (25)	4 (25)	3 (18,8)	5 (31,3)	0,958
	Positivo	5 (21,7)	5 (21,7)	6 (26,1)	7 (30,4)	
p53	Negativo	7 (26,9)	5 (19,2)	7 (26,9)	7 (26,9)	0,377
	Positivo	1 (16,7)	2 (33,3)	0 (0)	3 (50)	
Recorrência	Não	9 (28,1)	8 (25)	7 (21,9)	8 (25)	0,741
	Sim	1 (12,5)	2 (25)	3 (37,5)	2 (25)	

* p<0,05

A ausência da expressão de receptores de progesterona (PgR-) apresentou uma associação significativa (0,040) com a VCE do gene BRRN1. Assim, os 19 casos PgR- foram comparados às demais variáveis, porém não apresentaram associação estatisticamente significativa com nenhuma delas. Os resultados referentes à associação entre a relação VCE/EC do gene BRRN1 e as variáveis clinicopatológicas estão representados na Tabela 20.

Tabela 20 - Associação entre a relação VCE/EC do gene BRRN1 e as variáveis clinicopatológicas.

Variáveis clinicopatológicas		BRRN1cec				pvalor BRRN1c ec
		1º quartil n (%)	2º quartil n (%)	3º quartil n (%)	4º quartil n (%)	
Faixa etária	≤ 40 anos	1 (9,1)	2 9(18,2)	3 (27,3)	5 (45,5)	0,29
	> 40 anos	8 (29,6)	7 (25,9)	7 (25,9)	5 (18,5)	
Menarca	≤ 12 anos	5 (27,8)	3 (16,7)	5 (27,8)	5 (27,8)	0,799
	> 12 anos	4 (20)	6 (30)	5 (25)	5 (25)	
Status menopausal	Não	4 (23,5)	3 (17,6)	4 (23,5)	6 (35,3)	0,681
	Sim	5 (23,8)	6 (28,6)	6 (28,6)	4 (19)	
Estadio Clínico	I+II	6 (19,4)	8 (25,8)	8 (25,8)	9 (29)	0,544
	III+IV	3 (42,9)	1 (14,3)	2 (28,6)	1 (14,3)	
Grau SBR	1	2 (40)	1 (20)	1 (20)	1 (20)	0,065
	2	6 (35,3)	6 (35,3)	1 (5,9)	4 (23,5)	
	3	1 (6,3)	2 (12,5)	8 (50)	5 (31,3)	
Status nodal	Negativo	4 (16,7)	7 (29,2)	7 (29,2)	6 (25)	0,489
	Positivo	5 (35,7)	2 (14,3)	3 (21,4)	4 (28,6)	
Receptor de Estrógeno	Negativo	2 (16,7)	2 (16,7)	5 (41,7)	3 (25)	0,51
	Positivo	7 (26,9)	7 (26,9)	5 (19,2)	7 (26,9)	
Receptor de Progesterona	Negativo	3 (15,8)	3 (15,8)	8 (42,1)	5 (26,3)	0,133
	Positivo	6 (31,6)	6 (31,6)	2 (10,5)	5 (26,3)	
ERBB2	Negativo	5 (31,3)	1 (6,3)	4 (25)	6 (37,5)	0,185
	Positivo	4 (19)	7 (33,3)	6 (28,6)	4 (19)	
p53	Negativo	5 (20,8)	6 (25)	7 (29,2)	6 (25)	0,833
	Positivo	2 (33,3)	1 (16,7)	1 (16,7)	2 (33,3)	
Recorrência	Não	7 (23,3)	8 (26,7)	8 (26,7)	7 (23,3)	0,794
	Sim	2 (25)	1 (12,5)	2 (25)	3 (37,5)	

* p<0,05

A relação VCE/EC do gene BRRN1 não apresentou associação estatisticamente significativa com nenhuma das variáveis analisadas.

5 DISCUSSÃO

Splicing alternativo do mRNA é um dos mais importantes mecanismos de geração de variabilidade transcricional, influenciando o repertório de proteínas existentes em uma célula, podendo de certa forma justificar o número relativamente pequeno de genes humanos encontrados (LANDER et al. 2001; VENTER et al. 2001). Há relatos na literatura de que entre 60% e 81% dos genes humanos sofrem *splicing* alternativo (JOHNSON et al. 2003; KAMPA et al. 2004; HUI et al. 2004; KIM et al. 2007) e que as variações nos processos de *splicing* também ocorrem durante a tumorigênese, gerando muitas vezes variantes aberrantes que podem desempenhar um papel chave na progressão tumoral (CABALLERO et al. 2001; VENABLES 2004; KIM et al. 2008). Essas variantes podem ser utilizadas como marcadores tumorais em câncer, sendo muito promissora a identificação de exons alternativos superexpressos em células tumorais que, quando situados na região codificadora do gene, podem produzir um peptídeo que sirva como alvo terapêutico.

A identificação de alterações moleculares que distinguem células tumorais de normais ajuda a conhecer a natureza da célula tumoral e entender o seu comportamento patológico. O câncer de mama é bastante heterogêneo, com diferenças moleculares que requerem tratamentos específicos (VAN'T VEER e BERNARDS 2008). Alguns marcadores moleculares têm sido utilizados na rotina deste tipo de neoplasia, auxiliando

no prognóstico dos pacientes e na determinação da terapia adjuvante a ser adotada (MOLINA et al. 2005; DUFFY 2005; BURSTEIN 2005). Porém muitos pacientes com o mesmo *status* desses marcadores podem evoluir de maneiras bastante distintas, o que leva à necessidade da busca por novos marcadores moleculares. Muitas evidências mostram a ocorrência de *splicing* alternativo ou aberrante durante o desenvolvimento, progressão e metástase do câncer de mama, como por exemplo, a ocorrência de variantes de *splicing* em vários genes, incluindo receptores de estrógeno (ER), *BRCA1*, *AIB1*, *CD44*, *ERBB2* e VEGF em linhagens ou amostras deste tipo de neoplasia (VENABLES 2004). Assim, a exploração da variabilidade transcricional gerada por *splicing* alternativo torna-se promissora na identificação de novos marcadores moleculares em câncer de mama.

Recentes estudos têm sido realizados nos quais a associação entre *splicing* alternativo e câncer é investigada em larga escala (WANG Z et al. 2003; XIE et al. 2002; XU et al. 2002; HUI et al. 2004; KIRSCHBAUM-SLAGER et al. 2005). E mais recentemente alguns estudos têm aplicado a tecnologia de *microarray* com esse objetivo (RELOGIO et al. 2005; GARDINA et al. 2006; MILANI et al. 2006; LI et al. 2006; NAKAYA et al. 2007). Plataformas de *microarray* contendo exons pertencentes a variantes de *splicing* com expressão preferencial em tecido tumoral são significativamente poderosas na busca por marcadores moleculares para câncer, possibilitando a identificação de indicadores clínicos, tanto para diagnóstico, quanto para prognóstico dos tumores, do mesmo modo que poderiam ser utilizados como alvos terapêuticos.

No presente estudo foi empregada a tecnologia de *microarray* em dados obtidos por análises computacionais (KIRSCHBAUM-SLAGER et al. 2005) para identificar variantes de *splicing* superexpressas em câncer de mama. Para isso foram imobilizados em uma membrana de *nylon 270* exons pertencentes a variantes de transcritos identificados como superexpressos em bibliotecas de tecido tumoral, daí denominados de exons associados a tumores. Desses, 75 foram identificados em bibliotecas tumorais de mama, sendo portanto associados a tumor de mama. Essa plataforma foi hibridizada com amostras tumorais e normais de mama, incluindo amostras de pacientes e linhagens celulares.

Já está bem estabelecido o uso de cDNA *microarray* em membrana de *nylon* utilizando RNA total e mRNA (MEIRELES et al. 2003; LOPEZ et al. 2004; SHARMA et al. 2005). Neste trabalho, como a quantidade de RNA total obtido das amostras foi, na maioria das vezes, insuficiente para a realização dos experimentos, decidiu-se utilizar amostras de RNA amplificado. Para isso foi avaliada a correlação dos dados das intensidades médias dos 270 exons em membranas hibridizadas com RNA total e amplificado de linhagem celular de mama HB4a, apresentando valor de correlação aceitável (dados não mostrados). Há evidências demonstrando que o processo de amplificação leva a uma diminuição da representatividade da região 5' dos transcritos (NYGAARD e HOVIG 2006), devendo portanto as sondas imobilizadas se localizarem a no máximo 1.500 pb da região 3' do gene (Brentani et al. 2005). O fato de alguns exons desse estudo se posicionarem em regiões mais distantes da extremidade 3' (variando de 110

a 3.338 pb), levou-nos a preocupação com a reprodutibilidade e robustez dos dados de expressão. Para isso foram avaliadas as médias do sinal de intensidade dos exons obtidos por hibridização com RNA amplificado de acordo com sua posição no transcrito (distância 3'). O resultado obtido mostrou que o sinal de intensidade média foi independente da distância 3', não sugerindo nenhuma diferença que indicasse que as sondas mais distantes estariam sendo prejudicadas por uma diminuição de sinal (por estarem sub-representadas no RNA amplificado), o que acarretaria maiores possibilidades de variabilidade experimental desses exons. Dessa forma, decidiu-se utilizar RNA amplificado em 2 ciclos (GOMES et al. 2003) para todas as amostras.

Tem sido sugerido que a utilização da tecnologia de *microarray* requer vários controles experimentais para se alcançar resultados reprodutíveis e robustos (BRENTANI et al. 2005; ALLISON et al. 2006; CARRARO et al. 2007). Neste trabalho, com a finalidade de se controlar ao máximo as alterações decorrentes de variações experimentais, foram estabelecidos alguns critérios para monitoramento da qualidade das membranas. Os altos valores de correlação de *Pearson* (média: 0,90) obtidos entre as duplicatas experimentais revelaram uma boa reprodutibilidade entre os experimentos, garantindo a confiabilidade dos dados. Nos exons de menor intensidade foi observada uma maior dispersão entre esses valores, mesmo entre os pontos da mesma membrana, o que levou a se cogitar a sua remoção da análise, porém, como haveria perda de um número considerável de pontos, foi decidido não remover nenhum exon, mesmo aqueles com baixa intensidade

e utilizar outros mecanismos para se obter dados mais robustos. Assim, neste trabalho, várias comparações foram realizadas utilizando-se tanto amostras quanto linhagens normais e tumorais com o objetivo de identificar exons superexpressos em tumores em mais de uma comparação. Primeiramente, foi comparado o grupo total de amostras normais e tumorais (TxN), o que identificou superexpressão de 24% dos exons previamente selecionados pela análise computacional como associados a tumor de mama. Em relação à comparação de amostras pareadas (PTxPN), observou-se uma porcentagem surpreendentemente maior desses exons superexpressos (78,6%). Esse número revelou que uma grande parcela das variantes parece estar de fato superexpressa em tumor quando se controla a variabilidade biológica entre as amostras. No entanto, a representatividade dos exons associados à mama (75) em relação aos demais exons superexpressos não foi significativa em nenhuma dessas comparações. Tem sido sugerido que as alterações na geração de variantes no tumor não são eventos randômicos de processamentos incorretos do pré-mRNA e sim que a ampla maioria dessas alterações podem estar relacionadas ao processo tumorigênico de uma maneira não tecido-específica (XU e LEE 2003). Isso, de certa forma, justificaria a baixa representatividade daqueles exons previamente selecionados como associados à mama na maior parte das comparações. Já na comparação das linhagens tumorais em relação às linhagens normais (LTxLN), observou-se superexpressão de 24% dos exons associados a tumor de mama, havendo nesta comparação uma representatividade significativa dos 75 exons associados a tumor de mama

em relação aos demais (18 de 24; $p < 0,001$). Foi também observado que à medida em que se aumentou a estringência para a seleção dos candidatos, aumentou-se a representatividade desses exons. Isso foi visto entre os 20,7% de exons superexpressos que foram comuns entre 2 comparações, apresentando uma representatividade com significância estatística (24 de 56; $p < 0,001$) daqueles associados a mama. Também nos 2,6% de exons comuns entre as 3 comparações, 04 dos 07 superexpressos foram inicialmente selecionados como associados a mama ($p < 0,001$).

A concordância relativamente baixa entre as 2 abordagens, análise computacional baseada em dados de ESTs, e *microarray* através da análise de amostras e linhagens tumorais e normais, pode estar relacionada ao menor número de seqüências de ESTs referentes a amostras normais em bancos de dados (103.261 seqüências tumorais e 74.396 seqüências normais - KIRSCHBAUM-SLAGER et al. 2005), além do fato de que bancos de dados públicos não estão completos em sua cobertura de formas de *splicing* alternativo em humanos (JIN et al. 2004; XU e LEE 2003). Análises realizadas por ROY et al. (2005), mostraram que bancos de dados públicos não apresentam muitas formas de *splicing* alternativo, e este efeito é muito mais pronunciado em formas de *splicing* que estão associadas a tecidos normais, em relação às formas associadas a tumor, mostrando que algumas variantes de *splicing* tumor-associadas podem estar apenas super-representadas em relação às normais. Foi mostrado em seu estudo que aproximadamente 2/3 das variantes de *splicing* câncer-associadas

previamente desconhecidas em tecido normal apareceram expressas em amostras normais, das quais 40% foram dominantes neste tecido.

Abordagens computacionais e análises de expressão por *microarray* podem ser influenciadas por vários fatores. O fato de a plataforma desenvolvida neste trabalho apresentar exons imobilizados supostamente superexpressos em tecido tumoral impossibilitou a utilização do método de energia total para a normalização dos dados. Assim, a utilização de genes normalizadores, tais como GAPDH e ACTB pode ter introduzido um viés nos resultados, pois há relatos na literatura de que GAPDH, por exemplo, pode apresentar expressão diferencial em mama (REVILLION et al. 2000), o que de certa forma poderia explicar a baixa concordância entre os resultados experimentais e *in silico* (KIRSCHBAUM-SLAGER et al. 2005), pois as normalizações dos dados foram realizadas de maneiras diferentes.

Validação experimental através de RT-PCR desse mesmo grupo de variantes (KIRSCHBAUM-SLAGER et al. 2005) mostrou que em amostras tumorais houve um maior percentual de validação de candidatos (5 de 6: 83,3%) em relação às linhagens celulares utilizadas (4 de 10: 40%). Esses dados discordam em parte dos dados de *array* do presente estudo, uma vez que foi encontrado o mesmo número de exons (18) superexpressos associados a mama em amostras e linhagens celulares.

No estudo realizado por WANG Z et al. (2003) foi relatado que 59,2% das isoformas de *splicing* selecionadas eram específicas de tumor, ou seja, quando uma variante era detectada em um tumor, ela somente aparecia naquele tecido tumoral. Alguns estudos mostram que na maior parte dos

casos há expressão no tecido normal, mesmo que seja menor, e, portanto, essas variantes deveriam ser chamadas de variantes de *splicing* associadas a tumores (KIRSCHBAUM-SLAGER et al. 2005; ROY et al. 2005). Isso ocorre também no caso oposto, onde as variantes associadas a tecido normal podem apresentar expressão, mesmo que menor, no tecido tumoral (XU e LEE 2003).

Neste trabalho, de acordo com os critérios estabelecidos, foram selecionados 14 exons candidatos que apareceram superexpressos em amostras de câncer de mama, sendo 9 deles previamente selecionados por análise computacional como associados a tumor de mama, o que significa que 12% dos 75 apareceram superexpressos em mais de uma comparação através de análise experimental. Num primeiro momento foram selecionados a partir dos 14 exons aqueles cujos genes não apresentam superexpressão em tecido tumoral, analisados com dados disponíveis de SAGE e, em seguida, todos foram avaliados através de RT-PCR quantitativo (qRT-PCR) para seleção daqueles que apresentaram superexpressão da variante e não do gene como um todo.

Critérios foram estabelecidos para seleção das variantes superexpressas. Primeiramente, os dados de expressão da VCE deveriam apresentar *fold-change* ≥ 3 , obtido através da média do nível de expressão da VCE nas amostras tumorais em relação à média do nível de expressão da VCE nas amostras normais, para tornar mais estridente a seleção dos candidatos. Desse modo, após a realização dos experimentos de qRT-PCR foram feitas análises da expressão através da razão entre as médias de

expressão das amostras tumorais e das amostras normais (VCE tumor / VCE normal), e, seguindo-se o critério estrigente $fold \geq 3$, dos 12 genes avaliados, somente 3 (25%) mostraram superexpressão da VCE nas amostras tumorais em relação às normais, sendo eles: TRIM37 (5,69), MK-STYX (3,29) e BRRN1 (10,17). Outros genes, tais como LIP8 (2,92), PRKD2 (2,53), KIAA0676 (1,61), SND1 (1,59), BAP1 (1,56) e FLJ12528 (1,35) não preencheram o critério 1 ($fold \geq 3$), embora tenham se mostrado mais expressos em tecido tumoral em relação ao normal.

O baixo percentual de validação técnica dos dados de *microarray* (25%), comparado com outros índices de validação do nosso grupo, por volta de 60-70% (Maschietto et al. *submitted*; Castro et al. *submitted*), pode estar associado ao fato de estarmos quantificando variantes de *splicing*, e não genes como um todo. É visto na literatura taxa de validação de variantes de *splicing* selecionadas por *microarray* utilizando-se a técnica de qRT-PCR em torno de 30% (GARDINA et al. 2006), sendo similar ao percentual obtido. Se considerarmos os *fold*s resultantes da comparação TxN obtidas pelos dados de *array*, veremos que variou de 1,8 a 2,8 dentre os 14 genes selecionados para validação, e o $fold \geq 3$ (um dos critérios de seleção) foi garantido pelas demais comparações (PTXPN e LPXLN). Se o *fold* adotado na validação dos nossos candidatos fosse de 1,5, como é o caso de muitos trabalhos na literatura (SUNDE et al. 2006; MEMON et al. 2006; FROMER et al. 2006), teríamos uma validação técnica dos dados de *array* de aproximadamente 67%. Assim utilizando-se o critério de $fold \geq 3$, foram selecionadas três VCE.

O segundo critério seria: para aqueles genes que responderam ao primeiro critério (TRIM37, MK-STYX e BRRN1), o valor relativo de expressão (*fold change*) da VCE no tumor em relação ao normal deveria ser significativamente maior ($p < 0,05$) que o valor relativo de expressão da variante não contém o exon (VSE) na mesma comparação (T/N) [(VCE tumor / VCE normal) > (VSE tumor / VSE normal)]. Como foi detectada a amplificação inespecífica da VSE para os genes TRIM37 e MK-STYX, e como para o gene BRRN1 não foi possível a amplificação desta variante utilizando-se os iniciadores de junção exon1-exon3, o cálculo relativo da VCE foi realizado também em relação à exons constitutivos, isto é, presentes nas 2 variantes (VCE e VSE). Assim, para os 3 casos, foi avaliada a expressão constitutiva do gene (EC) através do desenho de iniciadores em exons constitutivos.

Partimos então para a avaliação do gene como um todo para verificar se a superexpressão se referia realmente àquela variante específica, ou ainda, mesmo no caso de haver aumento da expressão do gene como um todo no tumor, se a variante apresentava maiores diferenças, acarretando um desbalanço entre variantes do mesmo gene nas amostras tumorais em relação às normais. Para isso foi avaliada a expressão do gene como um todo nas mesmas amostras e avaliou-se a expressão relativa nas amostras tumorais em relação às normais (EC tumor / EC normal). Desse modo, as 3 VCE (TRIM37, MK-STYX e BRRN1) tiveram seus respectivos genes avaliados. Todas elas foram confirmadas como variantes de *splicing* superexpressas em câncer de mama, apresentando diferenças de expressão

significativamente maior ($p < 0,05$) nas VCEs em relação ao gene como um todo ($VCE \text{ tumor} / VCE \text{ normal} > (EC \text{ tumor} / EC \text{ normal})$), indicando um desbalanço entre elas no tumor e tecido normal de mama. Vale lembrar que o gene como um todo também teve sua expressão aumentada no tumor em relação ao normal, no entanto quando sua expressão foi comparada à expressão da VCE no tumor em relação ao normal, esta foi significativamente maior, mostrando um notável desbalanço entre elas. Tem sido demonstrado que alterações no balanço de transcritos alternativos são conhecidas por afetar a angiogênese (Woolard et al. 2004), a diferenciação celular (Sturla et al. 2003) e invasão (SINGH et al. 2004).

Uma vez validados tecnicamente, esses 3 genes foram avaliados em um grupo independente de 40 amostras tumorais. É importante destacar a grande robustez que se adquire ao se trabalhar com um grupo independente de amostras, pois demonstra que os dados obtidos não são dependentes daquele grupo específico. Validar em amostras independentes é especialmente importante em câncer de mama devido a sua alta heterogeneidade. Os dados obtidos no grupo inicial foram confirmados no grupo independente, demonstrando superexpressão da VCE dos 3 genes selecionados em amostras tumorais em relação às normais: TRIM37 (*fold change*: 6,20), MK-STYX (*fold change*: 9,57) e BRRN1 (*fold change*: 17,37). Além disso, as 3 VCE foram confirmadas como variantes de *splicing* superexpressas em câncer de mama em relação ao gene como um todo ($p < 0,05$) neste grupo de amostras, tornando os dados ainda mais confiáveis.

O gene *TRIM37* (*tripartite motife-containing 37*) apresenta 9 formas alternativas de *splicing* e está localizado na região 17q23.2 do cromossomo 17, codificando uma proteína de 964 aminoácidos com 108 kD. A proteína TRIM37 é um novo membro da subfamília RING-B-box-coiled-coil (RBCC) de proteínas de dedo-de-zinco (KALLIJÄRVI et al. 2002). Essas proteínas estão localizadas no núcleo e citoplasma, e medeiam diversos processos celulares, tais como apoptose, replicação viral, controle do ciclo celular e biogênese do peroxissomo (BORDEN e FREEMONT 1996; BORDEN 2000). Mutações gênicas que ocasionem defeitos em TRIM37 (muitas vezes resultando em uma alteração do quadro de leitura e levando a uma proteína truncada) estão associadas com nanismo *mulibrey*, também chamado de nanismo músculo-fígado-cérebro-olho, uma desordem autossomal recessiva que envolve vários tecidos de origem mesodérmica (AVELA et al. 2000). As principais características incluem falha severa no início do crescimento pré-natal e pericárdio constrictivo com conseqüente hepatomegalia. Além disso, são comuns: hipotonia do músculo, *J-shaped sella turcica*, pontos amarelados no fundo ocular, características dimórficas típicas e hipoplasia de várias glândulas endócrinas causando deficiência hormonal (*GeneCards*-<http://www.genecards.org>; *Swiss-Prot* - <http://ca.expasy.org/sprot>).

O gene *MK-STYX* está localizado na região 7q11.23 do cromossomo 7 e apresenta 10 formas alternativas de *splicing*. Codifica a proteína *phosphatase-dead dual specificity phosphatase-like* com 313 aminoácidos e 35818 Da, que por sua vez está implicada na regulação de MAP quinases. Embora sua função esteja ainda sob investigação, é uma provável

pseudofosfatase que contém um resíduo de serina ao invés de cisteína em sua alça catalítica e provavelmente interage com MAPKs de uma maneira análoga às MKPs (*Dual specificity protein phosphatases*) (WISHART e DIXON 1998), inativando suas quinases alvo por desfosforilar tanto os resíduos fosfoserina/treonina quanto os fosfotirosina. Essa proteína está superexpressa em ESFT (*Ewing's sarcoma family tumors*) (SILIGAN et al. 2005).

BRRN1 (*barren homolog-Drosophila*), está localizado na região 2q11.2 do cromossomo 2 e apresenta 8 formas alternativas de *splicing*. Codifica uma proteína de 741 aminoácidos e 82563 Da. É uma subunidade regulatória do complexo condensina, um complexo requerido para a conversão de cromatina na intérfase em cromossomos condensados na mitose, e está presente em células em proliferação. O complexo condensina introduz provavelmente *supercoils* no DNA relaxado na presença de topoisomerases tipo I e converte DNA *nicked* em formas *knotted* na presença de topoisomerases tipo II. Essa associação é requerida para a segregação cromossomal por facilitar a decatenação de cromátides na anáfase (BHAT et al. 1996).

Variantes de *splicing* de genes envolvidos em diferenciação, apoptose e proliferação celular existem freqüentemente em um equilíbrio delicado, o qual é encontrado perturbado em tumores (VENABLES 2006). O aumento das VCEs desses 3 genes e o conseqüente desbalanço entre variantes em tecidos tumorais e normais de carcinoma ductal de mama, podem ter um papel importante na tumorigênese e progressão do câncer, no entanto,

maiores investigações são necessárias para definir melhor o papel dessas variantes.

Diferentes isoformas de proteínas geradas por variantes de *splicing* podem desempenhar funções distintas ou até mesmo antagônicas nas células, podendo estar envolvidas em processos tumorigênicos de uma maneira geral, seja através da perda ou ganho de uma função, via mudanças na estrutura da proteína que podem esconder ou expor domínios funcionais, ou através da alteração da atividade de interação com alguma proteína ou complexos protéicos.

Uma variante de *splicing* do gene BRCA1, *exon 13A-containing transcript*, contém um exon adicional *in-frame* (FORTIN et al. 2005). O peptídeo adicional codificado pelo exon 13A parece modificar o domínio de transativação 1 de BRCA1 AD1 (HU et al. 2000; HU e LI 2002) além do domínio de interação de BRCA1 com as proteínas hMSH2 (*human mismatch repair protein*) (WANG et al. 2001), BRCA2 (CHEN et al. 1998) e AR (androgen receptor) (YEH et al. 2000). Sugere-se que a inserção desse exon modificaria os domínios de interação através de alterações no *folding* da proteína BRCA1 (FORTIN et al. 2005).

Outro exemplo de que variantes de *splicing* podem influenciar nas interações entre proteínas é o gene *Elk1*, pertencente à família de fatores de transcrição que apresentam o domínio ETS. Elk1 possui um domínio de interação ETS, o qual foi encontrado interagir com BRCA1. *Elk1* é um regulador do protooncogene *c-Fos* e parece que o complexo Elk1:BRCA1 age através da inibição da expressão de genes ativados por Elk1, tal como o

c-Fos. Foi observado que BRCA1 não interage com a variante de *splicing* de Elk1 (Δ Elk-1), influenciando a regulação de vários genes, dentre eles o protooncogene *c-Fos*. Sugere-se, desse modo, que Elk1, mas não a sua variante Δ Elk1, pode melhorar a atividade supressora de tumor de BRCA1 em células de câncer de mama humanas (CHAI et al. 2001).

Sabendo-se que possíveis mudanças na conformação de proteínas causadas por *splicing* alternativo, podem afetar de maneira significativa a estrutura das proteínas, sua função, localização e as interações proteína-proteína (BOUÉ et al. 2002; GARCIA et al. 2004), a investigação do efeito que *splicing* alternativo pode exercer em proteínas sabidamente envolvidas em processos celulares importantes, pode contribuir para uma melhor compreensão do envolvimento das variantes de *splicing* no processo tumorigênico. Assim, fica evidente não apenas a necessidade de se estudar a expressão de diferentes variantes de *splicing*, como também analisá-las do ponto de vista funcional.

Com o objetivo de identificar eventual presença de algum domínio funcional nas variantes que contêm o exon associado a tumor (VCE) identificadas como superexpressas neste trabalho, os domínios funcionais foram mapeados na seqüência de aminoácidos, utilizando a ferramenta *Pfam* (<http://www.sanger.ac.uk/Software/Pfam>). Essa análise mostrou que a variante que contém o exon associado a tumor de MK-STYX (de 42 aminoácidos) e de BRRN1 (de 45 aminoácidos) estavam contidos em domínios funcionais, e sugere-se que a presença do exon inserido pode, de certa forma, interferir na função desta proteína, o que poderá ser explorado

em estudos posteriores. Para o gene TRIM37, os domínios foram localizados fora da região do exon alternativo (de 45 aminoácidos), sugerindo que, se a presença do exon na variante tem um papel importante na função da proteína no câncer de mama via domínio funcional, deve ser por alteração na estrutura da proteína resultante, que pode estar, por exemplo, escondendo ou super-expondo esse domínio.

Procuramos também saber se as proteínas codificadas pelos genes TRIM37, MK-STYX e BRRN1 interagem com outras proteínas para especular se as variantes de *splicing* identificadas neste estudo como superexpressas em câncer de mama poderiam influenciar nesta interação.

TRIM37 interage com TRAF1, TRAF2, TRAF3, TRAF4, TRAF5, TRAF6, que são membros da família de proteínas de fatores associados a receptores de TNF (TNFR). Proteínas TRAF se associam com a superfamília de vários receptores de TNFR e medeiam a sua transdução de sinal. MK-STYX interage com EWS-FLI1, fator de transcrição expresso nos tumores de Ewing (ESFT). BRRN1 interage com CAPG, que codifica um membro de proteínas reguladoras de actina, as quais bloqueiam reversivelmente os filamentos de actina F, contribuindo para o controle de motilidade baseado em actina em células não musculares.

O estudo de potenciais marcadores moleculares tumorais associados ao conhecimento dos fatores prognósticos da doença é fundamental. Para isso neste trabalho foi feita a correlação dos achados com dados clínicos e histopatológicos dos pacientes. Para cada gene realizou-se comparação dos dados de expressão da variante que contém o exon associado a tumor

(VCE) e da relação entre a VCE e o fragmento correspondente à expressão constitutiva do gene (VCE/EC) com as variáveis clínicas e histopatológicas de cada paciente. Algumas associações foram encontradas, tais como a expressão de receptores de estrógeno (ER+) e progesterona (PgR+), que mostraram uma associação significativa ($p=0,009$ e $p=0,024$, respectivamente) com a VCE do gene TRIM37. Já a variável p53 negativo apresentou uma associação significativa ($p=0,040$) com a relação VCE/EC do gene TRIM37. A variável *status* menopausal negativo apresentou uma associação significativa ($p=0,018$) com a relação VCE/EC do gene MK-STYX e a ausência da expressão de receptores de progesterona (PgR-) apresentou uma associação significativa ($p=0,040$) com a VCE do gene BRRN1.

Porém os dados obtidos nesse estudo são insuficientes para inferir alguma explicação para as associações clínicas apresentadas. Isso se deve a vários motivos. Entre eles ao pequeno número de amostras avaliadas. Embora tenha sido um bom número do ponto de vista de confiabilidade dos dados obtidos, para associações confiáveis com características clínicas ainda é um número pequeno. O mais interessante seria análise por imunohistoquímica usando anticorpo específico para cada variante, podendo assim avaliar um grande número de amostras e estabelecer associações mais confiáveis.

Em resumo, embora um maior grupo de dados experimentais seja requerido para a completa exploração do estudo desenvolvido, os resultados

relatados aqui sugerem 3 variantes de *splicing* superexpressas em carcinoma ductal de mama, candidatos a marcadores moleculares.

6 CONCLUSÕES

- ✓ A tecnologia de *microarray* mostrou-se eficaz para rastrear candidatos a variantes de *splicing* superexpressas em câncer de mama, a partir de dados computacionais;
- ✓ Quanto mais estridentes foram as análises para seleção de exons superexpressos em tumor, maior a representatividade dos exons associados a câncer de mama;
- ✓ Esse estudo identificou 3 genes que apresentam variantes de *splicing* superexpressas em câncer de mama, TRIM37, MK-STYX e BRRN1;
- ✓ Os resultados de expressão apresentados foram validados em um grupo independente de amostras, o que confere grande confiabilidade dos dados;
- ✓ O estudo sugere variantes de *splicing* superexpressas em carcinoma ductal de mama candidatos a marcadores moleculares;
- ✓ Os dados obtidos nesse estudo são ainda insuficientes para inferir alguma explicação para as associações clínicas encontradas.

7 REFERÊNCIAS BIBLIOGRÁFICAS

Agrawal S, Eng C. Differential expression of novel naturally occurring splice variants of PTEN and their functional consequences in Cowden syndrome and sporadic breast cancer. **Hum Mol Genet** 2006; 15:777-87.

Albanell J, Baselga J. Unraveling resistance to trastuzumab (Herceptin): insulin-like growth factor-I receptor, a new suspect. **J Natl Cancer Inst** 2001; 93:1830-2.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. **Molecular biology of the cell**. 4th ed. New York: Garland Publishing; 2002. How Cells Read the Genome: From DNA to Protein; p.299-375.

Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. **Nat Rev Genet** 2006; 7:55-65.

Avela K, Lipsanen-Nyman M, Idänheimo N, et al. Gene encoding a new RING-B-box-Coiled-coil protein is mutated in mulibrey nanism. **Nat Genet** 2000; 25:298-301.

Barbati A, Cosmi EV, Sidoni A, et al. Value of c-erbB-2 and p53 oncoprotein co-overexpression in human breast cancer. **Anticancer Res** 1997; 17:401-5.

Bardou VJ, Arpino G, Elledge RM, Osborne CK, Clark GM. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. **J Clin Oncol** 2003; 21:1973-9.

Baruch A, Hartmann M, Yoeli M, et al. The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein. **Cancer Res** 1999; 59:1552-61.

Baudry D, Hamelin M, Cabanis MO, et al. WT1 splicing alterations in Wilm's tumors. **Clin Cancer Res** 2000; 6:3957-65.

Bhat MA, Philp AV, Glover DM, Bellen HJ. Chromatid segregation at anaphase requires the barren product, a novel chromosome-associated protein that interacts with Topoisomerase II. **Cell** 1996; 87:1103-14.

Black DL. Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. **Cell** 2000; 103:367-70.

Black DL. Mechanisms of alternative pre-messenger RNA splicing. **Annu Rev Biochem** 2003; 72:291-336.

Borden KL, Freemont PS. The RING finger domain: a recent example of a sequence-structure family. **Curr Opin Struct Biol** 1996; 6:395-401.

Borden KL. RING domains: master builders of molecular scaffolds? **J Mol Biol** 2000; 295:1103-12. Review. Erratum in: **J Mol Biol** 2000; 297:1027.

Boué S, Vingron M, Kriventseva E, Koch I. Theoretical analysis of alternative splice forms using computational methods. **Bioinformatics** 2002; 18 Suppl 2:S65-73.

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

Brett D, Pospisil H, Valcarcel J, et al. Alternative splicing and genome complexity. **Nat Genet** 2002; 30:29-30.

Brinkman BM. Splice variants as cancer biomarkers. **Clin Biochem** 2004; 37:584-94.

Buchhagen DL, Qiu L, Etkind P. Homozygous deletion, rearrangement and hypermethylation implicate chromosome region 3p14.3-3p21.3 in sporadic breast-cancer development. **Int J Cancer** 1994; 57:473-9.

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

Buzdar A, O'Shaughnessy JA, Booser DJ, et al. Phase II, randomized, double-blind study of two dose levels of arzoxifene in patients with locally advanced or metastatic breast cancer. **J Clin Oncol** 2003; 21:1007-14.

Caballero OL, de Souza SJ, Brentani RR, Simpson AJ. Alternative spliced transcripts as cancer markers. **Dis Markers** 2001; 17:67-75.

Caceres JF, Kornbliht AR. Alternative splicing: multiple control mechanisms and involvement in human disease. **Trends Genet** 2002; 18:186-93.

Camp RL, Dolled-Filhart M, King BL, Rimm DL. Quantitative analysis of breast cancer tissue *microarrays* shows that both high and normal levels of HER2 expression are associated with poor outcome. **Cancer Res** 2003; 63:1445-8.

Carraro DM, Brentani HP, Soares FA, Reis LFL, Brentani RR. From Tissue Samples to Tumor Markers. In: Krishanarao A, organizador. **From basics to diagnostics**. New Jersey; Humana Press; 2007. p.17-28.

Carninci P, Kasukawa T, Katayama S, et al. Genome science group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. **Science** 2005; 309:1559-63. Erratum in: **Science** 2006; 311:1713.

Carraway HE, Wang S, Blackford A, et al. Promoter hypermethylation in sentinel lymph nodes as a marker for breast cancer recurrence. **Breast Cancer Res Treat** 2008.

Castro NP, Osório CABT, Torres C, et al. Evidence of molecular program of ductal carcinoma progression taking place before morphological alterations. 2008 (*Submitted*).

Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. **Lancet** 2003; 362:362-9.

Chai Y, Chipitsyna G, Cui J, et al. c-Fos oncogene regulator Elk-1 interacts with BRCA1 splice variants BRCA1a/1b and enhances BRCA1a/1b-mediated growth suppression in breast cancer cells. **Oncogene** 2001; 20:1357-67.

Chen J, Silver DP, Walpita D, et al. Stable interaction between the products of the BRCA1 and BRCA2 tumor suppressor genes in mitotic and meiotic cells. **Mol Cell** 1998; 2:317-28.

Cho HS, Mason K, Ramyar KX, et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. **Nature** 2003; 421:756-60.

Church GM, Gilbert W. Genomic sequencing. **Proc Natl Acad Sci U S A** 1984; 81:1991-5.

Clarke R, Leonessa F, Welch JN, Skaar TC. Cellular and molecular pharmacology of antiestrogen action and resistance. **Pharmacol Rev** 2001; 53:25-71.

Clarke M, Collins R, Darby S, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). **Lancet** 2005; 366:2087-106. Review.

Compton CC, Fielding LP, Burgart LJ, et al. Prognostic factors in colorectal cancer, College of American Pathologists Consensus Statement 1999. **Arch Pathol Lab Med** 2000; 124:979-94.

Cooper DL. Retention of CD44 introns in bladder cancer: understanding the alternative splicing of pre-mRNA opens new insights into the pathogenesis of human cancers. **J Pathol** 1995; 177:1-3.

Corvello CM. Genética molecular do câncer de mama: predisposição e prognóstico. In: Rossi BM, Pinho M, editores. **Genética e biologia para o cirurgião**. São Paulo: Lemar; 1999. p.201-11.

Costa FF, Verbisck NV, Salim AC, et al. Epigenetic silencing of the adhesion molecule ADAM23 is highly frequent in breast tumors. **Oncogene** 2004; 23:1481-8.

Costa LJM, del Giglio A. Marcadores moleculares em oncologia. In: Brentani MM, Coelho FRG, Kowalski LP, editores. **Bases da oncologia**. São Paulo: Lemar; 2003. p.309-26.

Cragg MS, Chan HT, Fox MD, et al. The alternative transcript of CD79b is overexpressed in B-CLL and inhibits signaling for apoptosis. **Blood** 2002; 100:3068-76.

Cuperlovic-Culf M, Belacel N, Culf AS, Ouellette RJ. Microarray analysis of alternative splicing. **OMICS** 2006; 10:344-57.

de 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.

Ding, W.Q., Cheng, Z.J., McElhiney, J., Kuntz, S.M., Miller, L.J. Silencing of secretin receptor function by dimerization with a misspliced variant secretin receptor in ductal pancreatic adenocarcinoma. **Cancer Res** 2002; 18:5223-9.

Doussal VL, Tubiana-Hulin M, Friedman S, Hacene K, Spyrtos F, Brunet M. Prognostic value of histologic grade nuclear components of Scarff-Bloom-Richardson (SBR). **Cancer** 1989; 64:1914-21.

Duffy MJ. Predictive markers in breast and other cancers: a review. **Clin Chem** 2005; 51:494-503.

El Shamy WM, Livingston DM. Identification of BRCA1-IRIS, a BRCA1 locus product. **Nat Cell Biol** 2004; 6:954-67.

Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression based predictors for breast cancer. **N Engl J Med** 2006; 355:560-6.

Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. **Genes Dev** 2003; 17:419-37.

Feneley MR, Partin AW. Diagnosis of localized prostate cancer: 10 years of progress. **Curr Opin Urol** 2000; 10:319-32.

Ferreira EN, Galante PA, Carraro DM, de Souza SJ. Alternative splicing: a bioinformatics perspective. **Mol Biosyst** 2007; 3:473-7.

Fisher B, Dignam J, Bryant J, Wolmark N. Five versus more than five years of tamoxifen for lymph node-negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel Project B-14 randomized trial. **J Natl Cancer Inst** 2001; 93:684-90.

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

Folgueira MA, Carraro DM, Brentani H, et al. Gene expression profile associated with response to doxorubicin-based therapy in breast cancer. **Clin Cancer Res** 2005; 11:7434-43.

Fortin J, Moisan AM, Dumont M, et al. A new alternative splice variant of BRCA1 containing an additional in-frame exon. **Biochim Biophys Acta** 2005; 1731:57-65.

Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ. Detection and evaluation of intron retention events in the human transcriptome. **RNA** 2004; 10:757-65.

Garcia A, Cayla X, Caudron B, Deveaud E, Roncal F, Rebollo A. New insights in protein phosphorylation: a signature for protein phosphatase 1 interacting proteins. **C R Biol** 2004; 327:93-7.

Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. **Nat Biotechnol** 2004; 22:535-46.

Gardina PJ, Clark TA, Shimada B, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. **BMC Genomics** 2006; 7:325.

Ge K, DuHadaway J, Du W, Herlyn M, Rodeck U, Prendergast GC. Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. **Proc Natl Acad Sci U S A** 1999; 96:9689-94.

Glisin V, Crkvenjakov R, Byus C. Ribonucleic acid isolated by cesium chloride centrifugation. **Biochemistry** 1974; 13:2633-7.

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

Graveley BR. Sex, AGility, and the regulation of alternative splicing. **Cell** 2002; 109:409-12.

Greenblatt MS, Bennett WP, Hollstein M, Harris CC. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. **Cancer Res** 1994; 54:4855-78.

Gupta S, Zink D, Korn B, Vingron M, Haas SA. Genome wide identification and classification of alternative splicing based on EST data. **Bioinformatics** 2004; 20:2579-85.

Hanamura A, Caceres JF, Mayeda A, Franza BR Jr, Krainer AR. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. **RNA** 1998; 4:430-44.

Harris RA, Hiles ID, Page MJ, O'Hare MJ. The induction of apoptosis in human mammary luminal epithelial cells by expression of activated c-neu and its abrogation by glucocorticoids. **Br J Cancer** 1995; 72:386-92.

Hashizume R, Fukuda M, Maeda I, et al. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. **J Biol Chem** 2001; 276:14537-40.

Herrera-Gayol A, Jothy S. Adhesion proteins in the biology of breast cancer: contribution of CD44. **Exp Mol Pathol** 1999; 66:149-56.

Herrlich P, Morrison H, Sleeman J, et al. CD44 acts both as a growth- and invasiveness-promoting molecule and as a tumor-suppressing cofactor. **Ann N Y Acad Sci** 2000; 910:106-18; discussion 118-20.

Hoffman JD, Hallam SE, Venne VL, Lyon E, Ward K. Implications of a novel cryptic splice site in the BRCA1 gene. **Am J Med Genet** 1998; 80:140-4.

Howel L, Gandour-Edwards R, O'Sullivan D. Application of the Scarff-Bloom-Richardson tumor grading system to fine needle aspirates of the breast. **Am J Clin Pathol** 1994; 101:262-5.

Hsu FR, Chang HY, Lin YL, et al. AVATAR: a database for genome-wide alternative splicing event detection using large scale ESTs and mRNAs. **Bioinformatics** 2005; 1:16-18.

Hu GK, Madore SJ, Moldover B, et al. Predicting splice variant from DNA chip expression data. **Genome Res** 2001; 11:1237-45.

Hu YF, Li R. JunB potentiates function of BRCA1 activation domain 1 (AD1) through a coiled-coil-mediated interaction. **Genes Dev** 2002; 16:1509-17.

Hui L, Zhang X, Wu X, et al. Identification of alternatively spliced mRNA variants related to cancers by genome-wide ESTs alignment. **Oncogene** 2004; 23:3013-23.

Huggett J, Dheda K, Bustin S, Zumla A. Real-time RT-PCR normalisation; strategies and considerations. **Genes Immun** 2005; 6:279-84.

Jensen DE, Proctor M, Marquis ST, et al. BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. **Oncogene** 1998; 16:1097-112.

Jensen DE, Rauscher FJ 3rd. BAP1, a candidate tumor suppressor protein that interacts with BRCA1. **Ann N Y Acad Sci** 1999; 886:191-4.

Jin P, Fu GK, Wilson AD, et al. PCR isolation and cloning of novel splice variant mRNAs from known drug target genes. **Genomics** 2004; 83:566-71.

Johnson JM, Castle J, Garrett-Engle P, et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction *microarrays*. **Science** 2003; 302:2141-4.

Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. **Mol Cell** 2003; 12:5-14.

Kallijärvi J, Avela K, Lipsanen-Nyman M, Ulmanen I, Lehesjoki AE. The TRIM37 gene encodes a peroxisomal RING-B-box-coiled-coil protein: classification of mulibrey nanism as a new peroxisomal disorder. **Am J Hum Genet** 2002; 70:1215-28.

Kalnina Z, Zayakin P, Silina K, Linē A. Alterations of pre-mRNA splicing in cancer. **Genes Chromosomes Cancer** 2005; 42:342-57.

Kampa D, Cheng J, Kapranov P, et al. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. **Genome Res** 2004; 14:331-42.

Kim E, Magen A, Ast G. Different levels of alternative splicing among eukaryotes. **Nucleic Acids Res** 2007; 35:125-31.

Kim E, Goren A, Ast G. Alternative splicing and disease. **RNA Biol** 2008; 5.

Kirschbaum-Slager N, Parmigiani RB, Camargo AA, de Souza SJ. Identification of human exons over-expressed in tumors through the use of genome and expressed sequence data. **Physiol Genomics** 2005; 21:423-32.

Koduri S, Goldhar AS, Vonderhaar BK. Activation of vascular endothelial growth factor (VEGF) by the ER-alpha variant, ERDelta3. **Breast Cancer Res Treat** 2006; 95:37-43.

Kumar CC, Madison V. Drugs targeted against protein kinases. **Expert Opin Emerg Drugs** 2001; 6:303-15.

Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. **Nature** 2001; 409:860-921.

Li C, Kato M, Shiue L, Shively JE, Ares M Jr, Lin RJ. Cell type and culture condition-dependent alternative splicing in human breast cancer cells revealed by splicing-sensitive microarrays. **Cancer Res** 2006; 66:1990-9.

Lonning PE, Knappskog S, Staalesen V, Chrisanthar R, Lillehaug JR. Breast cancer prognostication and prediction in the postgenomic era. **Ann Oncol** 2007; 18:1293-306.

Lu Y, Yao HP, Wang MH. Multiple variants of the RON receptor tyrosine kinase: biochemical properties, tumorigenic activities, and potential drug targets. **Cancer Lett** 2007; 257:157-64.

Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. **Proc Natl Acad Sci U S A**. 2003; 100:5974-9.

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.

Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. **Nature** 2002; 418:236-43.

Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. **Nat Rev Mol Cell Biol** 2005; 6:386-98.

Matthews J, Gustafsson JA. Estrogen signaling: a subtle balance between ER alpha and ER beta. **Mol Interv** 2003; 3:281-92. Review.

Megha T, Ferrari F, Benvenuto A, et al. p53 mutation in breast cancer: correlation with cell kinetics and cell of origin. **J Clin Pathol** 2002; 55:461-6.

Meireles SI, Carvalho AF, Hirata R, et al. Differentially expressed genes in gastric tumors identified by cDNA array. **Cancer Lett** 2003; 190:199-211.

Memon AA, Sorensen SB, Nexø E. The epidermal growth factor family has a dual role in deciding the fate of cancer cells. **Scand J Clin Lab Invest** 2006; 66:623-30.

Milani L, Fredriksson M, Syvanen AC. Detection of alternatively spliced transcripts in leukemia cell lines by minisequencing on microarrays. **Clin Chem** 2006; 52:202-11.

Minneman KP. Splice variants of G protein-coupled receptors. **Mol Interv** 2001; 1:108-16.

Ministério da Saúde. Instituto Nacional do Câncer. **Câncer de mama**. Disponível em: <URL:http://www.inca.gov.br/conteudo_view.asp?id=336>. [2008 mar 13].

Modrek B, Lee C. A genomic view of alternative splicing. **Nat Genet** 2002; 30:13-9.

Mokbel K. Focus on anastrozole and breast cancer. **Curr Med Res Opin** 2003; 19:683-8.

Molina R, Barak V, van Dalen A, et al. Tumor markers in breast cancer-European Group on Tumor Markers recommendations. **Tumour Biol** 2005; 26:281-93.

Morrison TB, Wei, JJ e Wittwer, C.T. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. **Biotechniques** 1998; 954-62.

Nakamura S, Gomyo Y, Roth JA. *C-terminus of p53 is required for G(2) arrest*. **Oncogene** 2002; 27:2102-7.

Nakaya HI, Beckedorff FC, Baldini ML, Fachel AA, Reis EM, Verjovski-Almeida S. Splice variants of TLE family genes and up-regulation of a TLE3 isoform in prostate tumors. **Biochem Biophys Res Commun** 2007; 364:918-23.

Naor D, Nedvetzki S, Golan I, Melnik L, Faitelson Y. CD44 in cancer. **Crit Rev Clin Lab Sci** 2002; 39:527-79.

Nishidate T, Katagiri T, Lin ML, et al. Genome-wide gene-expression profiles of breast-cancer cells purified with laser microbeam microdissection: identification of genes associated with progression and metastasis. **Int J Oncol** 2004; 25:797-819.

Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? **Bio Essays** 2003; 1147-9.

Nygaard V, Hovig E. Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. **Nucleic Acids Res** 2006; 34:996-1014.

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

Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. **Nat Rev Genet** 2004; 5:389-96.

Pan Q, Bakowski MA, Morris Q, et al. Alternative splicing of conserved exons is frequently species-specific in human and mouse. **Trends Genet** 2005; 21:73-7.

Parl FF. Multiple mechanisms of estrogen receptor gene repression contribute to ER-negative breast cancer. **Pharmacogenomics J** 2003; 3:251-3.

Petit T, Borel C, Ghnassia JP, et al. Chemotherapy response of breast cancer depends on HER-2 status and anthracycline dose intensity in the neoadjuvant setting. **Clin Cancer Res** 2001; 7:1577-81.

Pettigrew CA, Brown MA. Pre-mRNA splicing aberrations and cancer. **Front Biosci** 2008; 13:1090-105.

Phillips KA. Current perspectives on BRCA1- and BRCA2-associated breast cancers. **Intern Med J** 2001; 31:349-56.

Posner MR, Mayer RJ. The use of serologic tumor markers in gastrointestinal malignancies. **Hematol Oncol Clin North Am** 1994; 8:533-53.

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

Religio A, Ben-Dov C, Baum M, et al. Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. **J Biol Chem** 2005; 280:4779-84.

Revillion F, Pawlowski V, Hornez L, Peyrat JP. Glyceraldehyde-3-phosphate dehydrogenase gene expression in human breast cancer. **Eur J Cancer** 2000; 36:1038-42.

Rolland P, Spendlove I, Madjid Z, Rakha EA, Patel P, Ellis IO, Durrant L. The p53 positive Bcl-2 negative phenotype is an independent marker of prognosis in breast cancer. **Int J Cancer** 2007; 120:1311-7.

Rosen EM, Fan S, Isaacs C. BRCA1 in hormonal carcinogenesis: basic and clinical research. **Endocr Relat Cancer** 2005; 12:533-48.

Roy M, Xu Q, Lee C. Evidence that public database records for many cancer-associated genes reflect a splice form found in tumors and lack normal splice forms. **Nucleic Acids Res** 2005; 33:5026-33.

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

Sampath J, Pelus LM. Alternative splice variants of survivin as potential targets in cancer. **Curr Drug Discov Technol** 2007; 4:174-91.

Schena M. **Microarray analysis: introduction to microarray analysis**. New Jersey: Johan Willey & Sons; 2003.

Scheper W, Zwart R, Baas F. Alternative splicing in the N-terminus of Alzheimer's presenilin 1. **Neurogenetics** 2004; 5:223-7.

Scheurlen WG, Senf L. Analysis of the GAP-related domain of the neurofibromatosis type 1 (NF1) gene in childhood brain tumors. **Int J Cancer** 1995; 64:234-8.

Shimizu D, Ishikawa T, Ichikawa Y, et al. Current progress in the prediction of chemosensitivity for breast cancer. **Breast Cancer** 2004; 11:42-8.

Shipp M, Harrington D, Anderson J, et al. A predictive model for aggressive non Hodgkin's lymphoma. **N Engl J Med** 1993; 329:987-4.

Shulzhenko N, Smirnova AS, Morgun A, Gerbase-DeLima M. Specificity of alternative splice form detection using RT-PCR with a primer spanning the exon junction. **Biotechniques** 2003; 34:1244-9.

Siligan C, Ban J, Bachmaier R, et al. EWS-FLI1 target genes recovered from Ewing's sarcoma chromatin. **Oncogene** 2005; 24:2512-24.

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

Singh A, Karnoub AE, Palmby TR, Lengyel E, Sondek J, Der CJ. Rac1b, a tumor associated, constitutively active Rac1 splice variant, promotes cellular transformation. **Oncogene** 2004; 58:9369-80.

Slamon D, Pegram M. Rationale for trastuzumab (Herceptin) in adjuvant breast cancer trials. **Semin Oncol** 2001; 28(1 Suppl 3):13-9.

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

Smart CR, Byrne C, Smith RA, et al. Twenty-year follow-up of the breast cancers diagnosed during the Breast Cancer Detection Demonstration Project. **CA Cancer J Clin** 1997; 47:134-49.

Sneath RJ, Mangham DC. The normal structure and function of CD44 and its role in neoplasia. **Mol Pathol** 1998; 51:191-200.

Sorek R, Shamir R, Ast G. How prevalent is functional alternative splicing in the human genome? **Trends Genet** 2004; 20:68-71.

Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. **Proc Natl Acad Sci U S A** 2003; 100:10393-8.

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.

Stickeler E, Kittrell F, Medina D, Berget SM. Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. **Oncogene** 1999; 18:3574-82.

Stimpfl M, Tong D, Fasching B, et al. Vascular endothelial growth factor splice variants and their prognostic value in breast and ovarian cancer. **Clin Cancer Res** 2002; 8:2253-9.

Sturla LM, Merrick AE, Burchill SA. FGFR3IIIS: a novel soluble FGFR3 spliced variant that modulates growth is frequently expressed in tumour cells. **Br J Cancer** 2003; 7:1276-84.

Styblo TM, Wood WC. Traditional prognostic factors in breast cancer In: Bland KI, Copeland III EM, editors. **The breast: comprehensive management of benign and malignant diseases**. 2nd ed. Philadelphia: W B Saunders; 1998. p.419-25.

Sunde JS, Donniger H, Wu K, et al. Expression profiling identifies altered expression of genes that contribute to the inhibition of transforming growth factor-beta signaling in ovarian cancer. **Cancer Res** 2006; 66:8404-12.

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

Thai TH, Kearney JF. Distinct and opposite activities of human terminal deoxynucleotidyltransferase splice variants. **J Immunol** 2004; 173:4009-19.

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

Thorsen K, Sørensen KD, Brems-Eskildsen AS. Alternative splicing in colon, bladder, and prostate cancer identified by exon-array analysis. **Mol Cell Proteomics** 2008.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW. A gene-expression signature as a predictor of survival in breast cancer. **N Engl J Med** 2002; 25:1999-2009.

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:RESEARCH0034.

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

Van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. **Nature** 2008; 452:564-70.

Venables JP. Aberrant and alternative splicing in cancer. **Cancer Res** 2004; 64:7647-54.

Venables JP. Unbalanced alternative splicing and its significance in cancer. **Bioessays** 2006; 28:378-86.

Vêncio RZ, Brentani H, Patrão DF, Pereira CA. Bayesian model accounting for within-class biological variability in Serial Analysis of Gene Expression (SAGE). **BMC Bioinformatics** 2004; 5:119.

Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. **Cell** 2002; 108:171-82.

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

Wagner KD, Wagner N, Schedl A. The complex life of WT1. **J Cell Sci** 2003; 116:1653-8.

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

Wang Q, Zhang H, Guerrette S, et al. Adenosine nucleotide modulates the physical interaction between hMSH2 and BRCA1. **Oncogene** 2001; 20:4640-9.

Wang L, Duke L, Zhang PS, et al. Alternative splicing disrupts a nuclear localization signal in spleen tyrosine kinase that is required for invasion suppression in breast cancer. **Cancer Res** 2003; 63:4724-30.

Wang Z, Lo HS, Yang H, et al. Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. **Cancer Res** 2003; 63:655-7.

Watahiki A, Waki K, Hayatsu N, et al. Libraries enriched for alternatively spliced exons reveal splicing patterns in melanocytes and melanomas. **Nat Methods** 2004; 1:233-9.

Weinert T. DNA damage and checkpoint pathways: molecular anatomy and interactions with repair. **Cell** 1998; 94:555-8.

Wieben ED. Primer on medical genomics. Part VII: The evolving concept of the gene. **Mayo Clin Proc** 2003; 78:580-7.

Wikstrand CJ, Hale LP, Batra SK, et al. Monoclonal antibodies against EGFRvIII are tumor specific and react with breast and lung carcinomas and malignant gliomas. **Cancer Res** 1995; 55:3140-8.

Wishart MJ, Dixon JE. Gathering STYX: phosphatase-like form predicts functions for unique protein-interaction domains. **Trends Biochem Sci** 1998; 23:301-6.

Wolff AC, Davidson NE. Primary systemic therapy in operable breast cancer. **J Clin Oncol** 2000; 18:1558-69.

Woolard J, Wang WY, Bevan HS, Qiu Y, Morbidelli L, Pritchard-Jones RO et al. VEGF165b, an inhibitory vascular endothelial growth factor splice variant: mechanism of action, in vivo effect on angiogenesis and endogenous protein expression. **Cancer Res** 2004; 21:7822-35.

Wooster R, Weber BL. Breast and ovarian cancer. **N Engl J Med** 2003; 348:2339-47.

Xie H, Zhu WY, Wasserman A, Grebinskiy V, Olson A, Mintz L. Computational analysis of alternative splicing using EST tissue information. **Genomics** 2002; 80:326-30.

Xin Y, Grace A, Gallagher MM, Curran BT, Leader MB, Kay EW. CD44V6 in gastric carcinoma: a marker of tumor progression. **Appl Immunohistochem Mol Morphol** 2001; 9:138-42.

Xu Q, Lee C. Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. **Nucleic Acids Res** 2003; 31:5635-43.

Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. **Nucleic Acids Res** 2002; 30:3754-66.

Yamashita H, Yando Y, Nishio M, et al. Immunohistochemical evaluation of hormone receptor status for predicting response to endocrine therapy in metastatic breast cancer. **Breast Cancer** 2006; 13:74-83.

Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. **Nat Rev Mol Cell Biol** 2001; 2:127-37.

Yeh S, Hu YC, Rahman M, et al. Increase of androgen-induced cell death and androgen receptor transactivation by BRCA1 in prostate cancer cells. **Proc Natl Acad Sci U S A** 2000; 97:11256-61.

Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. **Cancer Sci** 2004; 95:866-71.

Anexo 1 - cTNM- Classificação Clínica Baseada em evidências anteriores ao tratamento

O estadiamento pelo sistema **TNM** (*Classification of Malignant Tumors*) é o método mais utilizado na classificação de tumores malignos e a descrição de sua extensão anatômica, desenvolvido e publicado pela União Internacional Contra o Câncer - UICC. Para descrever a extensão anatômica da doença, o sistema TNM é baseado na avaliação de três componentes: **T**- extensão do tumor primário, **N**- ausência ou presença de metástases em linfonodos regionais e **M**- ausência ou presença de metástases à distância.

T- Características do tumor

TX- Tumor primário não pode ser avaliado

T0- Não há evidência de tumor primário.

Tis- Carcinoma *in situ*.

T1- Tumor de 2cm ou menos em sua maior dimensão.

T2- Tumor de mais de 2cm, porém não mais de 5cm em sua maior dimensão.

T3- Tumor de mais de 5cm em sua maior dimensão.

T4- Tumor de qualquer dimensão, com extensão direta à parede torácica ou à pele.

N- Linfonodos Regionais

NX- Linfonodos regionais não podem ser avaliados.

N0- Ausência de metástase em linfonodos regionais.

N1- Metástase em linfonodo(s) axilar(es), homolateral(ais), móvel(eis).

N2- Metástase em linfonodo(s) axilar(es), homolateral(ais), fixo(s) uns aos outros ou a outras estruturas.

N3- Metástase em linfonodo(s) mamário(s) interno(s), homolateral(ais).

M- Metástase à distância

MX- A presença de metástase à distância não pode ser avaliada.

M0- Ausência de metástase à distância.

M1- Metástase à distância.

TNM- Grupamento por Estádios

Estadio 0	Tis	N0	M0
Estadio I	T1	N0	M0
Estadio IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Estadio IIB	T2	N1	M0
	T3	N0	M0
Estadio IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1, N2	M0
Estadio IIIB	T4	Qualquer N	M0
	Qualquer T	N3	M0
Estadio IV	Qualquer T	Qualquer N	M1

Fonte: SOBIN e WITTEKIND (2002)

Anexo 2 - pTNM- Classificação Patológica

A classificação histopatológica pós-cirúrgica é baseada nas evidências observadas antes do tratamento, suplementada ou modificada pela evidência adicional alcançada através da cirurgia e do exame histopatológico. A avaliação histopatológica do tumor primário (**pT0**) exige a ressecção do tumor primário ou biópsia adequada para avaliar a maior categoria **pT**. A avaliação histopatológica dos linfonodos regionais (**pN**) exige a remoção representativa de nódulos para comprovar a ausência de metástases em linfonodos regionais (**pN0**) e suficiente para avaliar a maior categoria **pN**. A investigação histopatológica de metástases à distância (**pM0**) exige o exame macroscópico.

pT- Características do tumor

As categorias **pT** correspondem às categorias **T**.

pN- Linfonodos Regionais

pNX- Linfonodos regionais não podem ser avaliados (não removidos para estudo ou previamente removidos).

pN0- Ausência de metástase em linfonodos regionais.

pN1-Metástase em linfonodo(s) axilar(es), homolateral(ais), móvel(eis) [PN1a, PN1b (PN1bi, PN1bii, PN1biii, Pnbiv)].

pN2- Metástase em linfonodo(s) axilar(es), homolateral(ais), fixo(s) uns aos outros ou a outras estruturas.

pN3-Metástase em linfonodo(s) mamário(s) interno(s), homolateral(ais).

pM- Metástase à distância

As categorias **pM** correspondem às categorias **M**.

G- Graduação Histopatológica

GX- O grau de diferenciação não pode ser avaliado.

G1- Bem diferenciado.

G2- Moderadamente diferenciado.

G3- Pouco diferenciado.

G4- Indiferenciado.

R- Classificação R- ausência ou presença de tumor residual após tratamento

RX- A presença de tumor residual não pode ser avaliada.

R0- Ausência de tumor residual.

R1- Tumor residual microscópico.

R2- Tumor residual macroscópico.

(UICC- União Internacional Contra o Câncer- TNM- *Classification of malignant tumors*, SOBIN & WITTEKIND, 2002).

Anexo 3 - Parecer do Comitê de Ética em Pesquisa

CENTRO DE TRATAMENTO E PESQUISA
HOSPITAL
DO CANCER
A. C. CAMARGO

São Paulo, 01 de abril de 2004.

À
Dra. Maria Cristina Rodrigues Rangel

Ref.: Projeto de Pesquisa n.º 586/04
“ Identificação de marcadores moleculares em câncer de mama através da técnica de microarray utilizando uma plataforma de exons tumor-associado”.

Prezada Doutora:

Seu projeto de pesquisa, acima mencionado, foi apreciado pela Comissão de Ética em Pesquisa (CEP) do Hospital do Câncer em sua última reunião de 31.03.2004. Os membros desta comissão **aprovaram** a realização deste estudo.

Informações a respeito do andamento do referido projeto deverão ser encaminhados à secretaria do CEP dentro de 12 meses.

Atenciosamente,


Dr. Agnaldo Anelli
Presidente da Comissão de Ética em Pesquisa

C.C.
Orientadora: *Dra. Dirce Maria Carraro*
Co-orientadora: *Dra. Helena Brentani*

Anexo 4 - Consentimento informado



TERMO DE INFORMAÇÃO E CONSENTIMENTO

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

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

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

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

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

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

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

D E C L A R A Ç Ã O

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

São Paulo, de de

.....
Responsável ou Paciente

Nome:.....
RG:.....
RGH:.....

Anexo 5 - Gel de Agarose 1%

- 0,30 gr de agarose
- 30 ml de TBE (Tris;Acido Borico;EDTA)

Misturar os reagentes, esquentar no microondas até a agarose estar totalmente diluída (cerca de 1 minuto), aguardar esfriar e acrescentar 0,01% de brometo de etideo. Transferir para uma cuba de eletroforese.

Aguardar o gel solidificar e aplicar as amostras. Correr durante 60 minutos na voltagem de 60 mV.

Anexo 6 - Amostras Iniciais

Quadro 1 - Relação das 27 amostras utilizadas.

Nome	Estadiamento	Idade	TNM	LN	Grau	Marcadores Moleculares
1T	Ila	69 anos	T2N0M0	Negativo	grau II de SBR grau nuclear:3	ER -/ PR -/ p53 +/ ERBB2+ (3+)
2T	Ila	63 anos	T2N0M0	Negativo	grau I de SBR grau nuclear:1	ER+/ PR+/ p53 -/ ERBB2- (1+)
4T	Ila	67 anos	T1cN0M0	Negativo	grau II de SBR grau nuclear:2	ER +/ PR -/ p53 +/ ERBB2+ (2+)
2B	Ila	48 anos	T2N0M0	Negativo	grau II de SBR grau nuclear:3	ER+/ PR+/ p53 +/ ERBB2+
6T	Ila	78 anos	T1N1M0	Positivo	grau III de SBR grau nuclear:3	ER+/ PR+/ p53 -/ ERBB2- (1+)
7T	Ila	40 anos	T2N0M0	Negativo	grau II de SBR grau nuclear:3	ER+/ PR+/ p53 +/ ERBB2+ (2+)
8T	Ila	43 anos	T1cN1M0	Positivo	grau I de SBR grau nuclear:2	ER+/ PR+/p53 +/ ERBB2+ (3+)
9T	Ila	55 anos	T2N0M0	Negativo	grau I de SBR grau nuclear:2	ER +/ PR +/ p53 -/ ERBB2+ (3+)
3B	Ila	43 anos	T1cN0M0	Negativo	grau II de SBR grau nuclear:3	ER +/ PR -/ p53 -/ ERBB2+ (3+)
11T	Ila	46 anos	T3N0M0	Negativo	grau II de SBR grau nuclear:2	ER +/ PR +/ p53 -/ ERBB2+ (2+)
12T	Ila	78 anos	T2N0M0	Negativo	grau II de SBR grau nuclear:2	ER +/ PR +/ p53 -/ ERBB2+ (2+)
14T	Ila	70anos	T1N0M0	Negativo	grau II de SBR grau nuclear:3	ER +/ PR +/ p53 -/ ERBB2+ (2+)
17T	Ilb	36 anos	T2N1M0	Positivo	grau II de SBR grau nuclear:3	ER +/ PR+/ p53 +/ ERBB2+ (2+)
18T	Ilb	41 anos	T2N0M0	Negativo	grau II de SBR grau nuclear:2	ER +/ PR +/ p53 +/ ERBB2+ (3+)
19T	Ila	67 anos	T2N0M0	Negativo	grau II de SBR grau nuclear:3	ER -/ PR -/ p53 -/ ERBB2+ (3+)
20T	Ilb	87 anos	T2N0M0	Negativo	grau II de SBR grau nuclear:2	ER +/ PR -/ p53 -/ ERBB2+ (3+)
21T	Ilb	67 anos	T2N1M0	Positivo	grau II de SBR grau nuclear:2	ER+/ PR+/ p53/ ERBB2+
22T	Ilb	56 anos	T2N1M0	Positivo	grau III de SBR grau nuclear:3	ER +/ PR -/ p53 -/ ERBB2+ (2+/3+)
26T	Ilb	32 anos	T3N1M0	Positivo	grau II de SBR grau nuclear:3	ER +/ PR +/ p53+/ ERBB2+ (2+)
27T	Ilb	45 anos	T2N1M0	Positivo	grau nuclear:3	ER -/ PR -
28T	IIla	42 anos	T2N2M0	Positivo	grau II SBR grau nuclear III	ER +/ PR -/ p53 -/ ERBB2+ (+3)

30T	Ila	35 anos	T2N0M0	Negativo	grau III de SBR grau nuclear:3	ER -/ PR -/ p53 -/ ERBB2- (0)
32T	I	59 anos	T2N0M0	Positivo	grau I de SRB grau nuclear:2	ER +/- PR +/- p53 -/ ERBB2+ (2+)
33B	I	38 anos	T1cN0M0	Positivo	grau III de SBR grau nuclear:3	ER +/- PR -/ p53 -/ ERBB2-
36T	I	45 anos	T1N0M0	Negativo	grau III SBR grau nuclear 3	ER +/- PR -/ p53 -/ ERBB2+ (3)
37T	I	63 anos	T2N0M0	Negativo	grau I SRB / grau nuclear 1	ER +/- PR + / p53 -/ /ERBB2- (0)
38T	IIla	69 anos	T3N2M0	Positivo	grau III SBR grau nuclear 3	ER -/ PR- / p53+/ ERBB2-(+1)

Legenda: LN: *Status* linfonodal; ER- Receptor de Estrógeno; PR- Receptor de Progesterona.

Quadro 2 - Relação das 9 amostras normais utilizadas.

MN04	N	Tecido Normal- borda de fibroadenoma
MN06	N	Tecido Normal- borda de carcinoma invasor
MN08	N	Tecido Normal- borda de carcinoma invasor
MN12	N	Tecido Normal- borda de fibroadenoma
MNP12	N	Tecido Normal- borda de carcinoma invasor
MN13	N	Tecido Normal- borda de fibroadenoma
MN16	N	Tecido Normal- borda de fibroadenoma
MN32	N	Tecido Normal- borda de carcinoma invasor
MN46	N	Tecido Normal- mamoplastia

Legenda: N: Tecido Normal. Para as amostras 6T, 8T, 12T e 32T também foram utilizadas margem de tecido normal.

Anexo 7 - Linhagens Celulares

Quadro 3 - Relação de todas as linhagens utilizadas no estudo.

Linhagem	T/N	Características
MCF7	T	célula epitelial de adenocarcinoma, ER+, PgR+
MDA-MB-231	T	célula epitelial de adenocarcinoma, ER-, PgR-
MDA-MB-435S	T	célula epitelial de carcinoma ductal invasivo, ER-, PgR-
HB4aC5.2	T	célula epitelial de lúmen de mama normal transfectada com o oncogene <i>ERBB2</i>
HB4a	N	célula epitelial de lúmen de mama normal
MCF10A	N	célula epitelial de doença fibrocística

Legenda: T: Linhagem Tumoral; N: Linhagem Normal; ER+: Presença de receptor de estrógeno; ER-: ausência de receptor de estrógeno; PgR+: Presença de receptor de progesterona; PgR-: ausência de receptor de progesterona.

Anexo 8 - Extração de RNA por Cloreto de Césio

Utilizando-se o método de GLISIN et al. (1974), com algumas modificações, pipeta-se 2 ml de solução de lise nos tubos do Politron[®] 1600 E System (Kinematica AG). “Politronar” o tumor com a solução de lise. Em seguida dividir o material em outro tubo e adicionar 1ml de solução de lise em cada tubo. Preparar os tubos da ultracentrifuga adicionando 4ml da solução de Cloreto de Césio 5,7 M e ao final adicionar com cuidado ao colchão de Césio a solução de lise até completar 4ml para cada tubo, balanceando-os. Centrifugar 29000rpm, 20°C, 20horas. Com o auxílio de uma pipeta *pasteur* plástica retirar cuidadosamente o material do tubo, deixando aproximadamente 1ml da solução. Emborcar o tubo e deixá-lo invertido sobre um lenço de papel autoclavado. Com auxílio de uma pinça e lenço de papel autoclavado limpar o interior da parede do tubo retirando o excesso de Cloreto de Césio que pode ter ficado na parede, tendo sempre o cuidado de não tocar o fundo do tubo. Ressuspender o pellet em 30µl de H₂O DEPC e passar o material para um tubo *ependorf* de 2ml. Ressuspender novamente o pellet com mais 30µl de H₂O DEPC e passar o material para o mesmo tubo. Fazer a leitura espectrofotométrica (OD₂₆₀ e OD₂₈₀) e correr um gel de agarose 1% corado com Brometo de Etídeo para avaliação da qualidade do RNA.

Anexo 9 - Extração de RNA por TRIZOL[®]

De acordo com o fabricante, em tubos de Politron[®] Pt 1600 E System (Kinematica AG), pipeta-se 1ml de TRIZOL[®] (Invitogen[™]), insere-se o tumor e “politrona-o” (limpar ou trocar todo o material sempre que mudar de amostra). Deixar as amostras descansando no TRIZOL[®] entre 10 minutos e uma hora. Em seguida, centrifugar as amostras por 10 minutos a 4°C e 14000 rpm. Transferir o sobrenadante para um novo tubo eppendorf e adicionar 200µl de clorofórmio, agitar vigorosamente por 15 segundos e centrifugar 15 minutos a 4°C e 14000rpm. Transferir o sobrenadante para um novo tubo e adicionar 500µl de Isopropanol. Deixar precipitando por no mínimo 12 horas a -20°C. Centrifugar as amostras 15 minutos a 4°C e 14000rpm. Descartar o sobrenadante. Lavar o pellet com Etanol 70% e centrifugar 10 minutos a 4°C e 14000rpm duas vezes. Eluir em água tratada com DEPC (0,1%) e a seguir quantificar em espectrofotômetro (OD₂₆₀ e OD₂₈₀) e correr as amostras em gel de agarose 1% corado com Brometo de Etídeo para avaliação da qualidade do RNA.

Anexo 10 - Amplificação de RNA mensageiro (aRNA)

Protocolo para amplificação de RNA mensageiro Baseado em GOMES et al. 2003. Modificado.

Seqüências dos *primers* (Empresa IDT)

- **Oligo dT (18)-T7 primer:** (5'AAA CGA CGG CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG CGC T(15)3')[57-pb]
- **TS (template switch) oligo primer:** (5' AAG CAG TGG TAA CAA CGC AGA GTA CGC GGG 3') [30-pb]

Nota: As incubações são todas feitas em termocicladores.

São utilizadas de 3 a 5 µg de RNA total.

Obs: A integridade e qualidade do RNA dever ser verificado em gel de agarose analisando das bandas correspondentes ao RNA ribossômico 18S e 28S.

1 Primeiro *round* de amplificação

1.1 Síntese da primeira fita de cDNA:

Tubo de PCR 0,2 ml

Reação:

3-5ug de RNA total (máximo de 6 ul)

Água tratada com DEPC (H₂O/DEPC) q.s.p. 6ul

1ul de 0,5 ug/ul oligo dT(15)-T7 *primer*

Incubar 70 °C por 5 minutos

4 °C

Mix para reação da 1a. fita de cDNA:

4ul 5x *First strand buffer*

3ul 0,5 ug/ul TS (*template switch*) *primer*

2ul 0,1 M DTT

1ul Rnase IN

2ul 10mM dNTP

1ul Superscript II (SSII)

Spin

Incubar a 42°C por 1 hora no termociclador

Adicionar 1 ul de SSII

Incubar a 42°C por 1 hora

4 °C

1.2 Síntese da 2a. fita de cDNA:

Reação:

104ul água tratada DEPC

15ul Advantage PCR *buffer*

3ul 10mM dNTP *mix*

1ul RNase H

5ul Advantage Polymerase

Incubar:

37°C por 10 min para digestão do RNA

94°C por 3 min para denaturação

65°C por 5 min para a ligação dos primers

75°C por 30 min para extensão

Parar a reação com solução de 7,5ul 1M NaOH 2mM EDTA

Incubar a 65°C por 10 minutos para inativação enzimática

1.3 Purificação com Fenol : Clorofórmio : Álcool isoamílico

Adicionar 150ul de Fenol:Clorofórmio:Isoamílico (25:24:1) pH 8.0

Vortex bem e descansar 1 min

Centrifugar por 10 min a 14000 rpm (temperatura ambiente)

Cuidadosamente retirar aproximadamente 140ul da fase aquosa

Dispensar em um tubo novo

Adicionar 150ul de Clorofórmio

Vortex bem e descansar 1 min

Centrifugar por 10 min a 14000 rpm (temperatura ambiente)

Cuidadosamente retirar aproximadamente 140ul da fase aquosa

Dispensar em um tubo novo

1.4 Precipitação

Adicionar 70ul de 3M Acetato de sódio (Na₂OAC), pH 5.2

Adicionar 500 ul de etanol absoluto (-20C)

Vortex bem

Centrifugar a 14000 rpm a 4°C por 30min

Remover com a pipeta o sobrenadante e reservar

Lavar 3x com 1ml de etanol 70% preparado DEPC (-20C)

Centrifugar a 14000 rpm, a 4°C por 5min

Secar a 37°C por 10 minutos para evaporar o etanol residual

Diluir em 21ul H₂O/DEPC e quantificar o dscDNA em diluição 1:100

1.5 Transcrição *in vitro*

20μl dscDNA (entrar com máximo de 40μg de cDNA)

Reação: 20ul ds cDNA

15ul of 25mM rNTP (A,G,C and UTP, prepare apenas o necessário para a reação)

10ul 5x *reaction buffer*

5ul *enzyme mix (Rnase inhibitor and T7 phage polymerase)*

Incubar a 37°C por 5-6hs.

1.6 Purificação do RNA usando Trizol

(centrífuga refrigerada a 4°C e ETOH a -20°C)

Adicionar 1ml of Trizol em cada tubo

Adicionar 200 ul clorofórmio por 1 ml de solução de Trizol

Vortex

Incubar a temperatura ambiente por 5 minutos

Centrifugar 12.000g por 15 minutos a 4°C

Transferir a fase aquosa para um novo tubo e adicionar isopropanol na proporção de 1:1 em relação à fase aquosa. *Vortex*

Incubar a temperatura ambiente por 10 minutos

Centrifugar a 14.000 rpm por 15 minutos

Lavar os pellets 3 vezes com 1 ml de solução 70% ETOH preparado com água DEPC

Centrifugar a 14.000 rpm por 5 minutos

Secar os pellets e ressuspender em 10ul de água DEPC

Verificar a concentração e a pureza do RNA por espectrofotômetro através de OD 260 e OD 280

1 Segundo *round* de amplificação (fase exponencial)

2.1 Síntese da 1a. fita de cDNA:

Concentrar o RNA amplificado (0.5-1ug) para um volume de 7ul

Adicionar 1ul (2ug/ul) *random hexamer* (por exemplo: dN6)

Incubar 70 °C por 5 minutos

4° C

Reação: 4ul 5x *First strand buffer*

1ul de 0,5 ug/ul oligo dT(15)-T7 *primer*

2ul 0,1 M DTT

1ul Rnase IN

2ul 10mM dNTP

1ul Superscript II (SSII)

Incubar a 42°C por 1 hora no termociclador

Adicionar 1ul de SSII

Incubar a 42°C por 1 hora

4° C

2.2 Síntese da 2a. fita de cDNA

Mix:

104ul água tratada com DEPC

15ul Advantage PCR *buffer*

3ul 10mM dNTP *mix*

1ul RNase H

5ul Advantage Polymerase

Incubar à:

37°C por 10 min para digestão do RNA

94°C por 3 min para denaturação

65°C por 5 min para a ligação dos primers

75°C por 30 min para extensão

Parar a reação com solução de 7,5ul 1M NaOH 2mM EDTA

Incubar a 65°C por 10 minutos para inativação enzimática

2.3 Purificação com Fenol : Clorofórmio : Álcool isoamílico

Adicionar 150ul de Fenol:Clorofórmio:Isoamilico (25:24:1) pH 8.0

Vortex bem e descansar 1min

Centrifugar por 10 min a 14000 rpm à temperatura ambiente

Cuidadosamente retirar aproximadamente 140ul da fase aquosa

Dispensar em um tubo novo

Adicionar 150ul de Clorofórmio

Vortex bem e descansar 1min

Centrifugar por 10 minutos a 14000 rpm à temperatura ambiente

Cuidadosamente retirar aproximadamente 140ul da fase aquosa

Dispensar em um tubo novo

2.4 Precipitação

Adicionar 70ul de 3M Acetato de sódio (Na₂OAC), pH 5.2

Adicionar 500 ul de etanol absoluto (-20C)

Vortex bem

Centrifugar a 14000 rpm a 4°C por 30minutos

Remover com a pipeta o sobrenadante e reservar

Lavar 3x com 1ml de etanol 70% DEPC (-20C)

Centrifugar a 14000 rpm a 4°C por 5min

Secar a 37°C por 10 min para evaporar o etanol residual

Eluir em 21ul H₂O/DEPC e quantificar o dscDNA em diluição 1:100

2.5 Transcrição *in vitro*

20µl ds cDNA (entrar com máximo de 40µg de cDNA)

Reação: 20ul ds cDNA

15ul of 25mM rNTP (A,G,C and UTP, prepare somente o suficiente para a reação)

10ul 5x reaction *buffer*

5ul *enzyme mix* (*Rnase inhibitor and T7 phage polymerase*)

Incubar a 37°C por 5-6hs.

2.6 Purificação do RNA usando Trizol

(centrífuga refrigerada a 4°C e ETOH a -20°C)

Adicionar 1ml of Trizol em cada tubo

Adicionar 200 ul clorofórmio por 1 ml de solução de Trizol. *Vortex*

Incubar a temperatura ambiente por 5 minutos

Centrifugar 12.000g por 15 minutos a 4°C

Transferir a fase aquosa para um novo tubo e adicionar isopropanol na proporção de 1:1 em relação à fase aquosa

Vortex

Incubar a temperatura ambiente por 10 minutos

Centrifugar a 14.000 rpm por 15 minutos

Lavar os pellets 3 x com 1 ml de solução 70% ETOH preparado com DEPC

Centrifugar a 14.000 rpm por 5 minutos

Secar o *pellet* e ressuspender em 10ul de água DEPC

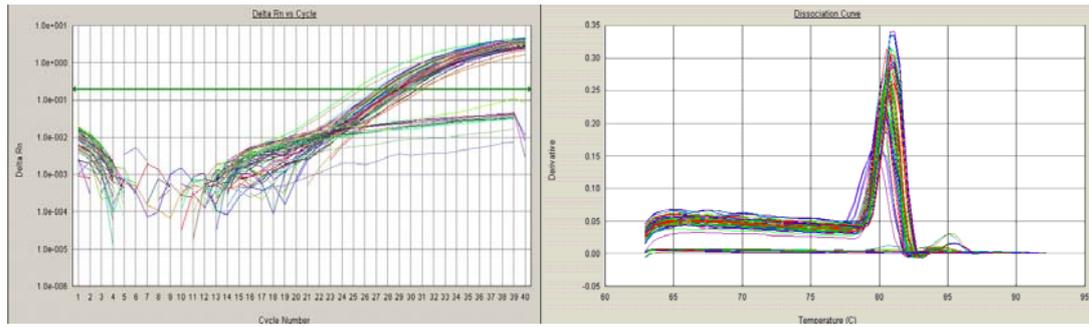
Verificar a concentração e a pureza do RNA por espectrofotômetro (OD 260) através OD260 e OD 280

3 Avaliar a integridade e qualidade do RNA amplificado em gel de agarose

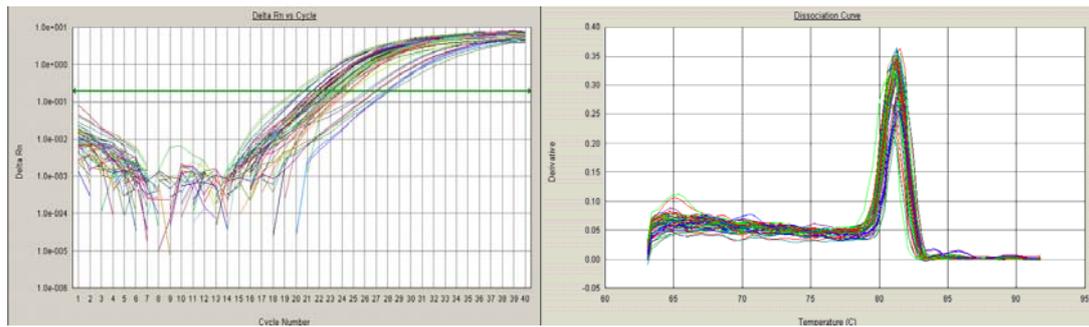
Aplicar 0,5 a 1 µg de RNA em gel de agarose 1% corado com brometo de etídeo.

Anexo 11 - Gráficos de amplificação das amostras por qRT-PCR e curvas de dissociação obtidas para todas as VCE estudadas

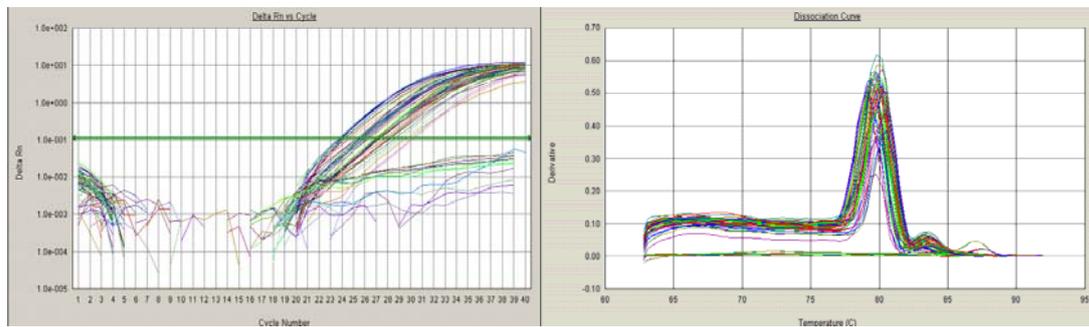
MK-STYX



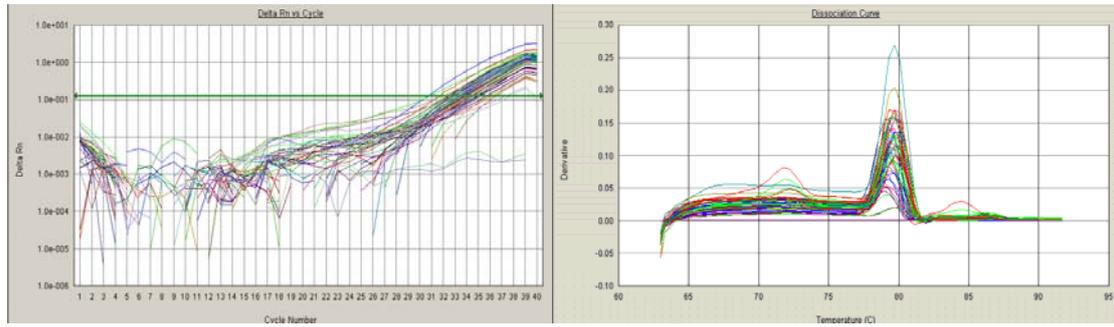
BAP1



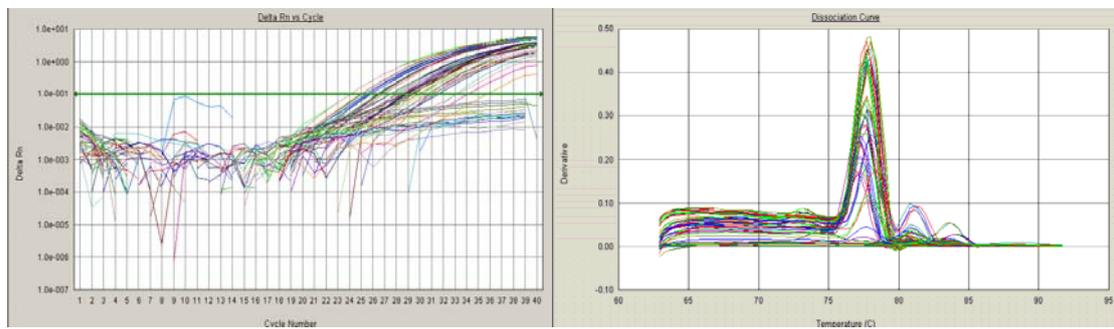
FLJ125-28



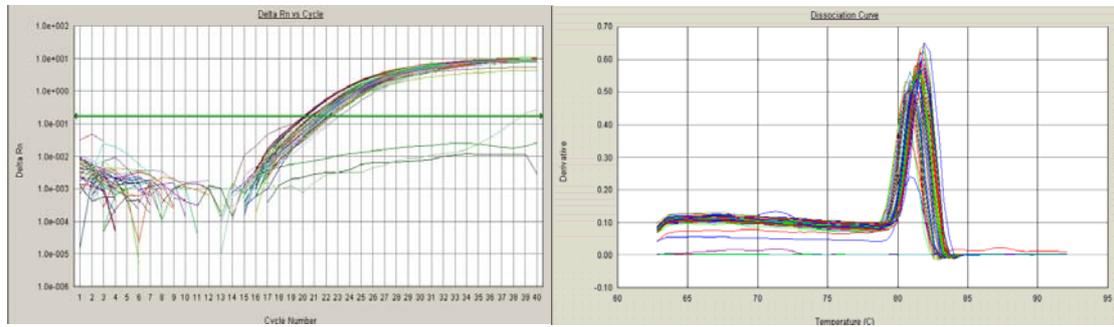
PPP1R8



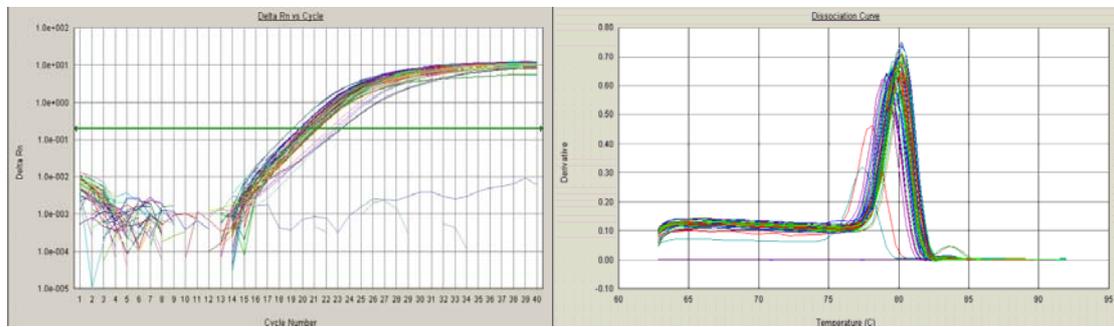
BRRN1



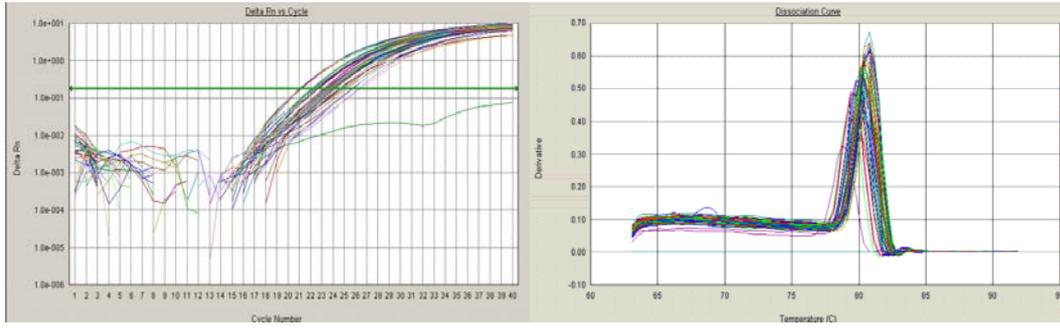
LIP8



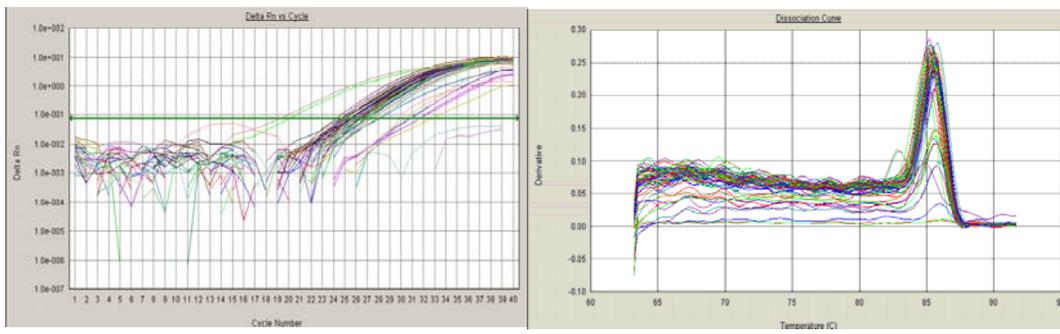
PRKD2



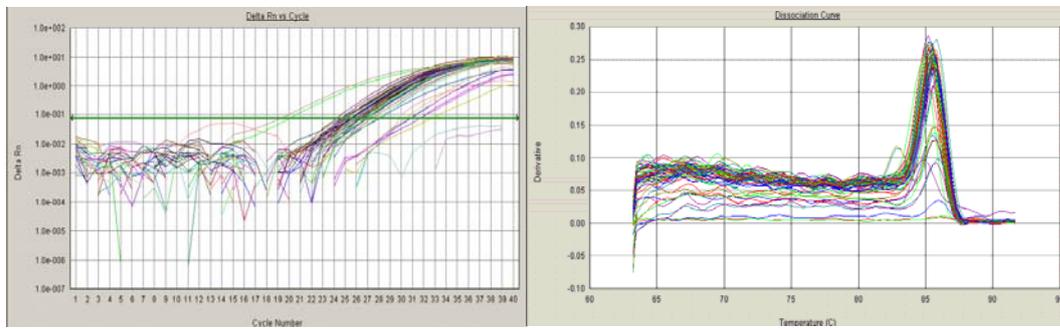
SND1



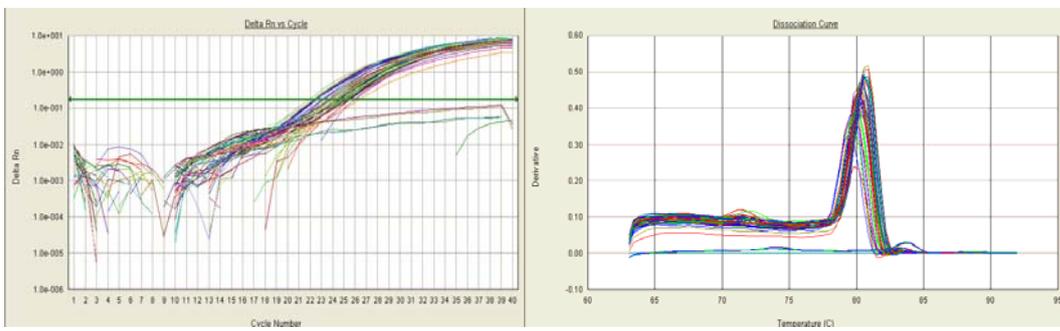
ASC1p100



KIAA0696



MBTPS1



Anexo 12 - Amostras Independentes

Quadro 4 - Relação das 40 amostras utilizadas.

Amostra	Estadiamento	Idade (anos)	TNM	LN	GrauSBR	Marcadores Moleculares
44 T	II	66	T3N0M0	Negativo	3	ER -/ PR -/ p53 +/ ERBB2- (1+)
50 T	II	55	T2N1M0	Negativo	2	ER+/ PR+/ ERBB2- (0)
51 T	II	47	T2N0M0	Positivo	Não informado	ER +/ PR +/ p53 -/ ERBB2+ (2+)
54 T	II	70	T2N0M0	Negativo	3	ER+/ PR-/ ERBB2- (0)
58 T	II	67	T2N1M0	Negativo	2	ER+/ PR+/ p53 -/ ERBB2+ (2+)
59 T	II	62	T2N1M0	Negativo	3	ER-/ PR-/ p53 -/ ERBB2+ (2+)
60 T	II	68	T2N0M0	Negativo	1	ER+/ PR+/ ERBB2- (1+)
61 T	I	39	T1N0M0	Negativo	1	ER +/ PR +/ p53 -/ ERBB2- (0)
62 T	II	51	T2N0M0	Positivo	3	ER +/ PR +/ ERBB2- (1+)
64 T	I	57	T1N0M0	Negativo	2	ER +/ PR +/ ERBB2- (1+)
65 T	II	88	T2N1M0	Negativo	3	ER -/ PR -
67 T	II	76	T2N1M0	Negativo	1	Não informado
68 T	II	47	T2N0M0	Negativo	3	ER -/ PR-/ p53 -/ ERBB2- (1+)
70 T	II	66	T2N1M0	Positivo	3	ER +/ PR -/ p53 -/ ERBB2+ (2+)
71 T	III	62	T2N1M0	Positivo	3	ER +/ PR +/ p53 -/ ERBB2+ (2+)
BIO QT 01	II	49	T2N0M0	Positivo	2	ER -/ PR -/ p53 -/ ERBB2+ (3+)
BIO QT 04	III	71	T4N1M0	Positivo	2	ER +/ PR +/ p53 -/ ERBB2- (1+)
BIO QT 05	II	58	T2N0M0	Positivo	2	ER -/ PR -/ p53 -/ ERBB2+ (2+)
BIO QT 15	III	56	T4N1M0	Positivo	2	ER+/ PR+/ p53 -/ ERBB2- (1+)
BIO QT 33	III	72	T4N1M0	Negativo	2	ER -/ PR -/ p53 -/ ERBB2+ (2+)
BIO QT 42	III	52	T4N1M0	Negativo	2	ER +/ PR +/ p53 +/ ERBB2+ (3+)

BIO QT 49	II	48	T2N1M0	Positivo	2	ER -/ PR-/ p53 -/ ERBB2- (1+)
BIO QT 71	III	48	T4N1M0	Não informado	2	ER +/ PR +/ p53 -/ ERBB2+ (2+)
BIO QT 83	III	47	T4N1M0	Negativo	2	ER+/ PR+/ p53 -/ ERBB2- (0)
MJ 2 T	II	30	T2N1M0	Positivo	2	ER+/ PR+/ p53 -/ ERBB2+ (3+)
MJ 3 T	II	26	T2N0M0	Negativo	3	ER-/ PR-/ ERBB2-
MJ 5 T	II	29	T2N0M0	Negativo	3	ER+/ PR-/ ERBB2- (1+)
MJ 7 T	II	29	T2N1M0	Positivo	3	ER +/ PR +/ ERBB2+ (2+)
MJ 12 T	I	29	T1N0M0	Negativo	3	ER -/ PR-/ ERBB2- (0)
MJ 13 T	II	29	T2N1M0	Positivo	3	ER+/ PR+/ p53 -/ ERBB2- (0)
MJ 14 T	II	35	T2N0M0	Negativo	3	ER -/ PR -/ p53 +/ ERBB2+ (2+)
MJ 16 T	II	29	T2N0M0	Negativo	2	ER +/ PR +/ p53 -/ ERBB2- (1+)
MJ 8T-2	II	33	T2N0M0	Positivo	3	ER +/ PR -/ p53 +/ ERBB2- (0)
53 T	IV	77	T4N1M1	Positivo	2	ER+/ PR-/ p53 -/ ERBB2+ (2+)
72 T	II	56	T2N0M0	Negativo	2	ER -/ PR-/ ERBB2+ (3+)
BIO QT 43	IV	57	T2N2M1	Positivo	2	ER+/ PR+/ p53 -/ ERBB2- (1+)
BIO QT 65	IV	43	T4N2M1	Positivo	3	ER +/ PR +/ p53 +/ ERBB2+ (2+)
MJ 11T-2	III	30	T2N1M0	Positivo	2	ER -/ PR -/ p53 +/ ERBB2+ (3+)
MJ-1T	IV	28	T1N1M1	Positivo	2	ER +/ PR +/ ERBB2+ (3+)
BIOQT64	IV	67		Positivo	3	ER +/ PR -/ p53 -/ ERBB2- (0)

Legenda: LN: *Status* linfonodal; ER: Receptor de Estrógeno; PR: Receptor de Progesterona.