

**ANÁLISE DO PERFIL DE EXPRESSÃO GÊNICA DE  
CARCINOMA DUCTAL *IN SITU* E INVASIVO EM  
TUMORES DE MAMA ATRAVÉS DA TÉCNICA DE  
*MICROARRAY***

**NÁDIA PEREIRA DE CASTRO**

**Tese apresentada à Fundação Antônio  
Prudente para obtenção do título de Doutor  
em Ciências**

**Área de Concentração: Oncologia**

**Orientadora: Dra. Dirce Maria Carraro**

**Co-Orientadora: Dra. Anamaria Aranha  
Camargo**

**São Paulo**

**2008**

## FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Castro, Nadia Pereira de

**Análise do perfil de expressão gênica de carcinoma ductal *in situ* e invasivo em tumores de mama através da técnica de *microarray*** / Nadia Pereira de Castro -- São Paulo, 2008.

174p.

Tese (doutorado)-Fundação Antônio Prudente.

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

Orientadora: Dirce Maria Carraro.

Descritores: 1. MICRODISSECÇÃO A LASER. 2. CARCINOMA DUCTAL DE MAMA/genética. 3. PERFIL DA EXPRESSÃO GÊNICA. 4. ANALISES MICROARRAY.

## DEDICATÓRIA

Dedico esta conquista...

À **Deus** pela vida plena e por tudo que sou. Por Seu imensurável amor e cuidado. Pela paz e esperança. Meu criador, mantenedor e amigo.

Aos **meus pais** queridos (Rubles e Neide) pela dedicação, amor, apoio incondicional e por nunca medirem esforços para a minha formação.

Aos melhores **irmãos** do mundo (Rômulo e Rener) pela atenção, companheirismo e cuidado.

Vocês são a base das minhas conquistas. Amo vocês!!

## AGRADECIMENTOS ESPECIAIS

Agradeço especialmente...

À minha avó (Maria) pelo apoio e cuidado.

Aos meus queridos tios (Vanderlei e Sônia) e primos (Vanderson e Emerson) por me acolherem com tanto carinho e dedicação.

Ao meu namorado (Rodrigo) pela paciência e compreensão.

Aos meus amigos pela atenção, disponibilidade em ajudar, confiança e amizade.

Obrigada por acreditarem em mim!

## AGRADECIMENTOS

À minha orientadora Dra. Dirce pela confiança, ensinamentos e oportunidade de crescer.

À minha co-orientadora Dra. Anamaria Camargo pela atenção, sugestões e leitura crítica da tese.

Ao Laboratório de Biologia Molecular do Hospital A.C. Camargo pelo bom ambiente de trabalho. À Maria Cristina Rangel (amiga de todas as horas, por sempre me ouvir, me ajudar, por se importar e acreditar. Muuuuito obrigada *friend* por tudo o que fez por mim, vou sempre me lembrar com carinho tudo o que passamos juntas), Sabrina Daniela (amiga do coração, muito obrigada pelo carinho, atenção, apoio, incentivo, conselhos, por acreditar em mim. Obrigada pela ajuda indispensável com as proteínas e o TMA. Você é única Sassá!), Mev (pelo companherismo, incentivo e convivência. Tu eres buena gente!!) Mariana Maschietto (pela amizade, atenção, sempre pronta a ajudar, obrigada Ma, você é muito querida!), Thiago F. Saraiva (pela amizade, risadas, bolos, corridas no parque e disponibilidade em ajudar nos experimentos. Obrigada Fo!), Jane Kaiano e Louise D. Mota (obrigada pela dedicação em ensinar e ajuda fundamental nos experimentos. Sou muito grata!), Elen Bastos (pela amizade, receitas vegetebas, momento esporte, viagens, atenção e ajuda com as “bolinhas“. Obrigada EB!), Paulo Pineda (pela atenção, paciência, pelas amostras e géis. Valeu Pol!), Reimar Padovani (pelas risadas e ajuda), Jamila (pela atenção e carinho), Elisa (pela prontidão em ouvir e ajudar), Adriana (pela atenção e ajuda), Laura (pelo sorriso e compreensão), Aderbal (pelas risadas), Waleska (Sempre prestativa e amável. Apesar da curta passagem, foi muito bom tê-la conosco), Wilson (pelos momentos de descontração e conselhos) Gustavo (obrigada pela companhia nas corridas, bom humor e atenção), Eloísa (sempre disposta a ajudar, sorridente e querida), Luana e Rafael (pela ajuda com o Bionalyser), Vera (pelos bolos!) Alexandre (o dermatologista do Lab, sempre prestativo).

À amiga e 2 vezes vizinha Fernanda Ayala, pela consideração, amizade, pelas conversas sempre produtivas. Gosto muito de você Fê!!

Obrigada pessoal por participarem desta etapa da minha vida.

Ao laboratório de Biologia Computacional do Hospital A. C. Camargo, à Dra Helena P. Brentani (pela disponibilidade em ensinar e ajudar), Luiz Paulo Camargo (obrigada pela amizade, atenção, momentos de degustação -que saudade daquele peixe- pela prontidão em ajudar sempre, até nas análises!), Artur Fabri (pela atenção, amizade, pela 'chavinha'), Diogo Patrão (pelos *clusters*, página da *web*, ajuda com as análises), e César Torres (pelas análises e atenção). Obrigada pela análise da qualidade das lâminas de cDNA microarray.

Aos pesquisadores do IME-USP, Eduardo Jordão Neves e Lucas (pela dedicação e prontidão em ajudar) por realizarem as análises estatísticas e computacionais dos dados de microarray. Obrigada pela paciência e valiosa colaboração nesse trabalho.

Ao Instituto Ludwig de Pesquisa Sobre o Câncer - São Paulo, na pessoa da diretora e querida professora Dra. Luisa Villa e à todos do Instituto pela ajuda, pelo convívio e amizades conquistadas.

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

Ao Dr. Fernando, Isa, Cynthia, Hugo, Renata, André Abreu do departamento de Anatomia Patológica do Hospital A.C. Camargo que se envolveram neste trabalho e me ajudaram com as microdissecções a *laser*, diagnóstico das lâminas e retirada das amostras no banco de tumores. À Miyuki Silva e Severino Silva pelo trabalho tão cuidadoso com as amostras.

Ao José Ivanildo Neves pela prontidão em ajudar e realizar as reações imunohistoquímicas. À Sabrina Daniela pelo envolvimento e dedicação na leitura dos TMAs e ao Dr. João Gonçalves pelas análises estatísticas.

Aos médicos do Departamento de Mastologia pelas amostras coletadas para o banco de tumores. Em especial, o Dr Mourão e a Dra Socorro pelos ensinamentos.

Aos responsáveis da Pós-Graduação, especialmente Dr. Luiz Fernando Lima Reis, Ana Maria Kurinari e Luciane Pitombeira. Pela paciência, compreensão e cuidado dispensados.

À Biblioteca da Fundação Antônio Prudente pelo acervo e auxílio nas pesquisas bibliográficas, em especial Suely pela ajuda na elaboração da versão final da tese. As meninas da biblioteca: Rosinéia Aguiar Carneiro e Francyne Pólen G. Lima sempre tão dispostas e eficientes. Não tenho palavras para agradecer !!

À todos os funcionários do SAME pela disposição e ajuda, em especial a Dona Hirde Contesini pela simpatia e atenção.

À FAPESP pela bolsa concedida e suporte financeiro para o desenvolvimento desse projeto.

Aos pacientes e seus familiares por permitirem a realização desta pesquisa.

E a todas as outras pessoas que, direta ou indiretamente, me auxiliaram e contribuíram na realização deste estudo.

Deixo aqui os meus sinceros agradecimentos.

Este trabalho é o resultado de todos vocês !!

## RESUMO

Castro NP. **Análise do perfil de expressão gênica de carcinoma ductal *in situ* e invasivo em tumores de mama através da técnica de *microarray*.** São Paulo; 2008. [Tese de Doutorado-Fundação Antônio Prudente].

O câncer de mama está entre as neoplasias de maior incidência e é responsável pela alta taxa de mortalidade entre as mulheres no mundo todo. O carcinoma ductal é o tipo histológico mais freqüente. O carcinoma ductal *in situ* (DCIS) inclui um grupo de tumores de mama pré-invasivos com potencial maligno distinto, podendo progredir rapidamente para carcinoma invasivo ou não apresentar evolução durante um longo período da doença. Atualmente, um dos maiores desafios na pesquisa molecular nessa área é identificar genes que possam predizer o risco de progressão para doença invasiva e de marcadores de prognósticos. Neste estudo foram analisados 40 casos de mama, sendo 5 amostras de tecido de mamário não neoplásico (N), 16 casos de amostras pareadas de carcinoma ductal (*in situ* e invasivo), 5 DCIS puro, 9 DCIS coexistindo com o componente invasor (DCIS/IDC) e 5 carcinomas ductais invasivos (IDC). Os RNAs provenientes das células epiteliais microdissecadas foram amplificados e hibridizados pelo sistema inverso de incorporação de fluoróforos (*Dye swap*) em duas plataformas distintas de cDNA *microarray*. Uma contendo 4.608 seqüências de cDNA correspondentes a genes humanos (4.8k) e outra contendo 390 genes pertencentes as vias de sinalização WNT, PI3K e processo EMT (Transição Epitélio-Mesênquima). Dois delineamentos experimentais foram usados: comparação entre lesões *in situ* e invasivo da mesma amostra (DCIS e IDC); e comparação de grupos de lesões que mimetizam a progressão de câncer ductal de mama, utilizando amostras independentes [células epiteliais de mama capturadas de amostra não neoplásica (N), células tumorais capturadas das lesões: DCIS puro, DCIS/IDC e IDC]. Na plataforma 4.8K, 16 amostras pareadas foram usadas para comparar o padrão de expressão de células tumorais do DCIS e IDC de uma mesma paciente. Foram

identificados 33 genes diferencialmente expressos ( $t$  de *Student – Fold* > |1,5| e  $p$ valor<0,01), sendo genes candidatos a estarem envolvidos na transição do DCIS para o IDC. Para 4 (*LUM*, *RDH-E2*, *CXCL13* e *POSTN*) de 8 genes selecionados foi confirmada expressão diferencial por RT-PCR quantitativo (qRT-PCR). Dois genes (*LUM* e *CRABP2*) com maior expressão nas amostras IDC foram selecionados para verificar associação com outros marcadores moleculares e/ou dados clinicopatológicos por *Tissue Microarray*. A expressão da proteína LUM se mostrou associada positivamente com as CKs 5/6, CK 14, CK8 e 18 e o grupo super-expressão do HER2/neu, enquanto que a expressão da proteína CRABP-2 se mostrou associada positivamente com ER, PR, CK8, CK18, luminal A, p53 e negativamente com a CK14. Buscando caracterizar os aspectos moleculares da progressão do carcinoma ductal *in situ* de mama, os padrões de expressão gênica entre 4 grupos que mimetizam a progressão foram avaliados (N, DCIS puro, DCIS/IDC e IDC) através do teste ANOVA ( $p$ FDR<0,01) seguido de Tukey, sendo caracterizado como mais divergente o grupo N, como esperado. Na comparação entre as 3 lesões (DCIS puro, DCIS/IDC e IDC), o DCIS puro mostrou maior divergência molecular, contradizendo os aspectos morfológicos. Para identificar genes com potencial de prever risco de invasão do DCIS, foi comparado o padrão de expressão de 2 lesões morfológicamente idênticas (DCIS puro e DCIS/IDC), identificando 147 genes (ANOVA - *Fold* > |2| e  $p$ FDR<0,01). Clusterização hierárquica baseada no padrão de expressão desses genes separou as amostras em dois grupos distintos, 100% das amostras normais e 60% DCIS puro permaneceram no mesmo grupo e separadas em 100% das DCIS/IDC. Cinco de 8 genes selecionados (*C16orf5*, *SULF-1*, *LOX*, *GOSR2* e *TXNL2*) foram confirmados por qRT-PCR, sendo candidatos a marcadores de predição de progressão do carcinoma ductal *in situ*. Para avaliar os aspectos funcionais do processo de invasão, foi utilizada a plataforma contendo genes das vias de sinalização WNT, PI3K e processo EMT. O mesmo delineamento experimental foi utilizado para 10 casos de lesões pareadas, resultando 32 genes diferencialmente expressos entre DCIS e IDC ( $t$  de

*Student* pareado - *p*valor <0,05). Esse resultado confirmou que o padrão de expressão entre essas lesões é semelhante mesmo avaliando genes sabidamente envolvidos no processo. Entre o DCIS puro e DCIS/IDC foram identificados 15 genes diferentemente expressos (Wilcoxon – *p*<0,05) sendo que o padrão de expressão foi capaz de separar as amostras de forma semelhante ao grupo de 147 genes. Trios classificadores foram construídos para discriminar DCIS puro de DCIS/IDC. Um trio de genes pertencentes a cada via WNT (*CSNK1A1L*, *LRP3* e *SDC2*), PI3K (*PLCG2*, *INPP1* e *DGKA*) e processo EMT (*HDGF*, *CDH13* e *TWIST1*), foi capaz de separar as amostras, constituindo-se candidatos de predição de progressão de DCIS. Assim, baseado nos dados desse estudo, demonstramos que as células do carcinoma ductal *in situ* que coexistem com as células do carcinoma invasivo apresentam alterações moleculares antes da modificação morfológica da lesão, e isso pode ser explorado para se identificar genes alterados que possam predizer a capacidade de invasão. Além disso, esse estudo identificou vários genes candidatos a marcadores moleculares de prognóstico e também preditivos do risco de progressão de doença não invasiva.

## SUMMARY

Castro NP. **[Gene expression profile analysis of ductal carcinoma *in situ* and invasive of the breast cancer by Microarray technology]**. São Paulo; 2008. [Tese de Doutorado-Fundação Antônio Prudente].

Breast carcinoma is one of the most incidence neoplasias and is responsible for a high death-rate among women worldwide. The ductal carcinoma is the most frequent histological type. Ductal carcinoma *in situ* (DCIS) includes a group of preinvasive breast tumors with distinct malignant potentials. DCIS can have different outcomes, progressing rapidly to invasive cancer or slowly changing over a long period of the disease. Lately, one of the most challenges in molecular research in this field is to identify genes able to predict the risk of progression to invasive disease and prognostic marker. In this study were used 40 breast cases, being 5 non-neoplastic mammary tissues samples (N), 16 matched pairs of ductal carcinoma (*in situ* and invasive), 5 pure DCIS, 9 DCIS coexisting with invasive ductal carcinoma (DCIS/IDC) e 5 IDC. The RNA from epithelium cells laser capture microdissected were amplified and hybridized using reference design with dye swap in two distinct customized cDNA microarrays platforms. One containing 4.608 cDNA that represent human genes (4.8K) and other containing 390 genes belonging to WNT, PI3K signaling pathways and EMT process (Epithelial-mesenchymal transition). Two designing assays were used: comparisons between lesions *in situ* and invasive from the same patient (DCIS e IDC) and comparison among lesion groups that mimics the progression of breast ductal carcinoma using independent samples (breast epithelium cells microdissected of non-neoplastic tissues (N), tumoral cells microdissected of lesions: DCIS puro, DCIS/IDC e IDC. In the 4.8k platform, 16 matched-pair samples were used to compare the tumor cells expression profile of DCIS and IDC from the same patient. It was identified 33 candidates differentially expressed ( $t$  de Student pairwise -  $Fold > |1,5|$  e

$p$ value $<0,01$ ), being candidates' genes to be involved in transition from DCIS to IDC. To 4 (*LUM*, *RDH-E2*, *CXCL13* e *POSTN*) of 8 selected genes was confirmed differential expression through quantitative RT-PCR (qRT-PCR). Two genes (*LUM* and *CRABP2*) over expressed in IDC were selected to verify association with others molecular markers and/or clinicopathological parameters through *Tissue Microarray*. The LUM protein expression showed positive association with CKs 5/6, CK 14, CK8 e 18 and HER2/neu positive, while CRABP2 showed positive association with ER, PR, CK8, CK18, luminal A, p53 and negatively with CK14. Seeking to characterize molecular aspects of ductal carcinoma *in situ* of the breast progression, the general gene expression profile among 4 groups that mimics the progression was analyzed (N, DCIS pure, DCIS/IDC e IDC) performing ANOVA test ( $p$ FDR $<0,01$ ) and followed by Tukey's test, being characterized as most diverge the N group, as expected. The comparison among the 3 lesions (DCIS pure, DCIS/IDC e IDC), the pure DCIS showed the most molecular divergence, contradicting the morphological aspects. To identify genes able to predict the potential risk of invasion of DCIS, it was compared the expression profile of two morphologically identical lesions (DCIS pure e DCIS/IDC), identifying 147 genes (ANOVA -  $Fold >|2|$  e  $p$ FDR $<0,01$ ). Hierarchical cluster based on expression profile of those genes could segregate the samples into two distinct groups in which 100% of non-neoplastic samples and 60% of pure DCIS remained in the same group and discriminated from DCIS/IDC (100%). Five (*C16orf5*, *SULF-1*, *LOX*, *GOSR2* and *TXNL2*) candidates were confirmed through qRT-PCR as predict markers of DCIS progression. To assess the functional aspects of invasion process, it was used a platform containing genes belonging to WNT, PI3K signaling pathways and EMT process. The same experimental designing was used for 10 matched-pair, resulting in 32 differentially expressed genes between DCIS e IDC ( $t$  de Student pairwise -  $p$ value  $<0,05$ ). This result confirmed that expression profile between the lesions is similar even assessing genes involved in this process. Between pure DCIS and DCIS/IDC were identified 15 differentially expressed genes (Wilcoxon -  $p<0,05$ ) whose expression profile could

segregate the samples in the same way as for the 147 genes. Triple classifiers were built seeking to segregate the DCIS pure and DCIS/IDC samples. One gene triple belonged to each signaling pathways, WNT (*CSNK1A1L*, *LRP3* and *SDC2*), PI3K (*PLCG2*, *INPP1* and *DGKA*) and EMT process (*HDGF*, *CDH13* e *TWIST1*) were able to segregate the samples being predictor candidates of DCIS progression. Hence, based on this study data, we showed that ductal carcinoma *in situ* coexisting with invasive ductal carcinoma presents molecular alteration before lesion morphological modifications, and this may be exploited to identify altered genes able to predict invasion capacity. Furthermore, this study identified many candidate genes to molecular markers for outcome and also predicting the risk of non-invasive disease progression.

## LISTA DE FIGURAS

<b>Figura 1</b>	Estrutura da mama feminina normal adulta.	3
<b>Figura 2</b>	Representação de carcinoma ductal <i>in situ</i> e invasivo de mama.	4
<b>Figura 3</b>	Processo de disseminação de células na formação de metástase.	18
<b>Figura 4</b>	Esquema da via de transdução de sinal associada ao processo de transição epitélio-mesenquima, PI3K e WNT.	28
<b>Figura 5</b>	Representação de microdissecção a <i>laser</i> de carcinoma ductal <i>in situ</i> .	43
<b>Figura 6</b>	Representação de cDNA amplificado para o gene <i>NOTCH 2</i> em gel de poliacrilamida 8%.	45
<b>Figura 7</b>	Qualidade de RNA total extraído a partir de amostras microdissecadas a <i>laser</i> avaliado através do aparelho Agilent 2100 Bioanalyser.	47
<b>Figura 8</b>	Representação da lâmina contendo genes pertencentes as vias de sinalização WNT, PI3K e processo EMT hibridizada com amostra de carcinoma ductal <i>in situ</i> marcada com Cy3, referência com Cy5 e imagem sobreposta respectivamente.	54
<b>Figura 9</b>	Representação da fluorescência de todos os genes de uma lâmina.	57
<b>Figura 10</b>	Gráficos representativos da normalização por <i>Lowess</i> .	59
<b>Figura 11</b>	Os gráficos de amplificação ( <i>Amplification Plot</i> ), especificidade da reação ( <i>Dissociation curve</i> ) e diluições de curva padrão respectivamente para os normalizadores <i>GAPDH</i> e <i>HPRT1</i> e para os genes <i>C16orf5</i> e <i>LOX</i> .	70
<b>Figura 12</b>	Gráfico da média dos valores de estabilidade de expressão dos genes <i>GAPDH</i> , <i>HPRT</i> e <i>BCR</i> .	73
<b>Figura 13</b>	Modelo matemático para quantificação relativa em RT-PCR em tempo real (PFAFFL 2001).	74
<b>Figura 14</b>	Esquema mostrando a montagem do bloco do <i>tissue microarray</i> e a lâmina para a análise.	76

<b>Figura 15</b>	Gráfico MMPlot mostrando a qualidade de correlação entre a lâmina controle e a lâmina <i>swap</i> avaliada através de correlação de <i>Pearson</i> . R- referência e T- Tumor.	83
<b>Figura 16</b>	Dendograma das lâminas controle e <i>swap</i> das amostras hibridizadas na plataforma 4.8k.	83
<b>Figura 17</b>	Genes diferencialmente expressos na comparação entre DCIS e IDC coexistindo na mesma amostra ( <i>fold</i> >1.5 e <i>p</i> <0.01).	85
<b>Figura 18</b>	A: Clusterização hierárquica com os 33 genes diferencialmente expressos entre DCIS e IDC (grupo pareado). B: Legenda do suporte do cluster avaliado pela técnica do <i>Bootstrap</i> .	86
<b>Figura 19</b>	<i>Box Plot</i> dos genes selecionados para confirmação através de qRT-PCR quanto a tendência crescente ou decrescente no nível de expressão entre as amostras normais, DCIS e IDC.	89
<b>Figura 20</b>	Proteína LUM detectada no núcleo através de reações imunohistoquímicas em TMA de mama.	93
<b>Figura 21</b>	Proteína CRABP2 detectada em núcleo e citoplasma através de reações imunohistoquímicas em TMA de mama.	94
<b>Figura 22</b>	Representação dos 4 grupos que mimetizam a progressão do carcinoma ductal.	97
<b>Figura 23</b>	Esquema de análise estatística comparando o perfil de expressão gênica e as características morfológicas entre os estágio que mimetizam a progressão do carcinoma ductal.	99
<b>Figura 24</b>	Representação de diagrama de Venn.	100
<b>Figura 25</b>	Clusterização hierárquica não supervisionada. A: Dendograma das amostras não-neoplásicas (n=4), DCIS puro (n=5) e DCIS/IDC (n=10).	102
<b>Figura 26</b>	Comportamento dos genes entre as amostras que mimetizam a progressão do carcinoma ductal.	106
<b>Figura 27</b>	Dendograma das lâminas controle e <i>swap</i> das amostras hibridizadas na plataforma contendo as vias de sinalização WNT, PI3K e processo EMT.	108
<b>Figura 28</b>	Clusterização hierárquica utilizando o grupo de 30 genes diferencialmente expressos entre DCIS e IDC (grupo pareado) mostrando a robustez da segregação com o TMEV.	110

<b>Figura 29</b>	Representação de diagrama de Venn.	112
<b>Figura 30</b>	Clusterização hierárquica não supervisionada.	114
<b>Figura 31</b>	Trios classificadores encontrados utilizando os grupo de genes pertencentes as vias de sinalização WNT, PI3K e processo EMT representadas na plataforma visando separar as amostras DCIS puro e DCIS/IDC.	115
<b>Figura 32</b>	Rede de relevância para o grupo de amostras não-neoplásico e DCIS puro versus DCIS/IDC.	116
<b>Figura 33</b>	Gráficos <i>scatter-plot's</i> para aqueles pares de genes que apresentaram coeficientes de correlação superiores a 0,9 entre amostras não-neoplásica e DCIS puro versus DCIS/IDC.	121

## LISTA DE TABELAS

<b>Tabela 1</b>	Características específicas de cada fluoróforo utilizado.	54
<b>Tabela 2</b>	Lista das seqüências dos iniciadores.	65
<b>Tabela 3</b>	Normalizadores e genes alvo selecionados com seus respectivos valores de <i>slope</i> e valores das eficiências aceitáveis calculadas.	69
<b>Tabela 4</b>	Processos biológicos com relevância estatística no grupo de genes diferencialmente expressos entre as amostras pareadas (DCIS x IDC).	87
<b>Tabela 5</b>	Via Metabólica com relevância estatística no grupo de genes diferencialmente expressos entre as amostras pareadas (DCIS x IDC).	87
<b>Tabela 6</b>	Genes avaliados através de qRT-PCR nas amostras pareadas.	91
<b>Tabela 7</b>	Características clínicas e epidemiológicas dos 400 casos analisados neste estudo através da metodologia de <i>Tissue Microarray</i> .	92
<b>Tabela 8</b>	Correlações entre a expressão imunohistoquímica de LUM e CRABP2 e as variáveis clinicopatológicas.	96
<b>Tabela 9</b>	Análise funcional dos 147 genes diferencialmente expressos entre amostras não-neoplásicas, DCIS puro e DCIS/IDC selecionados como genes envolvidos no processo de progressão do DCIS.	103
<b>Tabela 10</b>	Genes selecionados randomicamente e avaliados por qRT-PCR com suas respectivas diferenças no nível de expressão tanto no microarray como no qRT-PCR.	105
<b>Tabela 11</b>	Representação dos genes diferencialmente expressos (DE) pertencentes as vias de sinalização estudadas nas amostras pareadas (DCIS x IDC).	109
<b>Tabela 12</b>	Representação dos genes diferencialmente expressos (DE) pertencentes ao processo EMT e vias de sinalização PI3K e WNT estudadas nas amostras de progressão.	113

**Tabela 13** Coeficientes de correlação linear para os pares gênicos selecionados.

## LISTA DE ABREVIACOES

<b>ADH</b>	Do ingls, <i>Atypical ductal hyperplasia</i>
<b>AJCC</b>	Do ingls, <i>The American Joint Committee on Cancer</i>
<b>BSA</b>	Do ingls, <i>Bovine serum albumin</i>
<b>cDNA</b>	DNA complementar a fita de RNA
<b>cm</b>	centmetro
<b>cRNA</b>	RNA complementar a fita de DNA
<b>CV</b>	Do ingls, <i>Cross Validation</i>
<b>Cy3/Cy5</b>	Citosina marcada com fluorforos
<b>DCIS</b>	Do ingls, <i>Ductal Carcinoma in situ</i>
<b>DCIS/IDC</b>	Do ingls, <i>Ductal Carcinoma in situ of Invasive Ductal Carcinoma</i>
<b>dCTP</b>	Nucleotdeo Citosina
<b>DEPC</b>	Dietil Pirocarbonato
<b>DNA</b>	cido Desoxirribonuclico
<b>DNTP</b>	Nucleotdeos de DNA
<b>dscDNA</b>	DNA complementar dupla fita
<b>EC</b>	Estadiamento Clnico
<b>EMT</b>	Do ingls, <i>Epithelial-mesenchymal transition</i>
<b>ER</b>	Do ingls, <i>Estrogen Receptor</i>
<b>EST</b>	Do ingls, <i>Expressed Sequence Tag</i>
<b>ETOH</b>	Etanol
<b>FAPESP</b>	Fundao de Amparo  Pesquisa do Estado de So Paulo
<b>GO</b>	Do ingles, <i>Gene Ontology</i>
<b>HB4a</b>	Linhagem de clulas epitliais mamrias humanas
<b>HCl</b>	cido Cloridrico
<b>HE</b>	Hematoxilina-eosina
<b>IDC</b>	Do ingls, <i>Invasive Ductal Carcinoma</i>
<b>ILC</b>	Do ingls, <i>Invasive Lobular Carcinoma</i>
<b>KEGG</b>	Do ingls, <i>Kyoto Encyclopedia of Genes and Genomes</i>

<b>LCIS</b>	Do inglês, <i>Lobular Carcinoma in situ</i>
<b>LCM</b>	Do inglês, <i>Laser Capture Microdissection</i>
<b>LDA</b>	Do inglês, <i>Linear Discriminant Analysis</i>
<b>LOHs</b>	Do inglês, <i>Loss of heterozygosities</i>
<b>LOWESS</b>	Do inglês, <i>Local Weight Regression</i>
<b>M</b>	Molar
<b>mL</b>	Mililitro
<b>mM</b>	Milimolar
<b>mm</b>	milímetro
<b>nm</b>	nanometro
<b>MPSS</b>	Do inglês, <i>Massive Parallel Signature Sequencing</i> ,
<b>N</b>	Tecido Não Neoplásico
<b>NaOH</b>	Hidróxido de Sódio
<b>OD</b>	Densidade Óptica
<b>ORESTES</b>	Do inglês, <i>open reading frame expressed sequence tags</i>
<b>PALM</b>	Do inglês, <i>Laser-MicroBeam System</i>
<b>PCR</b>	Do inglês, <i>Polimerase Chain Reaction</i>
<b>pb</b>	par de base
<b>PBS</b>	Do inglês, <i>Protein Block Serum</i>
<b>pFDR</b>	Do inglês, <i>positive False Discovery Ratio</i>
<b>PI3K</b>	Do inglês, <i>phosphoinositide 3-kinase</i>
<b>PR</b>	Do inglês, <i>Progesterone Receptor</i>
<b>qRT-PCR</b>	Do inglês, <i>quantitative Reverse Transcriptase-PCR</i>
<b>RIN</b>	Do inglês, <i>RNA Integrity Number</i>
<b>RNA</b>	Ácido Ribonucléico
<b>RNAa</b>	RNA amplificado
<b>RNA<sub>m</sub></b>	RNA mensageiro
<b>RNTP</b>	Nucleotídeos de RNA
<b>rpm</b>	rotações por minuto
<b>RT-PCR</b>	Do inglês, <i>Reverse Transcriptase-PCR</i>
<b>SAGE</b>	Do inglês, <i>Serial Analysis of Gene Expression</i>
<b>SBR</b>	Do inglês, <i>Scarff-Bloom-Richardson</i>

<b>SVD</b>	Do inglês, <i>Singular Value Decomposition</i>
<b>TAE</b>	Do inglês, <i>Tris-acetate-EDTA</i>
<b>U</b>	Unidade
<b>µg</b>	microgramas
<b>µg/µL</b>	microgramas por microlitro
<b>UICC</b>	Do inglês, <i>Union Internationale contre le Cancer</i>
<b>µL</b>	microlitro
<b>µm</b>	micrometro
<b>µM</b>	micromolar
<b>&gt;</b>	Maior
<b>≥</b>	Maior ou igual
<b>&lt;</b>	Menor

# ÍNDICE

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
1.1	Câncer de mama	1
1.2	Mama normal, carcinoma ductal <i>in situ</i> e invasivo	2
1.3	Diagnóstico, prognóstico e tratamento	7
1.4	Genética do câncer de mama	13
1.5	Processo de invasão e progressão tumoral	17
1.6	Vias de sinalização WNT e processo EMT	19
1.7	Via de sinalização PI3K e associação com a via WNT e processo EMT	25
1.8	Perfil de expressão gênica em larga escala	29
1.9	Análise do perfil de expressão gênica em câncer de mama	32
<b>2</b>	<b>OBJETIVOS</b>	<b>38</b>
2.1	Objetivo geral	38
2.2.	Objetivos específicos	38
2.2.1	Plataforma (4.8K)	38
2.2.2	Plataforma Contendo Genes das Vias de Sinalização WNT, PI3K e Processo EMT (LV-01)	39
<b>3</b>	<b>METODOLOGIA</b>	<b>40</b>
3.1	Amostras	40
3.2	Microdissecção a <i>laser</i>	41
3.3	Extração de RNA total	43
3.4	Avaliação do RNA total extraído	43
3.5	Amplificação do RNA mensageiro (RNAm)	48

3.6	Avaliação da qualidade e quantidade do RNA amplificado	49
3.7	Microarrays de cDNA	49
3.7.1	Plataforma 4.8k	49
3.7.2	Plataforma das vias de sinalização WNT, PI3K e processo EMT (LV-01)	50
3.7.3	Marcação e Hibridização	52
3.7.4	Captura das imagens e quantificação de sinal	53
3.7.5	Pré-análise das lâminas hibridizadas	55
3.8	Análises matemática	58
3.8.1	Identificação dos genes diferencialmente expressos na plataforma 4.8k	59
3.8.2	Identificação dos genes diferencialmente expressos na plataforma das vias de sinalização WNT, PI3K e processo EMT (LV-01)	60
3.9	qRT-PCR ( <i>quantitative reverse transcriptase polymerase chain reaction</i> )	64
3.9.1	Síntese da primeira fita de cDNA	64
3.9.2	Desenho dos iniciadores	64
3.9.3	RT-PCR para verificação da amplificação específica dos iniciadores	66
3.9.4	RT-PCR quantitativo (qRT-PCR) ou RT-PCR em tempo real ( <i>real time PCR</i> )	66
3.9.5	Padronização dos iniciadores	68
3.9.6	Cálculo da eficiência de amplificação dos iniciadores	68
3.9.7	Seleção do gene normalizador (endógeno)	72
3.9.8	Análise dos dados de expressão	73
3.9.9	Determinação das diferenças de expressão	74
3.10	Construção do <i>tissue microarray</i> (TMA)	75
3.11	Imunohistoquímica	77
3.12	Análises estatísticas	79
<b>4</b>	<b>RESULTADOS</b>	<b>81</b>
4.1	Delineamento experimental	81

4.2	Resultados referentes a plataforma 4.8K	82
4.2.1	Amostras e qualidade das lâminas	82
4.2.2	Identificação de genes diferencialmente expressos entre os componentes DCIS e IDC presentes na mesma amostra	84
4.2.3	Caracterização funcional dos genes diferencialmente expressos	86
4.2.4	Seleção dos genes diferencialmente expressos entre as amostras pareadas para serem avaliados por RT-PCR em tempo real	88
4.2.5	Confirmação da expressão gênica por qRT-PCR dos genes diferencialmente expressos selecionados entre as amostras pareadas	90
4.2.6	<i>Tissue Microarray</i> (TMA)	91
4.3	Identificação de genes diferencialmente expressos entre amostras não-neoplásicas, DCIS puro e <i>in situ</i> associado ao IDC (DCIS/IDC)	97
4.3.1	Avaliação da similaridade do padrão molecular e morfológico em lesões de carcinoma ductal	98
4.3.2	Genes envolvidos na progressão do DCIS	99
4.3.3	Caracterização funcional dos genes diferencialmente expressos	103
4.3.4	Confirmação dos genes selecionados como supostamente envolvidos na progressão do DCIS por RT-PCR em tempo real	104
4.3.5	Resultados referentes a plataforma contendo genes das vias de sinalização WNT, PI3K e processo EMT (LV-01)	106
<b>5</b>	<b>DISCUSSÃO</b>	<b>122</b>
<b>6</b>	<b>CONCLUSÕES</b>	<b>141</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>143</b>

## ANEXOS

- Anexo 1** Quadro completo com informações das pacientes utilizadas no estudo.
- Anexo 2** Quadro com as informações da qualidade do RNA total e quantidade do RNA amplificado.
- Anexo 3** Protocolo RNA 6000 – *bioanalyzer* – guia de referência rápida.
- Anexo 4** Protocolo de Amplificação de RNA mensageiro.
- Anexo 5** Protocolo de Marcação e Hibridização em Lâmina de Vidro.
- Anexo 6** Amostras Utilizadas no Estudo.
- Anexo 7** Lista dos 33 genes diferencialmente expressos entre as amostras pareadas (DCIS x IDC) através do teste T de *Student* pareado ( $p < 0,01$  e  $Fold > 1,5$ ).
- Anexo 8** *Box Plot* dos 33 genes com expressão alterada entre as amostras pareadas (Plataforma 4.8k).
- Anexo 9** Valores de diferença de expressão para cada par de amostra analisado através de qRT-PCR.
- Anexo 10** Lista dos 90 genes diferencialmente expressos entre DCIS puro, DCIS/IDC e IDC ( $p < 0,01$ ).
- Anexo 11** Lista dos 147 genes supostamente envolvidos nos estágios iniciais do processo de malignização do DCIS na lâmina 4.8k ( $p < 0,01$  e diferença de expressão (*Fold*)  $> 2$ ).
- Anexo 12** Lista dos 30 genes diferencialmente expressos entre as amostras pareadas ( $p < 0,05$ ) pertencentes as vias de sinalização WNT, PI3K e processo EMT.
- Anexo 13** Listas geradas através do teste não paramétrico de Wilcoxon realizado para a combinação dois a dois dos tipos histológicos com nível de significância  $\alpha < 0,05$  e diferença de expressão (*Fold*)  $> 2$ .
- Anexo 14** Lista dos 57 genes diferencialmente expressos entre as amostras progressão ( $p < 0,05$  e  $Fold > 2$ ) pertencentes as vias de sinalização WNT, PI3K e processo EMT.
- Anexo 15** Gráficos *scatter-plot*'s para aqueles pares de genes com coeficientes de correlação superiores a 0,9.

# 1 INTRODUÇÃO

## 1.1 CÂNCER DE MAMA

Dentre os diversos tipos de câncer, o de mama está entre as neoplasias de maior incidência mundial, afetando em média uma em cada dez mulheres. Devido sua alta frequência é o mais temido, sobretudo pelos seus efeitos psicológicos. No Brasil, este tipo de câncer é o que mais causa mortes entre as mulheres. De acordo com as estimativas de incidência de câncer no Brasil para o ano de 2008, o câncer de mama será o mais incidente entre as mulheres, com 49.400 novos casos. O estado de São Paulo apresenta a terceira mais alta taxa estimada de incidência (72,52 casos para cada 100.000 mulheres) sendo estimado um maior número de casos no estado do Rio de Janeiro (92,77:100.000) (Ministério da Saúde 2007).

O câncer mamário é uma doença de progressão variável, sendo sua classificação histológica baseada no grau, estadio e tipo de tumor. 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, sendo discreta, moderada ou elevada, são atribuídas pontuações de 1 a 3, respectivamente. Assim, a soma dos 3 parâmetros fornece um escore, que corresponde ao grau histológico, sendo grau 1: escore 3 a 5, grau 2: escore 6 ou 7 e grau 3: escore 8 ou 9, de

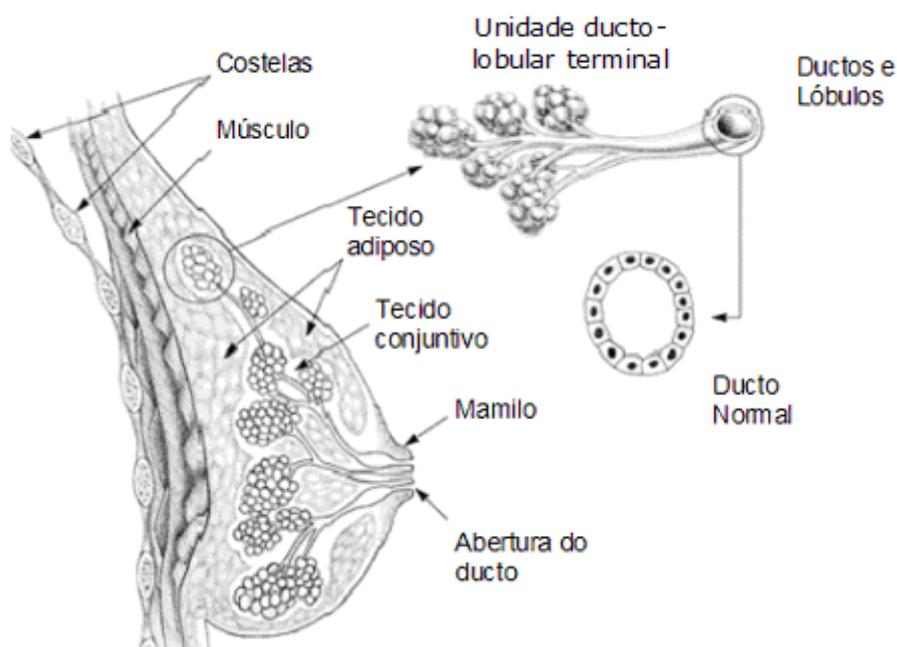
acordo com escore *Scarff-Bloom-Richardson* (SBR) (DOUSSAL et al. 1989; HOWEL et al. 1994).

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. Esse sistema foi criado pela *Union Internationale contre le Cancer-UICC* e pelo *The American Joint Committee on Cancer AJCC* e está baseado na extensão do tumor primário (T), presença e extensão de metastases em linfonodos regionais (N) e presença de metastases à distância (M). Os estadios são numerados de 0 a IV, em ordem decrescente de sobrevida. O estadio 0 – carcinoma *in situ*, estadio I – câncer 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 metástase à distância (STYBLO e WOOD 1998).

## **1.2 MAMA NORMAL, CARCINOMA DUCTAL *IN SITU* E INVASIVO**

A mama feminina normal adulta apresenta uma estrutura arborescente e ramificada, sendo composta por elementos epiteliais e estromais (Figura 1). O componente epitelial da mama é organizado em lóbulos que são compostos por grupos de ácinos. Estes se conectam em ductos terminais (unidade ducto-lobular terminal) que se unem formando ductos maiores. Há cerca de 15-20 ductos maiores cuja abertura é na papila. Já o estroma é composto de quantidade variável de tecido adiposo e

tecido conectivo fibroso, que consiste a maior parte do volume da mama no estado não lactente (SCHNITT 1995a). Quanto ao aspecto histológico, o epitélio ductal normal da mama consiste de camada interna (luminal) de células epiteliais cubóides e camada externa de células mioepiteliais (apoiada na membrana basal do ducto).

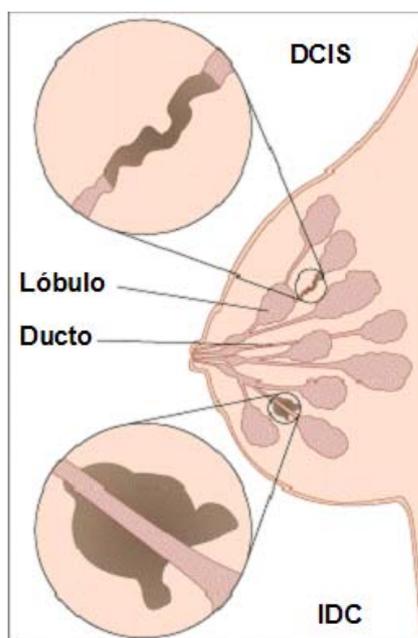


**Fonte:** The Cancer Council Victoria (2005)

**Figura 1** - Estrutura da mama feminina normal adulta.

Dentre os tipos histológicos de neoplasia de mama, o carcinoma ductal e lobular somam aproximadamente 90% dos casos, sendo o ductal o mais freqüente. O carcinoma *in situ* (não-invasivo) é o resultado de uma proliferação de células neoplásicas dentro dos ductos e/ou lóbulos. Há dois tipos de câncer de mama não-invasivo, o carcinoma ductal *in situ* (DCIS) e o lobular *in situ* (LCIS), caracterizados pela divisão anormal das células ductais ou lobulares respectivamente, permanecendo dentro dos ductos e/ou lóbulos. Já o carcinoma invasivo ocorre quando as células se espalham além

dos ductos (IDC) e lóbulos (ILC), com ruptura da membrana basal, apresentando invasão de células tumorais em tecidos adjacentes, sangue e linfa. Uma representação do DCIS e IDC pode ser visualizada na Figura 2.



**Fonte:** The Cancer Council Victoria (2005)

**Figura 2** - Representação de carcinoma ductal *in situ* e invasivo de mama.

O carcinoma ductal *in situ* é caracterizado por um heterogêneo espectro de lesões pré-invasivas originadas dentro do tecido normal da mama, variando em morfologia, extensão e apresentação clínica. É geralmente categorizado pela descrição arquitetural em 5 grupos: comedo (células malignas dentro do ducto com freqüente necrose na área central), cribriforme (padrão de crescimento em fenestra com proliferação relativamente homogênia), papilar (ducto contém tufos papilares fibrovasculares), sólido (ducto totalmente preenchido por células neoplásicas) e micropapilar (projeções papilíferas em forma de dedos em

espaços ductais dilatados) (SCHNITT 1995b). A classificação patológica do DCIS é baseada no grau nuclear (baixo, intermediário e alto), diferenciação arquitetural (polarização das células no lúmen, comparável a formação de túbulos) e necrose (presença ou ausência) (ROSSER 1998; ELLIS et al. 1999; BURSTEIN et al. 2004), aspectos estes, similares aos utilizados para avaliar o grau histológico em carcinomas invasivos (ELLIS et al. 1999).

Tem sido observado que a diferenciação nuclear das células tumorais é mais importante que o padrão de crescimento arquitetural, e novas classificações para o DCIS têm sido propostas. Entretanto, alguns estudos têm confirmado a importância de ambas as características mostrando que lesões de alto grau e lesões associadas à comedo-necrose estão relacionadas com maior risco de recorrência após cirurgia de conservação da mama (KERLIKOWSKA et al. 2003).

Um crescimento descontínuo dentro do sistema ductal da mama e a natureza irregular das calcificações associadas dificultam a determinação da extensão do carcinoma ductal *in situ*. Esta neoplasia se origina a partir de uma estrutura glandular única, mas pode se espalhar dentro da mama através do sistema ductal. Dois terços dos pacientes com DCIS de grau baixo à intermediário apresentam doença multifocal, caracterizada pelo crescimento intraductal descontínuo, com espaços de mais de 1 cm entre o foco tumoral. Em contraste, lesões de alto grau tendem a ser contínuas, não apresentando espaços maiores que 5 mm (HOLLAND et al. 1990; FAVERLY et al. 1994). O carcinoma ductal *in situ* pode estar associado ao tumor microinvasivo oculto (aquele que não excede 0,1 cm de diâmetro). Tais

casos são classificados como câncer de mama microinvasivo (SINGLETARY et al. 2002) e são geralmente tratados de acordo com diretrizes para doença invasiva.

Existe um consenso em literatura de que o desenvolvimento da maioria dos tumores ocorre a partir de estágios bem estabelecidos como hiperplasia (proliferação aumentada de células aparentemente normais), hiperplasia atípica (proliferação aumentada com células de aparência anormal), carcinoma *in situ* (proliferação localizada e delimitada de células anormais) e finalmente o carcinoma invasivo (invasão de células tumorais em tecidos adjacentes, sangue e linfa). Embora, em se tratando de câncer de mama, essas etapas não são bem estabelecidas.

Evidências patológicas e moleculares sugerem que o DCIS seja um precursor do carcinoma invasivo, porém não obrigatório (ZHUANG et al. 1995; CLAUS et al. 2001; AMARI et al. 2003; HWANG et al. 2004; MA et al. 2003; CHIN et al. 2004; BURSTEIN et al. 2004). No entanto, não é muito claro, qual a população de células progrediriam para a doença invasiva e quais as propriedades moleculares permitiriam a estas células a capacidade de se espalharem para os tecidos adjacentes.

### 1.3 DIAGNÓSTICO, PROGNÓSTICO E TRATAMENTO

Antes da introdução da mamografia, a maioria dos casos DCIS eram detectados por sintomas clínicos, incluindo massa palpável, descarga mamilar ou doença de Paget do mamilo. Atualmente no exame de mamografia, o DCIS é geralmente detectado por padrões típicos de microcalcificações (KESSAR et al. 2002). Este tipo de lesão representa 3-5% de todas as doenças malignas da mama detectadas clinicamente, e é responsável por aproximadamente 20% dos casos de câncer de mama detectados por rastreamento (Anonymus 1999; FRACHEBOUD et al. 2001; ERNSTER et al. 2002; LEONARD e SWAIN 2004).

A mastectomia sem quimioterapia adjuvante é o tratamento empregado em DCIS, sendo curativo em 98% dos casos e apresentando recorrência em 2% (BURSTEIN et al. 2004). Estudo recente (LEE et al. 2006) com 1.236 pacientes diagnosticadas com DCIS puro demonstrou que taxas de recorrência local invasiva após mastectomia foi menor que 1% e após tratamento de conservação da mama foi de 12% sendo que a probabilidade de recorrência sistêmica em 12 anos foi de aproximadamente 1%.

Esta neoplasia pode apresentar diferentes evoluções, progredindo rapidamente para carcinoma invasivo ou evoluindo lentamente durante um longo período da doença (PORTER et al. 2003).

O objetivo no tratamento desta neoplasia é a prevenção de recorrência local, em particular para o carcinoma invasivo. As opções de

tratamento cirúrgico incluem mastectomia simples ou cirurgia de conservação da mama, sendo extraídas apenas as áreas afetadas.

Além dos esforços em se prevenir recorrência é importante ressaltar que a mastectomia frequentemente representa um tratamento além do necessário em casos de DCIS não sintomáticos detectados por rastreamento, ainda mais, tendo em vista que a terapia de conservação da mama tem tornado-se uma alternativa geralmente aceitável em casos de câncer de mama invasivo detectados precocemente. Além disso, nem todos DCIS necessariamente progredem para carcinoma invasivo durante o período de vida da paciente, sendo o grande desafio identificar os casos que progrediriam para que medidas apropriadas possam ser tomadas.

Marcadores moleculares têm sido extensivamente investigados nos diferentes tipos de neoplasias com o objetivo de fornecer informações acuradas tanto em relação ao prognóstico, identificando pacientes com diferentes riscos de evolução da doença, quanto na predição de resposta a tratamento.

Nos casos de DCIS, o receptor de estrógeno (ER) é o primeiro marcador molecular com relevância na decisão do tratamento a ser empregado (LEONARD e SWAIN 2004). A presença deste receptor em câncer de mama não apenas prediz resposta à terapia endócrina, como se correlaciona com um bom prognóstico (GRUVBERGER et al. 2001; MAURIAC 2006). Alguns protocolos combinam radioterapia e cirurgia no tratamento de DCIS, reduzindo o risco de recorrência ipsilateral em 40-60% e/ou fazem uso de um modulador seletivo do receptor de estrógeno, o

tamoxifeno, como terapia adjuvante (FISHER et al. 2001a e b). Este hormonioterápico tem alta afinidade pelo receptor de estrógeno (ER), bloqueando a ação desse hormônio. O estrógeno age via seu receptor, estando envolvido na regulação de processos transcricionais relacionados ao crescimento e diferenciação celular da glândula mamária (HENDERSON et al. 1988; PIKE et al. 1993). Um estudo no qual pacientes com DCIS foram tratadas randômicamente com ou sem radiação e tamoxifeno, mostrou que as pacientes que receberam ambos tratamentos apresentaram a menor taxa de recorrência, porém benefícios adicionais do tamoxifeno foram pequenos e não apresentaram significância estatística (HOUGHTON et al. 2003). Apesar disso, a relativa baixa toxicidade apresentada pelo tamoxifeno têm garantido seu uso num período de 5 anos como profilaxia em tumores ER positivos (FISHER et al. 2001b). Outro marcador hormonal se refere a expressão do receptor de progesterona (PR) que parece ser similar a expressão do ER em DCIS (LEONARD e SWAIN 2004).

A família dos receptores de fatores de crescimento tem atraído a atenção de pesquisadores há alguns anos, especialmente após o desenvolvimento de terapias utilizando anticorpos monoclonais humanizados. Os quatro membros da família desses receptores denominados EGFR (HER1/ ERBB1), ERBB2 (HER2/neu), ERBB3 (HER3) e ERBB4 (HER4), são proteínas transmembrânicas com domínio citoplasmático tirosina quinase, sendo o ERBB2 (HER2/neu) aquele com maior impacto em pesquisas e tratamento de câncer de mama (LIU et al. 2004). Atualmente estes marcadores moleculares têm sido examinados em

DCIS, uma vez que sua expressão e implicações prognósticas são semelhantes entre as lesões DCIS e IDC (LEONARD e SWAIN 2004). Alguns estudos sobre ERBB2 em DCIS tem demonstrado uma maior expressão em 50% dos casos e está associada predominantemente ao tipo pouco diferenciado e pacientes com DCIS comedo e alto grau (LIU et al. 1992; MILLIS et al. 1996; VOS et al. 1999). A expressão desta proteína é encontrada em mais casos de DCIS do que em carcinoma invasivo, possivelmente indicando um papel para ERBB2 na tumorigênese (DIGIOVANNA et al. 2002). No entanto, a presença de ERBB2 em DCIS não está associada a alteração na conduta clínica. Quanto à quimioterapia, esta não desempenha nenhum papel no tratamento de carcinoma ductal *in situ* (BURSTEIN et al. 2004).

Já os casos de carcinoma invasivo, além dos marcadores ER e PR, que são utilizados com as mesmas condutas clínicas adotadas no DCIS, os membros da família dos receptores de fatores de crescimento são usados com o intuito de se aplicar tratamentos mais eficazes e garantir melhores resultados prognósticos.

A amplificação do gene *EGFR* tem sido descrita em vários tumores, inclusive no câncer de mama (AL-KURAYA et al. 2004; BHARGAVA et al. 2005; CORZO et al. 2005; REIS-FILHO et al. 2005). O ERBB2 aparece amplificado nos carcinomas invasivos (SLAMON et al. 1987; VAN DE VIJVER et al. 1987; ALLRED et al. 2001) em 20-40% (MORRIS e CAREY 2007), e representa um marcador de aumento de taxas de recorrência. Estudos relacionados a maior expressão da proteína codificada por este

gene tem sido realizados em câncer de mama (VAN DE VIJVER et al. 1988; KING et al.1989; TSUDA et al. 2001; KOBAYASHI et al. 2002) mostrando que a maior expressão de ERBB2 em carcinoma invasivo está correlacionada com pior prognóstico e sobrevida global mais curta (VARLEY et al. 1987; SLAMON et al. 1989; TSUDA et al. 1989; DONOVAN-PELUSO et al. 1991). O ERBB2 é um marcador preditivo de resposta a trastuzumab (Herceptina), que é um anticorpo monoclonal humanizado (mAb) contra ERBB2. Descoberto em 1998 para o tratamento do câncer de mama, tem um excelente perfil anti-tumoral, particularmente quando usado em combinação com doxorubicina e paclitaxol (KUMAR e MADISON 2001).

Em mama, uma vez que as células tumorais adquirem a habilidade de degradar a membrana basal, elas são capazes de escapar do sistema ductal e o tumor invasivo ganha o potencial de disseminar para órgãos distantes (ALLRED 1997), mais frequentemente acometendo fígado, pulmão e tecido ósseo.

A presença de metástase tumoral em linfonodos regionais é o fator prognóstico mais relevante em pacientes com tumores malignos de origem epitelial (MCGUIRE 1987; HARRISON et al. 1994; BALCH et al. 2001). Aproximadamente 90% de todas as mortes por câncer resultam de formação de metástase (CAVALLARO e CHRISTOFORI 2001). A mais ameaçadora é a metástase em órgãos viscerais e cérebro com taxas de sobrevida em 5 anos geralmente inferiores a 20% (LANDIS et al. 1998).

Em IDC, o *status* linfonodal desempenha um papel fundamental no prognóstico, estadiamento (FOSTER et al. 1996) e manejo desta doença

(ZANGHÌ et al. 2004). A biópsia de linfonodo sentinela foi introduzida na clínica e rapidamente aceita, uma vez que esta técnica tem confirmado a progressão anatômica das células tumorais a partir do sítio primário para o linfonodo regional através de capilares linfáticos (NATHANSON 2003). Desta forma, o *status* linfonodal é o fator que guia a seleção de pacientes elegíveis à terapia adjuvante, embora pacientes com câncer de mama linfonodo negativo com pelo menos um fator de risco (por exemplo, T>3, *status* ER negativo e alto grau) são indicados à terapia adjuvante (CARTER et al. 1989; ZUJEWSKI e LIU 1998; GOLDBIRSCHE et al. 2001). Assim, a quimioterapia sistêmica adjuvante está reservada para pacientes linfonodo-negativos de alto risco e todos os pacientes linfonodo-positivos.

A quimioterapia neoadjuvante em câncer de mama apresenta os mesmos benefícios de sobrevivência que a quimioterapia adjuvante, e ainda aumenta a probabilidade de conservação da mama em câncer avançado (FISHER et al. 1998; VAN DER HAGE et al. 2001). Outra vantagem se refere a possibilidade de ser usada na avaliação da responsividade ou resistência a um tratamento específico *in vivo*, permitindo alterar este tratamento após a cirurgia, se o mesmo mostrou-se inadequado. Em carcinoma de mama operável as taxas de resposta podem variar de 65 a 100% ou não apresentarem resposta para um regime de droga em particular. Vários regimes de drogas têm sido utilizados em uma variedade de ciclos, porém a predição da resposta ao tratamento permanece um desafio e muitos grupos tem se concentrado nesta proposta utilizando

técnicas moleculares em larga escala (WOLFF e DAVIDSON 2000; CHANG et al. 2003; SHIMIZU et al. 2004; FOLGUEIRA et al. 2005).

Neste contexto, apesar da classificação morfológica atual fazer uso de marcadores moleculares, ainda hoje muitas pacientes com mesmo diagnóstico e prognóstico clínico podem apresentar resultados muito diferentes em nível de progressão, recidiva e presença de metástase (NAKHLIS e MORROW 2003; SOTIRIOU et al. 2003).

#### **1.4 GENÉTICA DO CÂNCER DE MAMA**

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 idade avançada, doença benigna de mama, menarca precoce, menopausa tardia (após os 50 anos), e a nuliparidade ou a primeira gestação após os 30 anos de idade podem estar associados a uma maior suscetibilidade a essa doença (KERLIKOWSKA et al. 1997; TRENTHAM-DIETZ et al. 2000; CLAUS et al. 2001; Ministério da Saúde 2008).

Diversos genes e mecanismos envolvidos no desenvolvimento do câncer de mama têm sido identificados, como as proteínas codificadas pelos genes *BRCA1* e *BRCA2* que estão envolvidas na regulação da estabilidade genômica e reparo do DNA (FRANK 2002). Mutações nestes genes estão associadas a maior risco de câncer hereditário de mama e ovário (HILL et al. 1997; ROSENTHAL e PUCK 1999). Modelos baseados em populações que

fazem uso de dados demográficos e história familiar predizem que aproximadamente 3,7% e 1,9% das mulheres com carcinoma ductal carregam a mutação nos genes de câncer hereditário *BRCA1* e *BRCA2* respectivamente (CLAUS et al. 2003).

A proteína codificada pelo gene supressor de tumor *TP53* impede a progressão do ciclo celular permitindo que o DNA seja reparado ou levando a célula à apoptose. Uma célula que é incapaz de reparar o DNA com eficiência, gera assim, células geneticamente instáveis. Este evento genético é o mais freqüente nos cânceres humanos sendo que 50% dos casos apresentam mutação nessa proteína (SOUSSI 2000). Mutações neste gene são encontradas em aproximadamente 20-40% dos carcinomas invasivos de mama (PROSSER et al. 1990; VARLEY et al. 1991) e tem sido demonstradas em 40% dos DCIS alto grau, sendo muito raro em DCIS baixo grau ou intermediário (WALKER et al. 1997; DONE et al. 2001).

Receptores hormonais como o receptor de estrógeno (ER) e o receptor de progesterona (PR) são expressos em 50-60% dos pacientes com carcinoma ductal (LEONARD e SWAIN 2004). O estrógeno, via seu receptor (ER), desempenha um papel importante na regulação do crescimento e diferenciação do epitélio normal da mama (HENDERSON et al. 1988; PIKE et al. 1993). Este estimula a proliferação celular e regula a expressão de outros genes incluindo o receptor de progesterona (PR). O PR age como mediador do efeito mitogênico da progesterona estimulando a proliferação (HENDERSON et al. 1988; PIKE et al. 1993).

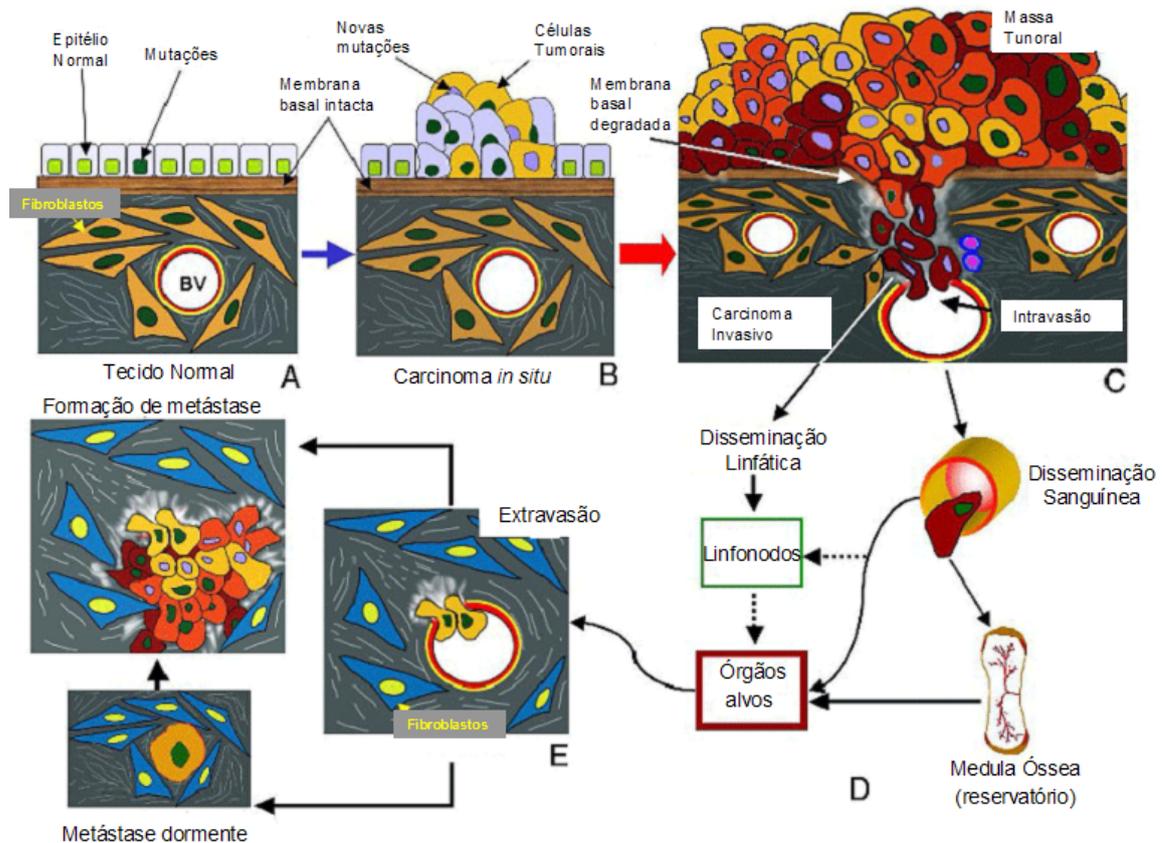
A família de receptor de fator de crescimento epidermal desencadeia sinais proliferativos e antiapoptóticos. A cascata de transdução de sinal frequentemente ativada no câncer de mama é iniciada pela família EGFR e seus 3 membros homólogos da família (ERBB2, ERBB3 e ERBB4) caracterizados por um sítio de ligação extracelular, uma região transmembrana e um domínio tirosino-quinase citoplasmático (HYNES e LANE 2005). A ativação do EGFR é decorrente da ligação de fatores relacionados ao EGF (*epidermal growth factor*), tais como TGF $\alpha$ , anfiregulina e o próprio EGF, que induzem a homo ou heterodimerização do receptor e a estimulação de vias tirosino-quinase intrínsecas (SCHLESSINGER 2000). Esse estímulo leva à auto-fosforilação dos resíduos de tirosina no domínio citoplasmático do receptor, criando sítios de ligação para proteínas adaptadoras e enzimas das cascatas da MAPK e PI3K, ligadas à proliferação celular, angiogênese, invasão, metástases e inibição da apoptose (SALOMON et al. 1995; HUTCHESON et al. 2006). O estudo da expressão de EGFR em câncer de mama foi observado em 48% dos 40 casos, mas sua expressão não foi associada com o grau histológico (SUO et al. 2001). O ERBB2 codifica este receptor localizado no braço longo do cromossomo 17 (17q21) e aparece amplificado em 15-25% nos carcinomas invasivos (SLAMON et al. 1987; VAN DE VIJVER et al. 1987; RAVDIN e CHAMNESS 1995; ALLRED et al. 2001), podendo promover a motilidade celular (DE POTTER e QUATACKER 1993; DE POTTER 1994) contribuindo assim na habilidade das células tumorais invadirem tecidos adjacentes e formarem metástases.

O gene da *ciclina D1* (*CCND1*) codifica uma proteína nuclear importante na regulação do ciclo celular, estando relacionada a estimulação do fator de crescimento na proliferação celular e apresenta um aumento de expressão em 50% dos tumores mamários humanos (BARTEK e LUKAS 2001). A expressão de *BCL2*, que codifica uma proteína que bloqueia apoptose, tem sido observada em lesões pré-invasivas de mama (SIZIOPIKOU et al. 1996). O gene *pRB* codifica uma proteína que age no controle da progressão do ciclo celular, apoptose e diferenciação (SIMIN et al. 2004). A angiogênese tem sido associada a maior agressividade (STYBLO e WOOD 1998). Podemos citar a glicoproteína VEGF (*vascular endothelial growth factor*) que é um potente mitógeno celular endotelial e fator angiogênico que desempenha um papel no desenvolvimento e crescimento do tumor (SALVEN et al. 1998; FINE et al. 2000). Além disso, está envolvida na manutenção da viabilidade celular em ambas as células, não-neoplásica e neoplásica (FERRARA et al. 1996; CARMELIET et al. 1996). Em carcinoma de mama, o VEGF tem se mostrado um fator prognóstico importante (RELF et al. 1997; ENGELS et al. 1997; GASPARINI 2000; LINDERHOLM et al. 2000), ocorre maior expressão em células tumorais do que em não-neoplásicas, e níveis significativamente mais altos em tumores de mama grau 2 do que grau 3 e tumores do tipo comedo no componente intraductal (YAMASAKI et al. 2005). Tem sido realizados estudos visando bloquear este potente fator pró-angiogênico (ELLIS et al. 2006; HAYES et al. 2007).

## 1.5 PROCESSO DE INVASÃO E PROGRESSÃO TUMORAL

As células normais se dividem sob instrução de um complexo processo de regulação de contacto célula a célula e do microambiente no qual está inserida. Essa regulação assegura que cada órgão mantenha seu tamanho e forma apropriada. De maneira geral, o câncer pode ocorrer por mutações que levam a célula a violar esta regulação, tornando-se independente do programa de reprodução celular no qual está inserida. Algumas das células do tumor primário podem perder a capacidade de adesão durante a progressão tumoral e invadir a membrana basal do tecido de origem por meio da produção de enzimas proteolíticas, caindo na circulação e formando áreas de proliferação em outros tecidos, sendo este processo denominado metástase. As 4 etapas de disseminação de células na formação de metástase foram bem caracterizadas por FIDLER (2003). O processo de disseminação de células envolve vários passos. Devido alterações moleculares o epitélio normal ganha potencial de proliferação anormal. No primeiro passo (**invasão**) as células epiteliais neoplásicas perdem a adesão célula-célula e ganham motilidade permitindo que invadam tecidos adjacentes. Durante o segundo passo (**intravasão**) células tumorais penetram através do endotélio sanguíneo ou vasos linfáticos para entrar na circulação sistêmica. Apenas algumas células tumorais circulantes são capazes de sobreviver a passagem através da circulação. Assim, estas sobreviventes se dirigem para completarem o próximo passo (**extravasão**) quando elas extravasam através do endotélio capilar em sítios distantes.

Finalmente, em um novo ambiente, um pequeno grupo de células obtém sucesso na colonização no tecido distante sendo capazes de proliferar e crescer (**micrometástase**) resultando em tumores malignos secundários (Figura 3).



Fonte: AGUIRRE-GHISO (2005)

**Figura 3** - Processo de disseminação de células na formação de metástase

Para espalharem dentro dos tecidos, as células tumorais usam mecanismos de migração que são similares, se não idênticos aqueles que ocorrem normalmente em células normais durante processos fisiológicos, tais como embriogênese, cicatrização de uma ferida e tráfico de células imunes (FRIEDL e BROCKER 2000). Os princípios de migração celular

foram inicialmente investigados em fibroblastos não neoplásicos, queratinócitos e mioblastos (ABERCROMBIE et al. 1977; LAUFFENBURGER E HORWITZ 1996), porém estudos adicionais em células tumorais mostraram que as mesmas estratégias básicas são mantidas (FRIEDL e WOLF 2003), sendo detectadas histologicamente características de migração tumoral em carcinoma mamário (BELL e WAIZBARD 1986).

Apesar de este mecanismo estar melhor caracterizado atualmente, um dos maiores desafios na pesquisa exploratória do câncer é a compreensão dos mecanismos moleculares envolvidos em progressão tumoral, invasão local e formação de metástases.

## **1.6 VIAS DE SINALIZAÇÃO WNT E PROCESSO EMT**

Alterações nas propriedades de adesão das células tumorais estão associadas com a progressão tumoral e formação de metástases e, freqüentemente, coincidem com a perda da adesão celular mediada por E-caderina (do inglês, *epithelial cadherin*), alteração na morfogênese e arquitetura do tecido epitelial, perda da polaridade celular e inibição por contato, crescimento desregulado e invasão de tecidos adjacentes (HANDSCHUH et al. 1999; CAVALLARO e CHRISTOFORI 2001). Este processo é denominado transição epitélio–mesênquima (EMT, do inglês, *epithelial-mesenchymal transition*), e envolve a diminuição da expressão das moléculas de *E-caderina* sendo relatado nas primeiras etapas da invasão e

metástases de células tumorais (BOYER et al. 2000). Este evento é essencial durante os movimentos de gastrulação e formação da crista neural no desenvolvimento embrionário (SHOOK e KELLER 2003). O processo EMT é caracterizado por uma diminuição da função da E-caderina que marca o surgimento de um programa complexo no qual as células epiteliais adotam um fenótipo fibroblasto-*like* e apresentam atividade invasiva (CONACCI-SORRELL et al. 2002; NELSON e NUSSE 2004) e recentemente tem sido associado com as mudanças moleculares observadas nas células neoplásicas metastáticas mais agressivas (THIERY 2003).

Esse processo tem sido investigado em vários tipos tumorais, incluindo câncer de mama (ACKLAND et al. 2003; VINCENT-SALOMON e THIERY 2003). Durante o desenvolvimento do fenótipo metastático, as células tumorais ativam programas morfogenéticos latentes visando adquirir a habilidade de executar múltiplos passos neste processo (YANG et al. 2004). Há evidências que mostram que essa ativação ocorre no início da programação de metástase, sendo que, a célula epitelial tumoral altera seu processo de capacidade proliferativa, adquirindo a capacidade de invadir (YANG et al. 2004), assim genes envolvidos no processo EMT e via de sinalização WNT podem ser candidatos a marcadores moleculares de metástase.

Entre os vários tipos de regulação da *E-caderina* está a repressão transcricional causada pelos repressores *Sip1* (*smad interacting protein 1*) (COMIJN et al. 2001), *Snail* e *Slug* (BOLOS et al. 2003; BATLLE et al. 2000; HEMAVATHY et al. 2000), os quais se ligam às seqüências *E-box* de seu

promotor e diminuem sua expressão, o que afeta a adesão célula-célula e permite a invasão dos tecidos. Recentemente, foi demonstrado que o *Snail* tem um domínio semelhante ao da  $\beta$ -*catenina* e que esse domínio sinaliza fosforilação por *GSK3 $\beta$*  (YOOK et al. 2005), sendo este último considerado o ponto de conexão dos dois processos - via de sinalização WNT e processo EMT. Também o gene *Twist*, um regulador importante na morfogênese embrionária, desempenha um papel essencial na metástase. Sua expressão ectópica resulta na perda de adesão célula-célula mediada por *E-caderina*, ativação de marcadores mesenquimais e indução de motilidade celular (YANG et al. 2004).

A  $\beta$ -*catenina*, efetor da via de sinalização WNT, apresenta localização na membrana e citoplasma no tecido epitelial maduro e interage na membrana com a *E-caderina*. Esse complexo está associado à junção aderente das células epiteliais por estar ancorado ao citoesqueleto de actina (JAWHARI et al. 1999). Esta interação é fundamental para que ocorra uma adesão célula-célula eficiente (HIROHASHI e KANAI 2003) e quando essa é alterada ou perdida, as células tumorais podem adquirir capacidade invasiva que representa a primeira etapa da metástase. A  $\beta$ -*catenina* tem, portanto, uma função dupla nesse processo: como membro da função aderente, ela é essencial na ligação das caderinas ao citoesqueleto, permitindo a adesão celular; e como membro efetor da via de sinalização WNT.

A via de sinalização WNT está ativada durante a embriogênese e desativada no tecido adulto. Estudos concentram esforços na compreensão do braço canônico da via WNT (MORIN 1999; PEIFER e POLAKIS 2000)

inferindo que, durante a ativação da WNT na embriogênese, a  $\beta$ -catenina é impedida de ser degradada e é transportada ao núcleo, regulando a expressão de diversos genes, e após o término da diferenciação celular esse braço é desativado. Esse mecanismo é altamente conservado entre as espécies (WALTZER e BIENZ 1999; MOON et al. 2002) controlando diversas funções durante o desenvolvimento, como a sinalização da proliferação, adesão, diferenciação, morfologia celular e apoptose (BRENNAN e BROWN 2004).

Assim a  $\beta$ -catenina atua como uma proteína importante para adesão celular e como fator de transcrição, ambos importantes na manutenção do tecido diferenciado e desenvolvimento de um órgão respectivamente, sendo que sua não regulação pode levar ao desenvolvimento do câncer (HARRIS et al. 2005). A regulação da localização celular da  $\beta$ -catenina, via fosforilação, ocorre nos resíduos de serina e treonina na porção amino-terminal pelo complexo formado pela APC e GSK3 $\beta$  (MORIN 1999; PEIFER e POLAKIS 2000). Essa fosforilação resulta na ubiquitinação da  $\beta$ -catenina levando-a para a degradação pelos proteossomos (ABERLE et al. 1997). Quando a  $\beta$ -catenina não é fosforilada, esta é transportada para o núcleo formando um complexo com a família de ligantes de DNA TCF/LEF e APC (APC/ $\beta$ -catenina/Tcf-Lef) que é responsável pela regulação da transcrição de diversos genes, tais como o *c-myc* (HE et al. 1998) e a *ciclina D1* (TETSU e McCORMICK 1999). Outros genes também regulados pela  $\beta$ -catenina participam na transição epitélio-mesênquima (EMT). Já foram descritas duas proteínas reguladas por esse complexo de transcrição como capazes de

estimular a motilidade celular: o fator de crescimento de hepatócitos (*HGF*) e o fator de crescimento epitelial (*EGF*) (MULLER et al. 2002).

Alterações na regulação desta via são reconhecidamente importantes no desenvolvimento de inúmeros tumores (SPARKS et al. 1998; MIYOSHI et al. 1998), uma vez que a  $\beta$ -*catenina* volta a ativar a transcrição de genes que deveriam permanecer silenciados.

As proteínas WNTs são sinalizadoras da ativação ou inibição da via WNT. Quando as células são expostas a essas proteínas, estas se ligam ao receptor de superfície celular da família *Frizzled*, ativando ou desativando a via. Se o processo for a ativação, o complexo formado – WNT/*FRIZZLED* – antagoniza o APC-Axina bloqueando a fosforilação da  $\beta$ -*catenina* por mecanismos ainda não conhecidos mas que requerem a participação das proteínas *Dishevelled* (PEIFER e POLAKIS 2000). Mutações nas próprias WNTs são raramente encontradas como responsáveis pela ativação da via durante a carcinogênese em vários órgãos, sendo essa ativação mais frequentemente associada com mutações de *APC* ou de  $\beta$ -*catenina*.

Há evidências que a ativação canônica da via de sinalização WNT promove tumorigênese em tecidos mamários de camundongos. Por exemplo, certas linhagens de camundongos carregando uma mutação no gene *APC* apresentam sensibilidade acentuada para tumores mamários induzidos por carcinógenos (MOSER et al. 2001), enquanto que animais transgênicos expressando  $\beta$ -*catenina* estabilizada em glândulas mamárias, isto é, não sofrendo degradação, desenvolvem carcinomas (IMBERT et al. 2001). A ativação intracelular da via WNT por mutações nestes genes e

também no *AXIN* é um dos sinais mais freqüentes de anormalidade conhecidos em câncer humano (HOWE e BROWN 2004), sendo que tais mutações são raramente encontradas em câncer de mama (CANDIDUS et al. 1996; JONSSON et al. 2000; SCHLOSSHAUER et al. 2000). Apesar disso, a sinalização da via WNT em câncer de mama humano é observada pelos níveis elevados de *β-catenina* nuclear ou citoplasmático detectados por imunohistoquímica na maioria das amostras de tecidos tumorais de mama (LIN et al. 2000; RYO et al. 2001) sugerindo que a ativação da via WNT ocorre por outros mecanismos. Níveis elevados de *β-catenina* localizada no núcleo e/ou citoplasma foi primeiramente demonstrado por LIN et al. (2000) em aproximadamente 60% das amostras clínicas analisadas e foi confirmado em um estudo independentemente realizado por RYO et al. (2001) que reportaram resultados similares em 63% das espécimes cirúrgicas, sendo que em nenhum desses estudos foi observada localização nuclear ou citoplasmática da *β-catenina* em tecidos normais de mama. Em análises de *Western blotting* foi demonstrado elevados níveis de *β-catenina* em lisados tumorais de mama (JONSSON et al. 2000; RYO et al. 2001), correlacionando com pior prognóstico e atuando como um fator prognóstico forte e independente em pacientes com esta doença (LIN et al. 2000). Também foi verificado em estudos de cultura de células e estudos clínicos em amostras tumorais, que níveis elevados de *β-catenina* desempenham um papel significativo em câncer de mama (LIN et al. 2000) podendo ativar constitutivamente componentes *downstream* da sinalização, tal como o gene *c-myc*, causando crescimento tumoral (JONSSON et al. 2000).

Defeitos genéticos no gene *APC* e *β-catenina* são incomuns em câncer de mama, portanto, mutações em outros componentes da via de sinalização WNT podem estar envolvidos na carcinogênese mamária (JONSSON et al. 2000).

Em conjunto, esses resultados sugerem que a caracterização da expressão de genes implicados nos dois processos – via de sinalização WNT e processo EMT – durante carcinogênese mamária e invasão pode contribuir para melhor compreensão destes mecanismos no tumor, além de revelar genes chaves que poderão ser testados como marcadores de prognóstico e preditivos de metástase.

## **1.7 VIA DE SINALIZAÇÃO PI3K E ASSOCIAÇÃO COM A VIA WNT E PROCESSO EMT**

A família das PI3K (*phosphoinositide 3-kinase*) são enzimas que quando ativadas por uma variedade de estímulos celulares, fosforila lipídeos de inositol para gerar 3-fosfoinositóis que desempenham o importante papel de mensageiros secundários na via de sinalização celular (BRADER e ECCLES 2004). Muitas proteínas de sinalização são recrutadas especificamente por lipídeos fosforilados. Estas proteínas incluem quinases serina-treonina, proteínas de tirosina quinase e fatores que regulam proteínas G, sendo encontradas normalmente no citosol na forma inativa, mas quando localizada na membrana se tornam ativas e participam na

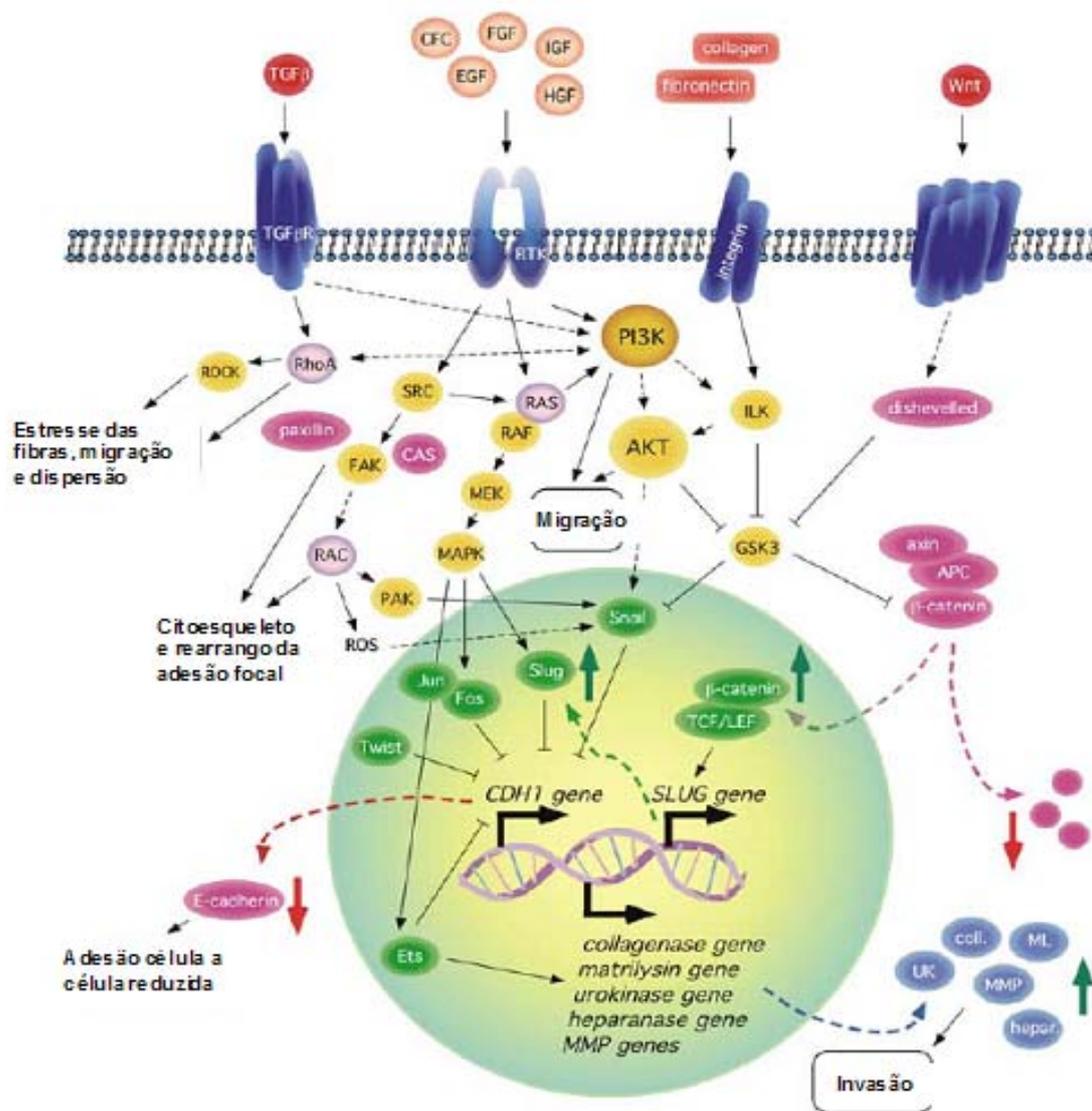
organização de complexos de sinalização e ativação de cascatas de quinase (BRADER e ECCLES 2004).

Efetores de sinalização *downstream* a PI3K incluem membros da família AKT serina-treonina quinase (também conhecida como PKB). Células produzindo constitutivamente formas ativas de AKT produzem o repressor transcricional, *Snail*, conhecido por reprimir a expressão do gene *E-caderina* (GRILLE et al. 2003). Esta repressão transcricional foi demonstrada em um estudo na qual linhagem de carcinoma de células escamosas que apresentaram maior expressão de AKT mutado induziram respostas celulares referentes a conversão de células epiteliais em células invasivas mesenquimais (GRILLE et al. 2003). Esta indução do processo de EMT através da AKT ativada envolve a perda de adesão célula a célula, mudanças na morfologia, perda da polarização celular, indução da motilidade celular, diminuição da adesão célula-matriz e mudanças na produção e distribuição de proteínas específicas (LARUE e BELLACOSA 2005).

A via PI3K/AKT apresenta-se frequentemente ativada no câncer epitelial humano (RINGEL et al. 2001; SUN et al. 2001; TESTA e BELLACOSA 2001). Além de estar envolvida na regulação da E-caderina, também se relaciona a produção de metaloproteinases (KIM et al. 2001; PARK et al. 2001), motilidade e invasão (ALTOMARE e TESTA 2005; MAJUMDER e SELLERS 2005).

Enquanto a via canônica de sinalização WNT e PI3K/AKT pode convergir no ponto de inibição da proteína *GSK3 $\beta$* , há indicações que as

duas vias possam afetar e fosforilar diferentes tipos de *GSK3* (WESTON e DAVIS 2001; GRILLE et al. 2003). Outra possível conexão inclui a estimulação da atividade AKT através da sinalização *Wnt/Disheveled* (FUKUMOTO et al. 2001) e a estimulação inversa da transcrição da  *$\beta$ -catenina* (e possivelmente fosforilação) pela AKT (TIAN et al. 2004). Porém, o papel da comunicação entre as vias de sinalização PI3K/AKT and *Wnt/ $\beta$ -catenina* no processo EMT precisa ser melhor estabelecido. Na Figura 4 pode ser visualizado um esquema da via de transdução de sinal associada ao processo de transição epitélio-mesênquima, PI3K e WNT.



**Legenda:** Os efeitos do processo EMT estão descritos nas caixas de texto. RTK: Receptor de tirosina quinase; ROS: Espécies reativas de oxigênio; ILK: Integrin-linked kinase; TGFβ: Fator de crescimento tumoral.

**Fonte:** LARUE e BELLACOSA et al. (2005)

**Figura 4** - Esquema da via de transdução de sinal associada ao processo de transição epitélio-mesênquima, PI3K e WNT.

## 1.8 PERFIL DE EXPRESSÃO GÊNICA EM LARGA ESCALA

Nos últimos 10 anos, os avanços tecnológicos tem resultado na habilidade de examinar alterações genômicas, transcricionais e proteômicas entre diferentes tipos de tecido em larga escala. Em termos de avaliação transcricional, métodos para detectar diferenças no nível de RNA mensageiro (RNAm) tem permitido acesso a uma visão global dos processos celulares por técnicas de análise da expressão gênica diferencial. Desta maneira, várias tecnologias baseadas em análise de seqüências podem ser citadas neste contexto, tais como EST (*Expressed sequence tag*), MPSS (*Massive Parallel Signature Sequencing*), SAGE (*Serial Analysis for Gene Expression*). Estas técnicas possibilitam a avaliação da expressão de um número muito grande de genes sem a necessidade do conhecimento prévio de suas seqüências, possuindo grande potencial analítico e sensibilidade, contudo sendo prejudicadas pela diferença genética individual até poucos anos atrás, uma vez que não eram apropriadas para um grande número de amostras e conseqüentemente apenas um pequeno número de casos de uma dada classe tumoral poderiam ser avaliados (CARRARO et al. 2007). Atualmente, com a introdução dos sequenciadores de nova geração, este problema está sendo resolvido, uma vez que várias amostras podem ser avaliadas (MEYER et al. 2008), contudo alguns ajustes estão sendo estabelecidos na aplicação destas metodologias.

Por outro lado, a tecnologia de *microarray* permite examinar a expressão de milhares de genes simultaneamente em um grande número de

amostras. Esta tecnologia é amplamente usada em estudos relacionados ao câncer. A plataforma de *microarray* pode ser confeccionada pela fixação de um segmento de DNA (cDNA, oligonucleotídeos ou pequenos fragmentos de produtos de PCR que correspondem ao RNAm) que deve representar um único gene, o qual é aderido a uma superfície sólida, frequentemente em lâmina de vidro (EISEN et al. 1998; KHAN et al. 1999) podendo ser utilizado plástico, silicone, acrílico ou substratos baseados em carbono (WU et al. 2008; PHILLIPS et al. 2008). Esta metodologia consiste na hibridização das seqüências imobilizadas com moléculas de cDNA ou cRNA provenientes de duas situações celulares distintas marcadas com diferentes fluoróforos e hibridizadas de forma competitiva. A intensidade de fluorescência emitida das moléculas hibridizadas por complementariedade à seqüência imobilizada na lâmina é proporcional ao nível de expressão do gene.

No estudo do câncer, a análise do perfil de expressão gênica tem se mostrado valiosa, uma vez que permite comparar o perfil de expressão de várias classes de tumores e tecidos normais, gerando informações importantes para a compreensão da transformação celular e progressão tumoral. Este método permite a análise rápida de múltiplos marcadores, sendo sua maior contribuição o reconhecimento de padrões de expressão gênica que possam ser associados a subgrupos com comportamento clínico específico (VAN´T VEER et al. 2005).

A expressão gênica célula-específica é frequentemente subestimada devido a composição heterogênea do tecido tumoral que é composto não apenas por células tumorais, mas também por células normais, estromais,

células infiltrativas do sistema imunológico e também células endoteliais dos vasos sanguíneos e linfáticos, interferindo assim, no perfil de expressão gênica (FULLER et al. 2003; CARRARO et al. 2004). Deste modo, técnicas de microdissecção permitem obter maior homogeneidade celular, aumentando a acurácia e a sensibilidade dos estudos moleculares, reproduzindo assim, resultados mais relacionados a célula do que ao tecido como um todo.

Em casos de microdissecção manual, dependendo do tipo de tecido, é possível obter mais de 80% de homogeneidade. Este processo é guiado por lâminas coradas por hematoxilina-eosina realizado com o auxílio de bisturi e observação em microscópio óptico comum. A retirada das áreas de tecido que se deseja excluir do fragmento congelado deve ser feita em condições livres de RNases e sem que haja descongelamento dos tecidos. Como exemplo, no caso de tecido neoplásico faz-se a retirada de tecido normal, fibrose e áreas de necrose.

Atualmente, com o desenvolvimento de LCM - *Laser Capture Microdissection* (EMMERT-BUCK et al. 1996; BONNER et al. 1997; SIMONE et al. 1998) e outras tecnologias similares de excisão de tecido tais como, PALM – *Laser-MicroBeam System* (SCHUTZE e LAHR 1998) e *Leica Laser Microdissection* (KOLBLE 2000) tem sido possível a obtenção de uma população de células mais pura a partir de regiões microscópicas específicas de um dado tecido. Isto tem permitido estudos mais acurados de análise de expressão gênica, possibilitando uma melhor correlação entre o perfil de expressão gênica e o tipo de célula em questão.

## 1.9 ANÁLISE DO PERFIL DE EXPRESSÃO GÊNICA EM CÂNCER DE MAMA

A análise do perfil de expressão gênica tem sido uma ferramenta poderosa na classificação prognóstica do câncer de mama. Além disso, tem-se demonstrado útil na subclassificação de tumores que apresentam a mesma classificação morfológica e características diferentes quanto a resposta à quimioterapia, à progressão tumoral e outros, permitindo estabelecer correlações com finalidades preditivas de comportamento tumoral (HIDALGO e SALCEDO 2001; FOLGUEIRA et al. 2005).

Alguns estudos têm aplicado a metodologia de *microarrays* para identificar uma 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), buscando um grupo de genes com valor preditivo de potencial metastático e prognóstico em câncer de mama (VAN'T VEER et al. 2002; SOTIRIOU et al. 2003; SETH et al. 2003) e compreensão da biologia molecular do desenvolvimento de diferentes tipos de DCIS e possível progressão para o IDC (HANNEMANN et al. 2006; SCHUETZ et al. 2006).

Apesar de grandes esforços em decifrar a base molecular da tumorigênese e progressão do câncer de mama (PEROU et al. 2000; ADEYINKA et al. 2002; VAN'T VEER et al. 2002; SETH et al. 2003; MA et al. 2003; PORTER et al. 2003; SOTIRIOU et al. 2003; NISHIDATE et al. 2004; NAGARAJA et al. 2005; SCHUETZ et al. 2006; NAMBA et al. 2006; DALGIN

et al. 2007; ALEXE et al. 2007) este assunto permanece não elucidado completamente.

VAN'T VEER et al. (2002) utilizaram essa metodologia para examinar o perfil de expressão gênica de 98 amostras de tumor primário de mama com *status* linfonodal negativo utilizando uma plataforma contendo 25.000 genes humanos, visando prever quais os casos de câncer mamário apresentariam potencial metastático. Dos 5.000 genes com expressão diferencial, 70 deles foram selecionados por terem capacidade de prever recorrência de doença. Usando esta classificação, os autores puderam prever corretamente o resultado de 65 dos 78 pacientes. SOTIRIOU et al. (2003) visando correlacionar fatores prognósticos com perfil de expressão gerado por *microarray*, estudaram um grupo de 99 pacientes, mostrando que o *status ER* (receptor de estrógeno) está fortemente associado a expressão gênica diferencial entre os diferentes subtipos de tumor, dado também observado por VAN'T VEER et al. (2002).

Neste sentido, tem sido realizados estudos visando classificar o carcinoma ductal de mama em subtipos baseado na variação do padrão de expressão gênica a partir de cDNA *microarrays* e análises de clusterização hierárquica (PEROU et al. 2000; SORLIE et al. 2001; SOTIRIOU et al. 2003; ALEXE et al. 2007; DALGIN et al. 2007). LIVASY et al. (2007) classificaram os tumores em 5 principais grupos baseado em características basais (predominantemente ER negativo, HER2 negativo, EGFR positivo e expressão de citoqueratina 5 e 6), super-expressão de HER2 (ER negativo e HER2 positivo), luminal A (ER positivo e HER2 negativo), luminal B (ER e

HER2 positivo) e não classificáveis. Esta subclassificação foi correlacionada com características tumorais e de prognóstico, revelando uma pior evolução para os tumores `basalóides` (RAKHA et al. 2006) e melhor prognóstico para o subgrupo luminal A.

Alguns poucos estudos utilizando amostras de carcinoma ductal *in situ* (DCIS) tem sido realizados buscando avaliar este tipo histológico que pode se apresentar associado ou não ao componente invasivo. SETH et al. (2003) utilizaram uma plataforma de cDNA de 9.200 genes para comparar um caso de DCIS com grau baixo à intermediário e um caso de DCIS que coexiste com o componente invasivo (DCIS/IDC) de alto grau. Trezentos e três cDNAs apresentaram sinal de expressão diferencial maior que 2 vezes em relação ao DCIS/IDC e DCIS, sendo que os genes induzidos por interferon apresentaram maior expressão em amostras de DCIS e genes que codificam uma imunoglobulina de cadeia pesada e calgranulina B, foram os mais abundantemente expressos no tumor DCIS/IDC.

HANNEMANN et al. (2006) analisaram 40 casos de DCIS e 40 de IDC utilizando classificação supervisionada, e identificaram um classificador de 35 genes que diferem entre DCIS e IDC e outro com 43 genes capaz de separar as amostras DCIS em bem ou pouco diferenciadas.

Estes trabalhos utilizaram microdissecção manual, sob a argumentação que o estroma localizado ao redor do tecido tumoral desempenha um papel fundamental no processo de migração, invasão e formação de metástase (BERARDO et al. 1998). No entanto, analisar separadamente cada tipo celular do tumor, seguramente permite

informações mais precisas sobre o padrão de expressão de um tipo celular em especial e seu papel em diferentes processos.

Recentemente, alguns autores têm utilizado microdisseção a *laser* (*Laser Capture Microdissection-LCM*) e contribuído para definição do perfil de expressão gênica de uma população celular mais pura. A combinação das duas metodologias (*microarrays* e LCM) tem sido utilizada na descoberta de novos marcadores de prognóstico e também na predição da evolução da doença (HUANG et al. 2003; MA et al. 2003; SOTIRIO et al. 2003; ZHU et al. 2003; NISHIDATE et al. 2004; SCHUETZ et al. 2006).

Como por exemplo, ADEYINKA et al. (2002) com o propósito de identificar alterações moleculares relacionadas à progressão do DCIS, compararam a expressão gênica de um grupo de 6 casos de DCIS com necrose e 4 casos sem necrose após microdisseção a *laser* (LCM) em uma plataforma de cDNA *microarray*. De 1.181 cDNAs, um grupo de 69 foram identificados como diferencialmente expressos. Entre estes, foram relatados o RNAm correspondente a proteína celular migratória angio-associada e de proteína quinase serina-treonina relacionada a regulação do ciclo celular.

ZHU et al. (2003) combinando microdisseção a *laser* PALM e cDNA *microarray*, realizaram uma análise comparativa de células da borda e da região central de DCIS (estadio 0) de 5 pacientes, que mostrou maior nível de expressão dos genes *Vimentina*, *Ciclina H e B1*, *RhoC (ras homolog gene family, member C)* e *TGF- $\beta$  (transforming growth factor, beta)* e menor nível dos genes que codificam proteínas de suscetibilidade tumoral (*TSG-101*), Beta-actina citoplasmática (*ACTB*) e fator 7 *splicing* rico em arginina-serina.

Já nas 5 amostras IDC a análise comparativa do padrão de expressão gênica das células epiteliais da borda em relação a região central mostrou aumento dos níveis de expressão dos genes que codificam o precursor de antígeno de histocompatibilidade classe II, *PIRIN* (*iron-binding nuclear protein*), *RhoC* e *EF-1 alpha* (*Elongation factor 1-alpha*) e com baixo nível de expressão dos genes que codificam *TSG-101*, *Ciclina D1*, *FCRN* (*Fc fragment of IgG, receptor, transporter, alpha*) e *CDKN1A* -inibidor 1 de quinase dependente de ciclina.

MA et al. (2003), combinando o uso de LCM e cDNA *microarray* analisaram o perfil de expressão gênica dos estágios pré-maligno (ADH), pré-invasivo (DCIS) e invasivo (IDC) de mama. Nesse estudo os autores utilizaram amostras de 36 pacientes, sendo 31 com dois ou mais estágios patológicos da progressão do câncer de mama, 5 com doença pré-invasiva somente e 3 amostras provindas de mamoplastia. Os resultados revelaram alta similaridade em nível transcricional entre as diferentes classes, uma vez que não foi possível encontrar um grupo de genes cujo padrão de expressão pudesse separar os diferentes grupos, sugerindo que alterações na expressão gênica que possa conferir um potencial de crescimento invasivo encontra-se presente em estágios iniciais. Também SCHUETZ et al. (2006), combinando o uso de LCM e oligo *microarray* analisaram amostras pareadas visando compreender a biologia molecular da transição do DCIS para o IDC e identificar genes cujo produto possa servir como marcador prognóstico e/ou alvo molecular para tratamento, identificando 445 genes com expressão diferencial entre os 2 tipos histológicos da mesma amostra.

A utilização de tecnologias de microdissecção combinadas a plataformas de expressão gênica reserva a grande promessa de gerar uma melhor compreensão, não apenas na pesquisa do câncer em geral, mas em particular, do processo de invasão e identificação de marcadores moleculares de predição da progressão da doença, melhorando o diagnóstico, prognóstico e promovendo uma terapia mais apropriada.

Nossa proposta neste estudo é avaliar o carcinoma ductal de mama sob dois aspectos combinando a tecnologia de microdissecção à *laser* e cDNA *microarrays*. Um dos aspectos se refere a identificação de genes envolvidos na transição do carcinoma ductal *in situ* para o invasivo, representando candidatos a marcadores de prognóstico e/ou terapêutico; e o outro, a caracterização molecular do carcinoma ductal *in situ* sem a presença do componente invasivo (DCIS puro) ou coexistindo com IDC (DCIS/IDC), identificando genes candidatos a marcadores moleculares de progressão do carcinoma ductal.

Para tanto, foram utilizadas duas plataformas diferentes. Uma delas contendo 4.608 genes humanos e não apresentando nenhuma classe funcional de genes preferencialmente, denominado neste trabalho de plataforma 4.8K (BRENTANI et al. 2005). A outra plataforma (LV-01) utilizada apresenta 1.854 genes, contendo genes pertencentes a vias de sinalização celular envolvidas no processo de transformação, invasão e metástase das células, tais como WNT, PI3K e processo EMT.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

- ◆ Identificar genes envolvidos no processo de transição do carcinoma ductal *in situ* (DCIS) para o ductal invasivo (IDC) e na progressão no carcinoma ductal *in situ* de mama.

### 2.2 OBJETIVOS ESPECÍFICOS

#### 2.2.1 Plataforma 4.8K

- ◆ Analisar o padrão de expressão gênica entre amostras pareadas de morfologia diferentes que coexistem na mesma lesão (DCIS e IDC), assim como entre diferentes amostras que mimetizam a progressão do carcinoma ductal de mama (tecido não neoplásico, DCIS puro, *in situ* associado ao IDC e IDC) pela técnica de *microarray* utilizando uma plataforma customizada contendo 4.608 genes humanos;
- ◆ Identificar genes diferencialmente expressos presentes entre amostras pareadas (DCIS e IDC) da mesma paciente; assim como, entre DCIS puro e *in situ* associado ao IDC capazes de diferenciar 2 lesões de mesma morfologia;
- ◆ Validar alguns dos genes diferencialmente expressos identificados na metodologia de *microarray* por RT-PCR quantitativo;

- ◆ Analisar a expressão da proteína para alguns genes alterados que apresentarem anticorpos comerciais por imunohistoquímica utilizando *tissue microarray*.

### **2.2.2 Plataforma Contendo Genes das Vias de Sinalização WNT, PI3K e Processo EMT (LV-01)**

- ◆ Avaliar o padrão de expressão dos genes pertencentes as vias de sinalização WNT, PI3K e processo EMT em lesões *in situ* e invasivo de mama que coexistem na mesma amostra (amostras pareadas), assim como em diferentes amostras que mimetizam a progressão do câncer de mama utilizando uma plataforma customizada contendo 1.854 genes, sendo 390 pertencentes as vias citadas;
- ◆ Identificar genes diferencialmente expressos, classificadores e genes linearmente correlacionados entre 2 lesões de mesma morfologia (DCIS puro e *in situ* associado ao IDC) utilizando uma plataforma customizada contendo genes pertencentes as vias de sinalização WNT, PI3K e processo EMT.

### 3 METODOLOGIA

#### 3.1 AMOSTRAS

Foram utilizadas neste estudo amostras de tecido não neoplásico, carcinoma ductal *in situ* e invasivo de mama de pacientes operadas no Hospital do Câncer e arquivadas no Banco de Tumores do Centro de Tratamento Médico e Pesquisa do Hospital do Câncer A. C. Camargo sob coordenação do Dr. Fernando Augusto Soares. Este estudo foi aprovado pelo Comitê de Ética do Centro de Tratamento e Pesquisa do Hospital do Câncer A. C. Camargo, no processo de número 587/04 e apresenta um termo de consentimento informado devidamente assinado por todas as participantes.

Inicialmente pretendia-se analisar todos os estadios separadamente. No entanto, amostras do estadio II apresentaram maior representatividade no banco de tumores; e além do mais, há dificuldade em se obter amostras no estadio III sem tratamento quimioterápico prévio. Além disso, foi observado pouca alteração em nível de expressão gênica entre os diferentes estadios (MA et al. 2003; FOLGUEIRA et al. 2006). Desta forma, devido a dificuldade em representar os estadios I e III, a análise foi realizada agrupando todos os estadios em uma única classe, aumentando o número de casos e o poder estatístico das análises.

Informações clínicas das pacientes foram coletadas a partir do prontuário. A idade das pacientes variou entre 25 e 75 anos, apresentando em média 49 anos, sendo que nenhuma delas recebeu quimio ou radioterapia como primeiro tratamento. Em relação ao envolvimento de linfonodos, 12 casos apresentaram envolvimento de linfonodo e 22 casos eram linfonodos-negativo. Quanto aos marcadores moleculares avaliados por imunohistoquímica, dados referentes ao receptor de estrógeno (ER), HER2/neu, receptor de progesterona (PR) e p53 também foram anotados. O quadro completo com informações das amostras utilizadas pode ser visualizado no Anexo 1.

Um total de 47 casos de mama foi selecionado para o estudo, sendo que destes, 40 foram analisados incluindo: 5 amostras de tecido não neoplásico (borda de fibroadenoma), 16 casos de amostras pareadas (16 DCIS e 16 IDC), 5 DCIS puro, 9 *in situ* associado ao IDC (DCIS/IDC) e 5 IDC. Dos 7 casos excluídos do estudo, 2 não atingiram o critério de qualidade adotado para o RNA total (item 3.4) e 5 apresentaram quantidade insuficiente de RNA amplificado para realização dos experimentos de *microarray*. O quadro com as informações da qualidade do RNA total e quantidade do RNA amplificado pode ser visualizado no Anexo 2.

### **3.2 MICRODISSECÇÃO A LASER**

O aparelho utilizado por nossa equipe, o *Pixcell Laser Capture Microdissection System (Arcturus Systems for Microgenomics)*, pertence ao

Departamento de Anatomia Patológica do Hospital do Câncer A.C. Camargo e atualmente está sob responsabilidade do Dr. Fernando Soares e Dra. Renata Coudry.

Este sistema consiste basicamente de um microscópio invertido, ao qual é acoplada uma fonte de raio *laser* infravermelho de baixa potência e uma câmera ligada a um monitor e um computador. A lâmina com o corte histológico é fixada ao *chariot* por um mecanismo de sucção a vácuo.

O feixe de raio *laser* incide sobre um *cap* revestido por um filme termoplástico. O *cap* fica posicionado entre o feixe de *laser* e o corte histológico. Com o disparo do feixe de *laser* o filme termoplástico expande e adere à célula escolhida. Quando o pulso cessa, o filme retrai levando consigo a célula aderida à superfície do *cap*. Desta forma repete-se a operação tantas vezes quantas forem necessárias. A quantidade de células necessárias varia de acordo com o tipo de tecido e com o estudo que será realizado. Ao final da operação retira-se o *cap* com as células e segue-se a extração de RNA.

Todas as amostras do projeto foram microdissecadas a *laser*, sendo este procedimento realizado pelas Dra. Cynthia Osório e Dra. Renata Coudry. Para a obtenção de quantidade suficiente de células para os experimentos de microarray foram necessárias a captura 4.000 células para cada caso. Uma representação de microdissecção a *laser* de carcinoma ductal *in situ* pode ser visualizada na Figura 5.



**Legenda** - **Antes**- representa o tecido antes da captura das células, **Depois**- após a captura e **CAP**- as células de interesse isoladas e aderidas ao cap.

**Figura 5** - Representação de microdissecção a *laser* de carcinoma ductal *in situ*.

### 3.3 EXTRAÇÃO DE RNA TOTAL

O RNA total das amostras microdissecadas a *laser* foi extraído utilizando o kit PicoPure™ RNA Isolation Kit (*Catalog # KIT0202 – Arcturus*) e tratado com DNase seguindo as recomendações do fabricante.

### 3.4 AVALIAÇÃO DO RNA TOTAL EXTRAÍDO

Todas as amostras extraídas a partir de células microdissecadas a *laser* no período de 2004 a 2006 tiveram a qualidade do RNA total avaliada por uma reação de RT-PCR do gene *NOTCH2*, cujo RNAm contém 11.433 pb e do gene *GAPDH* com 1.283 pb. Esta reação permite a avaliação não apenas do RNA extraído como da síntese da 1ª fita. As seqüências dos genes foram obtidas no banco de dados Ref Seq (NM\_024408) e (NM\_002046.3) respectivamente. Dois pares de *primers* para o gene

*NOTCH 2* e um par para o *GAPDH* foram desenhados manualmente em éxons diferentes e as sequências obtidas foram checadas para avaliação de estruturas secundárias com o auxílio do programa *OLIGOTECH* versão 1.00 (Copyright© 1995). Só foram submetidas a amplificação as amostras que apresentaram resultado positivo na PCR para pelo menos um dos *primers* do gene *NOTCH2*.

***NOTCH 2*** [*Notch homolog 2 (Drosophila)*]

*NOTCH2.1 Forward* – CCAACCAGTTCTCCTGC – 65,9 °C

*NOTCH2.1 Reverse* – GTGGTTAGGGCAGTCATC – 66,8 °C

Tamanho do fragmento: 317 pb

Localização: 742 da região 5'  
10.691 da região 3'

*NOTCH 2.2 Forward* – GGTGAACAAGAACAGGAG – 64,5 °C

*NOTCH2-2 Reverse* – GATGACAACAGCAACAGC– 64,5 °C

Tamanho do fragmento – 231 pb

Localização: 5.107 da região 5'  
6.326 pb da região 3'

***GAPDH*** (*Glyceraldehyde-3-phosphate dehydrogenase*)

*GAPDH Forward* – CTGCACCACCAACTGCTTA - 63 °C

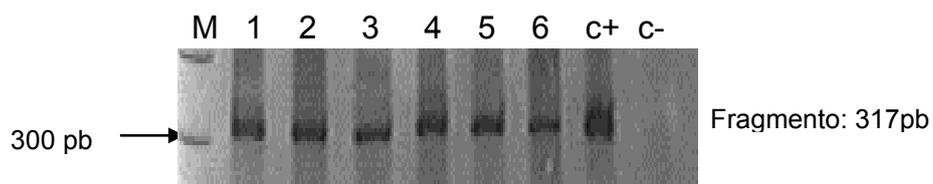
*GAPDH Reverse* – GACCTGACCTGCCGTCTAG - 63 °C

Tamanho do fragmento - 296 pb

Localização: 527 da região 5'  
460 da região 3'

Após a síntese da primeira fita de cDNA em volume total de 20,0 µL no processo de amplificação foram isolados 5 µL da reação, os quais foram diluídos em mais 15 µL de água tratada com DEPC. Dessa diluição, 2µL foram utilizados como molde na PCR. O aparecimento do amplificado garante a eficiência da reação de transcrição e a integridade do RNAm. As seguintes condições de amplificação utilizadas para os pares de *primers* foram: 1 mM MgCl<sub>2</sub>, 0,2µM dNTPs (*Invitrogen*), 0,4µM de cada *primer*, 1X

tampão de reação 10X, 1U de *Platinum® Taq DNA Polymerase* (Invitrogen) em uma reação de volume final de 25 $\mu$ L. O programa da PCR consistiu em dois minutos iniciais de desnaturação a temperatura de 94 °C, 40 ciclos repetidos com 1 minuto a 94° C, 1 minuto com a temperatura de 58°C e 1 minuto a 72°C, por fim, um ciclo único de 7 minutos a 72°C. O produto foi visualizado por meio de eletroforese em gel de poliacrilamida 8%, com posterior coloração por prata (SANGUINETTI et al. 1994). Um exemplo da qualidade do RNA extraído de amostras microdissecadas a *laser* avaliado por RT-PCR para o gene *NOTCH2* pode ser visualizado na Figura 6.

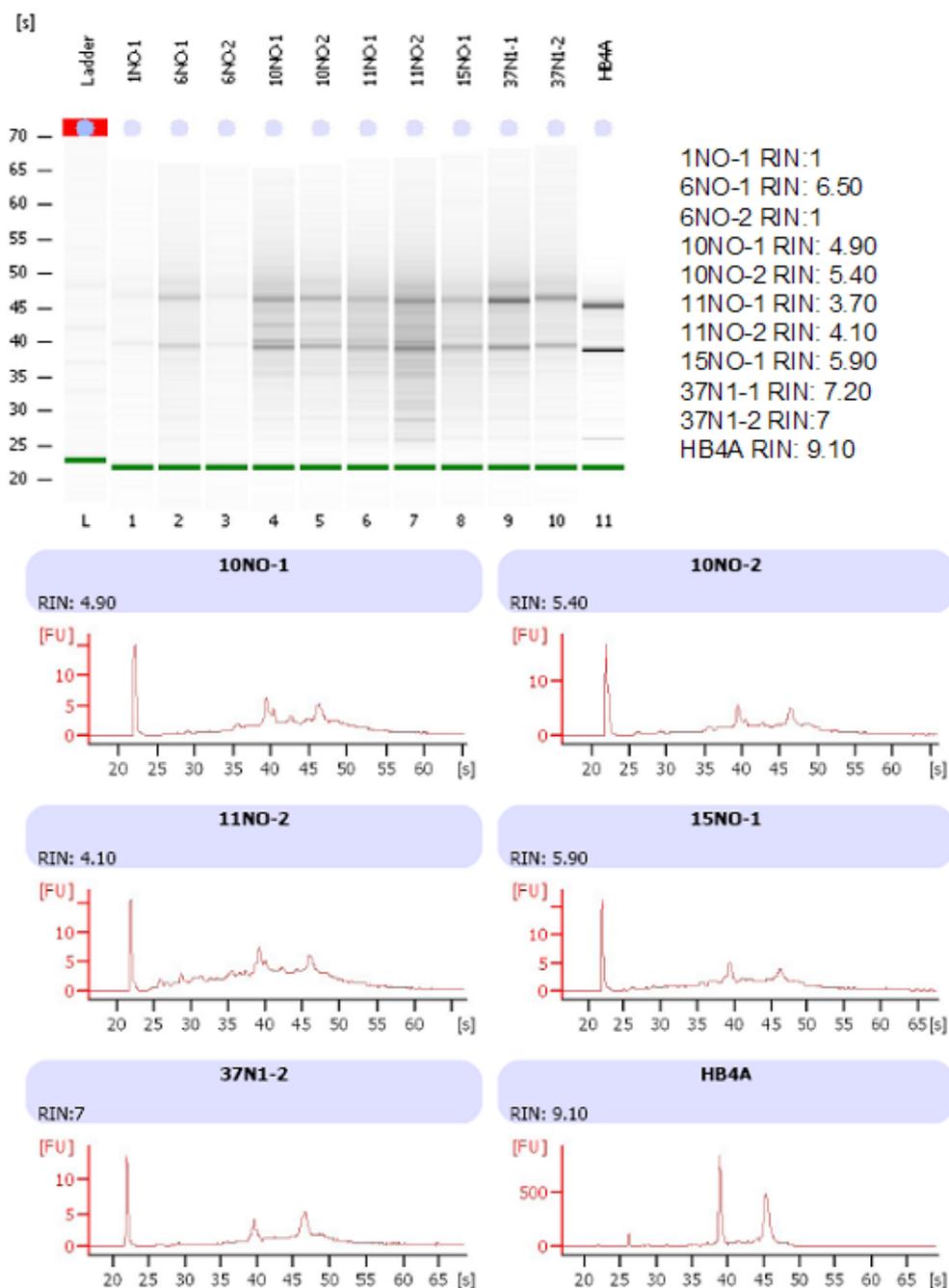


**Legenda** - RT-PCR para o gene *NOTCH 2* utilizando o *primer* 2.1. Fragmento amplificado de 317 pb. M- marcador de peso molecular de 100pb, 1 a 6- amostras tumorais microdissecadas a *laser*, c+: controle positivo, c-: controle negativo.

**Figura 6** - Representação de cDNA amplificado para o gene *NOTCH 2* em gel de poliacrilamida 8%.

A partir de 2007, a qualidade do RNA total das amostras microdissecadas a *laser* foi analisada usando o sistema *Agilent 2100 Bioanalyser* que usa detecção de fluorescência, monitorando-a entre 670nm e 700nm (SCHROEDER et al. 2006). Foi utilizado o RNA 6000 *Pico LabChip Kit*, no qual a quantidade de amostras a serem avaliadas deve ser extremamente pequena (de 200 pg a 5 ng). O *software* automaticamente calcula a proporção das áreas picos das bandas ribossomais 18S e 28S

(relação  $\geq 1$ ) para amostras eucarióticas. É construído um eletroferograma e calculado um valor referente à integridade do RNA (*RIN-RNA Integrity Number*). O RIN varia de 0 a 10 em ordem crescente de integridade (SCHROEDER et al. 2006). As amostras com índice inferior ao estabelecido como critério de qualidade (RIN>4) em pelo menos uma das duplicatas (2 *caps* por amostra) ou relação entre as bandas correspondentes aos RNA ribossomais 28S e 18S menor que 1 foram descartadas. O protocolo pode ser visualizado no Anexo 3. Um exemplo da qualidade do RNA total extraído de amostras microdissecadas a *laser* avaliado no sistema *Agilent 2100 Bioanalyser* pode ser visualizado na Figura 7.



**Legenda - A:** Eletroforese de RNA total mostrando as bandas 18 e 28S **B:** Número de integridade do RNA **C:** Eletroferograma de algumas amostras mostrando os picos das bandas ribossomais 18S e 28S repectivamente.

**Figura 7** - Representação da qualidade de RNA total extraído a partir de amostras microdissecadas a *laser* avaliado no aparelho *Agilent 2100 Bioanalyser*.

### 3.5 AMPLIFICAÇÃO DO RNA MENSAGEIRO (RNAm)

Todas as amostras foram submetidas a 2 ciclos de amplificação do RNAm para se obter quantidade de amostra suficiente para os experimentos de *microarray*. O protocolo utilizado é baseado em GOMES et al. (2003) com modificações. Resumidamente, o cDNA é sintetizado a partir do RNA total, utilizando *primers* que ancoram na cauda poli-A dos RNA mensageiros e possuem sítio para RNA polimerase II. Para uma produção exponencial de RNA, este cDNA é submetido à transcrição *in vitro* em dois ciclos de amplificação. Para as amostras microdissecadas a *laser* foram necessárias algumas alterações estabelecidas em nosso laboratório, que consistem em aumento do tempo de anelamento do *primer* Oligo-dT- T7 *primer* de 5 para 15 minutos, utilização de um *primer* oligo-dT (24)-T7 *primer* (5'GGC CAG TGA ATT GTA ATA CGA CTC ACT ATA GGG AGG CGG VTT TTT TTT TTT TTT TTT TTT TTT TTT T3')[64-pb], (sendo que a cauda poli T mais longa aumenta a eficiência de anelamento à cauda poli-A), ajustes nas quantidades utilizadas de *primer* oligo-dT (24)-T7 na síntese de primeira fita e adição de carreadores nas etapas de precipitação de síntese de cDNA, como o glicogênio e acrilamida (tanto na recuperação do RNA quanto do cDNA respectivamente). Essas alterações melhoraram a qualidade e quantidade do RNA amplificado recuperado (SARAIVA et al. 2006). O protocolo detalhado está descrito no Anexo 4.

### **3.6 AVALIAÇÃO DA QUALIDADE E QUANTIDADE DO RNA AMPLIFICADO**

A avaliação das amostras de RNA amplificado foi realizada em gel de agarose 1% em uma corrida com duração de 50 minutos na voltagem de 60 mV. 1µg de RNA amplificado foi adicionada a 5µl de tampão TAE (*Tris-acetate-EDTA*) com 7M uréia, 3X *Loading Buffer* azul de bromofenol e 2µl de brometo de etídio 0,5 µg/µl antes de ser aplicada no gel. As amostras devem apresentar um perfil de amplificação onde a maior concentração de RNA amplificado (RNAa) esteja entre 300-700 pb (critério de qualidade previamente estabelecido por nosso grupo, sendo aquelas que apresentam as lâminas hibridizadas de melhor qualidade). A concentração do RNA amplificado foi verificada em espectrofotômetro utilizando o comprimento de onda equivalente a 260 nm (1OD a 260 nm equivale a 40ng/µl de RNA) e a pureza pela relação entre as leituras OD 260 e 280.

### **3.7 MICROARRAYS DE cDNA**

#### **3.7.1 Plataforma 4.8k**

Esta lâmina contém 4.608 clones de ORESTES (*open reading frame expressed sequence tags*) (Dias-Neto et al. 2000) correspondentes a genes humanos que passaram por um *pipeline* desenvolvido pelo grupo da Dra. Helena Brentani (Laboratório de Biologia Computacional - Hospital do Câncer A C Camargo). Esses clones de cDNA foram isolados a partir de

tecidos tumorais, originados no Projeto Genoma Humano do Câncer (Instituto Ludwig/FAPESP).

A lâmina é dividida em 48 sub-*arrays*, cada um contendo 100 *spots*. Cada *spot* corresponde a um ponto na lâmina de *microarray* onde estão imobilizadas as seqüências de ORESTES, além de 192 controles positivos e negativos.

As seqüências imobilizadas na lâmina seguem os seguintes critérios:

i) clone com tamanho maior que 300 pb; ii) clone que não apresenta identidade maior que 85% em 100 pb com nenhum outro gene humano, que não ele mesmo; e iii) clone com seqüências localizadas próximas a região 3' (BRENTANI et al. 2005).

Dos 4.608 genes representados, 2.134 genes têm função conhecida divididos em 802 classes funcionais segundo o *Gene Ontology Consortium Annotation*. Dos genes com função conhecida, 287 genes distribuídos na classe regulação da transcrição, 202 genes na classe transdução de sinal, 177 genes na classe fosforilação de aminoácidos e proteínas e 100 genes na classe transporte, sendo estas classes as mais representadas. A classificação usada é redundante, ou seja, o mesmo gene pode ter sido colocado em mais de uma classe funcional, conforme sua caracterização.

### **3.7.2 Plataforma Das Vias de Sinalização WNT, PI3K e Processo EMT (LV-01)**

Esta plataforma foi construída a partir de um projeto de nosso grupo (Projeto FAPESP: Análise de expressão gênica de vários estágios do

desenvolvimento do rim e fígado e suas implicações em tumores embrionários. Processo número: 06/00054-0) contendo fragmentos de cDNA correspondentes a genes humanos envolvidos no processo de transição epitélio-mesênquima-EMT (84), via de sinalização celular WNT (197), via de sinalização celular PI3K (109), e mais 8 genes de expressão constitutiva. Para o projeto acima mencionado, as seqüências humanas deveriam compartilhar similaridade com as seqüências ortólogas de camundongo para permitir hibridação interespecífica. Os critérios para representação do gene foram: i) alinhamento de *e-value*  $-5$  e alinhamento mínimo 50% entre as seqüências ortólogas (humano e camundongo); ii) mínimo de 80% de identidade entre as seqüências humanas e de camundongos no trecho representado (80 a 98%). Outros 397 genes das vias do ácido retinóico, diferenciação neuronal e outros genes aleatórios também foram representados nessa lâmina além de 645 ORESTES de outros projetos e 414 fragmentos originados pela técnica RASH, somando 1.854 seqüências de cDNA. Para representação do gene na plataforma, foi inicialmente realizada uma busca no banco de ORESTES do Projeto Genoma Humano do Câncer (Instituto Ludwig/FAPESP). Para aqueles genes cujas seqüências de ORESTES não preencheram os critérios acima mencionados, iniciadores foram desenhados para amplificar um fragmento do gene a ser imobilizado na lâmina. O desenho dos iniciadores foi realizado pelo programa *Primer3* com posterior checagem manual pelo programa (*Oligotec*) para verificar a presença de estruturas secundárias seguindo alguns critérios tais como: as temperaturas dos pares de oligonucleotídeos não devem apresentar mais de

3°C de diferença, as temperaturas das estruturas secundárias devem ser menores 30°C, os iniciadores sempre que possível devem apresentar C ou G em suas extremidades e a temperatura de alinhamento deve ser maior que 60°C. A lâmina foi dividida em 48 sub-arrays dispostos em quatro colunas e doze fileiras. Cada sub-arrays contém 49 *spots* e entre estes, oito controles negativos e um positivo. Os amplicons foram imobilizados nas lâminas de vidro pelo grupo do Dr. Luiz Fernando Lima Reis.

### 3.7.3 Marcação e Hibridização

O cDNA marcado foi gerado pela reação de *transcriptase reversa* utilizando 4 µg de RNA amplificado (RNAa), *random primer* (Invitrogen Life Technologies, Carlsbad, CA), dCTP Cy3 e Cy5 marcado (Amersham, Biosciences) e *SuperScript II* (Invitrogen-Life Technology). Juntamente à reação foi adicionado um RNAm controle (Q-gene).

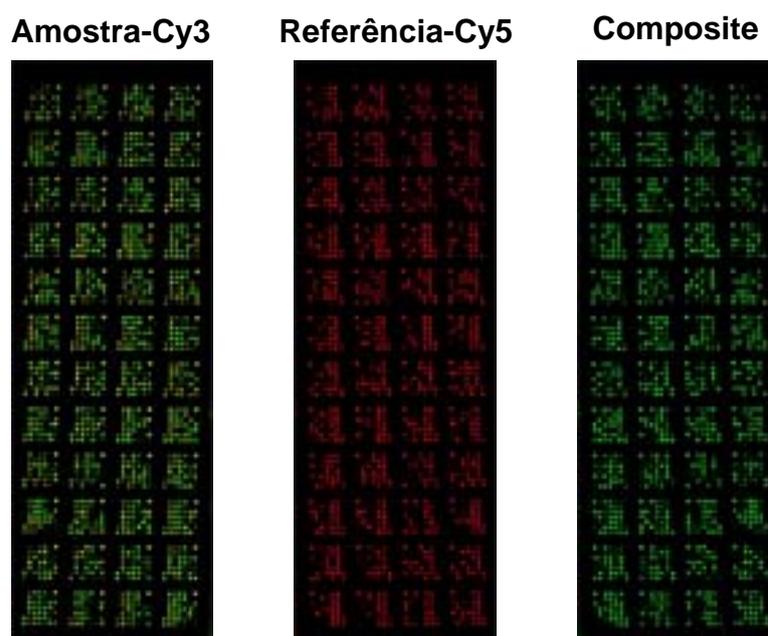
Foram utilizados 3,5 µg de RNA de linhagem de células normais epiteliais mamárias humanas da linhagem HB4a (gentilmente doadas pelos Drs. Mike O'Hare e Alan Mackay, *Ludwig Institute for Cancer Research-University College London, London, United Kingdom*- O'HARE et al. 1991) como referência. Como a hibridização é competitiva, os valores de fluorescência obtidos revelam níveis relativos de expressão de cada transcrito na amostra-teste comparada com a amostra-referência (NOVORADOVSKAYA et al. 2004). Cada amostra foi hibridizada em duplicata utilizando o sistema inverso de incorporação dos corantes Cy3 e

Cy5 (*Dye swap*). As lâminas foram previamente hibridizadas por 16 à 20 horas.

O processo de hibridização foi realizado a 42 °C por no mínimo 16 hs e no máximo 22hs na estação *GeneTac 2000 (Genomic Solutions)*. Após hibridização, as lâminas passaram por um processo de lavagem. Os detalhes de marcação, hibridização e lavagem das lâminas estão apresentados no Anexo 5.

#### **3.7.4 Captura das Imagens e Quantificação de Sinal**

Os sinais gerados pela hibridização competitiva das amostras foram capturados utilizando um *scanner a laser* confocal com canal duplo, o *ScanArray™ Express (Perkin Elmer Life Sciences, Inc., EUA)*, usando PMT 60% para ambos os canais (*Photomultiplier voltage*) (representação na Figura 8). O sistema tem dois *lasers* com comprimento de ondas diferentes, que estimulam especificamente o Cy3 ou o Cy5 (Tabela 1). Quando as moléculas são excitadas, emitem um sinal proporcional ao número de moléculas hibridizadas, que é captado pelo *scanner*. A quantificação desse sinal ocorre devido a conversão do comprimento de onda emitido pelos fluoróforos excitados em energia elétrica, produzindo um sinal mensurável. Os dados foram recuperados pelo *ScanArray (Perkin Elmer Life Sciences, Inc., EUA)* usando o método histograma.



**Figura 8** - Representação da lâmina contendo genes pertencentes às vias de sinalização WNT, PI3K e processo EMT hibridizada com amostra de carcinoma ductal *in situ* marcada com Cy3, referência com Cy5 e imagem sobreposta respectivamente.

**Tabela 1** - Características específicas de cada fluoróforo utilizado.

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

Esse sistema a *laser* produz dados com pequena variância, pois a leitura é sensível, com resolução de 5  $\mu\text{m}$ , e o sistema de filtros permite aumento de intensidade do sinal alvo enquanto um aparato bloqueia a passagem da luz refletida e proporciona a eliminação do ruído com maior eficiência. O programa realiza o cálculo da intensidade de cada ponto de hibridização para cada uma das amostras.

### 3.7.5 Pré-Análise das Lâminas Híbridizadas

As análises de qualidade dos sinais obtidos em ambas as plataformas foram realizadas sob a coordenação da Dra Helena Brentani (Laboratório de Biologia Computacional do Hospital AC Camargo).

Após a captura de sinal pelo *scanner*, os resultados foram submetidos a algumas análises de qualidade, que são denominadas de pré-análise e estão apresentadas a seguir.

#### A Dendograma

Refere-se a uma clusterização hierárquica com distância euclidiana não supervisionada. Essa clusterização leva em consideração os dados de expressão de todos os genes da lâmina.

Essa análise funciona como um controle da parte experimental, permitindo a visualização das similaridades das amostras. Assim, as lâminas controle e *swap* (réplica com inversão de corantes) devem apresentar maior semelhança no dendograma quando comparada a outra amostra do projeto.

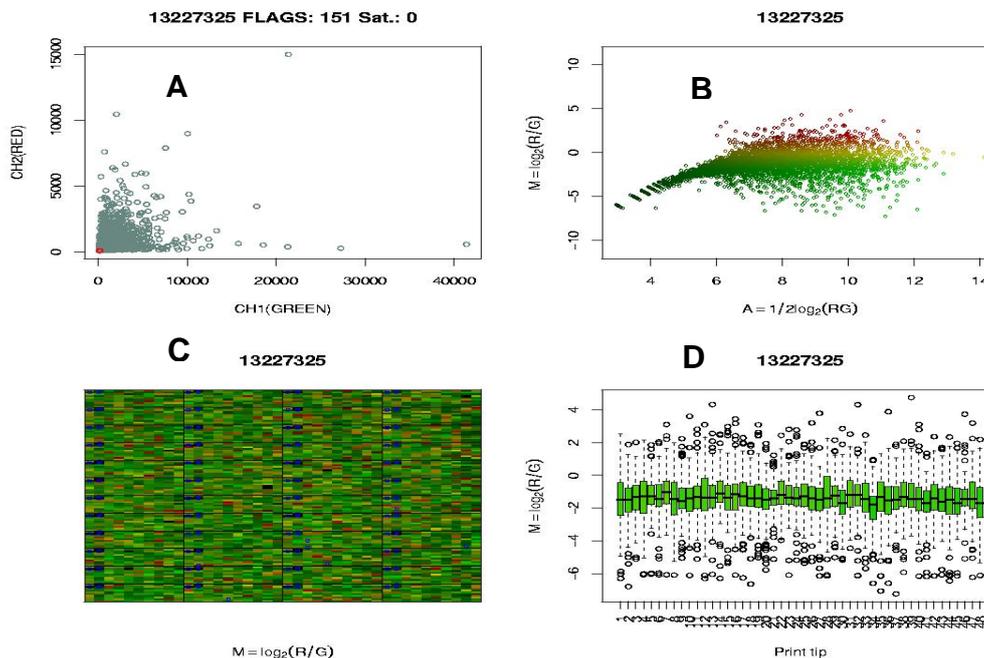
#### B Distribuição dos Valores de Coeficiente de Correlação de *Pearson*

A correlação entre a lâmina controle e a lâmina *swap* foi avaliada por correlação de *Pearson*, levando em consideração a eficiência de incorporação dos fluoróforos (Cy3 e Cy5). A correlação de *Pearson* determina uma medida do grau de correlação linear entre duas variáveis

aleatórias. Para uma lâmina ser considerada de boa qualidade, essa correlação deve ser superior a 80% após a normalização.

### **C Ma, Log X Log e Boxplot**

São medidos os valores de intensidade de fluorescência para cada gene, após a subtração de sinal do *background* usando escala *log* base 2. Um *plot* cartesiano é construído com *M* na coordenada e *A* na abcissa, onde  $M = \log_2 (R/G)$  e  $A = \log_2 (\sqrt{RxG})$ , ou seja, pode-se observar a variação da razão de fluoróforos dependente da intensidade dos pontos. Os *Box plots* consistem de caixas com uma linha central (mediana dos dados) e duas caudas, a superior representando o percentil 75 e a inferior, representando o percentil 25 (ALLISON et al. 2006). Os gráficos mostram as intensidades de Cy3 e Cy5 da mesma lâmina (Figura 9), sem nenhum tipo de normalização.



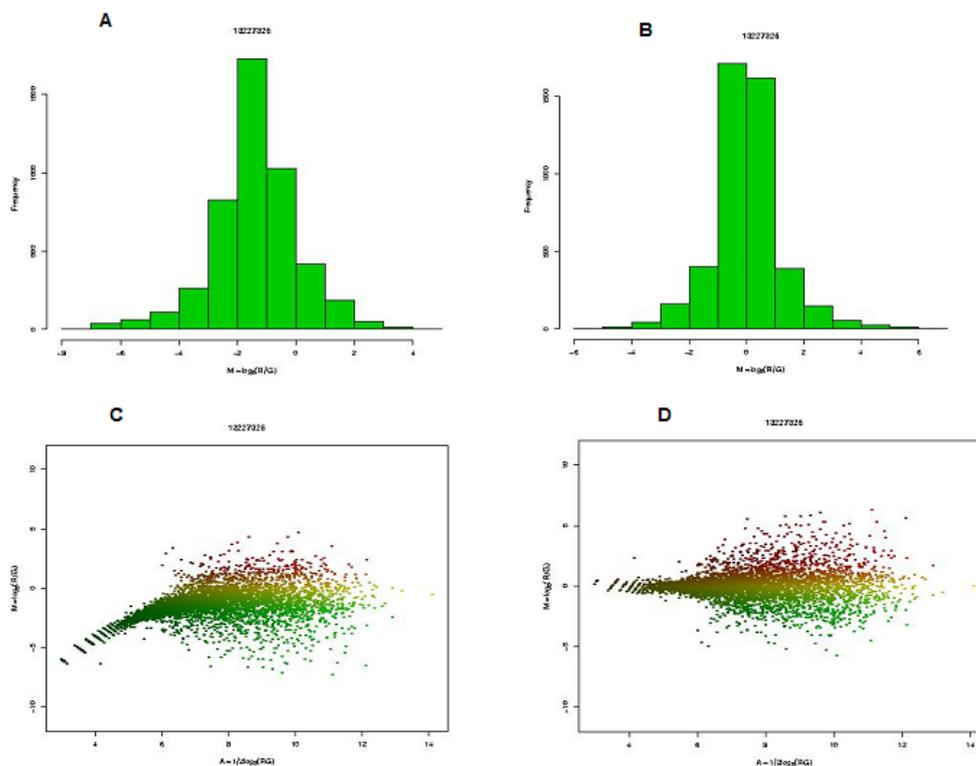
**Legenda - A:** Relação da intensidade bruta da lâmina (Cy3/Cy5). As cores em vermelho ou em verde representam pontos que foram retirados da análise. **B:** *MA plot*- gráfico representando a razão (Cy3/Cy5) no eixo y e média das intensidades dos corantes em escala logarítmica no eixo x. Os pontos em azul/cinza representam os *spots* que apresentaram sinal menor que o sinal do *background*. **C:** Gráfico representativo da lâmina. Pontos em azul representam genes que foram excluídos pelo programa *Scannarray* ou manualmente. **D:** *BoxPlot*-Representa a dispersão de pontos de cada sub-array (1 sub-array contém 100 spots). O bloco verde representa a concentração de 75% dos pontos, segundo a razão de intensidade.

**Figura 9** - Representação da fluorescência de todos os genes de uma lâmina.

## D Normalização das Lâminas

A proposta de normalização é minimizar variações sistemáticas introduzidas experimentalmente nas medidas da expressão gênica, para que as diferenças biológicas quantificadas possam ser distinguidas com menos artefatos. A normalização por *Lowess* (LOWESS – *Locally Weighted Scatter-*

*plot Smoothing*) (Cleveland 1974, citado por QUACKENBUSH 2001 p. 419) foi utilizada neste projeto. Para cálculo do *Lowess* definem-se janelas com intervalos pré-definidos, que variam de 0 a 100%. Nesse projeto, foi utilizado intervalo de 20%, ou seja, a análise de regressão foi realizada com 20% dos pontos ao redor do ponto em questão (Figura 10).



**Legenda** - **A** e **C**: representação dos dados antes de serem normalizados. **B** e **D**: representação dos dados normalizados.

**Figura 10** - Gráficos representativos da normalização por *Lowess*.

### 3.8 ANÁLISES MATEMÁTICA

As análises matemáticas e estatísticas referentes à plataforma 4.8k foram realizadas pelo grupo da Dra Helena Brentani (Laboratório de Biologia Computacional do Hospital AC Camargo).

Primeiramente foram eliminados todos os pontos da lâmina que apresentaram emissão de fluorescência igual ou menor (*outliers*) que o sinal emitido pelo ruído (*background*). Foram eliminados também os pontos que apresentaram saturação de hibridização, sendo esse dado baseado numa emissão de sinal maior ou igual a 63000 (65550 = 16 BIT). Nesse caso, o excesso de material hibridizado gera um sinal muito intenso que supera a capacidade do detector de medir a intensidade proporcional de fluorescência, invalidando a quantificação.

### **3.8.1 Identificação dos genes diferencialmente expressos na plataforma 4.8K**

Os valores relativos normalizados de cada gene foram submetidos ao teste T de *Student* permutado (10.000 permutações) Estes dados foram submetidos a correção de múltiplos testes pelo pFDR (*positive False Discovery Ratio*).

Para identificar os genes diferencialmente expressos entre as amostras pareadas, aquelas em que ambos componentes (DCIS e IDC) coexistiam no mesmo fragmento, foi aplicado o teste T de *Student* pareado. Foram considerados como diferencialmente expressos aqueles genes cuja intensidade apresentou diferença de expressão de pelo menos 1,5 vezes e  $p < 0,01$ .

Para o grupo de amostras que mimetizam a progressão do carcinoma ductal, denominado de amostras progressão (não-neoplásico, DCIS puro, *in situ* associado do IDC e IDC) foi aplicado o teste estatístico ANOVA (pFDR

<0.01) (STOREY 2002) seguido por teste de Tukey (NETER 1996). Este teste foi realizado através de comparação 2 a 2 entre os grupos distintos.

A clusterização hierárquica foi realizada baseada na distância Euclidiana usando os genes diferencialmente expressos. A robustez da clusterização foi avaliada pela técnica de *Bootstrap* usando o programa TMEV (SAEED et al. 2003).

Os genes foram classificados de acordo com processos biológicos através do *Gene Ontology* (GO) e com as vias metabólicas representativas usando o KEGG (*Kyoto Encyclopedia of Genes and Genomes* - [www.kegg.com](http://www.kegg.com)) (KHATRI et al. 2005). A avaliação da representatividade significativa dos processos biológicos e as vias metabólicas no grupo de genes diferencialmente expressos foi usado a distribuição qui-quadrada do *Onto-Tools* desenvolvida pelo Sistema Inteligente e Laboratório de Bioinformática do Departamento de Ciência Computacional da Universidade do Estado de Wayne (<http://vortex.cs.wayne.edu:8080>). Processos biológicos com  $pFDR < 0,05$  e vias com  $pvalor < 0,05$  foram considerados significantes, porém se representados por um único gene, não foram levados em consideração na tentativa de se evitar resultados artefatuais.

### **3.8.2 Identificação dos genes diferencialmente expressos na plataforma das vias de sinalização WNT, PI3K e processo EMT (LV-01)**

As análises matemáticas e estatísticas referentes a plataforma as vias de sinalização WNT, PI3K e processo EMT foram realizadas pelo grupo do

Professor Dr. Eduardo Jordão Neves (Instituto de Matemática e Estatística da Universidade São Paulo – IME/USP).

Na análise das lâminas contendo genes envolvidos nas vias de sinalização WNT, PI3K e processo EMT foi empregada uma linguagem computacional e gráfica R (IHAKA e GENTLEMAN 1996) do *Contributed R Archive Network* (CRAN, <http://cran.r-project.org>) que promove análises estatísticas e gráficas, incluindo modelos lineares e não-lineares, *clusters*, previsões, análises *resampling survival* e *time-series*. Além disso, utilizamos o pacote *maigesPack* (ESTEVES 2007), que integra diversas ferramentas de análise de dados de genômica e de expressão gênica (QUACKENBUSH 2001; DUDOIT et al. 2003; BALL et al. 2003) disponíveis através do projeto Bioconductor em um único pacote de análise de dados (<http://www.bioconductor.org>).

## **A Genes diferencialmente expressos**

Para se determinar a significância da diferença de expressão entre carcinoma ductal *in situ* e invasivo nas amostras pareadas foi utilizado o teste paramétrico T de *Student* pareado com hipótese  $H_0: d = 0$  e  $H_1: d \neq 0$ . O teste foi aplicado para a diferença da média de expressão de cada gene entre os dois tipos biológicos e a partir disto foi obtido um nível de significância *alpha* que indica a probabilidade deste evento ter ocorrido ao acaso. A partir dos níveis de significância foram geradas listas de genes diferencialmente expressos com *alpha* < 0.05.

Para as amostras de progressão optamos pelo teste não paramétrico de *Wilcoxon*. O teste foi realizado para a combinação dois a dois de 3 tipos histológicos de interesse. Dessa forma foram geradas 3 listas de genes diferencialmente expressos com nível de significância  $\alpha < 0.05$ .

A clusterização hierárquica foi realizada baseada na distância Euclidiana usando os genes diferencialmente expressos. A confiabilidade da clusterização foi avaliada pela técnica de *Bootstrap* usando o programa TMEV (SAEED et al. 2003).

## **B Classificadores**

Dentro do conjunto de dados disponíveis foi realizada uma busca por genes capazes de separar perfeitamente os grupos de amostras DCIS-puro e DCIS/IDC. Para isso foi utilizado uma variação do classificador linear de *Fisher* (LDA - *Linear Discriminant Analysis*) chamada de Busca e Escolha (CRISTO 2003). O objetivo deste método é reduzir a dimensão do espaço de busca de classificadores e assim diminuir o tempo computacional necessário para encontrá-los. Esse objetivo é atingido em duas etapas: primeiro o método faz uma busca pelos genes diferencialmente expressos entre os tipos histológicos que queremos classificar. Essa etapa foi feita através do teste de *Wilcoxon*. Depois disso, o espaço de busca foi reduzido para os 20 genes considerados como mais significativos de acordo com seu *pvalor* e todas combinações entre eles foi realizada.

Quando um trio é gerado, um teste de validação cruzada (CV - *Cross Validation*) é realizado para testar o número de acertos que a superfície

classificadora é capaz de realizar. Foi calculado também seu SVD (*Singular Value Decomposition*) que oferece uma medida de separação dos grupos. Portanto, quanto maior o SVD, melhor a separação dos grupos e melhor o classificador.

A busca por classificadores foi realizada dentro de grupos de genes pertencentes as vias de sinalização WNT, PI3K e processo EMT representadas na plataforma e, para cada grupo, foi gerada uma tabela de classificadores ordenada pelo valor de *cross validation*, assim como uma figura que mostra a separação dos grupos pelo melhor trio classificador de cada lista.

### **C Determinação de Redes de Relevância**

Redes de relevância foram construídas para cada grupo de genes de interesse das vias representadas na lâmina. A idéia original proposta por BUTTE et al. (2000) foi ligeiramente modificada para a busca por pares de genes com valores de correlação linear significativamente alterados entre determinados grupos amostrais.

Para se identificar estes pares, uma comparação entre pares de genes foi realizada, e a seguir, seus coeficientes de correlação determinados pelo método da correlação robusta de *Pearson*. A partir destes valores, uma rede foi construída por representação gráfica, atribuindo coloração verde para correlação positiva e, vermelha para correlação negativa. Para comparar as alterações entre os coeficientes de correlação, segundo o grupo de não-neoplásico e DCIS puro contra DCIS/IDC, primeiramente a transformação Z de *Fisher* foi aplicada, onde a diferença entre os valores de

correlação para cada par de genes, nos tipos histológicos estudados, foi transformada em uma nova variável Z, que apresenta distribuição normal aproximada. Esta informação foi usada para aferir a significância da quebra de correlação entre os grupos de amostras ao aplicarmos o teste de hipótese para esta variável Z.

### **3.9 QRT-PCR (QUANTITATIVE REVERSE TRANSCRIPTASE POLYMERASE CHAIN REACTION)**

#### **3.9.1 Síntese da primeira fita de cDNA**

A síntese da primeira fita de cDNA foi realizada a partir de 1 µg de RNA amplificado em volume final de 20µl, com iniciadores Oligo dT, utilizando o kit *Superscript<sup>TM</sup> II*, seguindo as instruções do fornecedor (Anexo 4 – item 1.1).

#### **3.9.2 Desenho dos iniciadores**

Foi utilizado o programa *PRIMER Express 3.0* na construção dos iniciadores segundo recomendações do fabricante (*Applied Biosystems*). Para verificar a presença de estruturas secundárias (*stem loop* e *homodimer*) foi utilizado o programa *OLIGOTECH* versão 1.00 (*Copyright*© 1995), onde foram permitidas temperaturas de no máximo 20°C. Na Tabela 2 estão relacionadas as sequências de todos os iniciadores.

Tabela 2 - Lista das seqüências dos iniciadores.

Nome do Gene	Seqüências	Tamanho do fragmento
<b>Genes Endógenos</b>		
<i>HPRT</i>	F: GAACGTCTTGCTCGAGATGTGA	101 pb
	R: TCCAGCAGGTCAGCAAAGAAT	
<i>GAPDH</i>	F: ACCCACTCCTCCACCTTTGA	102 pb
	R: CTGTTGCTGTAGCCAAATTCGT	
<i>BCR</i>	F: CCTTCGACGTCAATAACAAGGAT	67 pb
	R: CCTGCGATGGCGTTCAC	
<b>Genes Alvo - Amostras Pareadas</b>		
<i>KIAA0368</i>	F: GCCTGGC CGCGAAAC	78 pb
	R: GAGTTTTAGCCGTTACACATCAG	
<i>LUM</i>	F: CTATTACCTGGAGGTCAATCAACTTG	125 pb
	R: TGGTTTCTGAGATGCGATTGC	
<i>CRABP2</i>	F:CGAGGAATTGCTCAAAGTGCT	74 pb
	R:GCTGGCTTGGACGCTGC	
<i>RDH-E2</i>	F: CGATTGTGTGCCCTTTTTTA	80 pb
	R: CAGAATTGGCAACAGAGAAGGA	
<i>POSTN</i>	F: GACACACCCGTGAGGAAGTTG	71 pb
	R: CGACCTTCCCTTAATCGTCTTCTA	
<i>FBN1</i>	F:CGGAGAAGCACAAACGAAACT	99 pb
	R: GGCTGTCTTCTCAACATCCCA	
<i>LPL</i>	F: GGCACCTGCGGTATTTGTG	122 pb
	R: CTGTGAAGAATGAAGTGGAGGAAG	
<i>CXCL13</i>	F: TGAATGGATACAAAGAATGATGGAAG	102 pb
	R: TGGAAATATCAGCATCAGGGAATC	
<i>GDF15</i>	F: GCTACGAGGACCTGCTAACCA	150 pb
	R: CGAGAGATACGCAGGTGCAG	
<b>Genes Alvo - Amostras Progressão</b>		
<i>CGI-41 (C1orf66)</i>	F: CCAGGCGTGCAAGGTATC	64 pb
	R: GCCCCCGCTGCACAT	
<i>C16orf5</i>	F: CAGCCAGAGCAGTTAGCCAGTTA	92 pb
	R: CTGACTCCAGACAACTTACCCATTC	
<i>STK25</i>	F: ACCTGGTGGAGCGAGTGC	63 pb
	R: TTCAGCGGGTGGATGTCAG	
<i>LOX</i>	F: CAGGACATCATGCGTATGCC	87 pb
	R: CCAGGCACTGATTTATCCATTG	
<i>MARK3</i>	F: AGACACTCAGTGATTGAGAATGGC	68 pb
	R: GAAGCAACTGGAGTTCTCTGATCA	
<i>NEDD9</i>	F: CCACCATCAGTGCAGAGAAGC	84 pb
	R: GGACACTTGCCCATCTCTCTG	
<i>STX11</i>	F: TCGCACTCTCGCTCCCAG	89 pb
	R:CTGGGAACTGCTGGTCATATTG	
<i>GOSR2</i>	F: GCAGGAGAGACAGCGAGAAGA	90 pb
	R: TGCAGTGATTCTGCCATTGG	
<i>SULF1</i>	F: GGCATTTTGAATCAGCTACACGTA	152 pb
	R: TCCCATCCATCCCATAACTGTC	
<i>TXNL2</i>	F: GACCACAGGCGTGCACC	137 pb
	R: GATACCTTTCCTCATCCATCACAAG	

### **3.9.3 RT-PCR para verificação da amplificação específica dos iniciadores**

As RT-PCR foram realizadas em volume final de 20 $\mu$ l, sendo 2,0  $\mu$ l de tampão (10X); 1,5 mM de MgCl<sub>2</sub>; 0,4 $\mu$ l de iniciador *forward*; 0,4 $\mu$ l de iniciador *reverse*; 1 U de *Platinum Taq DNA Polymerase (Invitrogen Life Technologies)* e cDNA gerado a partir de 20ng de RNA total de HB4a. O programa apresentou a seguinte condição: 2 minutos a 94°C, seguidos de 40 ciclos de 45 segundos a 94°C, 30 segundos a 60°C e 30 segundos a 72°C de extensão, seguida de extensão final por 7 minutos a 72°C. Alíquotas de 5,0 $\mu$ L das reações de RT-PCR foram submetidas a eletroforese em gel de poliacrilamida 8,0% por 60 minutos em voltagem de 80 mV, a fim de se confirmar os tamanhos dos fragmentos esperados.

### **3.9.4 RT-PCR quantitativo (qRT-PCR) ou RT-PCR em Tempo Real (Real Time PCR)**

RT-PCR em tempo real foi realizado com o reagente *SYBR Green (Applied Biosystems)*, o qual emite fluorescência ao se ligar à moléculas de fita dupla de DNA. Uma vez que a cada ciclo da reação novas moléculas de fita dupla são formadas, o nível de fluorescência emitida aumenta gradualmente, 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).

Valores quantitativos são obtidos durante a ciclagem, no ponto em que a amplificação do produto é detectada, ao invés da quantidade do

produto acumulado ao final de um determinado número de ciclos. O parâmetro  $C_T$  (*Cycle Threshold* ou ciclo limite) é definido como o número do ciclo em que a fluorescência gerada atinge um limite acima da fluorescência de fundo que não é detectada pelo aparelho (LIVAK et al. 2001). Os dados de quantificação são analisados pelo *software* do aparelho (MORRISON et al. 1998). *ABI Prism® 7300 Sequence Detection System (Applied Biosystems)*. As reações foram reduzidas em um volume final de 20µl, contendo 10,0µl de *SYBR® Green PCR Master Mix (Applied Biosystems)*, 200nM ou 400nM de cada um dos iniciadores e aproximadamente 0,1 ng de cDNA segundo *round*. Como molde foram utilizados cDNA convertido de RNA amplificado de cada amostra. As reações foram realizadas nas condições universais de ciclagem que consiste 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 a 60°C para o pareamento dos iniciadores 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. As reações, incluindo suas duplicatas foram realizadas em placas de 96 *wells* - *MicroAmp® Optical (Applied Biosystems)* e valores de desvio padrão acima de 0,6 entre as replicatas foram desconsiderados e a PCR, quando possível, repetida. Foi utilizado uma

amostra de cDNA da linhagem HB4a (células normais epiteliais mamárias humanas), denominada calibradora, em cada placa.

### 3.9.5 Padronização dos iniciadores

Para todos os pares de oligonucleotídeos foi padronizada a RT-PCR usando diferentes concentrações dos iniciadores, sendo as quantidades ideais padronizadas individualmente. Considerou-se a utilização das menores concentrações possíveis dos oligonucleotídeos, a fim de evitar a formação de estruturas secundárias por excesso de iniciadores (*primer-dimer*) e se obter os menores valores de  $C_T$ s. A quantidade de iniciadores estabelecida para todos os genes foi de 200nM.

### 3.9.6 Cálculo da eficiência de amplificação dos iniciadores

Obtidas as concentrações ideais dos iniciadores, a eficiência para cada par foi calculada usando diluições de curva padrão como descrito nos protocolos da *Applied Biosystems*. Foram utilizadas 5 diferentes quantidades de cDNA obtidas a partir de diluição seriada da amostra original (10ng, 1ng, 0.1ng, 0.01ng e 0.001ng) de linhagem de mama normal HB4a. As reações foram realizadas em triplicata para cálculo da eficiência de amplificação ( $Ef = 10^{(-1/slope)}$ ), em que *slope* corresponde à inclinação da reta obtida quando se analisa a variação do  $C_T$  dos transcritos alvo e do gene endógeno em função do *log* das diferentes quantidades de cDNA. Os valores de *slope* e da eficiência de cada gene estão representados na Tabela 3 e somente foram

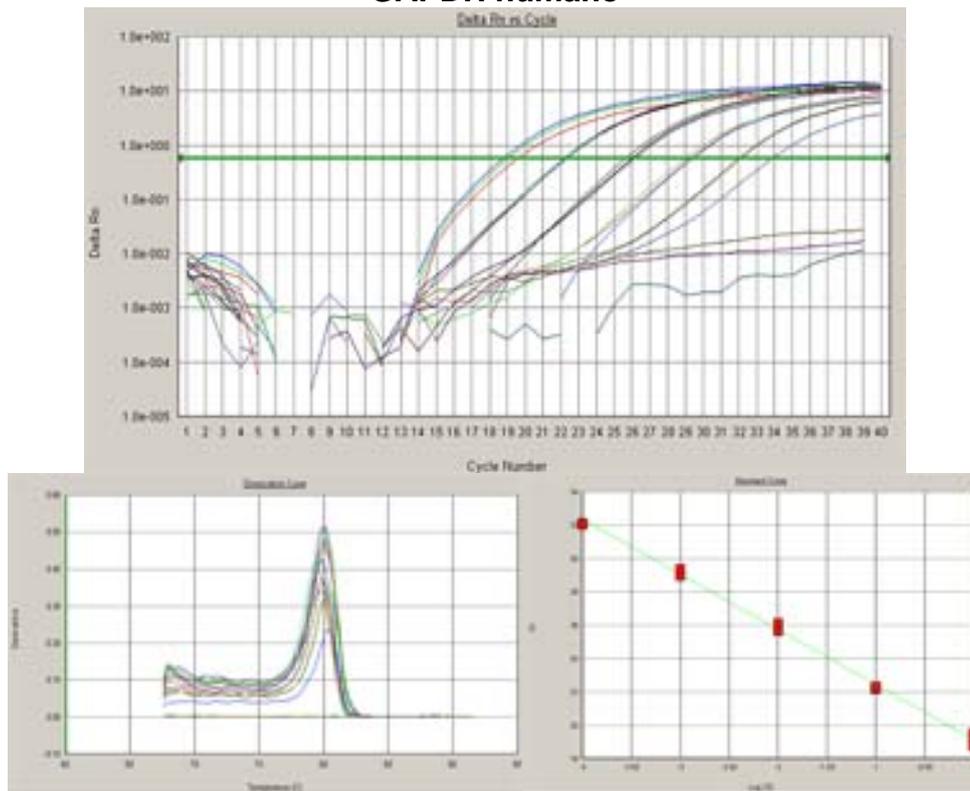
aceitos valores mínimos de eficiência de 70%, conforme descrito nos protocolos da *Applied Biosystems*.

**Tabela 3** - Normalizadores e todos os genes alvos selecionados com seus respectivos valores de *slope* e valores das eficiências calculadas.

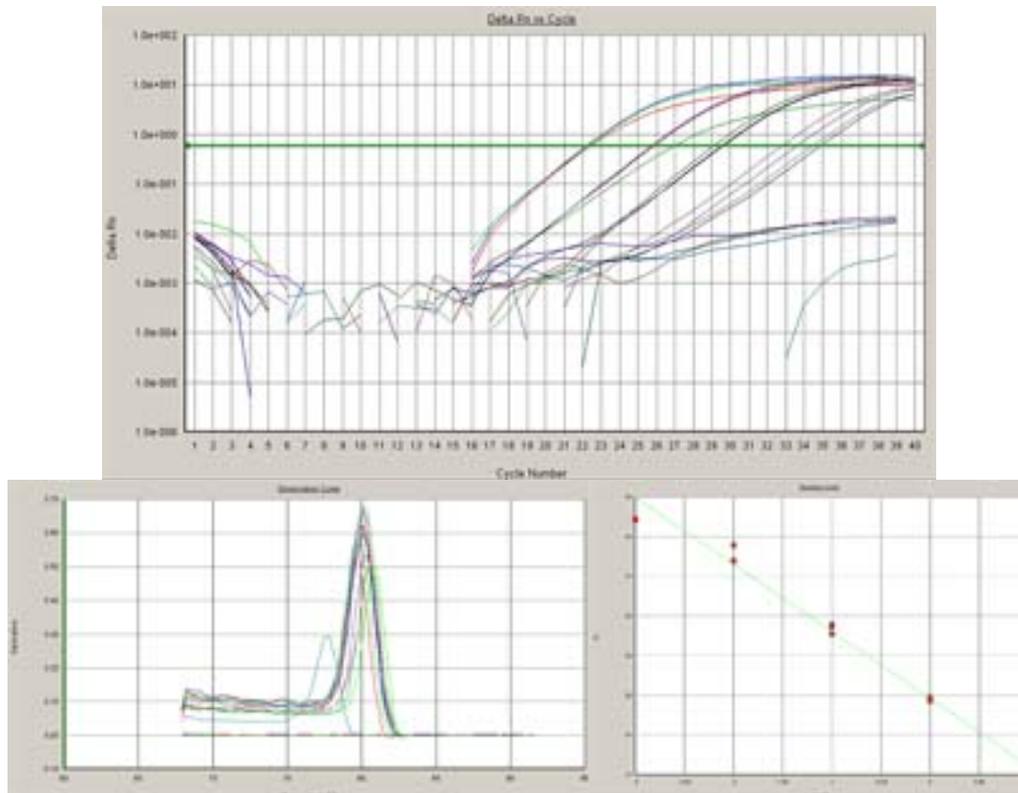
Gene	<i>Slope</i>	R2	Eficiência
<b>Genes Endógenos (Normalizadores)</b>			
<i>BCR</i>	-3,2	0,998	1,05
<i>HPRT</i>	-3,43	0,987	0,96
<i>GAPDH</i>	-3,301	0,996	1,01
<b>Amostras Pareadas</b>			
<i>COL10A1</i>	-2,877	0,9875	1,23
<i>CRABP-2</i>	-3,51	0,996	0,93
<i>CXCL13</i>	-3,25	0,98	1,03
<i>FBN1</i>	-3,58	0,985	0,9
<i>GDF15</i>	-3,88	0,997	0,81
<i>KIAA0368</i>	-3,224	0,98	1,04
<i>LUM</i>	-2,76	0,98	1,30
<i>MAGEH1</i>	-3,474	0,994	0,94
<i>POSTN</i>	-2,9	0,995	1,21
<i>PRLP</i>	-2,99	0,984	1,16
<i>RAB-10</i>	-3,456	0,997	0,95
<i>RDH-E2</i>	-2,99	0,978	1,16
<b>Amostras Progressão</b>			
<i>C16orf5</i>	-3,68	0,97	0,87
<i>CGI-41</i>	-3,63	0,99	0,89
<i>TXNL2</i>	-2,8	0,99	1,28
<i>GOSR2</i>	-3,86	0,99	0,82
<i>LOX</i>	-3,15	1,00	1,08
<i>MARK3</i>	-4,03	0,99	0,77
<i>SULF-1</i>	-3,39	0,098	0,97
<i>STK25</i>	-3,50	0,98	0,93

Os gráficos de amplificação (*Amplification Plot*), especificidade da reação (*Dissociation curve*) e diluições de curva padrão para os normalizadores *GAPDH* e *HPRT1* e para os genes *C16orf5* e *LOX* estão representados na Figura 11.

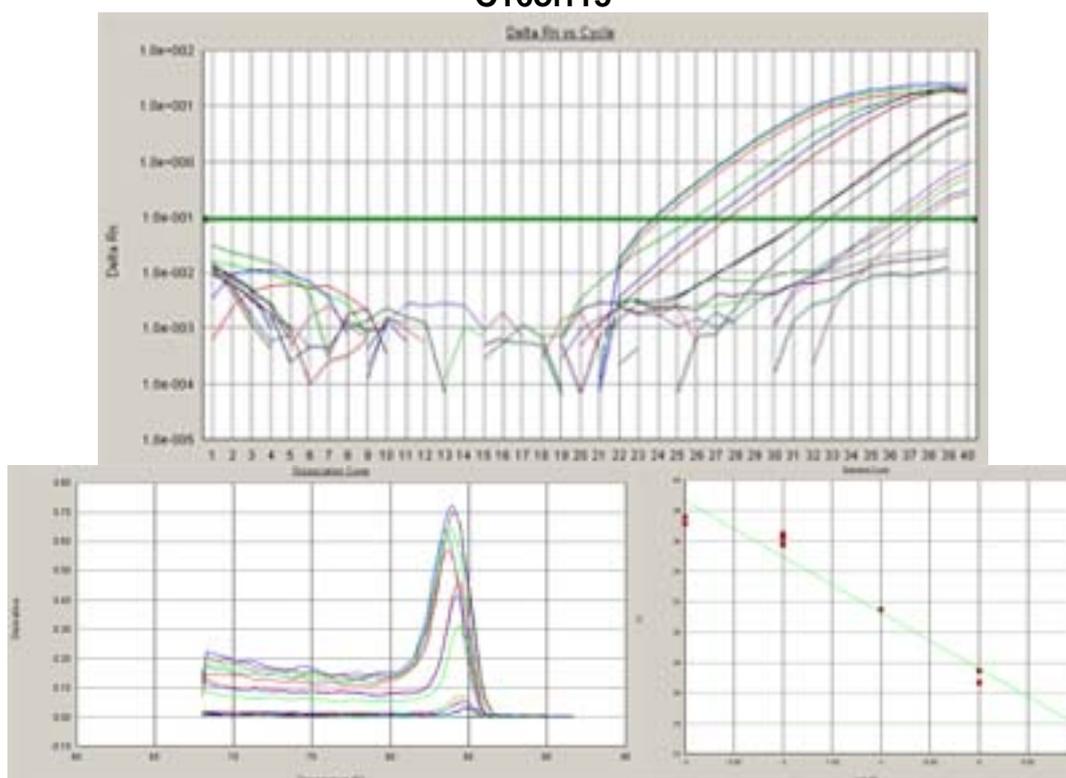
## GAPDH humano



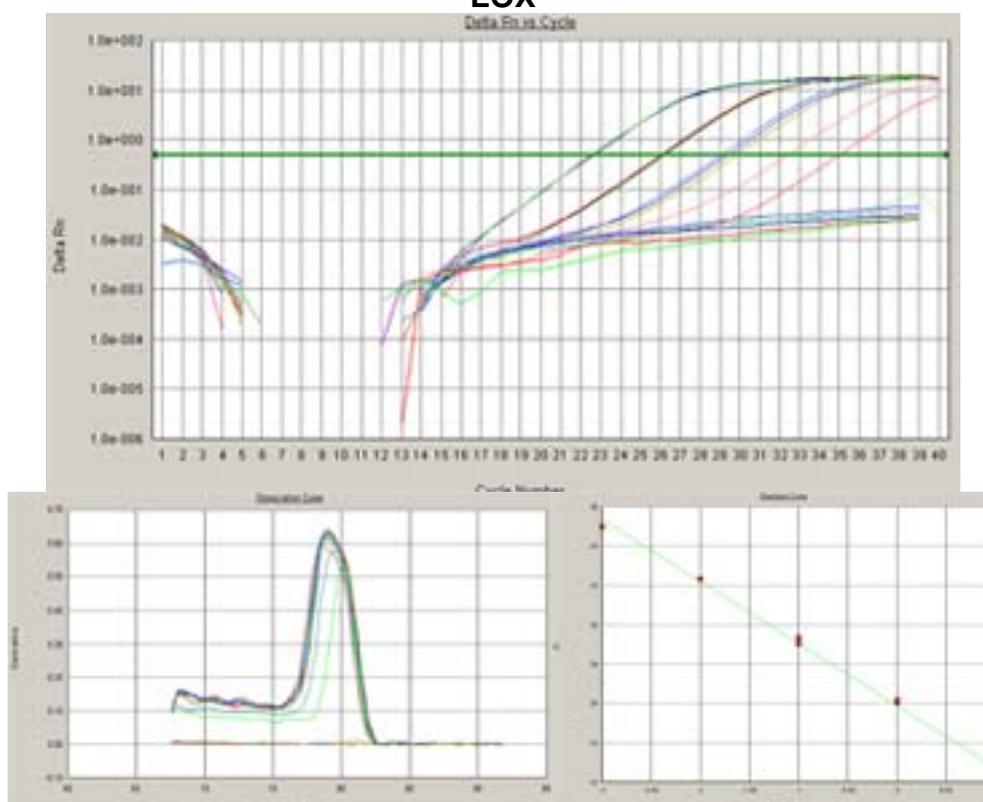
## HPRT



## C16orf15



## LOX

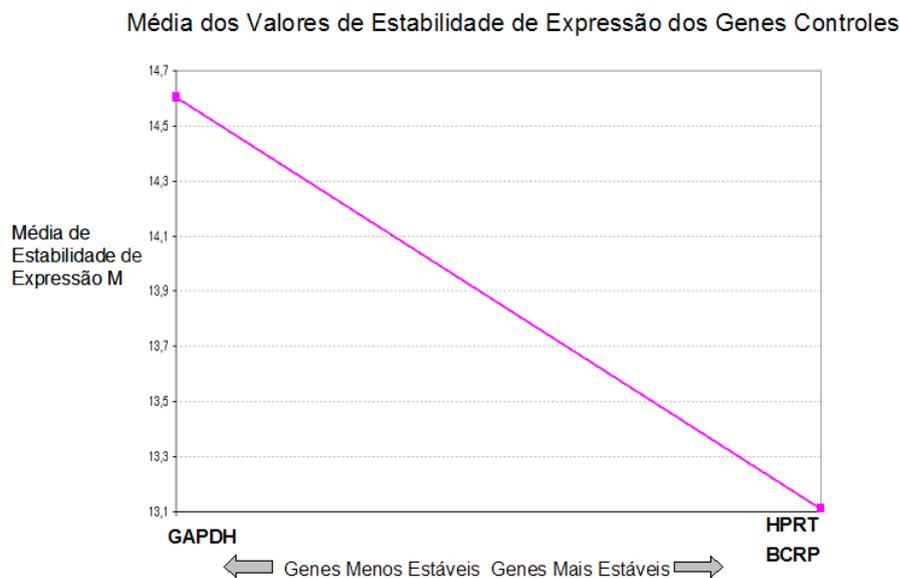


**Figura 11** - Os gráficos de amplificação (*Amplification Plot*), especificidade da reação (*Dissociation curve*) e diluições de curva padrão respectivamente para os normalizadores *GAPDH* e *HPRT1* e para os genes *C16orf5* e *LOX*.

### 3.9.7 Seleção do gene normalizador (endógeno)

Para identificar o gene normalizador mais estável neste grupo de amostras, foram avaliados 3 genes endógenos selecionados com base nos dados da literatura, o *GAPDH* (NM\_002046.2), *HPRT1* (NM\_000194) e *BCR* (NM\_004327) (VANDESOMPELE et al. 2002, de KOK et al. 2005, HUGGETT et al. 2005) pelo uso uma ferramenta disponível *online* (<http://medgen.urgent.be/~jvdesomp/genorm/>), denominada geNorm (VANDESOMPELE et al. 2002). O programa geNorm calcula uma medida de estabilidade gênica ( $M$ ) para os genes endógenos em um grupo de amostras indicando os mais estáveis para serem utilizados como normalizadores. O valor  $M$  sugerido pelo programa é de 1.5, abaixo do qual se tem os melhores genes endógenos. A estabilidade 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 das condições experimentais ou de tipos celulares.

Na primeira análise do geNorm os genes endógenos *GAPDH*, *HPRT1* e *BCR* foram verificados. Maior estabilidade foi observada para os genes *HPRT* e *BCR*, e menor para o gene *GAPDH* para este grupo de amostras, conforme resultado mostrado na Figura 12.



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

**Figura 12** - Gráfico da média dos valores de estabilidade de expressão dos genes *GAPDH*, *HPRT* e *BCR*.

Como o gene *HPRT* foi identificado por de Kok et al. (2005) como o melhor gene referência dentre 13 genes endógenos mais utilizados em normalização, decidimos fazer os cálculos de expressão relativa utilizando somente o gene *HPRT*. Outro motivo se deve ao fato de que os desvios-padrão de algumas das replicatas para o gene *BCR* foram maiores que 0,6, mesmo após repetição.

### 3.9.8 Análise dos dados de expressão

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 leva em consideração a eficiência de amplificação de cada par de iniciadores (Figura 13), em que ***E alvo*** corresponde à eficiência de amplificação dos iniciadores específicos para o transcrito alvo, ***E endógeno*** corresponde à eficiência de amplificação dos iniciadores específicos para o gene endógeno,  **$\Delta C_T$  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,  **$\Delta C_T$  endógeno** corresponde à diferença entre o  $C_T$  do gene endógeno obtido para a amostra em análise e o  $C_T$  do gene endógeno obtido para a amostra calibradora.

A linhagem de mama normal HB4a foi utilizada como amostra calibradora nos experimentos de RT-PCR em tempo real. Assim, o cDNA desta linhagem foi utilizado em duplicata em cada uma das placas de reação de RT-PCR em tempo real.

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

**Figura 13** - Modelo matemático para quantificação relativa em RT-PCR em tempo real (PFAFFL 2001).

### 3.9.9 Determinação das diferenças de expressão

A expressão diferencial dos genes selecionados na primeira fase do estudo foi determinada a partir dos dados normalizados pelo método de quantificação relativa considerando um gene endógeno (PFAFFL 2001), apresentado no item anterior. Para determinar as diferenças de expressão

entre as amostras foi estabelecido como critério uma concordância de pelo menos 50% entre os resultados das 2 metodologias.

Para determinar as diferenças de expressão entre amostras pareadas (DCIS e IDC) assim como entre os grupos de amostras - Normal e DCIS puro x Intraductal componente de IDC e IDC - foi utilizado o programa computacional *GraphPad Prism version 4.0* (*GraphPad Software Inc.*, San Diego, CA). Para cada expressão relativa de cada grupo foi aplicado o teste *T Student* que leva em conta a média e desvio padrão da medida relativa de cada grupo, sendo considerados como significativos valores com  $p < 0,05$  (nível de significância de 5%).

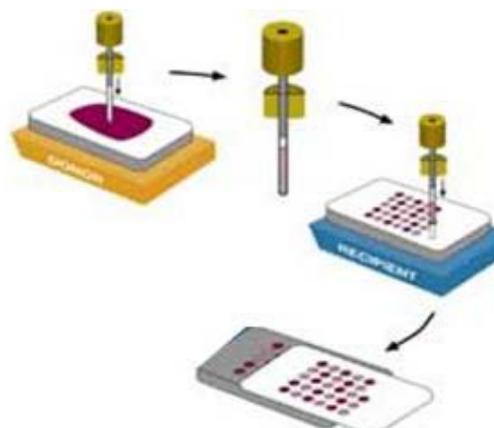
### **3.10 CONSTRUÇÃO DO *TISSUE MICROARRAY* (TMA)**

A metodologia de *tissue microarray* é uma técnica relativamente simples, cuja validação em relação às técnicas convencionais tem se consolidado cada vez mais, tendo se difundido rapidamente nos grandes centros mundiais de pesquisa (SCHRAML et al. 2003; VAN DE RIJN e GILKS 2004).

O TMA utilizado foi disponibilizado pelo Dr. Fernando Soares (Departamento de Anatomia Patológica do Hospital AC Camargo). Este foi construído utilizando-se o dispositivo técnico chamado de “*tissue microarrayer*” (*Beecher Instruments*, Silver Springs, EUA) em uso. Foram selecionadas as áreas apresentando representação qualitativa e quantitativa da neoplasia, suficiente para a definição diagnóstica em cada lâmina. Estas

áreas foram identificadas no bloco de parafina e puncionadas com agulha de 1mm (Figura 14). Os cilindros obtidos foram transferidos para o bloco de parafina receptor em posição bidimensional determinada e gravada pelo equipamento. Um mapa com identificação detalhada de cada cilindro foi construído em planilha de *Excel*® (*Microsoft*, EUA) permitindo a localização exata de cada caso.

Foram coletadas amostras em duas diferentes áreas do bloco original para confecção de duplicata no bloco receptor. Os blocos de TMA foram cortados no micrótomo rotativo na espessura de 3 µm e colhidos em lâminas adesivas da *Instrumedics* (*Hackensack NJ*, EUA).



**Figura 14** - Esquema mostrando a montagem do bloco do *tissue microarray* e a lâmina para a análise.

Foram utilizados 400 casos de carcinoma ductal invasivo de mama a partir de TMA construído no Departamento de Anatomia Patológica do Hospital A C Camargo.

### 3.11 IMUNOHISTOQUÍMICA

As reações de imunohistoquímica foram realizadas nas lâminas de TMA, usando o sistema de polímero (*Advance, DAKO*<sup>®</sup>). Os cortes foram desparafinizados através de três trocas de xilol de 5 minutos a temperatura ambiente. A seguir os cortes foram hidratados em soluções com concentrações decrescentes de etanol (100%, 90%, 70% e 50%) e posteriormente lavados em água corrente por 5 minutos. A recuperação antigênica foi realizada pelo uso do calor com irradiação por panela de pressão utilizando-se tampão citrato 10mM pH 6.0 por 30 minutos a 125°C, deixando mais 10 minutos a 90°C. A seguir foi feito o bloqueio da peroxidase endógena com solução de peróxido de hidrogênio a 3% (água oxigenada 10V) com três trocas de 5 minutos cada. Após o bloqueio da peroxidase, também se fez o bloqueio da avidina/biotina (*DAKO*<sup>®</sup> *Biotin Blocking System*) e proteína (*DAKO*<sup>®</sup> *Protein Block Serum-Free*) 20 minutos cada, com três lavagens de 5 minutos com PBS entre eles.

As lâminas foram incubadas com o anticorpo primário anti-CRABP2 (*MAB5488, Chemicon Inc., CA, 1:2000*) e anti-LUM (*ProteinTech Group. Inc., 1:100*) diluído em tampão PBS contendo albumina bovina (BSA) 1% (*SIGMA*<sup>®</sup>, A9647, USA) e azida sódica ( $\text{NaN}_3$ ), 0,1% por 2 horas em câmara úmida. Após a incubação as lâminas foram lavadas em tampão PBS com três trocas de 5 minutos cada. A seguir, foram incubadas com o pós-primário por 30 minutos a temperatura ambiente (*Advanced*<sup>™</sup> *HRP Link, DAKO A/S, K0690, Denmark*), seguidas por três lavagens de 5 minutos em PBS, e

incubação com o polímero (*Advanced<sup>TM</sup> HRP Link, DAKO A/S, K0690, Denmark*) por 30 minutos a temperatura ambiente. Em seguida, os cortes foram lavados com três trocas de PBS, reveladas com solução de Diaminobenzidina (*DAB, DAKO®*) a 60mg% e contra coradas com hematoxilina de Harris (*Merck®*).

As reações foram acompanhadas de controle positivo, em tecido sabidamente positivo para o anticorpo testado e dois controles negativos. O primeiro deles pelo não uso do anticorpo primário e o segundo pela retirada do anticorpo secundário durante os passos da reação.

Após a realização das reações imunohistoquímicas, os resultados foram analisados em microscópio óptico e classificados em relação a intensidade da reação, número de células positivas e padrão de marcação.

Em relação à positividade de marcação, as reações foram consideradas positivas se a coloração foi específica no tecido de interesse com frequência maior que 10% na lâmina analisada. Quando positivas, as lâminas foram classificadas de acordo com a intensidade e agrupadas nas categorias de reações com intensidade fraca ou forte, considerando a marcação com frequência maior que 50% no corte avaliado. A intensidade foi classificada forte quando a marcação foi facilmente observada em objetiva de menor aumento no microscópio óptico comum.

A quantificação das células com núcleos positivos foi feita em duplicata e o índice de positividade foi obtido através da leitura 1000 células (aumento de 400x) em cada amostra, sendo que a escolha da região foi determinada pela integridade estrutural do tecido e positividade mais

evidente (quando houve). As células positivas e negativas foram contadas e os resultados expressos em porcentagem de células por campo.

Duas lâminas em diferentes alturas de corte foram utilizadas para os marcadores. Assim, o marcador foi avaliado em quatro diferentes locais representativos do tumor.

### **3.12 ANÁLISES ESTATÍSTICAS**

Os dados coletados a partir dos prontuários das 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 inseridas no 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 para Windows (Statistical Package for Social Science, SPSS 15.0)*.

A caracterização da população de estudo e do tratamento efetuado foi realizada por meio de estatística descritiva, calculando-se, inicialmente, as freqüência de todas as variáveis estudadas. A associação entre os marcadores moleculares com as variáveis clínicas e histopatológicas foram avaliadas pelo teste estatístico qui-quadrado de *Pearson*. Quando as freqüências esperadas foram menores do que cinco ou o número da amostra foi pequeno, os resultados foram analisados por meio do teste exato de

*Fisher* com intuito de evitar distorções de significância estatística. As diferenças foram consideradas estatisticamente significativa quando o valor de  $p < 0,05$ .

## 4 RESULTADOS

### 4.1 DELINEAMENTO EXPERIMENTAL

Nesse estudo foram avaliados dois aspectos do carcinoma ductal de mama. Um deles se refere a identificação de genes possivelmente envolvidos na transição do carcinoma ductal *in situ* para o invasivo, representando genes candidatos a marcadores de prognóstico; e o outro, a caracterização molecular do carcinoma ductal *in situ* sem a presença do componente invasivo (DCIS puro) e a comparação com o ductal *in situ* coexistindo com o componente invasor (DCIS-IDC), identificando assim, genes que possam estar envolvidos na aquisição da capacidade de invadir tecidos adjacentes. Estes genes representariam candidatos a marcadores moleculares de progressão e prognóstico do carcinoma ductal *in situ*. Para isso foram realizados dois delineamentos experimentais. No primeiro propósito foi avaliado o padrão de expressão gênica de células tumorais de dois tipos histológicos presentes na mesma amostra, ductal *in situ* e invasivo, filtrando assim a variabilidade biológica entre os grupos de amostras. Esta abordagem aumenta a chance de identificar genes realmente envolvidos no processo de invasão de tecidos adjacentes. No segundo foi avaliado o padrão transcricional de células do mesmo tipo histológico (DCIS), mas que diferem pela presença ou não do componente invasor na mesma lesão, buscando assim identificar alterações moleculares que

possam prever a progressão da doença antes que ocorra alteração morfológica. Nesse delineamento os grupos de amostras utilizados foram distintos.

Para tanto, foram utilizadas duas plataformas diferentes. Uma delas contendo 4.608 genes humanos (4.8k) e não apresentando nenhuma classe funcional preferencial de genes (BRENTANI et al. 2005). A outra plataforma utilizada apresenta 1.854 seqüências, contendo genes pertencentes à vias de sinalização celular envolvidas no processo de transformação, invasão e metástase, tais como WNT, PI3K e processo EMT, denominada LV-01. Assim, nesse estudo, os resultados são apresentados separadamente a depender da plataforma utilizada.

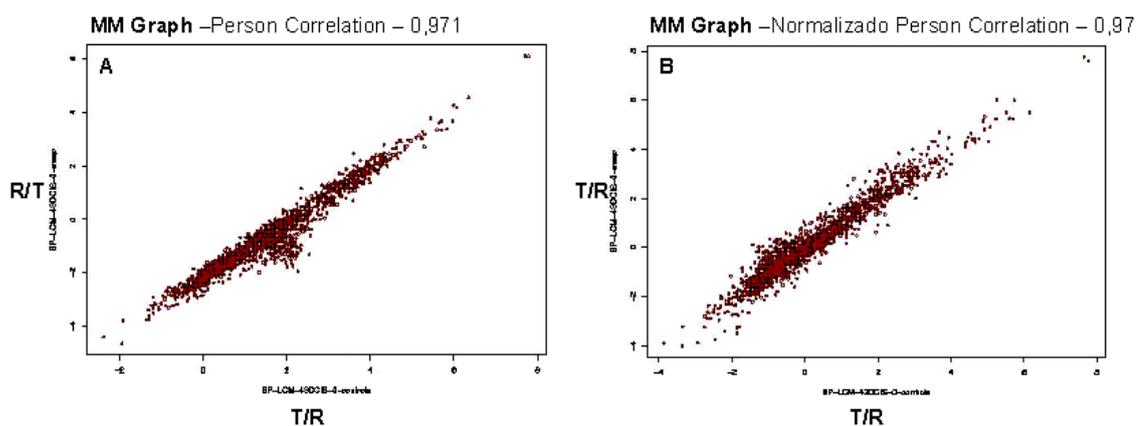
## **4.2 RESULTADOS REFERENTES A PLATAFORMA 4.8K**

### **4.2.1 Amostras e Qualidade das lâminas**

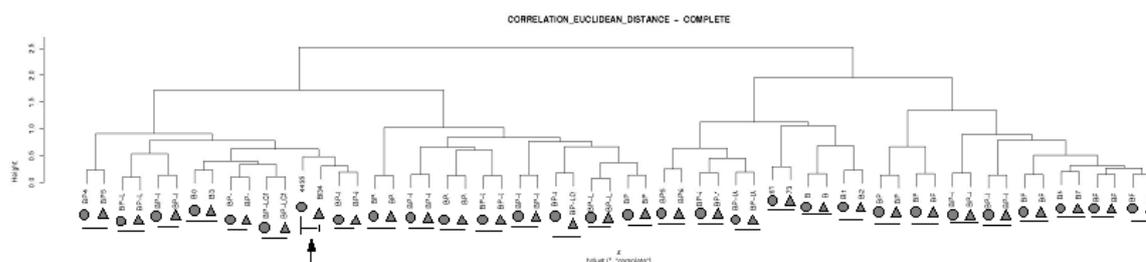
Neste estudo foram selecionados 30 casos (46 amostras), sendo 4 amostras de tecido não neoplásico (borda de fibroadenoma), 16 casos de amostras pareadas (16 DCIS e 16 IDC), 5 DCIS puro, 3 *in situ* associado ao IDC (DCIS/IDC) e 2 IDC somente (Anexo 6).

Todas as amostras marcadas e hibridizadas apresentaram boa qualidade com média de correlação de *Pearson* entre controle e *swap* de 0,91. Um exemplo pode ser visualizado na Figura 16 na qual está representado o *MMPlot* com os dados da intensidade de fluorescência para cada gene das lâminas controle e *swap* sem normalização (A) e depois da

normalização (B). As lâminas controle e *swap*, uma vez que são réplicas com inversão de corantes, clusterizaram juntas no dendograma reportando eficiência na correção de incorporação diferencial dos fluoróforos e alta reprodutibilidade experimental (BRENTANI et al. 2005) (Figura 16).



**Figura 15** - Gráfico MMPlot mostrando a qualidade de correlação entre a lâmina controle e a lâmina *swap* avaliada através de correlação de *Pearson*. R- referência e T- Tumor.



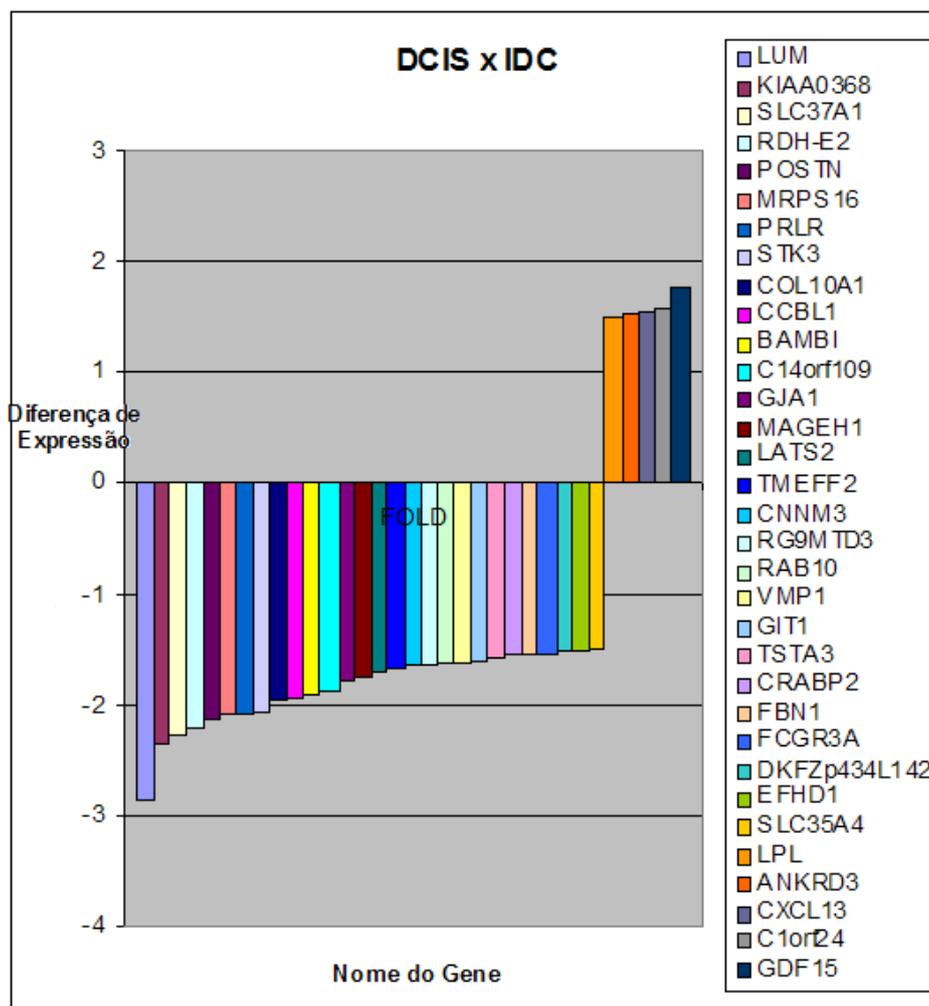
**Legenda:** Os círculos e triângulos com tracejado abaixo indicam lâminas controle e *swap* do mesmo caso respectivamente. Única amostra (controle e *swap*) não clusterizada dentro da mesma subdivisão aparece indicada pela seta.

**Figura 16** - Dendograma das lâminas controle e *swap* das amostras hibridizadas na plataforma 4.8k.

#### **4.2.2 Identificação de Genes Diferencialmente Expressos entre os Componentes DCIS e IDC Presentes na Mesma Amostra**

Para esta análise foram comparados 16 casos de amostras pareadas nas quais DCIS e IDC coexistem na mesma amostra.

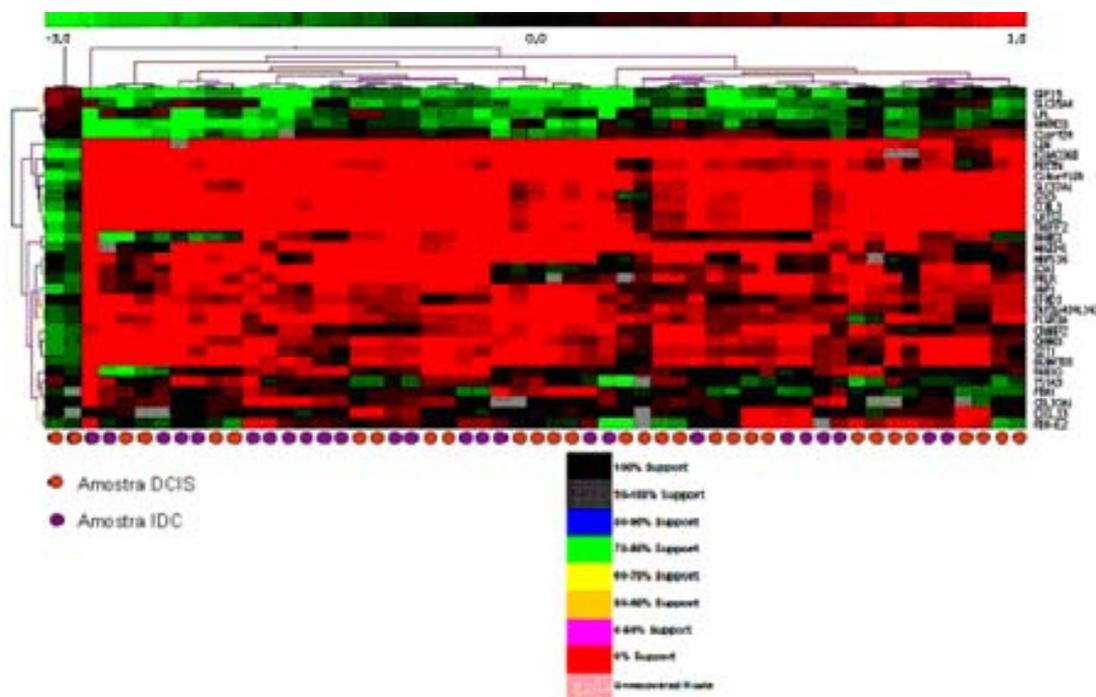
A comparação entre os componentes DCIS e IDC resultou na seleção de 33 genes diferencialmente expressos (segundo critério de diferença de expressão de pelo menos 1,5 vezes e  $p < 0,01$ ), denominado grupo de amostras pareadas, apresentando expressão diferencial não superior a 1,8 vezes para os 5 genes mais expressos em DCIS e 2,8 vezes para os 28 mais expressos em IDC (Figura 17). A lista dos 33 genes pode ser visualizada no Anexo 7.



**Figura 17** - Genes diferencialmente expressos na comparação entre DCIS e IDC coexistindo na mesma amostra ( $fold > |1,5|$  e  $p < 0,01$ ).

Para verificar se o padrão de expressão destes genes poderia discriminar as amostras segundo os diferentes componentes da lesão, uma clusterização hierárquica com distância euclidiana foi realizada baseada na expressão destes 33 genes (Figura 18). A clusterização não foi capaz de segregar as amostras, mostrando alta similaridade no perfil de expressão

gênica entre os dois tipos histológicos mesmo no padrão transcricional destes genes nas diferentes lesões.



**Figura 18 - A:** Clusterização hierárquica com os 33 genes diferencialmente expressos entre DCIS e IDC (grupo pareado). **B:** Legenda do suporte do cluster avaliado pela técnica do *Bootstrap*.

#### 4.2.3 Caracterização Funcional dos Genes Diferencialmente Expressos

Com a finalidade de identificar aspectos funcionais deste grupo de genes, foi investigado a representatividade dos genes anotados pelos processos biológicos através do *Gene Ontology Consortium Annotation* (GO) (<http://vortex.cs.wayne.edu:8080/index.jsp>), usando a distribuição qui-quadrada do *Onto-Tools* (KHATRI et al. 2005). Desenvolvimento do esqueleto, sinalização célula-célula, percepção visual, cascata de proteína quinase e transporte foram observados como processos biológicos com representatividade estatisticamente significantes no processo de transição

do DCIS para o IDC (Tabela 4). O mesmo foi realizado anotando os genes pelas vias metabólicas nas quais pertencem. Interação do receptor citosina-citosina foi a via metabólica encontrada com significância estatística para este mesmo grupo de genes (Tabela 5).

**Tabela 4** - Processos biológicos com relevância estatística ( $p < 0,05$ ) no grupo de genes diferencialmente expressos entre as amostras pareadas (DCIS x IDC).

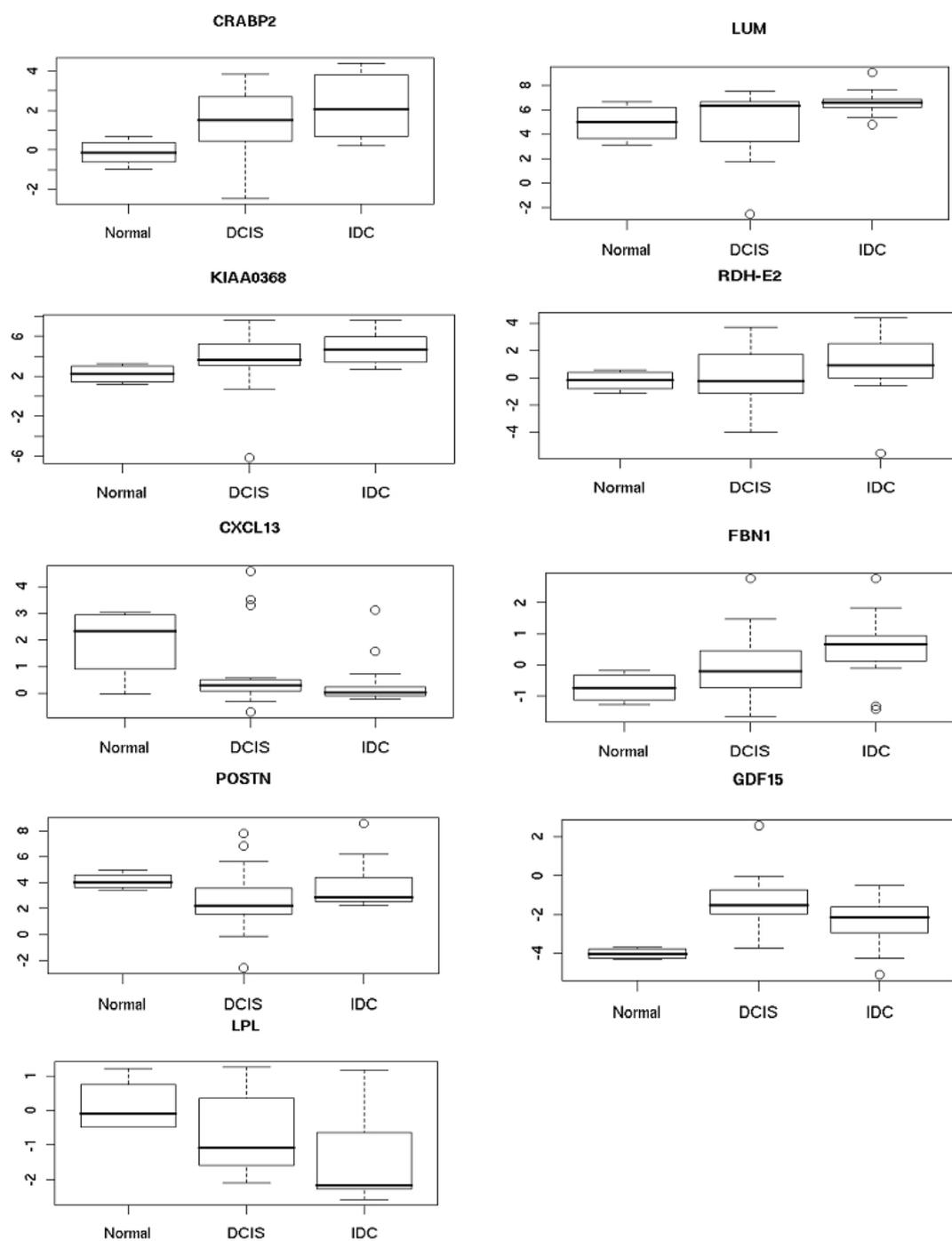
Processo Biológico	Nome do Gene	pFDR
Desenvolvimento do Esqueleto	<i>FBN1</i>	1,42E-09
	<i>COL10A1</i>	
	<i>POSTN</i>	
Sinalização Célula-Célula	<i>CXCL13</i>	3,57E-04
	<i>GDF15</i>	
	<i>GJA1</i>	
Percepção Visual	<i>FBN1</i>	6,39E-04
	<i>LUM</i>	
Cascata de Proteína Quinase	<i>STK3</i>	0,0012
	<i>LATS2</i>	
Transporte	<i>C1orf24</i>	0,0012
	<i>CRABP2</i>	
	<i>GJA1</i>	
	<i>SLC37A1</i>	

**Tabela 5** - Via Metabólica com relevância estatística no grupo de genes diferencialmente expressos entre as amostras pareadas (DCIS x IDC).

Via Metabólica	Genes da Via	Anotação
Interação do receptor citosina-citosina (P valor 0.039)	<i>CXCL13</i>	<i>chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)</i>
	<i>PRLR</i>	<i>prolactin receptor</i>

#### 4.2.4 Seleção dos genes diferencialmente expressos entre as amostras pareadas para serem avaliados por RT-PCR em tempo real

Para selecionar os genes candidatos à marcadores moleculares de prognóstico a serem validados por RT-PCR em tempo real ou qRT-PCR, o nível de expressão dos 33 genes selecionados nos experimentos de *microarray* foi comparado nas amostras normais (grupo independente), DCIS e IDC para avaliar a tendência de expressão nestes grupos de acordo com a progressão da doença (Normal – DCIS – IDC). Para isso, um gráfico (*Box Plot*) para cada um dos 33 genes diferencialmente expressos entre as amostras pareadas foi construído. Esta análise permite verificar a tendência no nível de expressão crescente ou decrescente entre os grupos de acordo com a progressão da doença. Segundo este critério, 13 genes foram classificados dos quais 7 (*CRABP2*, *LUM*, *KIAA0368*, *RDH-E2*, *CXCL13*, *FBN1* e *LPL*) foram selecionados randomicamente para serem confirmados por qRT-PCR. Outro critério estabelecido envolve a seleção dos genes que apresentaram o maior valor de diferença de expressão (*Fold*) na comparação entre DCIS e IDC (*LUM* e *GDF15*), sendo que um deles (*LUM*) preenche ambos os critérios citados acima. O gene *POSTN* foi selecionado para ser avaliado por qRT-PCR, uma vez que apresenta relato recente na literatura (SCHUETZ et al. 2006) como supostamente envolvido no processo de transição do DCIS para o IDC. Na Figura 20 estão representados os gráficos *Box Plot* dos genes selecionados para confirmação por qRT-PCR. Os gráficos de todos os 33 genes podem ser visualizados no Anexo 8.



**Figura 19** - *Box Plot* dos genes selecionados para confirmação por qRT-PCR quanto a tendência crescente ou decrescente no nível de expressão entre as amostras normais, DCIS e IDC.

#### **4.2.5 Confirmação da expressão gênica por qRT-PCR dos genes diferencialmente expressos selecionados entre as amostras pareadas**

Foram realizadas reações de RT-PCR para verificação de amplificação de fragmento de tamanho específico utilizando iniciadores desenhados em éxons distintos. Dos 9 genes (*LUM*, *RDH-E2*, *CRABP2*, *KIAA00368*, *CXCL13*, *FBN1*, *POSTN*, *GDF15* e *LPL*) selecionados como diferencialmente expressos entre as amostras pareadas, apenas um gene alvo (*LPL*) não apresentou amplificação específica.

Para determinar as diferenças de expressão entre as amostras pareadas (DCIS e IDC) foi estabelecida como critério uma concordância em pelo menos 50% dos pares de amostras que foram avaliados. Assim, 6 dos 8 genes (75%) analisados apresentaram expressão diferencial confirmada por qRT-PCR ( $Fold > |1,8|$ ) (Tabela 6). Os valores de diferença de expressão para cada par de amostra analisado podem ser visualizados no Anexo 9. Portanto, os genes *LUM*, *RDH-E2*, *CXCL13*, *FBN1*, *POSTN* e *GDF15* foram confirmados e podem estar envolvidos na transição do DCIS para o IDC, sendo assim, candidatos a marcadores de prognóstico no carcinoma ductal de mama.

**Tabela 6** - Genes avaliados através de qRT-PCR nas amostras pareadas. Valores positivos indicam maior expressão em DCIS e valores negativos maior expressão em IDC.

Gene	Microarray (Fold>1.5)	qR-PCR (Fold>1,8) Nº pares confirmados/Nº pares avaliados (%)
<i>LUM</i>	-2,85	5/10 (50%)
<i>RDH-E2</i>	-2,21	6/9 (66,7%)
<i>CRABP2</i>	-1,54	3/10 (30%)
<i>KIAA0368</i>	-2,34	2/10 (20%)
<i>CXCL13</i>	1,53	3/6 (50%)
<i>FBN1</i>	-1,53	5/10 (50%)
<i>POSTN</i>	-2,12	6/10 (60%)
<i>GDF15</i>	1,77	4/8 (50%)

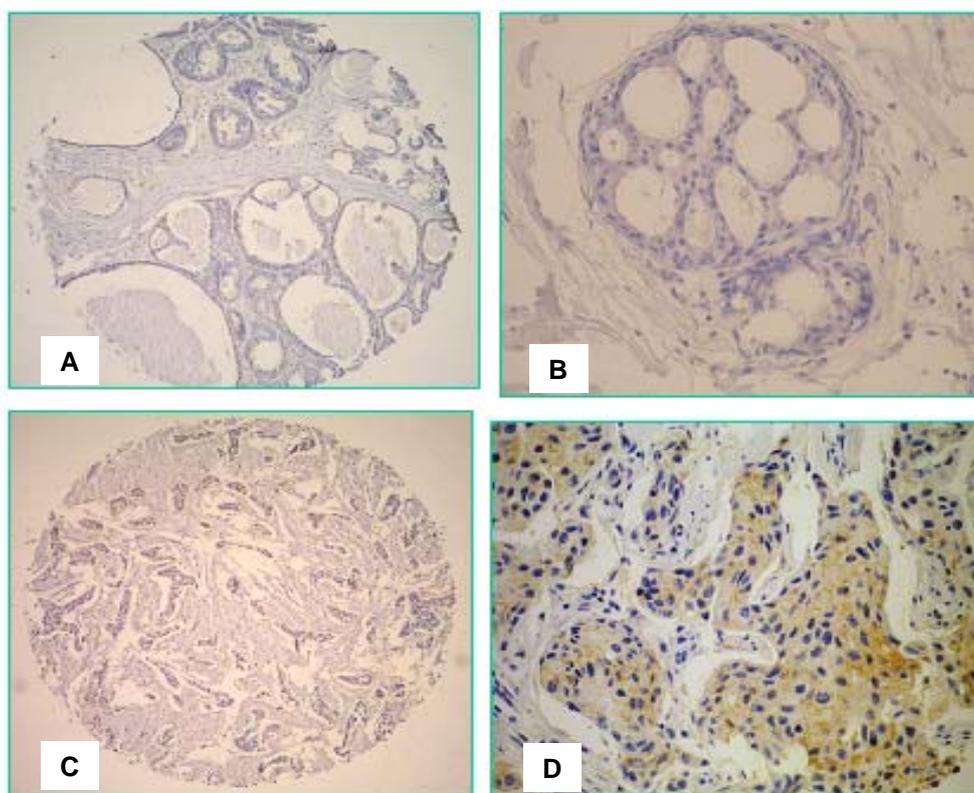
#### 4.2.6 Tissue Microarray (TMA)

Dos 33 genes encontrados como diferencialmente expressos nas amostras pareadas através de *microarray*, os genes *LUM* e *CRABP2* (com maior expressão de RNAm nas amostras IDC) foram selecionados para verificar a correspondência do RNAm em relação ao nível de proteína. O padrão de expressão das proteínas LUM e CRABP2 foi avaliado por reações imunohistoquímicas (IHQ) em duplicata em 400 casos de carcinoma ductal de mama. Este TMA não nos permitiu avaliar lesões pareadas (DCIS e IDC) da mesma amostra. Os dados clinicopatológicos dos casos avaliados neste estudo podem ser visualizados na Tabela 7.

**Tabela 7 - Característica clínica e epidemiológicas dos 400 casos analisados neste estudo através da metodologia de *Tissue Microarray*.**

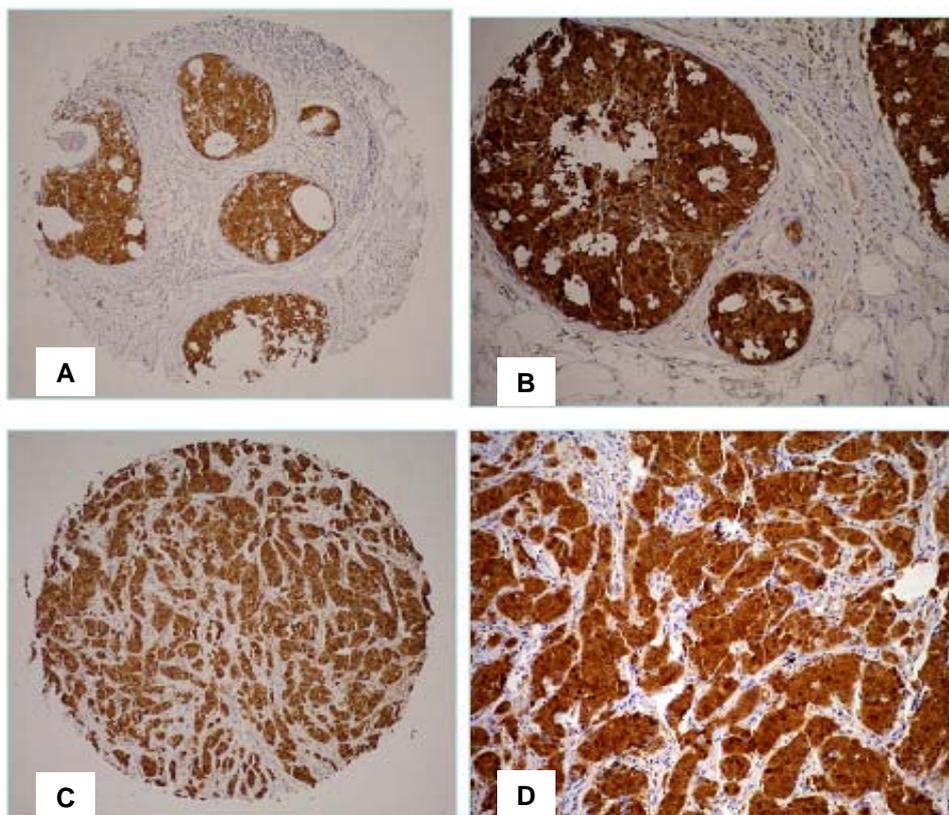
<b>Variáveis clinicopatológicas</b>		<b>nº de casos (%)</b>
Raça	Branca	348 (87)
	Não Branca	52 (13)
Idade	< 40 anos	51 (12,8)
	≤ 40 anos	349 (87,2)
Status Hormonal	Pré-menopausa	165 (41,3)
	Pós-menopausa	235 (58,8)
Paridade	Não	75 (18,8)
	Sim	325 (81,2)
História Familiar	Não	322 (80,5)
	Sim	78 (19,5)
Estadiamento T	T1	55 (14)
	T2	213 (54,3)
	T3	50 (12,8)
	T4	74 (18,9)
Status Linfonodal	Negativo	142 (35,5)
	Positivo	258 (64,5)
Estadio clínico	1	39 (9,8)
	2	184 (46)
	3	159 (39,8)
	4	18 (4,4)
SBR	1	73 (18,3)
	2	228 (57)
	3	99 (24,7)
DCIS associado	Não	223 (58,3)
	Sim	167 (41,8)
Receptor de Estrógeno	Negativo	106 (26,5)
	Positivo	279 (69,8)
Receptor de Progesterona	Negativo	168 (42,6)
	Positivo	200 (54,4)
ERBB2	0+1	303 (75,8)
	2+3	46 (11,5)
TP53	Negativo	15 (3,8)
	Positivo	385 (96,3)
Radioterapia adjuvante	Não	72 (18)
	Sim	328 (72)
Quimioterapia adjuvante	Não	182 (45,5)
	Sim	218 (54,5)
Hormônio terapia	Não	208 (52)
	Sim	192 (48)
Recorrência	Não	246 (61,5)
	Sim	154 (38,5)
Morte	Não	262 (65,5)
	Sim	138 (34,5)

Os resultados foram analisados em microscópio óptico e classificados em relação a intensidade da reação, número de células positivas e padrão de marcação conforme descrito em metodologia. As reações para detecção da proteína LUM apresentaram um padrão de marcação citoplasmática (Figura 20) e as reações para detecção da proteína CRABP2 apresentaram um padrão de marcação tanto nuclear quanto citoplasmática (Figura 21).



**Legenda** – **A e B**: Padrão de marcação citoplasmático de células DCIS, DCIS em aumento 10x e 20x respectivamente; **C e D**: IDC em aumento 10x e 20x respectivamente.

**Figura 20** - Proteína LUM detectada no citoplasma por reações imunohistoquímicas em TMA de mama.



**Legenda – A e B:** Padrão de marcação nuclear e citoplasmática de células DCIS, DCIS em aumento 10x e 20x respectivamente; **C e D:** IDC em aumento 10x e 20x respectivamente.

**Figura 21** - Proteína CRABP2 detectada em núcleo e citoplasma por reações imunohistoquímicas em TMA de mama.

Associações entre a expressão dos receptores utilizados na rotina clínica (ER, PR, HER2/neu e EGFR), diferentes marcadores moleculares, como citoqueratinas (CK) 5, 6, 8, 14 e 18, p53 e P63n (dados de IHQ já disponíveis para essas amostras), dados clínicos (Faixa etária, *status* hormonal, *status* linfonodal, estadio clínico, grau histológico, presença de DCIS), subtipos do carcinoma ductal e a expressão das proteínas LUM e CRABP2 foram realizadas.

Os subtipos do carcinoma ductal de mama foram descritos por Perou et al. (2000) e classificados mais detalhadamente segundo LIVASY et al.

(2007) em luminal A (classificação baseada nas características de expressão dos receptores ER positivo, HER2/neu negativo, CKs 8 e 18), luminal B (expressão dos receptores ER positivo, HER2/neu positivo, CKs 8 e 18), maior expressão do HER2/neu (expressão dos receptores ER negativo e HER2 positivo) e basal ou 'basalóide' (expressão dos receptores ER negativo, HER2 negativo e CKs 5/6 e 14).

Considerando estas variáveis, a expressão da proteína LUM associou-se positivamente com as CKs 5/6, CK 14, CK8, CK18 e o grupo maior expressão do HER2/neu. Enquanto que a expressão da proteína CRABP2 associou-se positivamente com a expressão de ER e PR positivo, luminal A, p53 e negativamente com a CK14. Apesar da proteína CRABP2 apresentar localização distinta na célula (núcleo ou citoplasma), as análises foram realizadas considerando presença ou ausência da proteína independente da localização, uma vez que resultados preliminares não mostraram diferença significativa relacionada a posição da proteína na célula para os casos estudados.

Para os subgrupos com associações estatisticamente significativas, uma nova análise foi realizada. Destas, o subgrupo classificado com luminal A (206 casos) apresentou uma correlação positiva entre as proteínas LUM e CRABP2 (pvalor <0,006).

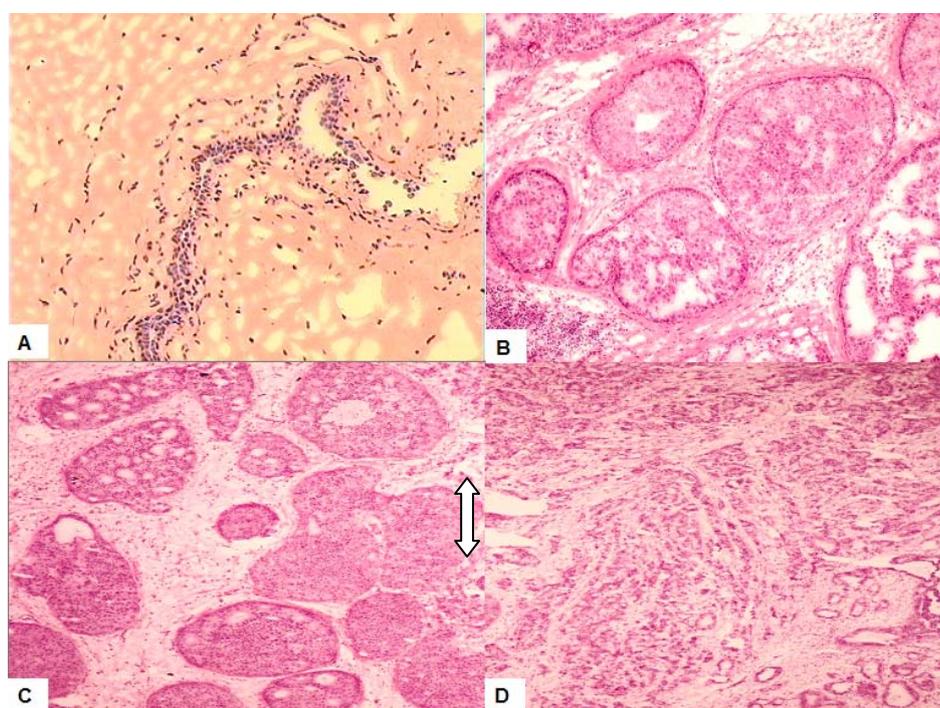
Apenas 3 casos foram classificados como pertencentes ao subtipo 'basalóide', não apresentando assim poder estatístico. As correlações observadas entre a expressão IHQ das proteínas LUM e CRABP2 e as variáveis estudadas podem ser visualizadas na Tabela 8.

**Tabela 8 - Correlações entre a expressão imunohistoquímica de LUM e CRABP2 e as variáveis clinicopatológicas.**

Variáveis Estudadas		LUM		pvalor LUM	CRABP2		pvalor CRABP2
		Negativo (%)	Positivo (%)		Negativo (%)	Positivo (%)	
Faixa etária	= ≤ 40 anos	23 (45,1%)	28 (54,9%)	0,86	21 (41,2%)	30 (58,8%)	0,184
	> 40 anos	162 (46,4%)	187 (53,6%)		111 (31,8%)	238 (68,2%)	
Menopausa	Não	78 (47,3%)	87 (52,7%)	0,731	56 (33,9%)	109 (66,1%)	0,738
	Sim	107 (45,5%)	128 (54,5%)		76 (32,3%)	159 (67,7%)	
Estadio Clínico	1+2	99 (44,4%)	124 (55,6%)	0,404	69 (30,9%)	154 (69,1%)	0,326
	3+4	86 (48,6%)	91 (51,4%)		63 (35,6%)	114 (64,4%)	
Status Linfonodal	Negativo	61 (42,7%)	82 (57,3%)	0,282	53 (37,1%)	90 (62,9%)	0,197
	Positivo	124 (48,2%)	133 (51,8%)		79 (30,7%)	178 (69,3%)	
SBR	1	33 (45,2%)	40 (54,8%)	0,871	28 (38,4%)	45 (61,6%)	0,376
	2	108 (47,4%)	120 (52,6%)		76 (33,3%)	152 (66,7%)	
	3	44 (44,4%)	55 (55,6%)		28 (28,3%)	71 (71,7%)	
DCIS	Ausente	107 (45,9%)	126 (54,1%)	0,877	83 (35,6%)	150 (64,4%)	0,188
	Presente	78 (46,7%)	89 (53,3%)		49 (29,3%)	118 (70,7%)	
ER	Negativo	54 (50,9%)	52 (49,1%)	0,228	60 (56,6%)	46 (43,4%)	0,000
	Positivo	123 (44,1%)	156 (55,9%)		63 (22,6%)	216 (77,4%)	
PR	Negativo	79 (47%)	89 (53%)	0,77	65 (38,7%)	103 (61,3%)	0,017
	Positivo	91 (45,5%)	109 (54,5%)		54 (27%)	146 (73%)	
HER2	Negativo	145 (47,9%)	158 (52,1%)	0,097	106 (35%)	197 (65%)	0,235
	Positivo	16 (34,8%)	30 (65,2%)		12 (26,1%)	34 (73,9%)	
EGFR	Negativo	139 (45,9%)	164 (54,1%)	0,79	95 (31,4%)	208 (68,6%)	0,216
	Positivo	46 (47,4%)	51 (52,6%)		37 (38,1%)	60 (61,9%)	
Luminal A	Não	96 (49,5%)	98 (50,5%)	0,208	86 (44,3%)	108 (55,7%)	0,000
	Sim	89 (43,2%)	117 (56,8%)		46 (22,3)	160 (77,7%)	
Luminal B	Não	178 (46,4%)	206 (53,6%)	0,838	130 (33,9%)	254 (66,1%)	0,075
	Sim	7 (43,8%)	9 (56,2%)		2 (12,5%)	14 (87,5%)	
Basal	Não	185 (46,6%)	212 (53,4%)	0,107	130 (32,7%)	267 (67,3%)	0,213
	Sim	0 (0%)	3 (100%)		2 (66,7%)	1 (33,3%)	
Super-expressão de HER2	Não	178 (47,7%)	195 (52,3%)	0,028	124 (33,2%)	249 (66,8%)	0,70
	Sim	7 (25,9%)	20 (74,1%)		8 (29,6%)	19 (70,4%)	
CK 5,6	Negativo	164 (49,4%)	168 (50,6%)	0,005	113 (34%)	219 (66%)	0,33
	Positivo	21 (30,9%)	47 (69,1%)		19 (27,9%)	49 (72,1%)	
CK 8	Negativo	5 (2,7%)	180 (97,3%)	0,015	5 (3,8%)	127 (96,2%)	0,001
	Positivo	0 (%)	215 (100%)		0 (%)	268 (100%)	
CK 18	Negativo	30 (16,2%)	155 (83,8%)	0,006	32 (24,2%)	100 (75,8%)	0,000
	Positivo	16 (7,4%)	199 (92,6%)		14 (5,2%)	254 (94,8%)	
CK 14	Negativo	182 (47,4%)	202 (52,6%)	0,024	123 (32%)	261 (68%)	0,044
	Positivo	3 (18,8%)	13 (81,3%)		9 (56,3%)	7 (43,8%)	
P63n	Negativo	174 (46,5%)	200 (53,5%)	0,677	120 (32,1%)	254 (67,9%)	0,14
	Positivo	11 (42,3%)	15 (57,7%)		12 (46,2%)	14 (53,8%)	
p53	Negativo	7 (46,7%)	8 (53,3%)	0,974	9 (60%)	6 (40%)	0,023
	Positivo	178 (46,2%)	207 (53,8%)		123 (31,9%)	262 (68,1%)	
CRABP2	Negativo	28 (31,5%)	61 (68,5)	0,006			
	Positivo	18 (15,4%)	99 (84,6%)				

### 4.3 IDENTIFICAÇÃO DE GENES DIFERENCIALMENTE EXPRESSOS ENTRE AMOSTRAS NÃO-NEOPLÁSICAS, DCIS PURO E *IN SITU* ASSOCIADO AO IDC (DCIS/IDC)

Nesta análise foram utilizadas 29 amostras distribuídas em 4 grupos que mimetizam a progressão do câncer de mama (Figura 23). Destas, 4 são provenientes de tecido não neoplásico (borda de fibroadenoma), 5 DCIS puro, 10 *in situ* associado ao IDC (DCIS/IDC) e 10 IDC somente (Anexo 6). Todos os componentes (tecido não neoplásico, DCIS puro, DCIS/IDC e IDC) foram microdissecados a partir de amostras distintas, ou seja, nenhum componente da mesma amostra está representado em mais de um grupo.



**Legenda** - **A:** Ducto mamário normal (20x); **B:** DCIS puro. A neoplasia está restrita à luz ductal (10x); **C:** *In situ* associado ao IDC. Pequenos blocos de IDC são observados no canto superior direito (ver seta) (Lupa); **D:** IDC (Lupa).

**Figura 22** - Representação dos 4 grupos que mimetizam a progressão do carcinoma ductal.

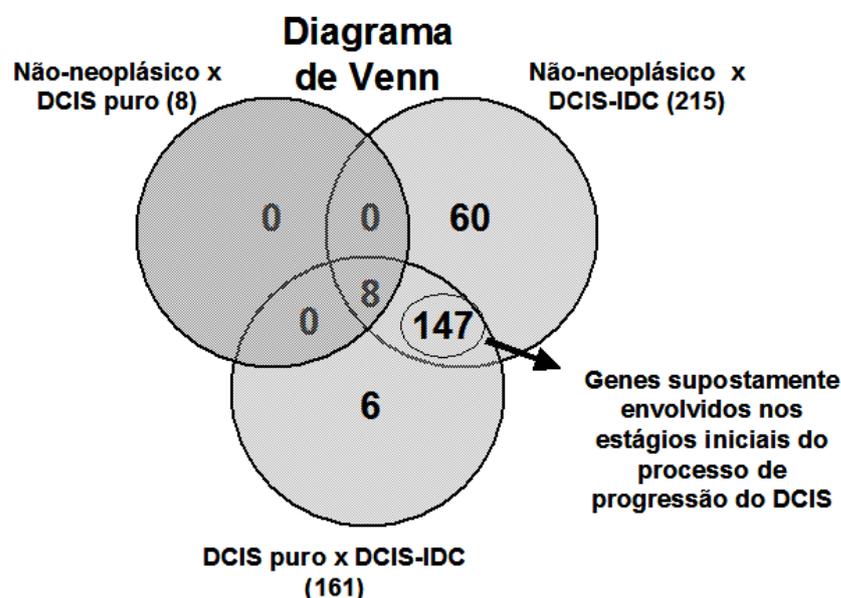
#### **4.3.1 Avaliação da Similaridade do Padrão Molecular e Morfológico em Lesões de Carcinoma Ductal**

Para avaliação do nível de similaridade molecular entre os diferentes grupos cujas lesões mimetizam a progressão do carcinoma ductal de mama, primeiramente foi analisado o padrão geral de expressão gênica nos 4 diferentes grupos de amostras (não-neoplásico, DCIS puro, DCIS/IDC e IDC) aplicando o teste ANOVA corrigido por pFDR ( $<0,01$ ) e 764 genes foram identificados que apresentaram alteração na expressão. Em seguida, o teste de Tukey foi realizado por comparações entre os grupos distintos. A distância dos quatro grupos foi medida usando o número de genes diferencialmente expressos como a medida da distância. Em outras palavras, quanto maior o número de genes diferencialmente expressos entre um tipo de amostra e todas as outras três, mais distante a amostra foi situada e conseqüentemente mais molecularmente divergente. Estes resultados mostraram que entre os quatro grupos de amostras, como esperado, as células epiteliais do tecido não-neoplásico apresentou um perfil de expressão gênica mais distinto. Duzentos e vinte e um genes dos 764 foram diferencialmente expressos entre amostras não neoplásicas e tumorais, enquanto DCIS puro, DCIS/IDC e IDC comparado com todos os outros três reportaram 12, 37 e 6 genes dos 764, respectivamente.

Para analisar o nível de similaridade molecular entre os três tipos de lesão de carcinoma ductal, uma nova análise foi realizada excluindo o grupo de amostra não-neoplásico. O teste ANOVA corrigido por pFDR ( $<0,01$ ) identificou 90 genes entre DCIS puro, DCIS/IDC e IDC. A lista dos 90 genes

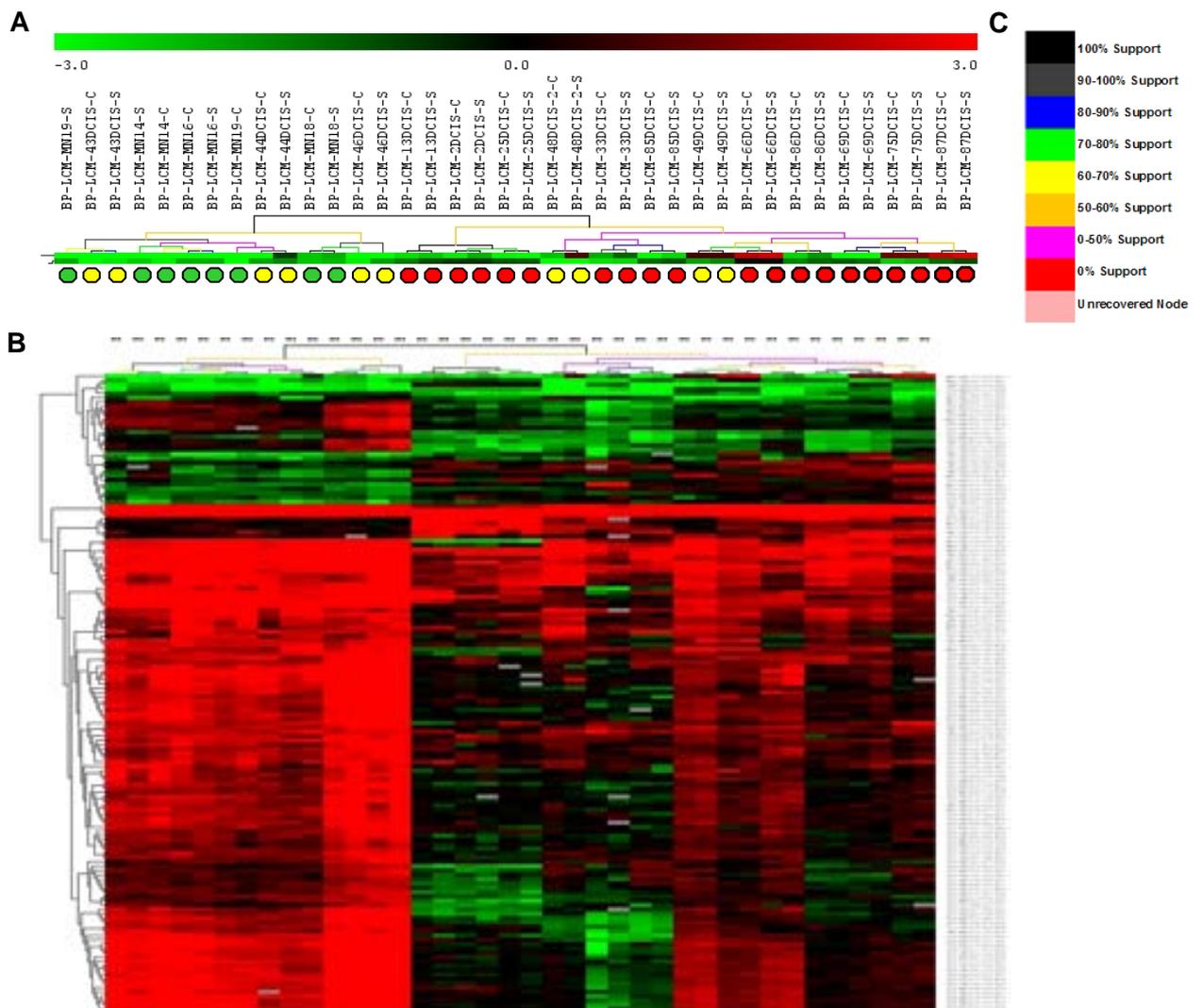


um critério mais estridente foi adotado incluindo uma diferença no nível de expressão de no mínimo 2 vezes. Comparações 2 a 2 entre os grupos de amostras mostraram 8, 215 e 161 genes diferencialmente expressos entre N x DCIS puro, N x DCIS/IDC e DCIS puro x DCIS/IDC, respectivamente. Em seguida, os genes foram representados em um Diagrama de Venn para identificar aqueles diferencialmente expressos comuns em mais de uma comparação. Foram considerados separadamente os genes com maior e menor expressão (Figura 24). Um grupo de 147 genes foi selecionado, seguindo a suposição citada acima e classificado como genes que podem estar envolvidos nos estágios iniciais do processo de progressão do DCIS podendo predizer invasão antes da alteração morfológica. A lista dos 147 genes pode ser visualizada no Anexo 11.



**Figura 24** - Representação de diagrama de Venn. Identificação dos genes supostamente envolvidos nos estágios iniciais do processo de progressão do DCIS.

O agrupamento hierárquico baseado no padrão de expressão dos 147 genes resultou em dois ramos principais. Amostras não-neoplásicas e 60% das DCIS puro foram discriminadas do grupo DCIS/IDC em 100% (Figura 25). A habilidade deste grupo de genes em gerar um perfil de expressão gênica capaz de segregar relativamente bem as duas lesões morfológicamente similares (DCIS puro e DCIS/IDC) e agrupar lesões DCIS puro juntamente com amostras não-neoplásicas, sugere que neste grupo de genes há representantes importantes envolvidos nas alterações moleculares precoces do processo de progressão do carcinoma ductal de mama.



**Legenda:** **A:** Dendrograma das amostras não-neoplásicas (n=4), DCIS puro (n=5) e DCIS/IDC (n=10). Os círculos verdes indicam amostras não neoplásicas; amarelas indicam DCIS puro e vermelhas DCIS/IDC. **B:** Representação do *cluster* mostrado em (A) baseada no grupo de 147 genes **C:** Legenda do suporte do cluster.

**Figura 25** - Clusterização hierárquica não supervisionada.

### 4.3.3 Caracterização Funcional dos Genes Diferencialmente Expressos

Com a finalidade de identificar aspectos funcionais do grupo de genes diferencialmente expressos, este grupo de genes foi classificado quanto aos processos biológicos através do *Gene Ontology Consortium Annotation* (GO) (<http://vortex.cs.wayne.edu:8080/index.jsp>) e a representatividade foi avaliada usando a distribuição qui-quadrada do *Onto-Tools* (KHATRI et al. 2005) desenvolvido pelo Laboratório de Bioinformática e Sistema Inteligente do Departamento de Ciências da Computação da Universidade do Estado de Wayne (<http://vortex.cs.wayne.edu:8080>). Os processos biológicos resposta de defesa celular, fusão de membrana, endocitose mediada por receptor, via de sinalização de receptor transmembrânico de proteína tirosina quinase foram observados como processos biológicos estatisticamente significantes no grupo de genes classificados como relacionados ao processo de progressão do DCIS (Tabela 9).

**Tabela 9** - Análise funcional dos 147 genes diferencialmente expressos entre amostras não-neoplásicas, DCIS puro e DCIS/IDC selecionados como genes envolvidos no processo de progressão do DCIS.

GO	Processos	pFDR	Genes	Anotação
Processo Biológico	Resposta de defesa celular	0,006	<i>CXCL9</i>	<i>chemokine (C-X-C motif) ligand 9</i>
			<i>MAPRE2</i>	<i>microtubule-associated protein, RP/EB family, member 2</i>
			<i>C3AR1</i>	<i>complement component 3a receptor 1</i>
	Fusão de Membrana	0,014	<i>STX11</i>	<i>syntaxin 11</i>
			<i>GOSR2</i>	<i>golgi SNAP receptor complex member 2</i>
			<i>RABEPK</i>	<i>Rab9 effector protein with kelch motifs</i>
	Endocitose mediada por receptor	0,032	<i>CLTCL1</i>	<i>clathrin, heavy chain-like 1</i>
			<i>KDR</i>	<i>kinase insert domain receptor (a type III receptor tyrosine kinase)</i>
	Via de sinalização receptor transmembrânico de proteína tirosina quinase	0,047	<i>NCOA4</i>	<i>nuclear receptor coactivator 4</i>
			<i>LOX</i>	<i>lysyl oxidase</i>

#### 4.3.4 Confirmação dos genes selecionados como supostamente envolvidos na progressão do DCIS por RT-PCR em tempo real

Visando avaliar o rigor dos resultados encontrados pela metodologia de *microarray* e evitar a escolha de falsos candidatos, foram selecionados aleatoriamente 10 genes da lista dos 147 (*CGI-41*, *C16orf5*, *GOSR2*, *MARK3*, *LOX*, *STK25*, *SULF-1*, *TXNL2*, *NEDD9* e *STX11*). Destes, dois genes foram excluídos das análises por não apresentarem amplificação de fragmento de tamanho específico, restando assim, oito para serem avaliados por qRT-PCR. Utilizando o critério de diferença de expressão  $\geq 2$ , 5 dos 8 genes avaliados por qRT-PCR (62,5%) mostraram concordância em ambas metodologias (Tabela 10). Maior expressão no DCIS puro em comparação ao DCIS/IDC foi encontrado para os genes *C16orf5*, *GOSR2* e *TXNL2*, enquanto que menor expressão foi vista para o gene *LOX* e *SULF-1*.

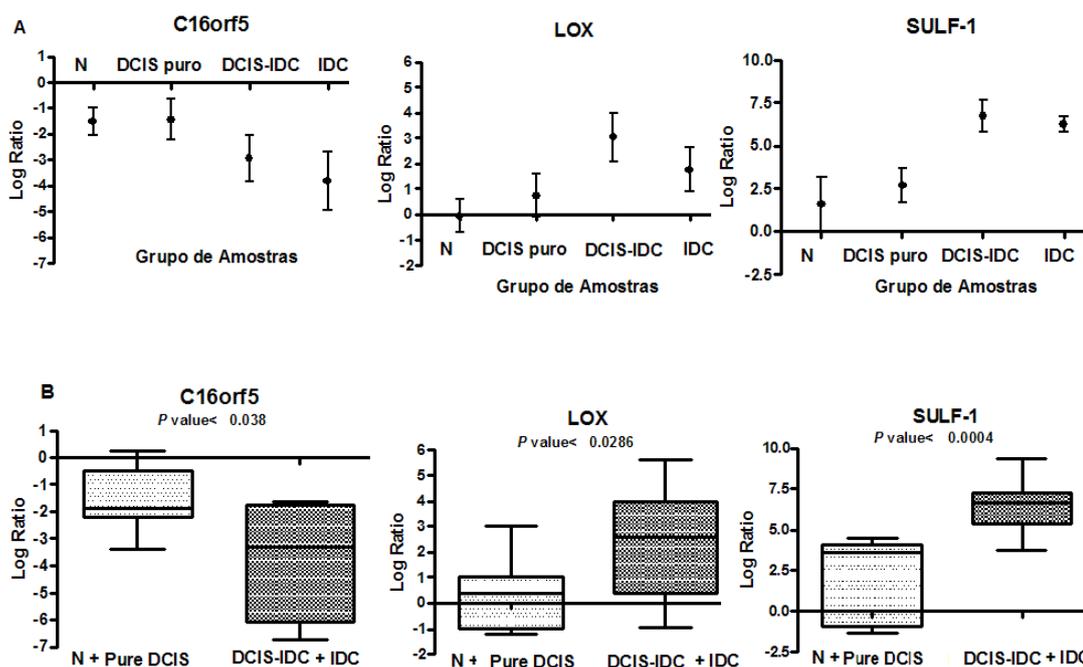
O nível de expressão dos 5 genes foi avaliado e o comportamento da expressão observado nos 4 grupos de amostras (Não-neoplásico, DCIS puro, DCIS/IDC e IDC). Destes 5 genes, 3 (*C16orf5*, *LOX* e *SULF1*) apresentaram um comportamento de expressão interessante, mostrando maior diferença apenas entre as lesões morfológicamente similares (DCIS puro e DCIS/IDC) (Figura 26a). Estes 3 genes supostamente envolvidos no processo de progressão do DCIS apresentaram as maiores diferenças de expressão entre DCIS puro e DCIS/IDC e pouca ou nenhuma entre não-neoplásico e DCIS puro, assim como entre DCIS/IDC e IDC. Apesar das diferenças morfológicas entre os dois primeiros e os dois últimos tipos histológicos, o nível de expressão gênica mostrou-se semelhante. Para os 3

genes, o grupo de amostras não-neoplásica e DCIS puro e o grupo DCIS/IDC e IDC foram agrupados e o nível de expressão gênica comparado, mostrando diferença estatística significativa entre eles (Figura 26b).

**Tabela 10** - Genes selecionados randomicamente e avaliados por qRT-PCR com suas respectivas diferenças no nível de expressão tanto no *microarray* como no qRT-PCR.

	<b>Nome do gene</b>	<b>Microarray (<i>Fold</i>)</b>	<b>qRT-PCR (<i>Fold</i>)</b>
<b>Genes Alvo</b>	<i>CGI-41</i>	5,3	1,0
	<i>C16orf5</i>	5,5	2,4
	<i>GOSR2</i>	4,0	2,0
	<i>MARK3</i>	4,9	-1,5
	<i>LOX</i>	-2,0	-7,0
	<i>STK25</i>	2,0	-2,4
	<i>SULF1</i>	-4,0	-11
	<i>TXNL2</i>	4,2	2,0

Valores positivos indicam maior expressão no DCIS puro, enquanto que valores negativos demonstram maior expressão no DCIS/IDC.



**Legenda:** **A:** Comportamento dos genes mostrando maior diferença apenas entre as lesões morfológicamente similares (DCIS puro e DCIS/IDC) pelo teste ANOVA. **B:** Significância estatística entre os 2 grupos [Não-neoplásica (N) + DCIS puro e *In situ* associado ao IDC (DCIS/IDC) + IDC] por qRT-PCR.

**Figura 26** - Comportamento dos genes entre as amostras que mimetizam a progressão do carcinoma ductal.

Parte desses resultados foram submetidos e aceitos com correções na revista Breast Cancer Research.

#### 4.3.5 Resultados Referentes a Plataforma Contendo Genes das Vias de Sinalização WNT, PI3K e Processo EMT (LV-01)

##### A Amostras e Qualidade das Lâminas

Neste estudo foi utilizado praticamente o mesmo grupo de amostras não-neoplásicas, carcinoma ductal *in situ* puro, *in situ* associado ao IDC e

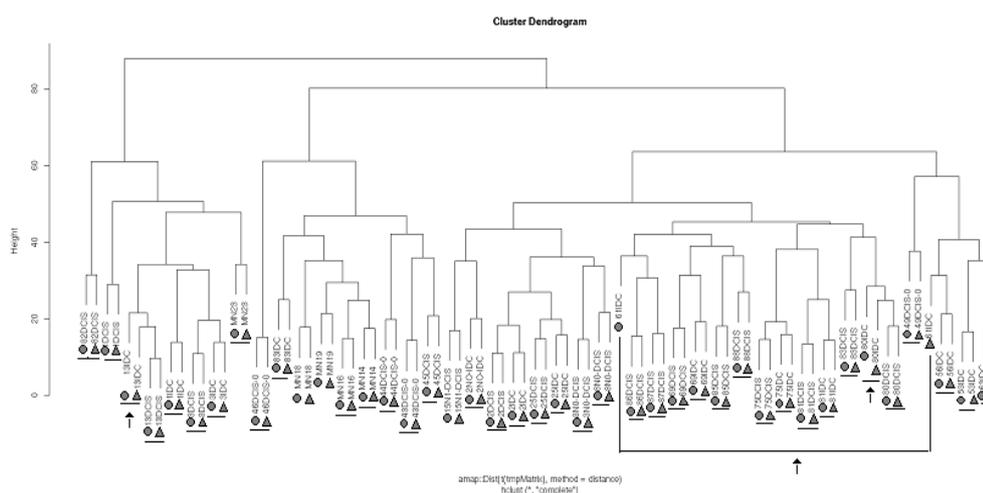
carcinoma ductal invasivo (IDC) de mama selecionado na primeira etapa do projeto.

Das 46 amostras utilizadas na primeira parte, 14 (6 pares e 2 DCIS puro) não apresentaram quantidade suficiente de RNA amplificado para hibridização. Para 9 destas 14, a quantidade de tecido para nova captura de células foi insuficiente não podendo ser microdissecadas novamente, restando 32 amostras para hibridização nessa plataforma (LV-01) e 5 para nova microdissecção.

Dessa forma, nova busca foi realizada no Banco de Tumores. Treze amostras foram selecionadas para nova microdissecção, sendo 6 componente intraductal de invasivo e 6 IDC e somente 1 amostra de tecido não neoplásico. Embora tenham sido realizadas buscas periódicas por carcinoma ductal *in situ* puro no banco, nenhuma amostra foi encontrada até o final desse trabalho, o que é de certa forma esperado por se tratar de uma lesão muito pequena e todo material ser usado para fins diagnósticos.

Das 18 amostras (5 antigas e 13 novas) microdissecadas a *laser*, 2 não apresentaram RNA total de qualidade aceitável e a quantidade de RNA amplificado foi insuficiente para 5 das amostras, restando 11 para marcação e hibridização (Anexo 2). Das amostras hibridizadas apenas uma foi retirada na análise, totalizando 42 amostras (32 antigas e 10 novas). Assim, nesta segunda fase, o grupo de amostras foi dividido em: 5 amostras de tecido não neoplásico (borda de fibroadenoma), 10 casos de amostras pareadas (10 DCIS e 10 IDC), 4 DCIS puro, 9 *in situ* associado ao IDC (DCIS/IDC) e 4 IDC somente (Anexo 6).

Todas as amostras marcadas e hibridizadas apresentaram boa qualidade com média de correlação de *Pearson* entre controle e *swap* de 0,90, exceto uma que foi excluída da análise. O dendograma das lâminas controle e *swap* podem ser visualizados na Figura 27.



**Legenda:** Os círculos e triângulos com tracejado abaixo indicam lâminas controle e *swap* do mesmo caso respectivamente. Três amostras (controle e *swap*) não clusterizadas dentro da mesma subdivisão aparecem indicadas pelas setas, sendo que somente uma foi excluída da análise por apresentar correlação de *Pearson* inferior a 0,9.

**Figura 27** - Dendograma das lâminas controle e *swap* das amostras hibridizadas na plataforma contendo as vias de sinalização WNT, PI3K e processo EMT.

## B Identificação de Genes Diferencialmente Expressos entre os Componentes DCIS e IDC Presentes na Mesma Amostra

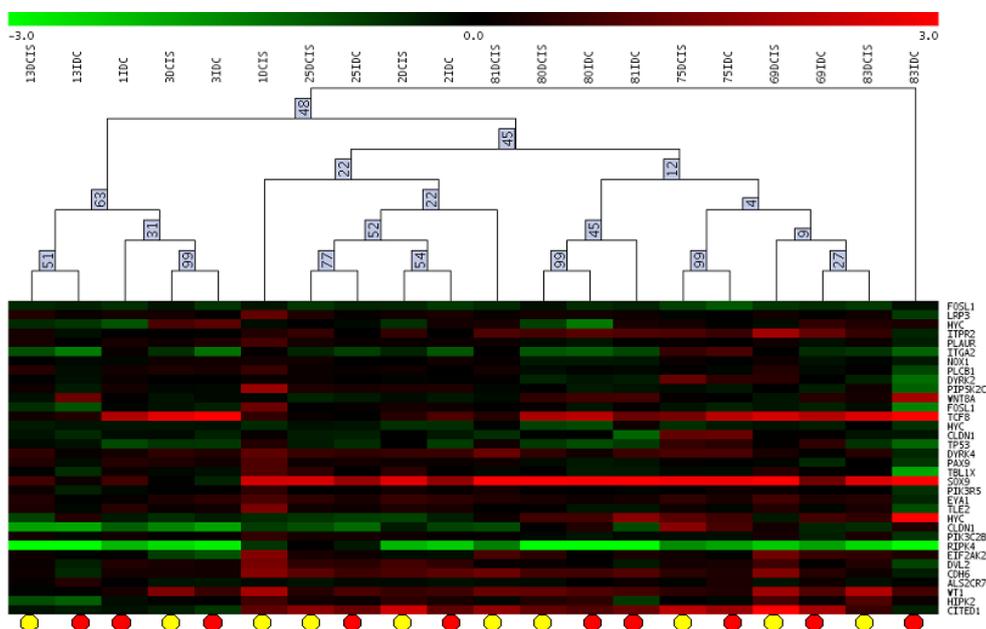
Para se determinar a diferença de expressão entre carcinoma ductal *in situ* e invasivo (DCIS x IDC) que coexistem na mesma amostra, comparamos 10 casos de amostras pareadas nesta plataforma. Aplicamos o teste T de *Student* pareado para todos os genes do conjunto de dados utilizando como critério  $p < 0,05$ .

A comparação entre os componentes DCIS e IDC resultou na seleção de 32 genes diferencialmente expressos pertencentes as vias de sinalização WNT, PI3K e processo EMT, apresentando expressão diferencial não superior a 1,61 vezes para os 26 genes mais expressos em DCIS e 1,64 vezes para os 4 mais expressos em IDC, sendo que 2 deles (*TCF8* e *PLCB1*) foram redundantes nas vias de sinalização. A lista dos 30 genes pode ser visualiza no Anexo 12. Dentre as vias presentes na lâmina, aquela que apresentou um maior percentual de genes diferencialmente expressos (11/109) foi a PI3K (Tabela 11).

**Tabela 11** - Representação dos genes diferencialmente expressos (DE) pertencentes as vias de sinalização estudas nas amostras pareadas (DCIS x IDC).

VIAS	EMT	Fold	PI3K	Fold	WNT	Fold
Genes	<i>CDH6</i>	1,35	<i>DYRK4</i>	1,29	<i>CITED1</i>	1,61
	<i>EYA1</i>	1,25	<i>ALS2CR7</i>	-1,14	<i>PLCB1</i>	1,24
	<i>NOX1</i>	1,11	<i>HIPK2</i>	1,26	<i>LRP3</i>	1,20
	<i>TCF8</i>	-1,36	<i>PIK3R5</i>	1,17	<i>WT1</i>	1,41
	<i>PAX9</i>	1,22	<i>PLCB1</i>	1,24	<i>FOSL1</i>	1,32
	<i>ITGA2</i>	1,31	<i>EIF2AK2</i>	1,35	<i>SOX9</i>	1,61
			<i>PIP5K2C</i>	1,36	<i>TP53</i>	1,30
			<i>ITPR2</i>	1,31	<i>TBL1X</i>	1,43
			<i>RIPK4</i>	1,57	<i>TCF8</i>	-1,36
			<i>DYRK2</i>	1,20	<i>TLE2</i>	1,31
			<i>PIK3C2B</i>	1,22	<i>WNT8A</i>	-1,43
					<i>CLDN1</i>	1,12
					<i>DVL2</i>	1,34
					<i>MYC</i>	-1,64
				<i>PLAUR</i>	1,14	
<b>Total de genes na plataforma:</b>	84		109		197	
<b>Genes DE na via:</b>	6		11		15	
<b>Representatividade (%):</b>	7,143		10,092		7,614	

Um agrupamento hierárquica baseado no nível de expressão desses 30 genes foi realizado visando verificar se o padrão de expressão baseado neste grupo de genes poderia discriminar as amostras de morfologia diferente. O cluster obtido não apresentou separação entre as amostras mostrando alta similaridade no perfil de expressão gênica entre os dois tipos histológicos, mesmo considerando que os genes representados nessa plataforma estão mais intimamente envolvidos com o mecanismo de invasão e se mostraram diferencialmente expressos entre os dois grupos (Figura 28). Esses dados confirmam os resultados obtidos com a plataforma 4.8K, mostrando uma grande semelhança do padrão molecular entre as 2 lesões de morfologia diferentes (DCIS e IDC).



**Legenda:** Os círculos amarelos e vermelhos indicam amostras DCIS e IDC respectivamente.

**Figura 28** - Clusterização hierárquica utilizando o grupo de 30 genes diferencialmente expressos entre DCIS e IDC (grupo pareado) mostrando a robustez da segregação com o TMEV.

## **C Identificação de Genes Candidatos a Estarem Envolvidos nas Primeiras Etapas de Progressão de Carcinoma Ductal *in situ***

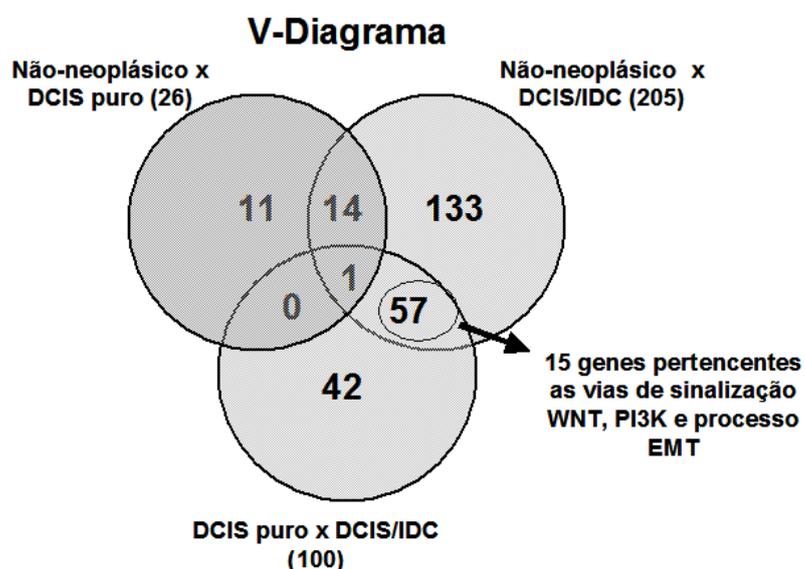
Para identificar os genes candidatos a estarem envolvidos nas primeiras etapas de progressão do DCIS foi avaliado a expressão gênica de amostras não-neoplásicas, DCIS puro, DCIS/IDC.

Para seleção de genes que pertencem às vias de sinalização WNT, PI3K e processo EMT foi aplicado dessa vez o teste não paramétrico de *Wilcoxon* considerando como genes diferencialmente expressos aqueles que apresentaram nível de significância  $alpha < 0,05$  (Anexo 13). O teste foi realizado para a combinação dois a dois de 2 tipos histológicos e amostras não-neoplásicas de interesse, gerando 3 listas de genes diferencialmente expressos (N x DCIS puro; N x DCIS/IDC e DCIS puro x DCIS/IDC). Baseados na mesma suposição citada anteriormente de que genes diferencialmente expressos entre células epiteliais que apresentam capacidade de invadir tecidos adjacentes (DCIS/IDC) subtraindo aqueles diferencialmente expressos entre lesões que não apresentam essa capacidade (N ou DCIS puro), representariam os genes supostamente envolvidos no processo de progressão do DCIS, que é a aquisição da capacidade de invadir.

Assim, para selecionar aqueles mais robustos, foi adotado como critério uma diferença no nível de expressão de no mínimo 2 vezes entre as comparações. Comparações 2 a 2 entre os grupos de amostras mostraram 26, 205 e 100 genes diferencialmente expressos entre N x DCIS puro, N x DCIS/IDC e DCIS puro x DCIS/IDC, respectivamente, considerando todos os

genes presentes na plataforma. Em seguida, uma análise de Diagrama de Venn foi realizada, considerando separadamente os genes menos e mais expressos visando identificar os genes comuns diferencialmente expressos (Figura 30). Um grupo de 57 genes foi selecionado, seguindo a suposição citada acima e classificado como genes candidatos que podem estar envolvidos nos estágios iniciais do processo de progressão do DCIS. Destes 57 (plataforma LV-01), somente um gene (*ITPKC* - *inositol 1,4,5-trisphosphate 3-kinase C*) apresentou sobreposição com os 147 genes (plataforma 4.8k)

Dessa lista de 57 genes (Anexo 14), foram selecionados aqueles que pertencem as vias de sinalização WNT, PI3K e processo EMT, resultando em 15 genes apresentados na Tabela 12. Essa comparação mostrou um maior percentual de genes pertencentes ao processo EMT com expressão alterada.



**Legenda:** Identificação dos genes supostamente envolvidos nos estágios iniciais do processo de progressão do DCIS.

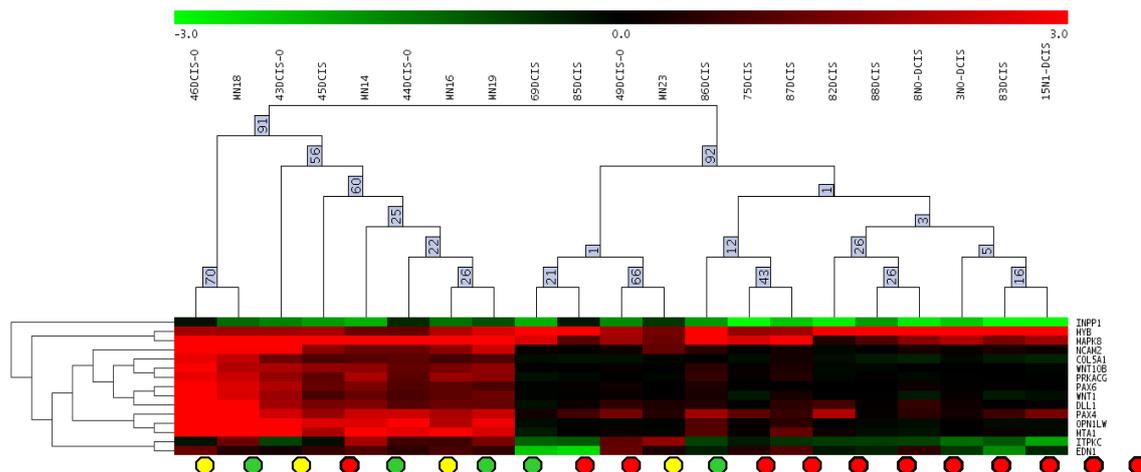
**Figura 29** - Representação de diagrama de Venn.

**Tabela 12** - Representação dos genes diferencialmente expressos (DE) pertencentes ao processo EMT e vias de sinalização PI3K e WNT estudadas nas amostras de progressão.

Vias	EMT	Fold	PI3K	Fold	WNT	Fold
Genes	<i>COL5A1</i>	2,52	<i>ITPKC</i>	4,95	<i>MAPK8</i>	3,71
	<i>EDN1</i>	2,48	<i>INPP1</i>	2,62	<i>MTA1</i>	4,43
	<i>NCAM2</i>	3,04			<i>OPN1LW</i>	5,28
	<i>PAX4</i>	2,76			<i>WNT1</i>	2,63
	<i>DLL1</i>	2,61			<i>PRKACG</i>	2,67
	<i>PAX6</i>	3,17			<i>WNT10B</i>	2,63
					<i>MYB</i>	-3,16
<b>Total de genes DE no grupo:</b>	6		2		7	
<b>Total de genes no grupo:</b>	84		109		197	
<b>Representatividade (%):</b>	7,14		1,84		3,55	

Valores positivos indicam maior expressão no tecido não-neoplásico ou DCIS puro, enquanto que valores negativos demonstram maior expressão no DCIS/IDC.

Um agrupamento hierárquico baseado no padrão de expressão dos 15 genes resultou em dois ramos principais. Oitenta por cento das amostras não-neoplásicas e 80% das DCIS puro foram discriminadas do grupo DCIS/IDC, sendo que este último agrupou corretamente para 91,7% das amostras (Figura 30). A habilidade deste grupo menor de genes (15 genes) em gerar um perfil de expressão gênica capaz de segregar relativamente bem as duas lesões morfológicamente similares evidenciam o potencial de marcador molecular para prever presença de componente invasor em lesões *in situ*.



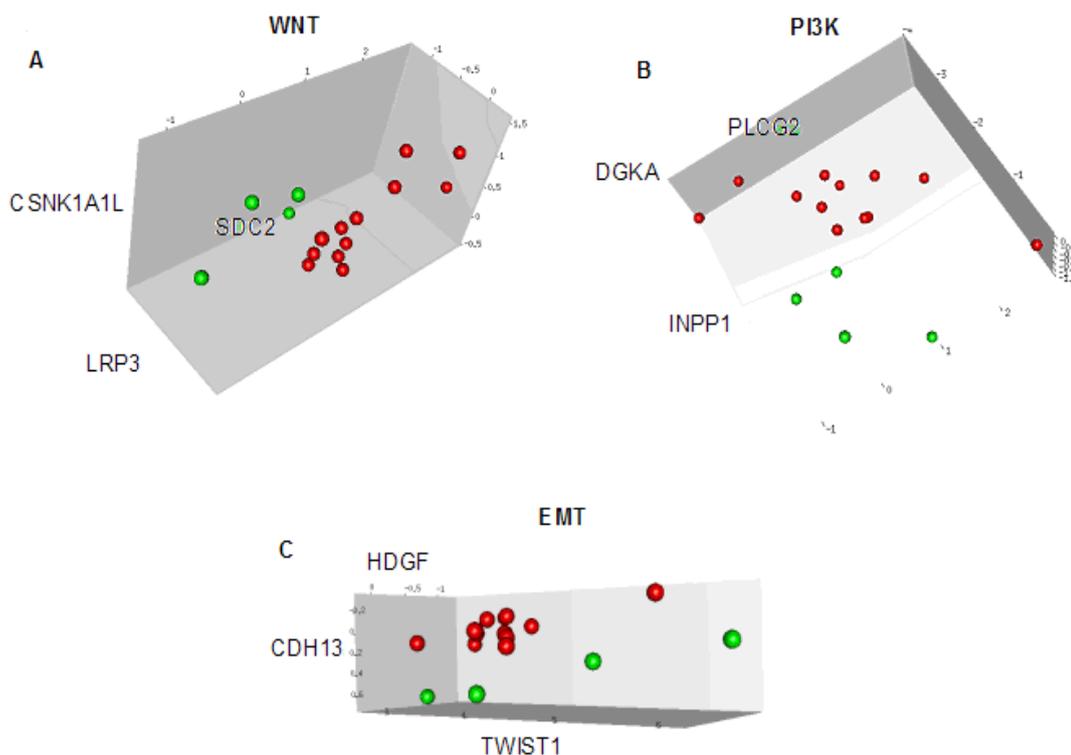
**Legenda:** Dendograma das amostras não-neoplásicas (n=5), DCIS puro (n=4) e DCIS/IDC (n=12) utilizando o grupo de 15 genes mostrando a robustez da segregação com o TMEV. Os círculos verdes indicam amostras não neoplásicas; amarelas indicam DCIS puro e vermelhas DCIS/IDC.

**Figura 30** - Clusterização hierárquica não supervisionada.

## D Classificadores

A busca por classificadores foi realizada dentro de cada grupo de genes pertencentes às vias de sinalização WNT, PI3K e processo EMT representadas na plataforma LV-01, visando separar perfeitamente os grupos de amostras DCIS puro e DCIS/IDC.

A separação dos grupos pelo melhor trio classificador construído utilizando genes pertencentes a via de sinalização WNT, via PI3K e o processo EMT, apresentaram os trios de genes *CSNK1A1L*, *LRP3*, *SDC2*, o trio *PLCG2*, *INPP1*, *DGKA* e o *HDGF*, *CDH13*, *TWIST1*, respectivamente. Todos eles podem ser visualizados na Figura 31.



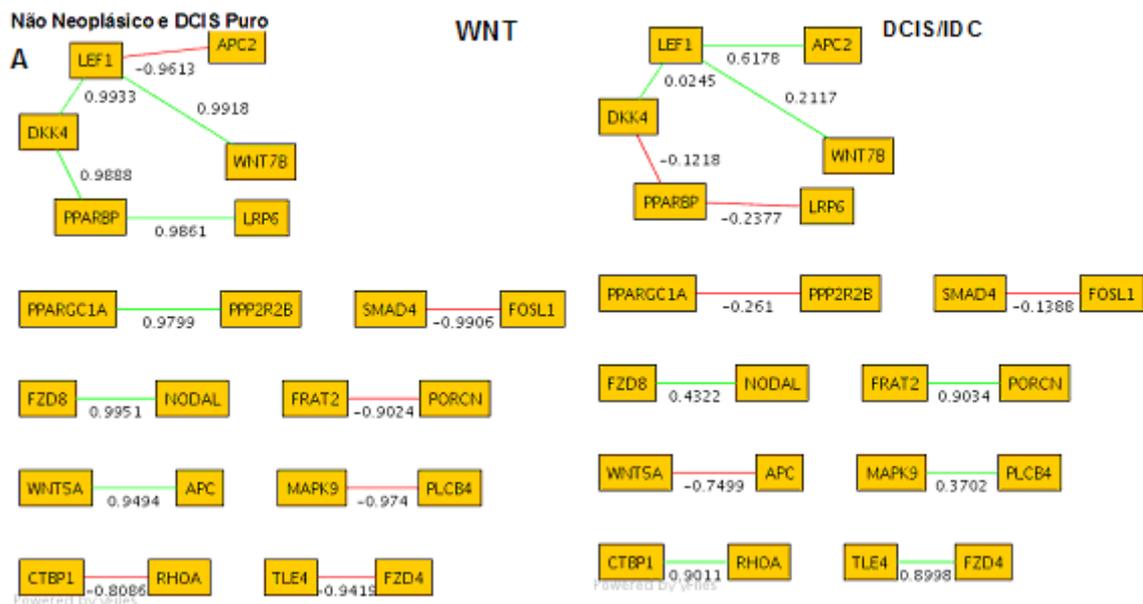
**Legenda** - **A:** Trio classificador da via de sinalização WNT; **B:** Trio classificador da via de sinalização PI3K **C:** Trio classificador do processo EMT. Os círculos verdes indicam amostras DCIS puro e as vermelhas DCIS/IDC.

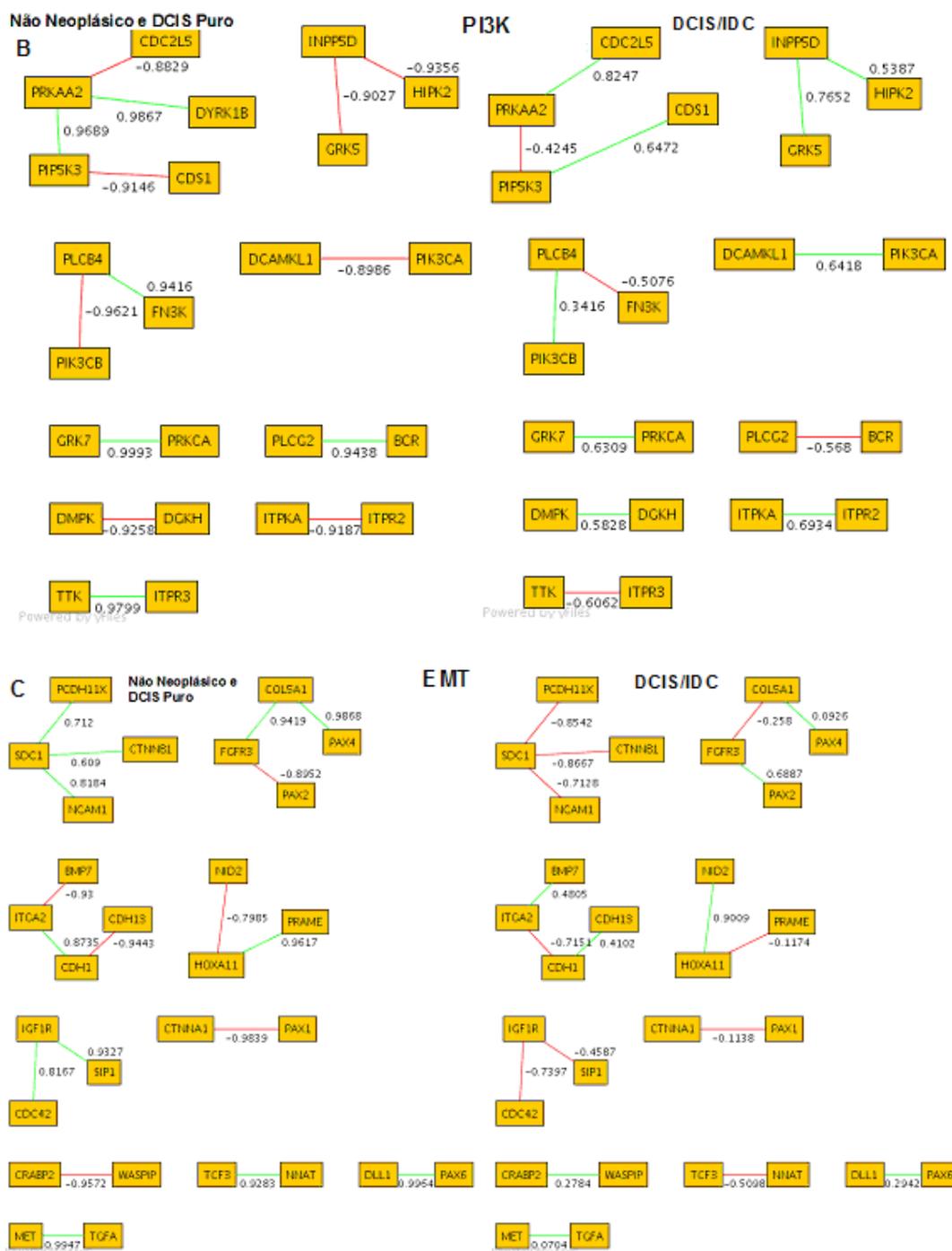
**Figura 31** - Trios classificadores encontrados utilizando os grupo de genes pertencentes às vias de sinalização WNT, PI3K e processo EMT representadas na plataforma visando separar as amostras DCIS puro e DCIS/IDC.

## E Redes de Relevância

Análise de dados baseada em hipóteses se relaciona a muitas das abordagens empregadas em análise de expressão gênica. Isto significa que, uma hipótese com base no conhecimento disponível pode ser postulada se um gene candidato mostra expressão alterada. Analisando-se hipóteses baseadas em dados, é possível identificar genes ou vias metabólicas até

então não associadas ao fenômeno biológico em estudo. Para isso, o conceito de redes de relevância foi empregado e pares de genes cujo padrão de expressão gênica mostrava-se linearmente correlacionado foram identificados. Após avaliar a significância das diferenças de correlação, foram identificados pares de genes cujos coeficientes de correlação foram significativamente alterados ( $p \leq 0,0005$ ) na comparação entre os grupos não-neoplásico e DCIS puro versus DCIS/IDC (Figura 32).





**Legenda:** Pares de genes representados cuja alteração na correlação entre as amostras Não-neoplásica e DCIS puro contra o *in situ* associado ao IDC foi significativa ( $p$ -valor  $\leq 0,05$ ). Correlação positiva está representada por linhas verdes, e correlação negativa por linhas vermelhas. **(A):** Via de Sinalização WNT. **(B):** Via de Sinalização PI3K e **(C):** Processo EMT.

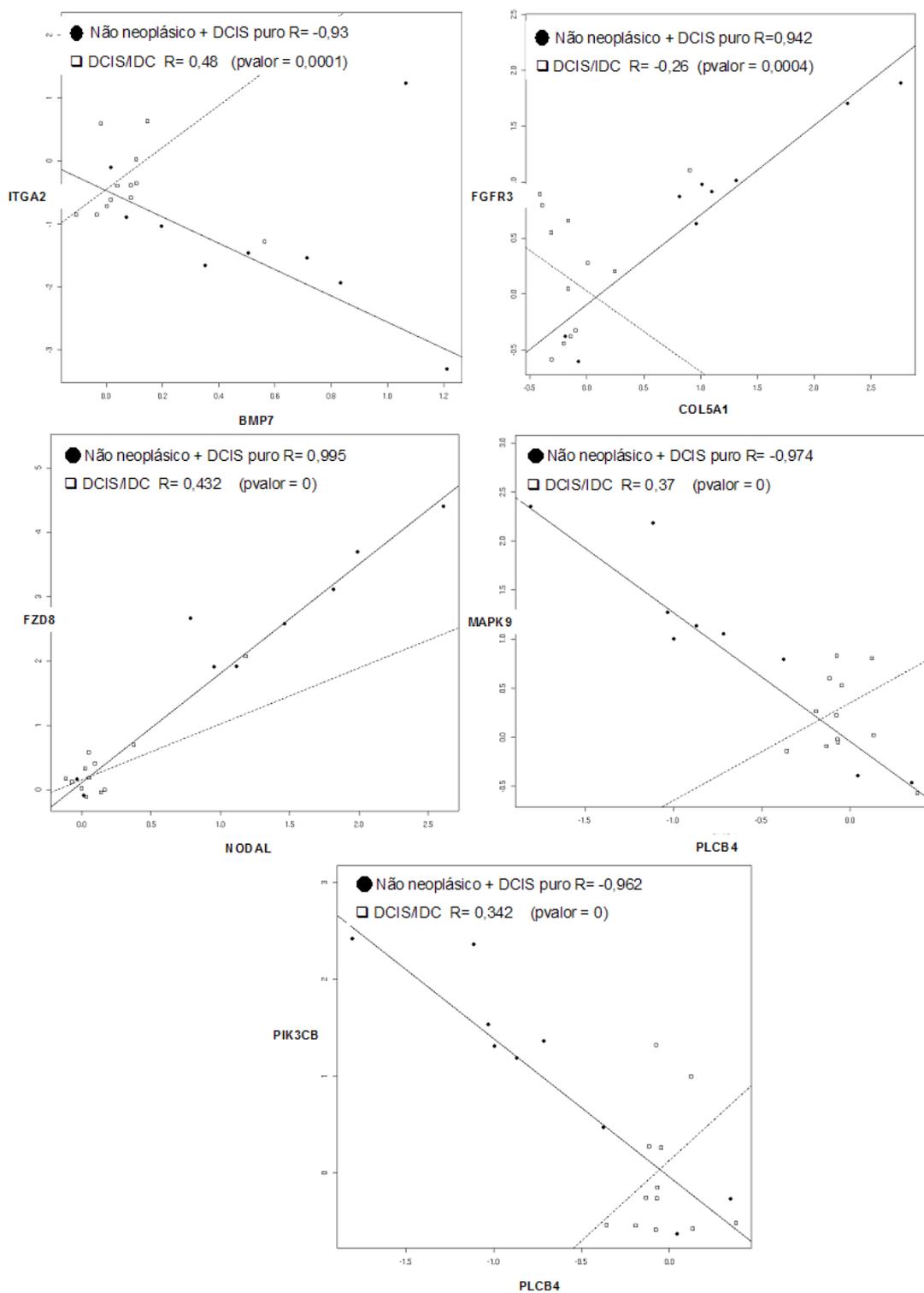
**Figura 32** - Rede de relevância para o grupo de amostras não-neoplásico e DCIS puro versus DCIS/IDC.

Para se evidenciar melhor estas alterações de correlação entre os pares gênicos, foi construído um gráfico *scatter-plot's* para aqueles pares que apresentaram coeficientes de correlação superiores a 0,9 em pelo menos um dos grupos para a busca das combinações de genes mais promissoras. Nestes gráficos é possível visualizar a distribuição das amostras não neoplásicas e DCIS puro em relação às DCIS/IDC de acordo com o valor de expressão dos pares gênicos. Os gráficos podem ser visualizados no Anexo 15 e os coeficientes de correlação na Tabela 13.

**Tabela 13** - Pares gênicos apresentando coeficientes de correlação linear superiores a 0,9.

<b>WNT</b>				
Pares Gênicos		Coeficiente de Correlação R		pvalor
		Não-neoplásico e DCIS puro	DCIS/IDC	
<i>FOSL1</i>	<i>SMAD4</i>	-0,991	-0,139	0,0000
<i>PLCB4</i>	<i>MAPK9</i>	-0,974	0,37	0,0000
<i>PPP2R2B</i>	<i>PPARGC1A</i>	0,98	-0,261	0,0000
<i>APC</i>	<i>WNT5A</i>	0,949	-0,75	0,0000
<i>PORCN</i>	<i>FRAT2</i>	-0,902	0,903	0,0000
<i>DKK4</i>	<i>PPARBP</i>	0,989	-0,122	0,0000
<i>DKK4</i>	<i>LEF1</i>	0,993	0,024	0,0000
<i>NODAL</i>	<i>FZD8</i>	0,995	0,432	0,0000
<i>PPARBP</i>	<i>LRP6</i>	0,986	-0,238	0,0000
<i>APC2</i>	<i>LEF1</i>	-0,961	0,618	0,0000
<i>RHOA</i>	<i>CTBP1</i>	-0,81	0,90	0,0000
<i>LEF1</i>	<i>WNT7B</i>	0,992	0,212	0,0000
<b>PI3K</b>				
Pares Gênicos		Coeficiente de Correlação R		pvalor
		Não-neoplásico e DCIS puro	DCIS/IDC	
<i>PLCB4</i>	<i>PIK3CB</i>	-0,962	0,342	0,0000
<i>PLCB4</i>	<i>FN3K</i>	0,942	-0,508	0,0001
<i>ITPR2</i>	<i>ITPKA</i>	-0,919	0,693	0,0000
<i>INPP5D</i>	<i>GRK5</i>	-0,903	0,765	0,0000
<i>INPP5D</i>	<i>HIPK2</i>	-0,936	0,539	0,0001
<i>PIP5K3</i>	<i>PRKAA2</i>	0,969	-0,425	0,0000
<i>PIP5K3</i>	<i>CDS1</i>	-0,915	0,647	0,0000
<i>DYRK1B</i>	<i>PRKAA2</i>	0,987	0,033	0,0000
<i>ITPR3</i>	<i>TTK</i>	0,98	-0,606	0,0000
<i>BCR</i>	<i>PLCG2</i>	0,944	-0,568	0,0000
<i>DGKH</i>	<i>DMPK</i>	-0,926	0,583	0,0001
<i>PRKCA</i>	<i>GRK7</i>	0,999	0,631	0,0000
<b>EMT</b>				
Pares Gênicos		Coeficiente de Correlação R		pvalor
		Não-neoplásico e DCIS puro	DCIS/IDC	
<i>WASPIP</i>	<i>CRABP2</i>	-0,957	0,278	0,0001
<i>BMP7</i>	<i>ITGA2</i>	-0,93	0,48	0,0001
<i>NID2</i>	<i>HOXA11</i>	-0,798	0,901	0,0000
<i>NNAT</i>	<i>TCF3</i>	0,928	-0,51	0,0001
<i>CDH13</i>	<i>CDH1</i>	-0,944	0,41	0,0001
<i>COL5A1</i>	<i>FGFR3</i>	0,942	-0,258	0,0004
<i>COL5A1</i>	<i>PAX4</i>	0,987	0,093	0,0000
<i>IGF1R</i>	<i>SIP1</i>	0,933	-0,459	0,0001
<i>PAX6</i>	<i>DLL1</i>	0,996	0,294	0,0000
<i>HOXA11</i>	<i>PRAME</i>	0,962	-0,117	0,0003
<i>TGFA</i>	<i>MET</i>	0,995	0,07	0,0000
<i>PAX1</i>	<i>CTNNA1</i>	-0,984	-0,114	0,0001

De todos os pares de genes apresentando coeficientes de correlação linear superiores a 0,9, cinco foram selecionados por apresentar uma melhor clusterização das amostras ao longo das retas. Para os pares *BMP7* e *ITGA2*, *PLCB4* e *MARK9*, *PLCB4* e *PIK3CB* observou-se uma correlação negativa de expressão do grupo não-neoplásico (N) e DCIS puro em relação ao DCIS/IDC, já para os pares *COL5A1* e *FGFR2*, *NODAL* e *FZD8* uma correlação negativa de expressão foi encontrada (Figura 33).



**Figura 33** - Gráficos *scatter-plot's* para os pares de genes que apresentaram uma melhor clusterização das amostras ao longo das retas entre amostras não-neoplásica e DCIS puro versus DCIS/IDC.

## 5 DISCUSSÃO

A invasão de células neoplásicas em tecidos adjacentes é o primeiro passo para a formação de metástase, promovendo uma alta taxa de mortalidade entre as mulheres com câncer de mama, de modo que a identificação destas alterações gênicas potencialmente envolvidas no processo de progressão tumoral permanece um desafio.

Nesse estudo, combinamos cDNA *microarray* e microdissecção a *laser* (LCM) para caracterizar o padrão de expressão gênica das células tumorais de diferentes tipos histológicos de carcinoma ductal *in situ* (DCIS) e invasivo (IDC) de mama e identificar alterações no perfil de genes envolvidos na transição do DCIS para IDC. Também analisamos o padrão geral de expressão de lesões que mimetizam a progressão do carcinoma ductal de mama, focando nas duas lesões de mesma morfologia (DCIS puro e DCIS coexistindo com o IDC). Essa avaliação foi realizada sob o aspecto morfológico das lesões, sendo que as diferenças na expressão desses dois tipos de lesões, foram observadas para identificar genes candidatos a prever progressão do carcinoma ductal *in situ*.

Para tanto, 2 plataformas diferentes foram utilizadas. Uma delas (4.8k) não apresentando nenhuma classe funcional de genes preferencialmente (Brentani et al. 2005). A outra plataforma contendo uma representação quase completa de genes pertencentes a vias de sinalização WNT, PI3K e processo EMT estando envolvidas no processo de

diferenciação celular e no câncer participam dos mecanismos de transformação, invasão e metástase.

Os carcinomas ductal *in situ* e invasivo são morfológicamente distintos, uma vez que o DCIS permanece restrito aos ductos e o IDC apresenta invasão em tecidos adjacentes. Neste estudo foi comparado o perfil de expressão gênica destas lesões presentes simultaneamente no mesmo tecido (amostras pareadas), sendo os genes selecionados como diferencialmente expressos candidatos importantes, devido a ausência, neste caso, da variabilidade biológica.

O uso da plataforma 4.8k permitiu a seleção de 33 genes diferencialmente expressos, sendo que a clusterização hierárquica baseada no padrão de expressão destes genes não foi capaz de separar os 2 grupos de lesões fenotipicamente distintos, demonstrando notável similaridade no aspecto geral do padrão de expressão gênica na transição do DCIS para o IDC. Este resultado corrobora com dados gerados em outros estudos que utilizaram abordagens diferentes, tais como perda de heterozigose (ZHUANG et al. 1995; AMARI et al. 2003), imunohistoquímica de alguns marcadores moleculares (OTTESEN et al. 2000; WARNBERG et al. 2001) e SAGE (PORTER et al. 2003), para avaliar o DCIS e o IDC. Também MA et al. (2003) e SCHUETZ et al. (2006) utilizando a tecnologia de *microarrays* e LCM, assim como, DALGIN et al. (2007) por análises de clusterização utilizando dados públicos de *microarrays* observaram resultados semelhantes. Portanto, a similaridade no perfil de expressão gênica entre estes tipos histológicos distintos presentes na mesma paciente, sugere uma

programação genética para invasão já nos estágios iniciais da doença antes da alteração morfológica.

Dentre os 33 genes com alterações no perfil de expressão gênica entre as amostras pareadas (DCIS e IDC), foram selecionados 8 genes para serem avaliados por análise quantitativa dos transcritos (qRT-PCR). Destes, 4 apresentaram expressão diferencial confirmada, sendo eles os genes *LUM*, *RDH-E2*, *CXCL13*, *FBN1*, *POSTN* e *GDF15*.

Dentre os encontrados como mais expressos no IDC, podemos citar o *LUM*, *RDH-D2*, *FBN1* e *POSTN*. Há evidências que o gene *LUM* (*lumican*) codifica a proteína mais abundante da família de proteínas proteoglicanas ricas em leucinas em câncer de mama (LEYGUE et al. 2000). Vários estudos têm mostrado que esta família de proteínas pode modular o comportamento celular, incluindo migração da célula e proliferação durante o crescimento tumoral (LEYGUE et al. 1998; LEYGUE et al. 2000). Estes resultados corroboram com os achados de PORTER et al. (2003) e SETH et al. (2003) sugerindo que o *lumican* pode desempenhar um importante papel na progressão tumoral de mama. Também o gene *RDHE2* (*epidermal retinal dehydrogenase 2*) que pertence a família SDR (*short-chain dehydrogenases/reductases*) e converte retinol em retinaldeído como o primeiro e limitante passo na via sintética do ácido retinóico, sendo esta via crítica na manutenção da homeostase epidérmica *in vivo* (MATSUZAKA et al. 2002; MATSUZAKA et al. 2004). O gene *FBN1* (*fibrillin 1*), codifica uma glicoproteína de matriz extracelular que serve como um componente estrutural das microfibrilas ligantes de cálcio e foi observada associação

deste gene com câncer de mama (CHEN et al. 2007). E o gene *POSTN* (*periostin*), também encontrado por SCHUETZ et al. (2006) com maior expressão no tipo histológico IDC, codifica para periostina que parece modular motilidade celular em carcinoma de ovário ao se ligar às integrinas  $\alpha\beta$ -3 e  $\alpha\beta$ -5 (GILLAN et al. 2002). Pode promover angiogênese em tumores de mama via maior regulação do *VEGF-R2* (SHAO et al. 2004) e sobrevivência em câncer de cólon pela ativação da via AKT ao se ligar à integrina  $\alpha\beta$ 3 (BAO et al. 2004). Polimorfismos avaliados no gene *POSTN* mostraram associação com suscetibilidade ao câncer de mama e associação a alguns fatores de prognóstico relacionados a doença (FÖRSTI et al. 2007).

Já os genes *CXCL13* e *GDF15* foram encontrados com maior expressão no grupo DCIS. O *CXCL13* [*chemokine (C-X-C motif) ligand 13*], codifica uma quimiocina envolvida em quimiotaxia, resposta inflamatória e via metabólica relacionada a interação de receptor citocina-citocina (via encontrada com significância estatística entre as amostras pareadas) (<http://www.genecards.org/index.shtml>). Quimiocinas (citocinas quimiotáticas) são pequenas proteínas ligantes de heparina que orientam o movimento circulatório de leucócitos em sítios de inflamação e injúria (CHARO e RANSOHOFF 2006). Mais recentemente, tem sido estabelecido que células tumorais fazem uso da sinalização por receptores de quimiocinas, em etapas chaves na iniciação e progressão do câncer primário e metastático (LUKER e LUKER 2005). Na via de sinalização mediada por integrinas, estas moléculas de adesão podem interagir e modular a função

de sinalização de receptores de tirosina quinase. Estes receptores por sua vez, podem afetar diretamente suas funções adesivas (CAVALLARO e CHRISTOFORI 2004). E o gene *GDF15* (*growth differentiation factor 15*) regula a manutenção e diferenciação do tecido e promove a supressão da proliferação celular em linhagem de carcinoma mamário (Abd El-Aziz et al. 2007).

Portanto, sugerimos que os genes *LUM*, *RDH-E2*, *CXCL13*, *FBN1*, *POSTN* e *GDF15* podem estar envolvidos na transição do DCIS para o IDC, sendo assim, candidatos a marcadores de prognóstico no carcinoma ductal de mama.

Dentre os 33 genes com alterações no perfil de expressão gênica entre as amostras pareadas (DCIS e IDC), foram selecionados dois deles com maior expressão nas amostras IDC para se verificar o nível de proteína pela metodologia de imunohistoquímica em *Tissue Microarray*.

Uma vez que ambos os genes (*LUM* e *CRABP2*) apresentaram maior expressão de RNAm em amostras invasivas nos dados de *microarray*, esperava-se que as proteínas codificadas por estes genes apresentassem correlações significativas com marcadores moleculares e/ou dados clínicos relacionados a um pior prognóstico.

Nossos resultados mostraram que a expressão da proteína *LUM* associou-se positivamente com a expressão das CKs 5/6, CK 14, CK8 e CK18 assim como a maior expressão do HER2.

A composição das citoqueratinas (CKs) é produto específico do tipo de epitélio e do grau de diferenciação celular. Cada tipo de epitélio

apresenta um padrão de expressão particular de citoceratinas. O epitélio do ducto mamário é complexo, constituído por uma camada de células mioepiteliais basais e número variável de camadas de células epiteliais luminais. As células mioepiteliais basais expressam CK 5/6, CK 14 e CK 17, enquanto as células luminais expressam CKs próprias dos epitélios simples, tais como CK 8, CK 18, CK 19 e CK 7. Nos carcinomas de mama as CKs expressas são geralmente aquelas do epitélio luminal, mas a expressão de CK de padrão mioepitelial (ou basal) indica diferenciação tumoral, associando-se a comportamento biológico mais agressivo (PEROU et al. 2000; SORLIE et al. 2001; VAN DE RIJN et al. 2002; NIELSEN et al. 2004; ABD EL-REHIM et al. 2005; FAN et al. 2006; STIEPCICH 2007).

A expressão aumentada da proteína HER2 nas pacientes com carcinomas linfonodo-negativos está associada à pior prognóstico, negatividade para os receptores de estrógeno e progesterona e menor resposta à quimio e hormonioterapia (YAMAUCHI et al. 2001a e b). Como este receptor está intimamente envolvido com a transmissão de sinais relacionados à proliferação celular, a divisão celular é estimulada, resultando em um crescimento celular acelerado, que contribui para o desenvolvimento e progressão do câncer.

Portanto, a expressão da proteína LUM apresentou associação positiva de expressão com marcadores moleculares relacionados a um pior prognóstico, assim como correspondência com os dados de expressão do RNAm encontrados por *microarrays*. Mas o fato de também apresentar

associação com CKs próprias de células luminais (bom prognóstico) dificulta fazer qualquer inferência.

Nossos dados mostraram que a expressão da proteína CRABP-2 associou-se positivamente com a expressão dos receptores hormonais ER e PR, CK8, CK18, com o subtipo luminal A (classificação baseada nas características de expressão dos receptores ER positivo e HER2 negativo), proteína p53 e negativamente com a CK14.

Estudos têm demonstrado maior sobrevivência entre mulheres com receptores hormonais (ER e PR) positivos quando comparadas aquelas que apresentam tumores que não expressam estes receptores (ANDERSON et al. 2001; COSTA et al. 2002; BARDOU et al. 2003; FISHER et al. 2004). Outro estudo observou que a expressão de ER e CRABP2 foi coordenadamente regulada, mostrando que a ativação de ER leva a transcrição de CRABP2 em células de câncer de mama ER positivas (LU et al. 2005; MCPHERSON et al. 2007). Da mesma maneira, BERTUCCI et al. (1999) encontraram maior expressão de *CRABP2* em linhagem celular de carcinoma mamário caracterizada pela presença de ER positivo. Contudo, apesar dos nossos resultados de IHQ não concordarem com os dados de *microarray*, maior expressão no componente invasor, o *CRABP2* parece ser um fator de bom prognóstico baseado na literatura e nos dados de proteína.

O carcinoma ductal *in situ* pode apresentar diferentes evoluções, progredindo rapidamente para carcinoma invasivo ou evoluindo lentamente durante um longo período da doença. Assim, o manejo clínico de pacientes

com DCIS permanece controverso e o desafio é prever o risco de progressão para a doença invasiva.

Dentro deste mesmo objetivo em buscar genes que apresentem alterações no nível de expressão e que possam estar relacionados ao processo de progressão de câncer de mama, comparamos o padrão transcricional de células do mesmo tipo histológico, mas que diferem pela presença ou não do componente invasor na mesma lesão. Foi encontrado um perfil de expressão gênica mais distinto nas amostras DCIS puro quando comparadas as outras lesões tumorais (DCIS/IDC e IDC), mesmo a lesão DCIS/IDC que apresenta mesma morfologia do DCIS puro. Este resultado sugere que alterações moleculares ocorrem de forma anterior as alterações morfológicas.

Na seleção de supostos genes envolvidos na progressão do DCIS, 147 genes foram considerados e utilizados nas análises de clusterização. Uma maior semelhança foi observada entre o perfil de expressão gênica da célula normal com DCIS puro do que em relação ao DCIS/IDC, tanto na comparação do padrão geral de expressão gênica como em relação ao número de genes diferencialmente expressos, sugerindo que essas lesões possuam características moleculares distintas. De fato, PARK et al. (2006) demonstraram que o DCIS puro difere do DCIS/IDC segundo o *status* de amplificação do HER2, sugerindo que a biologia das duas lesões é diferente. Outro estudo (STEINMAN et al. 2007) demonstrou que há um padrão de expressão diferencial para ER-*alpha*, PR e EGFR nos diferentes subtipos luminal (CK8, CK18) e basal (CK5/6, CK14, CK17) de carcinoma ductal de

mama, sugerindo que pelo menos algumas lesões DCIS puro são molecularmente distintas das DCIS/IDC.

Visando avaliar o rigor dos resultados encontrados na análise de *microarray*, oito genes da lista dos 147 supostamente envolvidos no processo de progressão do carcinoma ductal de mama foram selecionados randomicamente e avaliados por qRT-PCR.

Dentre os genes avaliados com expressão confirmada em ambas as metodologias nas amostras DCIS puro em relação as DCIS/IDC podemos citar com maior expressão os genes *C16orf5*, *GOSR2* e *TXNL2*, enquanto que menor expressão foi vista para os genes *LOX* e *SULF-1*. O *C16orf5* (*chromosome 16 open reading frame 5*) também conhecido como CDIP (*Cell death-inducing protein*) apresenta pouca informação disponível. O gene *GOSR2* (*golgi SNAP receptor complex member 2*) codifica uma proteína de tráfego de membrana e está envolvida em transdução de sinal e atividade de transporte. O *TXNL2* (*thioredoxin-like 2*) desempenha sua função em ligação de proteínas e está envolvido no processo de homeostase redox da célula. O *LOX* (*Lysyl Oxidase*) é um mediador da metástase das células humanas de câncer de mama em modelo de camundongo (ERLER et al. 2006) e também regula a migração de células de câncer de mama *in vitro* e formação de adesão célula-matriz pela regulação das quinases Scr e FAK (PAYNE et al. 2005). Nossos resultados corroboram com estes achados, uma vez que este gene apresentou maior expressão em DCIS/IDC quando comparado ao DCIS puro. Em relação ao gene *SULF-1* (*Sulfatase 1*), NISHIDATE et al. (2004) verificaram maior expressão em lesões IDC

quando comparado ao DCIS. Este gene inibe sinalização através de fatores de crescimento dependentes de heparina, diminui proliferação e facilita apoptose em resposta a estimulação exógena, estando também envolvido com a regulação do crescimento celular. A supressão da apoptose tem sido reconhecida como um contribuidor chave na tumorigênese em modelos animais e câncer humano. Nossos resultados mostraram uma maior expressão do *SULF-1* no DCIS/IDC em relação ao DCIS puro.

Buscando avaliar com mais detalhe estes cinco genes supostamente envolvidos no processo de progressão do DCIS, levantamos a suposição que estes não deveriam apresentar diferença de expressão entre o tecido normal e DCIS puro, uma vez que em ambos os grupos a capacidade de invasão não está estabelecida. Também não deveriam apresentar diferença de expressão entre DCIS/IDC e IDC, pois nossos dados mostram que apesar das diferenças morfológicas o DCIS/IDC coexiste com o carcinoma invasivo e deve, sustentado pelos nossos dados, apresentar perfil molecular semelhante onde a capacidade de invasão já está estabelecida no DCIS/IDC antes da alteração morfológica para o IDC.

Neste contexto, três genes (*C16orf5*, *LOX* e *SULF1*) confirmados por qRT-PCR mostraram diferença estatisticamente significantes ( $p < 0.05$ ) entre os grupos de amostras sem invasão (não-neoplásicas e DCIS puro) e aqueles com programação de invasão (DCIS/IDC e IDC).

Portanto, estes três genes descritos acima apresentam características que indicam seu suposto envolvimento na progressão do câncer de mama

sendo sugeridos como genes candidatos na predição de risco de progressão do carcinoma ductal *in situ*.

Em relação aos resultados obtidos entre as amostras pareadas (DCIS e IDC) utilizando a plataforma contendo genes pertencentes as vias de sinalização WNT, PI3K e processo EMT, temos uma lista de 30 genes ( $p < 0.05$ ). Dentre estas vias, a PI3K foi aquela com maior representatividade de genes entre os diferencialmente expressos.

Os efeitos biológicos da ativação da via PI3K relevantes na progressão do câncer podem ser de maneira geral categorizados em crescimento tumoral (proliferação celular, apoptose e senescência); angiogênese (produção e resposta a citocinas angiogênicas) e metástase (plasticidade do citoesqueleto, adesão celular, motilidade celular e invasão), segundo BRADER e ECCLES (2004). Esta via tem sido associada à sinalização de sobrevivência desempenhando um papel na tumorigênese mamária (SCHEID e WOODGETT 2001).

Para verificar o padrão de expressão destes genes nas amostras pareadas foi realizada uma clusterização hierárquica baseada na distância euclidiana. A clusterização mostrou alta similaridade entre o perfil de expressão gênica nos dois tipos histológicos, confirmando assim, grande semelhança entre eles. Estes dados também foram observados na plataforma 4.8K e em outros trabalhos com abordagens similares (MA et al. 2003; SCHUETZ et al. 2006; DALGIN et al. 2007) sugerindo fortemente que estes resultados não se referem apenas as amostras deste estudo, mas a uma característica biológica do câncer de mama onde a programação

molecular relacionada ao processo de invasão já está estabelecida em lesões precursoras do IDC.

Dentro deste mesmo objetivo em buscar alterações no nível de expressão de genes que representam vias de sinalização e que possam estar relacionados ao processo de progressão em câncer ductal de mama, comparamos o padrão transcricional de células do mesmo tipo histológico, mas que diferem pela presença ou não do componente invasor na mesma lesão. Foi encontrado um perfil de expressão gênica mais distinto entre as amostras DCIS puro quando comparadas as outras lesões tumorais (DCIS/IDC e IDC), resultados estes também observados na plataforma 4.8K.

Na seleção de supostos genes envolvidos na progressão do DCIS pertencentes a vias de sinalização WNT, PI3K e processo EMT, 15 genes foram considerados e utilizados nas análises de clusterização. Uma maior semelhança foi observada entre o perfil de expressão gênica da célula normal com DCIS puro do que em relação ao DCIS que coexiste com o IDC (DCIS/IDC), baseado no padrão de expressão dos genes diferencialmente expressos pertencentes as vias de sinalização WNT, PI3K e processo EMT, confirmando que essas lesões possuam características moleculares distintas nos genes pertencentes a estes processos.

Na busca por classificadores visando separar os grupos de amostras DCIS-puro e DCIS/IDC dentro de cada grupo de genes pertencentes as vias de sinalização WNT, PI3K e processo EMT representadas na plataforma, foi encontrado os trios classificadores *CSNK1A1L*, *LRP3*, *SDC2*, o trio *PLCG2*, *INPP1*, *DGKA* e o *HDGF*, *CDH13*, *TWIST1* respectivamente.

Para a via de sinalização WNT, o gene *CSNK1A1L* (*casein kinase 1, alpha 1-like*) está envolvido em atividade de proteína quinase serina/treonina, atividade transferase, assim como, em processos de fosforilação de aminoácidos. Outra via de sinalização na qual este gene desempenha um papel se refere a *hedgehog*. Esta via pode apresentar potencial em promover a carcinogênese mamária e progressão do câncer (HATSELL e FROST 2007). O gene *LRP3* (*low density lipoprotein receptor-related protein 3*) apresenta envolvimento em atividade de receptor e no processo de endocitose mediada por receptor. E o gene *SDC2* (*syndecan 2*) codifica uma proteína proteoglicana de transmembrana de sulfato de heparina (tipo 1) e membro da família proteoglicana *syndecan*. As *syndecans* mediam ligação e sinalização da célula e organização do citoesqueleto. A proteína SDC2 participa na proliferação e migração celular e interação célula-matriz. Expressão alterada desta proteína tem sido detectada em diferentes tipos de tumor. MUNESUE et al. (2007) verificaram que SDC2 pode suprimir a ativação da metaloproteinase de matriz 2 (MMP-2) levando a supressão da metatástase. Esta proteína também pode regular a atividade tumorigênica pela regulação da adesão e proliferação em células de carcinoma de cólon (PARK et al. 2002).

Para a via de sinalização PI3K, o gene *PLCG2* (*phospholipase C, gamma 2*) que codifica enzimas da família fosfolipase C. As enzimas C-gamma-2 são controladas por receptores de tirosina quinase e são compostas por domínios fosfolipase que flanqueiam regiões de homologia a domínios não catalíticos do produto do oncogene *SRC*, o SH2 e SH3.

Também estão envolvidas na via de sinalização *VEGF*, cálcio e *ERBB*. O gene *INPP1* (*inositol polyphosphate-1-phosphatase*) codifica uma enzima fosfatase inositol polifosfato-1, uma das enzimas envolvidas na via de sinalização fosfatidilinositol. Também se relaciona ao processo metabólico de fosfatos e transdução de sinal. A proteína codificada pelo gene *DGKA* (*diacylglycerol kinase, alpha*) pertence à família quinase diacilglicerol eucariótica. Ela atua como um modulador que compete com a proteína quinase C para o segundo mensageiro diacilglicerol na via de sinalização intracelular. *DGKalpha* tem sido descrito como um regulador positivo do *NF-kappaB* levando a supressão da apoptose de células de melanoma induzida por *TNF-alpha* (YANAGISAWA et al. 2007).

Para o processo EMT, o gene *HDGF* (*hepatoma-derived growth factor*) está envolvido na atividade de fator de crescimento e ligação a heparina e relacionado ao processo de proliferação celular e transdução de sinal. O HDGF é uma proteína altamente expressa em células vasculares endoteliais e um regulador da migração da célula endotelial por mecanismos distintos de *VEGF* (EVERETT et al. 2004). O gene *CDH13* (*cadherin 13*) é um membro da superfamília de caderinas. A proteína codificada é uma glicoproteína de adesão célula a célula dependente de cálcio. O locus gênico está hipermetilado ou deletado em câncer de mama, ovário e pulmão. Também conhecida por *T-caderina* (TAKEUCHI et al. 2001), esta proteína tem sido demonstrada protegendo células endoteliais da apoptose induzida por estresse oxidativo através da ativação da via de sinal de sobrevivência PI3K/Akt/mTOR e concomitante supressão da via proapoptótica p38 MAPK

(JOSHI et al. 2005). A proteína codificada pelo gene *TWIST1* (*twist homolog 1*) tem sido implicada em determinação de linhagem celular e diferenciação. Alta expressão deste RNAm é observada em tecido placentário, em adultos e tecidos derivados de mesoderma. Este fator de transcrição é considerado o regulador principal da morfogênese embriônica e tem sido relatado como apresentando um papel importante na metástase. A maior expressão de *TWIST* em células de câncer de mama induz angiogênese e promove o processo EMT que desempenha um papel fundamental na transformação das células mamárias em um fenótipo mais agressivo (MIRONCHIK et al. 2005). A supressão da expressão do *TWIST* em células de carcinoma mamário altamente metastáticas especificamente inibe a capacidade destas células metastizarem da glândula mamária para o pulmão (YANG et al. 2004). Este gene também foi descrito como um regulador transcricional positivo da expressão de um proto-oncogene (*AKT2*), estando a sinalização *TWIST-AKT2* envolvida na promoção da capacidade invasiva e sobrevivência das células do câncer de mama (CHENG et al. 2007).

Estes trios de genes mostram-se promissores uma vez que foram capazes de separar os grupos de amostras de interesse (DCIS puro e DCIS/IDC), No entanto, faz-se necessária a validação em amostras independentes.

Visando buscar interações gênicas dentro de cada grupo das vias de sinalização WNT, PI3K e processo EMT foram construídos redes de relevância entre as amostras não-neoplásicas e DCIS puro contra DCIS/IDC. A aplicação desta abordagem permitiu mostrar mudanças de correlação

entre pares de genes com função associada à migração (*NODAL* e *FZD8*) e transdução de sinal (*PLCB4* e *PIK3CB*), por exemplo.

Para a via de sinalização WNT, a diferença de correlação entre *NODAL* e *FZD8* mostrou-se positiva tanto no grupo não-neoplásico e DCIS puro quanto no grupo DCIS/IDC. A proteína codificada pelo gene *NODAL* (*nodal homolog*) é um membro da superfamília *TGF-beta* e está envolvida na proliferação celular e migração relacionada à gastrulação. Estudos têm demonstrado que a inibição da sinalização de *NODAL* promove a reversão de células de melanoma metastático para um fenótipo mais diferenciado, menos invasivo e não tumorigênico (POSTOVIT et al. 2007). E o gene *FZD8* (*frizzled homolog 8*) sem intron é um membro da família *frizzled*. A maioria dos receptores *frizzled* estão acoplados a via de sinalização canônica da beta-catenina.

Foi observada uma correlação negativa entre os pares *PLCB4-MARK9* e *PLCB4-PIK3CB* no grupo não-neoplásico e DCIS puro versus o grupo DCIS/IDC. A proteína codificada pelo gene *PLCB4* (*phospholipase C, beta 4*) desempenha um importante papel na transdução intracelular de vários sinais extracelulares na retina e está envolvida tanto na via de sinalização WNT como PI3K. O gene *PLCB4* se correlaciona negativamente com os genes *MARK9* e *PIK3CB*. O gene *PIK3CB* (*phosphoinositide-3-kinase, catalytic, beta polypeptide*) compõe as PI3Ks (fosfoinosítideo-3 quinases) e tem sido implicado como participantes na via de sinalização regulando o crescimento celular em virtude de sua ativação em resposta a vários estímulos mitogênicos. Apresenta-se envolvido na apoptose, via de

sinalização ErbB, Jak-STAT, mTOR e VEGF, adesão focal e regulação de citoesqueleto de actina. Relacionado a regulação de adesão célula-matriz e transdução de sinal.

Para o processo EMT, houve quebra de correlação entre *BMP7* e *ITGA2*, *COL5A1* e *FGFR2*. Muitas *BMPs* são parte da superfamília de fator-beta de crescimento transformante (*TGFB*). Baseado na sua expressão precoce na embriogênese, a proteína *BMP7* pode desempenhar um papel no desenvolvimento inicial e apresenta envolvimento na interação de receptor citocina-citocina, via de sinalização *Hedgehog* e *TGF-beta*, no processo de diferenciação celular, transição epitélio-mesênquima, crescimento e regulação negativa do ciclo celular. Também pode inibir e estimular o crescimento das células tumorais e tem sido relatado como apresentando um possível papel funcional no desenvolvimento de câncer de mama (ALARMO et al. 2006). Tem sido descrito como um novo gene alvo regulado pela família *p53* e que media a função de sobrevivência celular de nível basal fisiologicamente relevante de *p53* (YAN et al. 2007). O produto do gene *ITGA2* (*integrin, alpha 2*) pertence a família integrina de cadeia *alfa*. As integrinas são glicoproteínas de membrana integral heterodiméricas compostas por uma cadeia alfa distinta e uma cadeia beta comum. São encontradas em uma variedade de células incluindo, células T, fibroblastos e plaquetas. As integrinas estão envolvidas na adesão celular e célula-matriz e também participam na sinalização mediada por superfície celular e regulação do citoesqueleto de actina. Tem sido relatado que polimorfismos nos genes da integrina *ITGA2* influencia no desenvolvimento do câncer de

mama (LANGSENLEHNER et al. 2006). O gene *COL5A1* (*collagen, type V, alpha 1*) codifica uma cadeia alfa para um dos colágenos fibrilares de baixa abundância. Está envolvido em vias de comunicação das células, interação de receptor com a matrix extracelular e adesão focal. A proteína codificada pelo gene *FGFR2* (*fibroblast growth factor receptor 2*) é um membro da família de receptor de fator de crescimento de fibroblasto. Envolvida em vias de sinalização MAPK e regulação de citoesqueleto de actina. Estudos identificaram 4 SNPs no íntron 2 do gene *FGFR2*. Este gene codifica um receptor de tirosina quinase e está amplificado ou mais expresso em alguns tumores de mama (HUNTER et al. 2007).

Portanto, a avaliação das redes de relevância mostrou uma correspondência de correlação com a função dos genes e permitiu identificar interações de genes até então não associadas ao fenômeno biológico em estudo.

A caracterização molecular de carcinoma ductal *in situ* pode auxiliar na predição do comportamento biológico destas lesões direcionando a aplicação de um tratamento mais apropriado. A identificação de fatores prognósticos é de extrema valia, uma vez que o sucesso em identificar e tratar doenças pré-malignas de alto risco tem o potencial de prevenir carcinomas invasivos de mama.

Neste estudo alguns genes com alteração no perfil de expressão foi identificado como supostamente envolvidos na transição do DCIS para o IDC, assim como, na progressão do carcinoma ductal *in situ*. Estes genes, apresentam um potencial de aplicação prática, podendo ser testado para

predição do DCIS que adquira capacidade de invadir o tecido adjacente. Para isso será essencial aumentar o número de casos para avaliar as proteínas nos respectivos anticorpos em amostras parafinadas para que possam ser aplicados na prática clínica predizendo quais as lesões DCIS puro poderiam eventualmente adquirir capacidade de invadir o tecido adjacente, tornando assim o tratamento de lesões pré-maligna mais eficaz.

## 6 CONCLUSÕES

1. Os dados gerados neste estudo mostraram grande similaridade observada no padrão geral de expressão gênica entre tipos histológicos distintos presentes na mesma amostra (DCIS e IDC), sugerindo alterações moleculares anteriores as morfológicas.
2. Foi possível identificar os genes *LUM*, *RDH-E2*, *CXCL13*, *FBN1*, *POSTN* e *GDF15* como candidatos a participarem na transição do DCIS para o IDC.
3. Análises de IHQ mostraram uma associação positiva entre a expressão das proteína LUM e CRABP2. A proteína LUM associou-se positivamente com as CKs 5/6, CK 14, CK8 CK18 e o grupo maior expressão do HER2/neu. Enquanto que a expressão da proteína CRABP-2 associou-se positivamente com ER e PR positivos, CK8, CK18, Luminal A, p53 e negativamente com a CK14.
4. Foi observado padrão transcricional de células do mesmo tipo histológico que diferem pela presença ou não do componente invasor na mesma lesão. Foi verificado um perfil de expressão gênica mais distinto entre as amostras DCIS puro quando comparadas as outras lesões tumorais (DCIS/IDC e IDC), contradizendo os aspectos morfológicos e sugerindo que alterações moleculares possam estar presentes antes das alterações morfológicas.

5. Os nossos resultados mostraram que a via de sinalização PI3K parece desempenhar um papel importante no processo de transição do DCIS para o IDC.
6. Os nossos dados identificaram trios classificadores capazes de separar os grupos de amostras DCIS-puro e DCIS/IDC dentro de cada grupo de genes pertencentes as vias de sinalização WNT (*CSNK1A1L*, *LRP3* e *SDC2*), PI3K (*PLCG2*, *INPP1* e *DGKA*) e processo EMT (*HDGF*, *CDH13* e *TWIST1*).
7. Identificamos correlações entre pares de genes funcionalmente relacionados na via de sinalização WNT (*NODAL* e *FZD8*; *PLCB4* e *MARK9*), na via de sinalização PI3K (*PLCB4* e *PIK3CB*) e no processo EMT (*BMP7* e *ITGA2*; *COL5A1* e *FGFR2*) na comparação entre amostras não-neoplásicas e DCIS puro (sem capacidade de invadir) contra DCIS/IDC (com capacidade).

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

[Anonymus]. 16 years mortality from breast cancer in the UK trial of early detection of breast cancer in the UK Trial of early Detection of Breast Cancer. **Lancet** 1999; 353:1909-14.

Abd El-Aziz SH, Endo Y, Miyamaori H, Takino T, Sato H. Cleavage of growth differentiation factor 15 (GDF15) by membrane type 1-matrix metalloproteinase abrogates GDF15-mediated suppression of tumor cell growth. **Cancer Sci** 2007; 98:1330-5.

Abd El-Rehim DM, Ball G, Pinder SE, et al. High-throughput protein expression analysis using tissue microarray technology of a large well characterized series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. **Int J Cancer** 2005; 116:340-50.

Abercrombie M, Dunn GA, Heath JP. The shape and movement of fibroblasts in culture. **Soc Gen Physiol Ser** 1977; 32:57-70.

Ackland ML, Newgreen DF, Fridman M, et al. Epidermal growth factor-induced epithelio-mesenchymal transition in human breast carcinoma cells. **Lab Invest**. 2003; 83:435-48.

Aberle H, Bauer A, Stappert J, Kispert A, Kemler R. beta-catenin is a target for the ubiquitin-proteasome pathway. **EMBO J** 1997; 16:3797-804.

Adeyinka A, Emberley E, Niu Y, et al. Analysis of gene expression in ductal carcinoma *in situ* of the breast. **Clin Cancer Res** 2002; 8:3788-95.

Aguirre-Ghiso J. **Metastasis research areas**. Available from: <URL:<http://www.albany.edu/cancer/genomics/faculty/jaguirre/research.html>> [2005 ago 12].

Alarmo EL, Rauta J, Kauraniemi P, Karhu R, Kuukasjärvi T, Kallioniemi A. Bone morphogenetic protein 7 is widely overexpressed in primary breast cancer. **Genes Chromosomes Cancer** 2006; 45:411-9.

Alexe G, Dalgin GS, Ganesan S, Delisi C, Bhanot G. Analysis of breast cancer progression using principal component analysis and clustering. **J Biosci**. 2007; 32:1027-39.

Al-Kuraya K, Schraml P, Torhorst J, et al. Prognostic relevance of gene amplifications and coamplifications in breast cancer. **Cancer Res** 2004; 64:8534-40.

Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. **Nat Rev Genet** 2006; 7(1):55-65.

Allred DC: Biologic and genetics features of *in situ* breast cancer. In: Silverstein MJ, editor. **Ductal carcinoma *in situ* of the breast**. Baltimore: Williams and Wilkins; 1997. p.37.

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

Altomare DA, Testa JR. Perturbations of the AKT signaling pathway in human cancer. **Oncogene** 2005; 24:7455-64.

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

Anderson WF, Chu KC, Chatterjee N, Brawley O, Brinton LA. Tumor variants by hormone receptor expression in white patients with node-negative breast cancer from the surveillance, epidemiology, and end results database. **J Clin Oncol** 2001; 19:18-27.

Balch CM, Soong SJ, Gershenwald JE, et al. Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. **J Clin Oncol** 2001; 19:3622-34.

Ball CA, Chen Y, Panavally S, et al. An introduction to microarrays bioinformatics. In: Bowtell D, Sambrook J, editors. **DNA microarrays: a molecular cloning manual**. New York: Cold Spring Harbor Laboratory Press; 2003. p.509-94.

Bao S, Ouyang G, Bai X, et al. Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. **Cancer Cell** 2004; 5:329-39.

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.

Battle E, Sancho E, Franci C, et al. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. **Nat Cell Biol** 2000; 2:84-9.

Bell CD, Waizbard E. Variability of cell size in primary and metastatic human breast carcinoma. **Invasion Metastasis** 1986; 6:11-20.

Berardo MD, Allred DC, O'Connell P. Breast cancer. In: Jameson JL, Totowa NJ, editors. **Principles of molecular medicine**. New Jersey: Human Press, 1998. p.625-32.

Bertucci F, Van Hulst S, Bernard K, et al. Expression scanning of an array of growth control genes in human tumor cell lines. **Oncogene** 1999; 18:3905-12.

Bhargava R, Gerald WL, Li AR, et al. EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. **Mod Pathol** 2005; 18:1027-33.

Bolos V, Peinado H, Perez-Moreno MA, Fraga MF, Esteller M, Cano A. The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. **J Cell Sci** 2003; 116:499-511.

Bonner RF, Emmert-Buck M, Cole K, et al. Laser capture microdissection: molecular analysis of tissue. **Science** 1997; 278:1481-3.

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

Brennan KR, Brown AM. Wnt proteins in mammary development and cancer. **J Mammary Gland Biol Neoplasia** 2004; 9:119-31.

Brader S, Eccles SA. Phosphoinositide 3-kinase signalling pathways in tumor progression, invasion and angiogenesis. **Tumori** 2004; 90:2-8.

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.

Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM. Ductal carcinoma *in situ* of the breast. **N Engl J Med** 2004; 350:1430-41.

Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. **Proc Natl Acad Sci USA** 2000; 97:12182-6.

Candidus S, Bischoff P, Becker KF, Hofler H. No evidence for mutation in the alpha- and beta-catenin genes in human gastric and breast carcinomas. **Cancer Res** 1996; 56:49-52.

Carraro DM, Sredni ST, Maschietto M, Castro NP, De Camargo B, Brentani H. Applications of microarray technology in clinic oncology. **Bioscience J** 2004; 20(Suppl):53-8.

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.

Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. **Cancer** 1989; 63:181-7.

Cavalieri D, Dolara P, Mini E, et al. Analysis of gene expression profiles reveals novel correlations with the clinical course of colorectal cancer. **Oncol Res** 2007; 16:535-48.

Cavallaro U, Christofori G. Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough. **Biochim Biophys Acta** 2001; 1552:39-45.

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.

Charo IF, Ransohoff RM. The many roles of chemokines and chemokine receptors in inflammation. **N Engl J Med** 2006; 354:610-21.

Chen W, Salto-Tellez M, Palanisamy N, et al. Targets of genome copy number reduction in primary breast cancers identified by integrative genomics. **Genes Chromosomes Cancer** 2007;46:288-301.

Cheng GZ, Chan J, Wang Q, Zhang W, Sun CD, Wang LH Twist transcriptionally up-regulates AKT2 in breast cancer cells leading to increased migration, invasion, and resistance to paclitaxel. **Cancer Res** 2007; 67:1979-87.

Chin K, de Solorzano CO, Knowles D, et al. *In Situ* Analyses of genome instability in breast cancer. **Nat Genet** 2004, 36:984-8.

Claus EB, Stowe M, Carter D. Breast carcinoma *in situ*: risk factors and screening patterns. **J Natl Cancer Inst** 2001; 93:1811-7.

Christofori G, Semb H. The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. **Trends Biochem Sci** 1999; 24:73-6.

Comijn J, Berx G, Vermassen P, et al. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. **Mol Cell** 2001; 7:1267-78.

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

Corzo C, Tusquets I, Salido M, et al. Characterization of HER1 (c-erbB1) status in locally advanced breast cancer using fluorescence *in situ* hybridization and immunohistochemistry. **Tumour Biol** 2005; 26:25-30.

Costa SD, Lange S, Klinga K, Merkle E, Kaufmann M. Factors influencing the prognostic role of oestrogen and progesterone receptor levels in breast cancer – results of the analysis of 670 patients with 11 years of follow-up. **Eur J Cancer** 2002; 38:1329-34.

Cristo EB. **Métodos estatísticos na análise de experimentos microarray**. São Paulo; 2003. [Dissertação de Mestrado-Instituto de Matemática e Estatística da Universidade de São Paulo].

Dalgin GS, Alexe G, Scandfeld D, et al. Portraits of breast cancer progression. **BMC Bioinformatics** 2007; 8:291.

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.

Dias NE, Correa RG, Verjovski-Almeida S, et al. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. **Proc. Natl. Acad. Sci. U. S. A.** 2000; **97**:3491-3496.

DiGiovanna MP, Chu P, Davison TL, et al. Active signaling by HER-2/neu in a subpopulation of HER-2/neu-overexpressing ductal carcinoma *in situ*: clinicopathological correlates. **Cancer Res** 2002; 62:6667-73.

Donovan-Peluso M, Contento AM, Tobon H, Ripepi B, Locker J. Oncogene amplification in breast cancer. **Am J Pathol** 1991; 138:835-45.

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.

Dudoit S, Gentleman RC, Quackenbush J. Open source software for the analysis of microarray data. **Biotechniques** 2003; Suppl:45-51.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. **Proc Natl Acad Sci U S A** 1998; 95:14863-8.

Ellis LM, Rosen L, Gordon MS. Overview of anti-VEGF therapy and angiogenesis. Part 1: Angiogenesis inhibition in solid tumor malignancies. **Clin Adv Hematol Oncol** 2006; 4(suppl):1-10; quiz 11-2.

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

Erler JT, Bennewith KL, Nicolau M, Dornhöfer N, Kong C, Le QT, Chi JT, Jeffrey SS, Giaccia AJ. Lysyl oxidase is essential for hypoxia-induced metastasis. **Nature** 2006; 440:1222-6.

Ernster VL, Barclay J. Increases in ductal carcinoma *in situ* (DCIS) of the breast in relation to mammography: a dilemma. **J Natl Cancer Inst Monogr** 1997; 22:151-6.

Ernster VL, Barclay J, Kerlikowske K, Grady D, Henderson C. Incidence of and treatment for ductal carcinoma *in situ* of the breast. **JAMA** 1996; 275:913-8.

Ernster VL, Ballard-Barbash R, Barlow WE, et al. Detection of ductal carcinoma *in situ* in women undergoing screening mammography. **J Natl Cancer Inst** 2002; 94:1546-54.

Esteves GH. **Métodos estatísticos para a análise de dados de cDNA microarray em um ambiente computacional integrado**. São Paulo; 2007. [Tese de Doutorado- Instituto de Matemática e Estatística da Universidade de São Paulo].

Everett AD, Narron JV, Stoops T, Nakamura H, Tucker A Hepatoma-derived growth factor is a pulmonary endothelial cell-expressed angiogenic factor. **Am J Physiol Lung Cell Mol Physiol** 2004; 286:L1194-201.

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.

Faverly DR, Burgers L, Bult P, Holland R. Three dimensional imaging of mammary ductal carcinoma *in situ*: clinical implications. **Semin Diagn Pathol** 1994; 11:193-8.

Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. **Nat Rev Cancer** 2003; 3:453-8.

Fidler IJ, Hart IR. Biological diversity in metastatic neoplasms: origins and implications. **Science** 1982; 217:998-1003.

Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. **J Clin Oncol** 1998; 16:2672-85.

Fisher B, Jeong JH, Bryant J, et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. **Lancet** 2004; 364:858-68.

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.

Folgueira MA, Brentani H, Katajama ML, et al. Gene Expression Profiling Of Clinical Stages Ii And Iii Breast Cancer. **Braz J Med Biol Res** 2006; 39:1101-13.

Foster RS Jr. The biologic and clinical significance of lymphatic metastases in breast cancer. **Surg Oncol Clin N Am** 1996; 5:79-104.

Försti A, Jin Q, Altieri A, Johansson R, et al. Polymorphisms in the KDR and POSTN genes: association with breast cancer susceptibility and prognosis. **Breast Cancer Res Treat** 2007; 101:83-93.

Fracheboud J, de Koning HJ, Boer R, et al. Nationwide breast cancer screening programme fully implemented in The Netherlands. **Breast** 2001; 10:6-11.

Friedl P, Brocker EB. The biology of cell locomotion within three-dimensional extracellular matrix. **Cell Mol Life Sci** 2000; 57:41-64.

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

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

Fukumoto S, Hsieh CM, Maemura K, et al. Akt participation in the Wnt signaling pathway through Dishevelled. **J Biol Chem** 2001; 276:17479-83.

Gillan L, Matei D, Fishman DA, Gerbin CS, Karlan BY, Chang DD. Periostin secreted by epithelial ovarian carcinoma is a ligand for alpha(V)beta(3) and alpha(V)beta(5) integrins and promotes cell motility. **Cancer Res** 2002; 62:5358-64.

Goldhirsch A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. **J Clin Oncol** 2001; 19:3817-27.

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.

Gomes LI, Esteves GH, Carvalho AF, et al. Expression profile of malignant and nonmalignant lesions of esophagus and stomach: differential activity of functional modules related to inflammation and lipid metabolism. **Cancer Res** 2005; 65:7127-36.

Grille SJ, Bellacosa A, Upson J, et al. The protein kinase Akt induces epithelial mesenchymal transition and promotes enhanced motility and invasiveness of squamous cell carcinoma lines. **Cancer Res** 2003; 63:2172-8.

Gruvberber S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. **Cancer Res** 2001, 61:5979-84.

Handschuh G, Candidus S, Luber B, et al. Tumour-associated E-cadherin mutations alter cellular morphology, decrease cellular adhesion and increase cellular motility. **Oncogene** 1999; 18:4301-12.

Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma *in situ* by gene expression profiling. **Breast Cancer Res** 2006; 8:R61.

Harris TJ, Peifer M. Decisions, decisions: beta-catenin chooses between adhesion and transcription. **Trends Cell Biol** 2005; 15:234-7.

Harrison JC, Dean PJ, el-Zeky F, Vander Zwaag R. From Dukes through Jass: pathological prognostic indicators in rectal cancer. **Hum Pathol** 1994; 25:498-505.

Hatsell S, Frost AR. Hedgehog signaling in mammary gland development and breast cancer. **J Mammary Gland Biol Neoplasia** 2007; 12:163-73.

Hayes DF, Miller K, Sledge G. Angiogenesis as targeted breast cancer therapy. **Breast** 2007; 16 Suppl 2:S17-9.

Hemavathy K, Ashraf SI, Ip YT. Snail/slug family of repressors: slowly going into the fast lane of development and cancer. **Gene** 2000; 257:1-12.

Henderson BE, Ross R, Bernstein L. Estrogens as a cause of human cancer: the Richard and Hinda Rosenthal Foundation award lecture. **Cancer Res** 1988; 48:246-53.

Hidalgo A, Salcedo M. Global analysis strategies: toward the genetic management of neoplasias. **Rev Invest Clin** 2001; 53:430-43.

Hill AD, Doyle JM, McDermott EW, O'Higgins NJ. Hereditary breast cancer. **Br J Surg** 1997; 84:1334-9.

Hirohashi S, Kanai Y. Cell adhesion system and human cancer morphogenesis. **Cancer Sci** 2003; 94:575-81.

Holland R, Hendriks JH, Vebeek AL, Mravunac M, Schuurmans Stekhoven JH. Extent, distribution, and mammographic/histological correlations of breast ductal carcinoma *in situ*. **Lancet** 1990; 335:519-22.

Houghton J, George WD, Cuzick J, et al. Radiotherapy and tamoxifen in women with completely excised ductal carcinoma *in situ* of the breast in the UK, Australia, and New Zealand: randomised controlled trial. **Lancet** 2003; 362:95-102.

Howe LR, Brown AM. Wnt signaling and breast cancer. **Cancer Biol Ther** 2004; 3:36-41.

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.

Huang E, Cheng SH, Dressman H, et al. Gene expression predictors of breast cancer outcomes. **Lancet** 2003; 361:1590-6.

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

Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. **Nat Genet** 2007; 39:870-4.

Hutcheson IR, Knowlden JM, Jones HE, et al. Inductive mechanisms limiting response to anti-epidermal growth factor receptor therapy. **Endocr Rel Cancer** 2006; 13:s89-97.

Hwang ES, DeVries S, Chew KL, et al. Patterns of chromosomal alterations in breast ductal carcinoma *in situ*. **Clin Cancer Res** 2004, 10:5160-7.

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

Ihaka R, Gentleman R. R: a language for data analysis and graphics. **J Comput Graph Statist** 1996; 5:299-314.

Imbert A, Eelkema R, Jordan S, Feiner H, Cowin P:  $\Delta N89\beta$ -catenin induces precocious development, differentiation, and neoplasia in mammary gland. **J Cell Biol** 2001; 153:555-68.

Jawhari AU, Farthing MJ, Pignatelli M. The E-cadherin/epidermal growth factor receptor interaction: a hypothesis of reciprocal and reversible control of intercellular adhesion and cell proliferation. **J Pathol** 1999; 187:155-7.

Jonsson M, Borg A, Nilbert M, Andersson T. Involvement of adenomatous polyposis coli (APC)/beta-catenin signalling in human breast cancer. **Eur J Cancer** 2000; 36:242-8.

Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. **CA Cancer J Clin** 2007; 57:43-66.

Joshi MB, Philippova M, Ivanov D, Allenspach R, Erne P, Resink TJ. T-cadherin protects endothelial cells from oxidative stress-induced apoptosis. **FASEB J** 2005; 19:1737-9.

Kerlikowske K, Barclay J, Grady D, Sickles EA, Ernster V. Comparison of risk factors for ductal carcinoma *in situ* and invasive breast cancer. **J Natl Cancer Inst** 1997; 89:76-82.

Kerlikowske K, Molinaro A, Cha I, et al. Characteristics associated with recurrence among women with ductal carcinoma *in situ* treated by lumpectomy. **J Natl Cancer Inst** 2003; 95:1692-702.

Kessar P, Perry N, Vinnicombe SJ, Hussain HK, Carpenter R, Wells CA. How significant is detection of ductal carcinoma *in situ* in a breast screening programme? **Clin Radiol** 2002; 57:807-14.

Khan J, Saal LH, Bittner ML, Chen Y, Trent JM, Meltzer PS. Expression profiling in cancer using cDNA microarrays. **Electrophoresis** 1999; 20:223-9.

Kim D, Kim S, Koh H, et al. Akt/PKB promotes cancer cell invasion via increased motility and metalloproteinase production. **FASEB J** 2001; 15:1953-62.

King CR, Swain SM, Porter L, Steinberg SM, Lippman ME, Gelmann EP. Heterogeneous expression of erbB-2 messenger RNA in human breast cancer. **Cancer Res** 1989; 49:4185-91.

Kobayashi M, Ooi A, Oda Y, Nakanishi I. Protein overexpression and gene amplification of c-erbB-2 in breast carcinomas: a comparative study of immunohistochemistry and fluorescence *in situ* hybridization of formalin-fixed, paraffin-embedded tissues. **Hum Pathol** 2002; 33:21-8.

Kolble K. The LEICA microdissection system: design and applications. **J Mol Med** 2000; 78:B24-5.

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

Kurochkin IV, Yonemitsu N, Funahashi SI, Nomura H. ALEX1, a novel human armadillo repeat protein that is expressed differentially in normal tissues and carcinomas. **Biochem Biophys Res Commun** 2001; 280:340-7.

Landis SH, Murray T, Bolden S, Wingo PA. Cancer Statistics 1998. **CA Cancer J Clin** 1998; 48: 6-29. Erratum in: **CA Cancer J Clin** 1998; 48:192. **CA Cancer J Clin** 1998; 48:329.

Langsenlehner U, Renner W, Yazdani-Biuki B, et al. Integrin alpha-2 and beta-3 gene polymorphisms and breast cancer risk **Breast Cancer Res Treat** 2006; 97:67-72.

Larue L, Bellacosa A. Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. **Oncogene** 2005; 24:7443-54.

Lauffenburger DA, Horwitz AF. Cell migration: a physically integrated molecular process. **Cell** 1996; 84:359-69.

Lee LA, Silverstein MJ, Chung CT, et al. Breast cancer-specific mortality after invasive local recurrence in patients with ductal carcinoma-in-situ of the breast. **Am J Surg** 2006; 192:416-9.

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

Leygue E, Snell L, Dotzlaw H, et al. Expression of lumican in human breast carcinoma. **Cancer Res** 1998; 58:1348-52.

Leygue E, Snell L, Dotzlaw H, et al. Lumican and decorin are differentially expressed in human breast carcinoma. **J Pathol** 2000; 192:313-20.

Lin SY, Xia W, Wang JC, et al. Beta-catenin, a novel prognostic marker for breast cancer: its roles in cyclin D1 expression and cancer progression. **Proc Natl Acad Sci U S A** 2000; 97:4262-6.

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

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. **Methods** 2001; 25:402-8.

Livasy CA, Perou CM, Karaca G, et al. Identification of a basal-like subtype of breast ductal carcinoma *in situ*. **Hum Pathol** 2007; 38:197-204.

Lu M, Mira-y-Lopez R, Nakajo S, Nakaya K, Jing Y. Expression of estrogen receptor alpha, retinoic acid receptor alpha and cellular retinoic acid binding protein II genes is coordinately regulated in human breast cancer cells. **Oncogene** 2005; 24:4362-9.

Luker KE, Luker GD. Functions of CXCL12 and CXCR4 in breast cancer. **Cancer Lett** 2006; 238:30-41.

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

Majumder PK, Sellers WR. Akt-regulated pathways in prostate cancer. **Oncogene** 2005; 24:7465-74.

Matsuzaka Y, Okamoto K, Tsuji H, et al. Identification of the hRDH-E2 gene, a novel member of the SDR family, and its increased expression in psoriatic lesion. **Biochem Biophys Res Commun** 2002; 297:1171-80.

Matsuzaka Y, Okamoto K, Yoshikawa Y, et al. hRDH-E2 gene polymorphisms, variable transcriptional start sites, and psoriasis. **Mamm Genome** 2004; 15:668-75.

McGuire WL. Prognostic factors for recurrence and survival in human breast cancer. **Breast Cancer Res Treat** 1987; 10:5-9.

Mauriac L. Aromatase inhibitors: Effective endocrine therapy in the early adjuvant setting for postmenopausal women with hormone-responsive breast cancer. **Best Pract Res Clin Endocrinol Metab** 2006; 20(Suppl 1):S15-29.

McPherson LA, Woodfield GW, Weigel RJ. AP2 transcription factors regulate expression of CRABP II in hormone responsive breast carcinoma. **J Surg Res** 2007; 138:71-8.

Meireles SI, Cristo EB, Carvalho AF, et al. Molecular classifiers for gastric cancer and nonmalignant diseases of the gastric mucosa. **Cancer Res** 2004; 64:1255-65.

Meyer M, Stenzel U, Hofreiter M. Parallel tagged sequencing on the 454 platform. **Nat Protoc** 2008; 3:267-78.

Millis RR, Bobrow LG, Barnes DM. Immunohistochemical evaluation of biologic markers in mammary carcinoma *in situ*: correlation with morphological features and recently proposed schemes for histological classification. **Breast** 1996; 5:113-22.

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

Miyoshi Y, Iwao K, Nawa G, Yoshikawa H, Ochi T, Nakamura Y. Frequent mutations in the beta-catenin gene in desmoid tumors from patients without familial adenomatous polyposis. **Oncol Res** 1998; 10:591-4.

Mironchik Y, Winnard PT Jr, Vesuna F, et al. Twist overexpression induces in vivo angiogenesis and correlates with chromosomal instability in breast cancer. **Cancer Res** 2005; 65:10801-9.

Moon RT, Bowerman B, Boutros M, Perrimon N. The promise and perils of Wnt signaling through beta-catenin. **Science** 2002; 296:1644-6.

Morin PJ. Beta-catenin signaling and cancer. **Bioessays** 1999; 21:1021-30.

Morris SR, Carey LA. Molecular profiling in breast cancer. **Rev Endocr Metab Disord** 2007; 8:185-98.

Morrison TB, Wei JJ, Wittwer CT. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. **Biotechniques** 1998; 24:954-62.

Moser AR, Hegge LF, Cardiff RD: Genetic background affects susceptibility to mammary hyperplasias and carcinomas in *Apc(min)/+* mice. **Cancer Res** 2001, 61:3480-5.

Muller T, Bain G, Wang X, Papkoff J. Regulation of epithelial cell migration and tumor formation by beta-catenin signaling. **Exp Cell Res** 2002; 280:119-33.

Munesue S, Yoshitomi Y, Kusano Y, et al. A novel function of syndecan-2, suppression of matrix metalloproteinase-2 activation, which causes suppression of metastasis. **J Biol Chem** 2007; 282:28164-74.

Nagaraja GM, Othman M, Fox BP, et al. Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. **Oncogene** 2006; 25:2328-38.

Nakhlis F, Morrow M. Ductal carcinoma *in situ*. **Surg Clin North Am** 2003; 83:821-39.

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

Nathanson SD. Insights into the mechanisms of lymph node metastasis. **Cancer** 2003; 98:413-23.

Nielsen TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. **Clin Cancer Res** 2004; 10:5367-74.

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

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Tukey, multiple comparison procedure. In: **Applied Linear Statistical Models**. 4<sup>th</sup> ed. Irwin: Boston; 1996. p.725-32.

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.

Novoradovskaya N, Whitfield ML, Basehore LS, et al. Universal reference RNA as a standard for microarray experiments. **BMC Genomics** 2004; 5:20.

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

Ottesen GL, Christensen IJ, Larsen JK, et al. Carcinoma *in situ* of the breast: correlation of histopathology to immunohistochemical markers and DNA ploidy. **Breast Cancer Res Treat** 2000; 60:219-26.

Pike MC, Spicer DV, Dahmouch L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. **Epidemiol Rev** 1993; 15:17-35.

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

Park BK, Zeng X, Glazer RI. Akt1 induces extracellular matrix invasion and matrix metalloproteinase-2 activity in mouse mammary epithelial cells. **Cancer Res** 2001; 61:7647-53.

Park H, Kim Y, Lim Y, Han I, Oh ES. Syndecan-2 mediates adhesion and proliferation of colon carcinoma cells. **J Biol Chem** 2002; 277:29730-6.

Park K, Han S, Kim HJ, Kim J, Shin E. HER2 status in pure ductal carcinoma *in situ* and in the intraductal and invasive components of invasive ductal carcinoma determined by fluorescence *in situ* hybridization and immunohistochemistry. **Histopathology** 2006; 48:702-7.

Payne SL, Fogelgren B, Hess AR, et al. Lysyl oxidase regulates breast cancer cell migration and adhesion through a hydrogen peroxide-mediated mechanism. **Cancer Res** 2005; 65:11429-36.

Peifer M, Polakis P. Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 2000; 287:1606-9.

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

Phillips MF, Lockett MR, Rodesch MJ, Shortreed MR, Cerrina F, Smith LM. *In situ* oligonucleotide synthesis on carbon materials: stable substrates for microarray fabrication. **Nucleic Acids Res** 2008; 36:e7.

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

Postovit LM, Seftor EA, Seftor RE, Hendrix MJ. Targeting Nodal in malignant melanoma cells. **Expert Opin Ther Targets** 2007; 11:497-505.

Quackenbush J. Computational analysis of microarray data. **Nature** 2001; 2:418-27.

Rakha EA, El-Rehim DA, Paish C, et al. Basal phenotype identifies a poor prognostic subgroup of breast cancer of clinical importance. **Eur J Cancer** 2006; 42:3149-56.

Reis-Filho JS, Milanezi F, Carvalho S, et al. Metaplastic breast carcinomas exhibit EGFR, but not HER2, gene amplification and overexpression: immunohistochemical and chromogenic *in situ* hybridization analysis. **Breast Cancer Res** 2005; 7:R1028-35.

Ringel MD, Hayre N, Saito J, et al. Overexpression and overactivation of Akt in thyroid carcinoma. **Cancer Res** 2001; 61:6105-11.

Ryo A, Nakamura M, Wulf G, Liou YC, Lu KP. Pin1 regulates turnover and subcellular localization of beta-catenin by inhibiting its interaction with APC. **Nat Cell Biol** 2001; 3:793-801.

Rosenthal TC, Puck SM. Screening for genetic risk of breast cancer. **Am Fam Physician** 1999; 59:99-16.

Rosser RJ. Consensus conference on the classification of ductal carcinoma *in situ*. **Cancer** 1998; 82:2293-5.

Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology in human cancer. **Oncogene** 2003; 22:6497-507.

Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. **Biotechniques** 2003; 34:374-8.

Salomon DS, Bradt R, Ciardiello F, Normanno N. Epidermal growth factor related peptides and their receptors in human malignancies. **Crit Rev Oncol Haematol** 1995; 19:183-232.

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

Saraiva TF, Castro NP, Baldan-Pineda PH, et al. Effects of oligo dT-T7 RNA primer in RNA amplification from paraffin-embedded tissue for microarray experiments. **Applied Cancer Res** 2006; 26:14-20.

Scheid MP, Woodgett JR. Phosphatidylinositol 3' kinase signaling in mammary tumorigenesis. **J Mammary Gland Biol Neoplasia** 2001; 6:83-99.

Schlosshauer PW, Brown SA, Eisinger K, et al. APC truncation and increased beta-catenin levels in a human breast cancer cell line. **Carcinogenesis** 2000; 21:1453-6.

Schnitt SJ. Normal anatomy and development. In: Hayes DF, editor. **Atlas of breast cancer**. London: Mosby; 1995a. p.3.2-3.4.

Schnitt SJ. Pathology of breast cancer. In: Hayes DF, editor. **Atlas of breast cancer**. London: Mosby; 1995b. p.8.1-8.11.

Schlessinger J. Cell signaling by receptor tyrosine kinases. **Cell** 2000; 103:211-25.

Schraml P, Bucher C, Bissig H, et al. Cyclin E overexpression and amplification in human tumours. **J Pathol** 2003; 200:375-82.

Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. **BMC Mol Biol** 2006; 7:3.

Schutze K, Lahr G. Identification of expressed genes by laser-mediated manipulation of single cells. **Nat Biotechnol** 1998; 16:737-42.

Schuetz CS, Bonin M, Clare SE, et al. Progression-specific genes identified by expression profiling of matched ductal carcinomas *in situ* and invasive

breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. **Cancer Res** 2006; 66:5278-86.

Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. **Nat Genet** 2004; 36:1090-8.

Segal E, Wang H, Koller D. Discovering molecular pathways from protein interaction and gene expression data. **Bioinformatics** 2003; 19(Suppl 1):264-71.

Seth A, Kitching R, Landberg G, Xu J, Zubovits J, Burger AM. Gene expression profiling of ductal carcinomas *in situ* and invasive breast tumors. **Anticancer Res** 2003; 23:2043-51.

Shao R, Bao S, Bai X, et al. Acquired expression of periostin by human breast cancers promotes tumor angiogenesis through up-regulation of vascular endothelial growth factor receptor 2 expression. **Mol Cell Biol** 2004; 24:3992-4003.

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

Shook D, Keller R. Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. **Mech Dev** 2003; 120:1351-83.

Simone NL, Bonner RF, Gillespie JW, Emmert-Buck MR, Liotta LA. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. **Trends Genet** 1998; 14:272-6.

Singletary SE, Allred C, Ashley P, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. **J Clin Oncol** 2002; 20:3628-36.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. **Science** 1987; 235:177-82.

Slamon DJ, Godolphin W, Jones LA, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. **Science** 1989; 244:707-12.

Smith CA, McClive PJ, Sinclair AH. Temporal and spatial expression profile of the novel armadillo-related gene, Alex2, during testicular differentiation in the mouse embryo. **Dev Dyn** 2005; 233:188-93.

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.

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

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 USA** 2003; 100:10393-8.

Sparks AB, Morin PJ, Vogelstein B, Kinzler KW. Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. **Cancer Res** 1998; 58:1130-4.

Steinman S, Wang J, Bourne P, Yang Q, Tang P. Expression of cytokeratin markers, ER-alpha, PR, HER-2/neu, and EGFR in pure ductal carcinoma *in situ* (DCIS) and DCIS with co-existing invasive ductal carcinoma (IDC) of the breast. **Ann Clin Lab Sci** 2007; 37:127-34.

Stiepcich MMA. **Perfil de citoceratinas e sua relação com diferenciação e fatores clínicomorfológicos em carcinomas ductais soe de mama estudados em “array de tecido” (tissue microarray – TMA)**. São Paulo; 2007. [Tese de Doutorado-Fundação Antônio Prudente].

Stolf BS, Santos MM, Simao DF, et al. Class distinction between follicular adenomas and follicular carcinomas of the thyroid gland on the basis of their signature expression. **Cancer** 2006; 106:1891-900.

Storey J.D. A direct approach to false discovery rates. **J R Statist Soc B** 2002; 64:479-98.

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**. 2 ed. Philadelphia: W B Saunders; 1998. p.419-25.

Sun M, Wang G, Paciga JE, et al. AKT1/PKBAalpha kinase is frequently elevated in human cancers and its constitutive activation is required for oncogenic transformation in NIH3T3 cells. **Am J Pathol** 2001; 159, 431-7.

Takeuchi T, Ohtsuki Y. Recent progress in T-cadherin (CDH13, H-cadherin) research. **Histol Histopathol** 2001; 16:1287-93

Testa JR, Bellacosa A. AKT plays a central role in tumorigenesis. **Proc Natl Acad Sci U S A** 2001; 98:10983-5.

Tetsu O, McCormick F. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. **Nature** 1999; 398:422-6.

The Cancer Council Victoria. **Breast health**. Available from: <URL:<http://www.cancervic.org.au/cancer1/prevent/breasthealth.htm>> [2005 ago 12].

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

Tian Q, Feetham MC, Tao WA, et al. Proteomic analysis identifies that 14-3-3zeta interacts with beta-catenin and facilitates its activation by Akt. **Proc Natl Acad Sci USA** 2004; 101:15370-5.

Trentham-Dietz A, Newcomb PA, Storer BE, Remington PL. Risk factors for carcinoma *in situ* of the breast. **Cancer Epidemiol Biomarkers Prev** 2000; 9:697-703.

Tsuda H, Hirohashi S, Shimosato Y, et al. Correlation between long-term survival in breast cancer patients and amplification of two putative oncogene-coamplification units: hst-1/int-2 and c-erbB-2/ear-1. **Cancer Res** 1989; 49:3104-8.

Tsuda H, Akiyama F, Terasaki H, et al. Detection of HER-2/neu (c-erb B-2) DNA amplification in primary breast carcinoma. Interobserver reproducibility and correlation with immunohistochemical HER-2 overexpression. **Cancer** 2001; 92:2965-74.

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

van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. **Am J Pathol** 2002; 161:1991-6.

van de Rijn M, Gilks CB. Applications of microarrays to histopathology. **Histopathology** 2004; 44:97-108.

van de Vijver MJ, van de Bersselaar R, Devilee P, Cornelisse C, Peterse J, Nusse R. Amplification of the neu (c-erbB-2) oncogene in human mammary tumors is relatively frequent and is often accompanied by amplification of the linked c-erbA oncogene. **Mol Cell Biol** 1987; 7:2019-23.

van de Vijver MJ, Peterse JL, Mooi WJ, et al. Neu-protein overexpression in breast cancer: association with comedo-type ductal carcinoma *in situ* and limited prognostic value in stage II breast cancer. **N Engl J Med** 1988; 319:1239-45.

van der Hage JA, van de Velde CJ, Julien JP, Tubiana-Hulin M, Vandervelden C, Duchateau L. Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. **J Clin Oncol** 2001; 19:4224-37.

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.

Varley JM, Swallow JE, Brammar WJ, Whittaker JL, Walker RA. Alterations to either c-erbB-2(neu) or c-myc proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. **Oncogene** 1987; 1:423-30

Vincent-Salomon A, Thiery JP. Host microenvironment in breast cancer development: epithelial-mesenchymal transition in breast cancer development. **Breast Cancer Res** 2003; 5:101-6.

Vos CB, Ter Haar NT, Rosenberg C, et al. Genetic alterations on chromosome 16 and 17 are important features of ductal carcinoma *in situ* of the breast and are associated with histologic type. **Br J Cancer** 1999; 81:1410-8.

Waltzer L, Bienz M. The control of beta-catenin and TCF during embryonic development and cancer. **Cancer Metastasis Rev** 1999; 18:231-46.

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

Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. **Lancet** 2005; 365: 671-9.

Warnberg F, Nordgren H, Bergkvist L, Holmberg L. Tumour markers in breast carcinoma correlate with grade rather than with invasiveness. **Br J Cancer** 2001; 85:869-74.

Weston CR, Davis RJ. Signal transduction: signaling specificity- a complex affair. **Science** 2001; 292:2439-40.

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

Wu Q, Ma W, Zhu L et al. Oligo-microarray based on oligonucleotide immobilization on glass surface modified with activated acrylic acid-co-acrylamide copolymer. **J Biomed Mater Res B Appl Biomater** 2008 May 13.

Yamauchi H, Stearns V, Hayes DF. The Role of c-erbB-2 as a predictive factor in breast cancer. **Breast Cancer** 2001a; 8:171-83.

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

Yan W, Chen X. Targeted repression of bone morphogenetic protein 7, a novel target of the p53 family, triggers proliferative defect in p53-deficient breast cancer cells. **Cancer Res** 2007; 67:9117-24.

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

Yanagisawa K, Yasuda S, Kai M, et al. Diacylglycerol kinase alpha suppresses tumor necrosis factor-alpha-induced apoptosis of human melanoma cells through NF-kappaB activation. **Biochim Biophys Acta** 2007; 1771:462-74.

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

Zanghì G, Di Stefano G, Furci M, et al. Sentinel-node biopsy for breast cancer: our experience. **Ann Ital Chir** 2004; 75:325-8; discussion 328-30.

Zhu G, Reynolds L, Crnogorac-Jurcevic T, et al. Combination of microdissection and microarray analysis to identify gene expression changes between differentially located tumour cells in breast cancer. **Oncogene** 2003; 22:3742-8.

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

Zujewski J, Liu ET. The 1998 St. Gallen's Consensus Conference: an assessment. **J Natl Cancer Inst** 1998; 90:1587-9.

***ANEXOS***

---

## Anexo 1 - Quadro com informações das amostras utilizadas no estudo.

Identidade da amostra	EC	Idade (anos)	Peça	TNM	Grau	Marcador Molecular
1DCIS e 1IDC	Ila	45	IDC com componente intraductal associado	T2N0M0	grau II de SBR grau nuclear:2	ER - / PR - / p53 - / c-erB2 + (3+)
2DCIS e 2IDC	Ila	48	IDC com componente intraductal associado	T2N0M0	grau II de SBR grau nuclear:3	ER + / PR + / p53 + / c-erB2 +
3DCIS e 3IDC	Ila	43	IDC Multifocal com componente intraductal associado	T1cN0M0	grau II de SBR grau nuclear:3	ER + / PR - / p53 - / c-erB2 + (3+)
13DCIS e 13IDC	Ila	75	IDC com componente intraductal associado	T2N0M0	grau II de SBR grau nuclear:2	ER + 70% / PR + 75% / p53 + / c-erB2 + (2+)
15DCIS e 15IDC	IIb	47	IDC com componente intraductal associado	T2N1M0	grau I de SBR grau nuclear:2	ER + 40% / PR + 97% / p53 - / c-erB2 + (3+)
24DCIS e 24IDC	Ila	54	IDC com componente intraductal associado	T2N0M0	grau II de SBR grau nuclear:3	ER + 90% / PR + 80% / p53 + em menos de 10% / c-erB2 + (3+)
25DCIS e 25IDC	Ila	34	IDC com componente intraductal associado	T2N0M0	grau II de SBR grau nuclear:3	ER + 90% / PR + 80% / p53 + / c-erB2 + (3+)
33DCIS e 33IDC	I	38	IDC com componente intraductal associado	T1cN0M0	grau III de SBR grau nuclear:3	ER + 50% / PR - / p53 - / cerB2 -
50DCIS e 50IDC	III b	71	IDC com componente intraductal associado comedo solido	T4N2M0	grau III SBR grau nuclear 3	ER + / PR + / p53 + / c-erB2 + (+2)
51DCIS e 51IDC	II a	43	IDC com componente intraductal associado comedo solido	T2N0M0	grau III SBR grau nuclear 3	ER - / c-erB2 + (+3)
66DCIS e 66IDC	IIIb	44	IDC com componente intraductal associado tipo comedo	T4bN1M0	grau III SBR grau nuclear: 3	ER +(40%) / PR +(30%) / p53 + / cerbB2 -(1+)
69DCIS e 69IDC	IIb/IIIa	57	IDC com componente intraductal associado	T2N1M0/ pT2N2M0	grau II SBR grau nuclear 2	Biópsia-03/01/05-ER +(80%) / PR +(70%) / cebB2 +(2+)
75DCIS e 75IDC	IIb	43	IDC com componente intraductal associado tipo sólido	PT2N0M0	grau II SBR grau nuclear: 2/ DCIS grau nuclear:3	ER -/PR -/p53 +/c-erB2 + (3+)/Ki-67 19,4%
80DCIS e 80IDC	Ila	44	IDC + DCIS solido e cribriforme grau nuclear 2	pT1cpN1M0	grau II SBR grau nuclear: 3	ER + 80% / PR + 40% p53 - cerbB2: -

<b>81DCIS e 81IDC</b>	IIb	31	IDC com componente intraductal associado tipo sólido e carcinoma mucinoso	T2N1M0	grau II SBR grau nuclear 3	ER -/PR -/p53 +/c-erB2 + (3+)
<b>83DCIS e 83IDC</b>	IIa	48	IDC + com componente intraductal associado tipo cribriforme e sólido	T2N0M0	grau II SBR grau nuclear 2	ER + (98,5%) / PR +(96%) / p53 - / c-erB2 + (2+)
<b>43DCIS-0</b>	0	37	Carcinoma ductal <i>in situ</i> – Comedo carcinoma	T3N0M0	#	ER - / PR - / p53 +
<b>44DCIS-0</b>	0	44	DCIS multifocal, sólido e cribriforme	T0N0M0	grau nuclear:2 e áreas com grau nuclear:3	ER +(90%) / PR +(90%) / p53 - / cebB2 +
<b>46DCIS-0</b>	0	43	DCIS padrões sólido, cribriforme, micropapilar e comedo	T2N0M0	grau nuclear:3	ER +(100%) / PR +(100%) / p53 - / cebB2 +(3+)
<b>48DCIS-0</b>	0	52	Carcinoma intraductal sólido e cribriforme	T2N0M0	grau nuclear:3	ER +(100%) / PR +(50%) / p53 - (<5%) / cebB2 +(3+)
<b>49DCIS-0</b>	0	58	Carcinoma intraductal tipos sólido e comedo	T1cNoMo	grau nuclear:3	ER - / PR - / p53 - / c-erB2 + (3+)
<b>45DCIS</b>	0	55	DCIS residual tipo sólido e comedo com microcalcif determinantes /IDC	T1N0M0	grau nuclear III	ER - / PR - / p53 + / cebB2 + (3+)
<b>82DCIS</b>	IIIb	45	IDC c/ extenso comp. DCIS tipo sólido e cribriforme	T3N1Mo	grau II SBR grau nuclear 2	ER +(99%) / PR - / p53 - / cebB2 +(2+)
<b>85DCIS</b>	I	73	IDC + DCIS tipo cribriforme, sólido e papilífero	T1cNoMo pT3NoMo	grau II SBR grau nuclear 2	#
<b>86DCIS</b>	#	46	IDC c/ extenso DCIS cribriforme, sólido e comedo	T2N1Mo	grau II SBR grau nuclear 3	ER +(80%) / PR +(100%) / p53 + / cebB2 +(2+)
<b>87DCIS</b>	IIa	48	IDC + DCIS comedo	T2N1Mo	grau II SBR grau nuclear 3	ER -/PR -/p53 -/c-erB2 + (2+)
<b>88DCIS</b>		63	IDC multicêntrico com DCIS associado tipo comedo, sólido e cribriforme	T2N0M0	grau nuclear I	ER +(55%) / PR - (30%) / p53 - / cebB2 +(2+)

<b>3NO-DCIS</b>		39	IDC cm carcinoma intraductal associado	T2N0M0	#	ER +(50%) / PR +(10%) / p53 + /cebB2 (-)
<b>8NO-DCIS</b>		49	IDC cm carcinoma intraductal associado	T2N0M0	#	ER +(100%) / PR +(100%) / cebB2 -
<b>15N1-DCIS</b>	IIb	69	IDC cm carcinoma intraductal associado	T2N1M0	#	ER +(70%) / PR +(70%) /cebB2 + (+2)
<b>53IDC</b>	III a	43	IDC cm carcinoma intraductal associado tipo comedo	T2N2M0	grau III SBR grau nuclear 3	ER +(90%) / PR +(90%) / p53 - / cebB2 + (+2)
<b>56IDC</b>	I	66	IDC	T1cN0M0	grau nuclear 2	ER - / PR - / p53 - / c-erB2 -
<b>63IDC</b>	IIIa	54	IDC + DCIS comedo grau 3	T3N1Mo	grau III SBR grau nuclear 3	ER - / PR - / p53 - / c-erB2 - (1+)
<b>2NO-IDC</b>		39	IDC	T2N0M0	#	ER + 100%
<b>MN14</b>	#	25	Tec. Normal	#	#	#
<b>MN16</b>	#	29	Tecido fibroadiposo	#	#	#
<b>MN18</b>	#	29	Tec. Normal	#	#	#
<b>MN19</b>	#	44	Tec. Normal	#	#	#
<b>MN23</b>	#	39	Tec. Normal	#	#	ER - / PR - / p53 -

NE – Não encontrado

**Anexo 2** - Quadro com as informações da qualidade do RNA total e quantidade do RNA amplificado.

Identidade da Amostra	Tipo Histológico	Tipo de Célula Capturada	Qualidade do RNA Total		Quantidade do RNA amplificado (ug)
			Presença do fragmento GAPDH	Bionalyser-RIN ou razão 28 e 18S	
1DCIS e 1IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	12,88/27,44
2DCIS e 2IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	50,08/53,44
3DCIS e 3IDC	IDC Multifocal com componente intraductal associado	DCIS e IDC	SIM	----	109,76/126,46
13DCIS e 13IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	79,84/109,6
15DCIS e 15IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	40,4/22,84
24DCIS e 24IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	44,24/106,4
25DCIS e 25IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	54,08/43,12
33DCIS e 33IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	34,96/55,36
50DCIS e 50IDC	IDC com componente intraductal associado comedo solido	DCIS e IDC	SIM	----	23,12/37,28
51DCIS e 51IDC	IDC com componente intraductal associado comedo solido	DCIS e IDC	SIM	----	38,4/27,84
66DCIS e 66IDC	IDC com componente intraductal associado tipo comedo	DCIS e IDC	SIM	----	36,96/18,48
69DCIS e 69IDC	IDC com componente intraductal associado	DCIS e IDC	SIM	----	24/25,92
75DCIS e 75IDC	IDC com componente intraductal associado tipo sólido	DCIS e IDC	SIM	----	49,92/40,24
80DCIS e 80IDC	IDC + DCIS solido e cribriforme	DCIS e IDC	SIM	----	22,14/55,42
81DCIS e 81IDC	IDC com componente intraductal associado tipo sólido e carcinoma mucinoso	DCIS e IDC	SIM	----	34,05/52,48
83DCIS e 83IDC	IDC com componente intraductal associado tipo cribriforme e solido	DCIS e IDC	SIM	----	42,88/81,12
43DCIS-0	Carcinoma ductal <i>in situ</i> – Comedo carcinoma	DCIS	SIM	----	24,4
44DCIS-0	DCIS multifocal, sólido e cribriforme	DCIS	SIM	----	70,8
46DCIS-0	DCIS padrões sólido, cribriforme, micropapilar e comedo	DCIS	SIM	----	38,08
48DCIS-0	Carcinoma intraductal sólido e cribriforme	DCIS	SIM	----	17,68/26,96

<b>49DCIS-0</b>	Carcinoma intraductal tipos sólido e comedo	DCIS	SIM	---	26,24/15,8
<b>45DCIS</b>	DCIS residual tipo sólido e comedo com microcalcif determinantes /IDC	DCIS	SIM	---	78,48
<b>82DCIS</b>	IDC c/ extenso comp. DCIS tipo solido e cribriforme	DCIS	SIM	---	28
<b>85DCIS</b>	IDC + DCIS tipo cribriforme, solido e papilifero	DCIS	SIM	---	49,68
<b>86DCIS</b>	IDC c/ extenso DCIS cribriforme, solido e comedo	DCIS	SIM	---	25,72
<b>87DCIS</b>	IDC + DCIS tipo comedo	DCIS	SIM	---	61,28
<b>88DCIS</b>	IDC multicêntrico com DCIS associado tipo comedo, solido e cribriforme	DCIS	SIM	---	65,92
<b>1NO-DCIS</b>	IDC com componente intraductal associado	DCIS	---	Bandas 28 e 18S 1:1	---
<b>3NO-DCIS</b>	IDC com componente intraductal associado	DCIS	---	Bandas 28 e 18S 2:1	88,14
<b>8NO-DCIS</b>	IDC com componente intraductal associado	DCIS	---	Bandas 28 e 18S 1:1	37,05
<b>15N1-DCIS</b>	IDC com componente intraductal associado	DCIS	---	RIN 5,9	44,85
<b>17N0-DCIS</b>	IDC com componente intraductal associado	DCIS	---	RIN 2,4	---
<b>37N1-DCIS</b>	IDC com componente intraductal associado	DCIS	---	RIN 7,2/ 7	1,54
<b>53IDC</b>	IDC cm carcinoma intraductal associado tipo comedo	IDC	SIM	---	52,99
<b>56IDC</b>	IDC	IDC	SIM	---	40,6
<b>61IDC</b>	IDC	IDC	SIM	---	26,98
<b>63IDC</b>	IDC + DCIS tipo comedo	IDC	SIM	---	34,48
<b>2NO-IDC</b>	IDC	IDC	---	RIN 2,4/7,0	72,15
<b>10N0-IDC</b>	IDC	IDC	---	RIN 4,9/5,4	2,74
<b>11N0-IDC</b>	IDC	IDC	---	RIN 3,7/4,1	2,86
<b>15N0-IDC</b>	IDC	IDC	---	RIN 5,9	1,68
<b>19N1-IDC</b>	IDC	IDC	---	Bandas 28 e 18S 1:1	2,8
<b>MN14</b>	Tec. Normal	Células Normais	SIM	---	61,04
<b>MN16</b>	Tec. Normal	Células Normais	SIM	---	70,88
<b>MN18</b>	Tec. Normal	Células Normais	SIM	---	37,92
<b>MN19</b>	Tec. Normal	Células Normais	SIM	---	50,56
<b>MN23</b>	Tec. Normal	Células Normais	---	RIN 6,8/4,2	61,6

### Anexo 3 - Protocolo rna 6000 – bioanalyzer – guia de referência rápida

#### PREPARAÇÃO DO GEL (reagentes estocados na geladeira)

1. Colocar **550 µL** de *RNA 6000 Pico Gel* (matriz) num tubo e dar um **spin** rápido.
2. Centrifugar a **1500g** por **10 minutos** à temperatura ambiente.
3. Fazer alíquotas de **65 µL** do gel colocando em tubos livres de *Rnase* com capacidade para **0,5 mL**. Os tubos estão incluídos no kit.

#### PREPARAÇÃO DO MIX GEL-CORANTE

1. Deixar o reagente *RNA 6000 Pico* (corante) e o gel à temperatura ambiente durante **30 min**.
2. Dar um vortex no corante durante **10 seg** e em seguida dar um *spin* para depois adicionar **1 µL** do corante numa alíquota de **65 µL** do gel já preparado.
3. Vortexar bem a solução. Dar um *spin* no tubo a **13000g** durante **10 min** à temperatura ambiente.

#### LAVAGEM DA ESTAÇÃO

1. Lavagem do *chip* transparente presente dentro do aparelho com ~800uL de água Sigma
2. Descartar
3. Adicionar ~300uL de água Sigma e espalhar em todos os poços

#### TRANSFERÊNCIA DO MIX GEL-CORANTE

1. Colocar um chip novo de RNA na Estação de Chip (certificar se a localização do chip está na posição C)
2. Pipetar **9 µL** de mix Gel-corante no poço ■ (Cuidado para não formar bolhas).
3. Fechar a estação.
4. Apertar a seringa até que seja retida por todo o chip.
5. Aguardar exatamente **30 seg** e então liberar a trava.
6. Pipetar mais **9 µL** de mix Gel-corante nos poços **G**.
7. Descartar o restante do mix Gel-corante.

#### TRANSFERÊNCIA DA SOLUÇÃO CONDICIONANTE E MARCADOR

1. Pipetar **9 µL** do reagente condicionante no poço ● / ○ **CS**.
2. Pipetar **5 µL** do reagente marcador em todos os 12 poços de amostra e na cavidade marcada.
- 3.

#### TRANSFERÊNCIA DA AMOSTRA PADRÃO (LADDER) E DAS AMOSTRAS

1. Pipetar **1 µL** de amostra padrão – *LADDER* – no poço marcado.
2. Pipetar **1 µL** de amostra em cada uma dos 11 poços de amostra. Pipetar **1 µL** de marcador em cada poço de amostra não utilizada.
3. Colocar o *chip* no adaptador e vortexar durante **1 min a 2400 rpm**.
4. Analisar o *chip* no *Bioanalisador Agilent 2100*.

## Anexo 4 - Protocolo de Amplificação de RNA mensageiro.

### Método DNA Polimerase\ Amostras Microdissecadas a Laser

Baseado em GOMES et al. 2003. Modificado.

**Oligo dT (24)-T7 primer:** (5'AAA CGA CGG CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG CGC T(15)3')[57-pb]

**Nota:** As incubações são todas feitas em termocicladores.

É utilizado todo o material (RNA total) microdissecado a laser

**Obs:** A integridade e qualidade do RNA total dever ser verificado em BioAnalyser analisando as 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:

RNA total (máximo de 7 ul)

Água tratada com DEPC (H<sub>2</sub>O/DEPC) q.s.p. 7ul.

2ul de 0,1 ug/ul oligo dT(24)-T7 *primer*

Incubar 70 °C por 10 minutos

Gelo

##### Mix para reação da 1a. fita de cDNA:

4ul 5x *First strand buffer*

2ul 0,1 M DTT

2ul 10mM dNTP

1ul Rnase IN

2ul Superscript II

Volume total: 20 ul

*Spin*

Incubar a 42°C por 2 horas no termociclador

4 °C

#### 1.2 Síntese da 2a. fita de cDNA:

##### Reação:

20 ul cDNA (primeira fita)

20 ul Second strand buffer 5X

1 ul 10mM dNTP

1 ul E.coli DNA ligase

4 ul E. coli DNA polimerase I

1 ul RNase H

53 ul água tratada com DEPC

Volume total: 100ul

**Incubar:** à 16°C por 2 horas

Adicionar 2 ul de T4 DNA polimerase I (5U/ul)

**Incubar** : à 16°C por 5 min em seguida colocar no gelo

**Parar a reação:**

10 ul de 0,5M EDTA pH 8,0

**1.3 Purificação com Fenol : Clorofórmio : Álcool isoamílico (CLEAN-UP)**

Adicionar 100ul de Fenol:Clorofórmio:Isoamílico (25:24:1) pH 8.0

Vortex 30seg - 1 min

Incubar a temperatura ambiente por 5 min

Centrifugar por 15 min a 14000 rpm – (temperatura ambiente)

Cuidadosamente retirar fase aquosa e transferir para um tubo novo

Adicionar 100ul de Clorofórmio

Vortex 30seg - 1 min

Incubar a temperatura ambiente por 5 min

Centrifugar por 15 min a 14000 rpm (temperatura ambiente)

Cuidadosamente retirar fase aquosa e transferir para um tubo novo

**1.4 Precipitação**

Adicionar 40 ul de 3M Acetato de sódio (Na<sub>2</sub>OAC), pH 5.2

Adicionar 300 ul de etanol absoluto (-20C)

1 ul de acrilamida (5mg/mL)

Vortex 30seg - 1 min

Incubar a 80°C por 30 min

Centrifugar a 14000 rpm a 4°C por 30min

Remover com a pipeta o sobrenadante e reservar

Lavar 3x o *pellet* com 1ml de etanol 70% (-20C)

Centrifugar a 14000 rpm, a 4°C por 2 min

Secar a 37-42°C por 10 minutos para evaporar o etanol residual

Ressuspender em 14ul H<sub>2</sub>O/DEPC e colocar em banho-maria a 55°C por 10 min

Quantificar o dsDNA em diluição 1:100

**PS:** Reservar 3 ul de dsDNA e adicionar 17 ul de água para validação - Estocar a -20°C

**1.5 Transcrição *in vitro***

10µl dsDNA (entrar com máximo de 20µg de cDNA)

**Reação:**

7,5ul of 25mM rNTP (A,G,C and UTP, prepare apenas o necessário para a reação)

5 ul 5x *reaction buffer*

2,5 ul *enzyme mix (Rnase inhibitor and T7 phage polymerase)*

Incubar a 37°C por 6hs.

**1.6 Purificação do RNA usando Trizol**

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

Adicionar 25 ul de água/DEPC á reação

Prepare a centrifuga a 4°C

Descongelar a reação no gelo

Adicionar 1ml of Trizol em cada tubo

Incubar a temperatura ambiente por 5 min

Adicionar 200 ul clorofórmio para cada 1 ml de solução de Trizol.

*Vortex.*

Incubar a temperatura ambiente por 10 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.

Adicionar: ~ 500 ul de isopropanol

60 ul de NaAc

0,5 ul de glicogênio (20mg/mL)

*Vortex*

Incubar a temperatura ambiente por 30 minutos.

Centrifugar a 14.000 rpm por 30 minutos.

Lavar o pellet 3 X com 1 ml de solução 70% ETOH preparado com água DEPC

Centrifugar a 14.000 rpm por 2 minutos.

Secar o pellet a 37-42°C por alguns min (Max. 5 min)

Ressuspender em 7ul de água DEPC e colocar em banho-maria a 55°C por 10 min

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

## **2 Segundo round de amplificação (fase exponencial)**

### **2.1 Síntese da 1a. fita de cDNA:**

6ul aRNA-1ºround.

Adicionar 3ul (3ug/ul) *random hexamer* (por exemplo: dN6).

Incubar 70 °C por 10 minutos

4° C

#### **Reação:**

4ul 5x *First strand buffer*

2ul 0,1M DTT

2ul 10mM dNTP

1ul Rnase IN

2ul Superscript II (SS II)

Incubar a 42°C por 2 horas no termociclador

### **2.2 Síntese da 2a. fita de cDNA**

#### **Reação:**

20 ul de cDNA (reação)

10 ul Advantage PCR *buffer*

2ul 10mM dNTP

2ul de 0,5 ug/ul oligo dT(24)-T7 *primer*

0,7 ul RNase H

2 ul Advantage Polymerase

63,3 ul água tratada com DEPC

Volume final: 100 ul

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 5 ul 1M NaOH 2mM EDTA.

Incubar a 65°C por 10 minutos para inativação enzimática

### **2.3 Purificação com Fenol : Clorofórmio : Álcool isoamílico**

Adicionar 100ul de Fenol:Clorofórmio:Isoamílico (25:24:1) pH 8.0

Vortex bem

Incubar a temperatura ambiente por 5 min

Centrifugar por 15 min a 14000 rpm (temperatura ambiente)

Cuidadosamente retirar fase aquosa

Dispensar em um tubo novo

Adicionar 100ul de Clorofórmio

Vortex bem e incubar a temperatura ambiente por 5 min

Centrifugar por 15 min a 14000 rpm (temperatura ambiente)

Cuidadosamente retirar fase aquosa e dispensar em um tubo novo

Adicionar 40 ul de 3M Acetato de sódio (Na<sub>2</sub>OAC), pH 5.2

Adicionar 300 ul de etanol absoluto (-20C)

1 ul de acrilamida (5mg/mL)

Incubar a -80°C por 30 min

Centrifugar a 14000 rpm a 4°C por 30min

Remover com a pipeta o sobrenadante e reservar

Lavar 3x o *pellet* com 1ml de etanol 70% (-20C)

Centrifugar a 14000 rpm, a 4°C por 2min

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

Ressuspender em 14ul H<sub>2</sub>O/DEPC

Colocar em banho-maria a 55°C por 10 min

Quantificar o dsDNA em diluição 1:100

**PS:** Reservar 3 ul de dsDNA e adicionar 17 ul de água para validação -Estocar a -20°C

### **2.5 Transcrição *in vitro***

10µl dsDNA (entrar com máximo de 20µg de cDNA)

#### **Reação:**

7,5ul of 25mM rNTP (A,G,C and UTP, prepare apenas o necessário para a reação)

5 ul 5x *reaction buffer*

2,5 ul *enzyme mix (Rnase inhibitor and T7 phage polymerase)*

Incubar a 37°C por 6hs.

### **1.6 Purificação do RNA usando Trizol**

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

Adicionar 25 ul de água/DEPC á reação

Descongelar a reação no gelo

Adicionar 1ml of Trizol em cada tubo

Incubar a temperatura ambiente por 5 min

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

*Vortex.*

Incubar a temperatura ambiente por 10 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.

Adicionar: ~ 500 ul de isopropanol

60 ul de NaAc

0,5 ul de glicogênio (20mg/mL)

Incubar a temperatura ambiente por 30 minutos.

Centrifugar a 14.000 rpm por 30 minutos 4°C.

Lavar o pellet 3 X com 1 ml de solução 70% ETOH preparado com água DEPC

Centrifugar a 14.000 rpm por 2 minutos.

Secar o pellet a 37-42°C por alguns min

Ressuspender em 11ul de água DEPC.

Colocar em banho-maria a 55°C por 5 min

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

## Anexo 5 - Protocolo de Marcação e Hibridização em Lâmina de Vidro

### 1. SÍNTESE DO cDNA MARCADO COM FLUOROCROMOS.

RNA/oligo mix:

- 4 ug de aRNA
- 3 ul de dN6 ( 3 ug/ul)
- 1,2 ul controle<sup>1</sup> 4.5x (ou 1 ul controle 5x)

qsp: 15,9 ul

Incubar a 70°C por 10min em seguida colocar no gelo

Adicionar :

- 15,9 ul do RNA/oligo mix
  - 6 ul first strand buffer 5x
  - 3,0 ul 0,1M DTT
  - 0,6 ul Low C<sup>2</sup>
  - 0.5 ul Rnasin (Promega, cat # N251B, 40 U / ul).
  - 2ul dCTP<sup>3</sup> (25nmol; CY3 ou CY5, cat# PA53021 ou PA55021 respectivamente, amersham pharmacia).
  - 2,0 ul SSII reverse transcriptase (Life Technologies)
- 2 Volume final = 30,0 ul

Incubar a reacao por 2hs a 42°C

<sup>1</sup> 1 ul da mistura do RNAm controles 1X contém:

Q do fago λ "landmark" na concentração de 45 ng/ul.

Clone de *B. subtilis* trp gene (ATCC, 87485) 30 ng/ul

Clone de *B. subtilis* dapB gene (ATCC, 87485) 15 ng/ul

Clone de *B. subtilis* thrB gene (ATCC, 87485) 3 ng/ul

Clone de *B. subtilis* pheB gene (ATCC, 87485) 1,5 ng/ul

Clone de *B. subtilis* lysA gene (ATCC, 87485) 0,3 ng/ul

Esta mistura de genes está diluida em etanol 75%.

Adicionando 1ul desta mistura 1X em 25 a 50 ug de RNA total resulta em uma contaminação do RNA total de mamíferos com RNAs que correspondem a genes com níveis de transcrição de aproximadamente 1:20.000 (lysA), 1:4000 (pheB), 1:2000 (thrB), 1:400 (dapB), 1:200 (trp).

<sup>2</sup> dNTPs mix (Low C dDNTp mix): uso com dCTP marcado

Conc. Inicial	Vol.	Conc. final
dATP (100 mM)	5 µL	25 mM
dCTP (100 mM)	2 µL	10 mM
dGTP (100 mM)	5 µL	25 mM
dTTP (100 mM)	5 µL	25 mM
H <sub>2</sub> O DEPC q.s.p 20,0µL	3 µL	
Total	20 µL	

\*Incluídos no kit 18064-014, SSII reverse transcriptase, Life Technologies

<sup>3</sup> dNTPs mix (Low T dNTP mix): uso com dUTP marcado

Conc. Inicial	Vol.	Conc. final
dATP (100 mM)	5 µL	25 mM
dCTP (100 mM)	5 µL	25 mM
dGTP (100 mM)	5 µL	25 mM
dTTP (100 mM)	2 µL	10 mM
H <sub>2</sub> O DEPC q.s.p 20,0µL	3 µL	
Total	20 µL	

<sup>3</sup> O dCTP deve ser aliquotado em 3ul/tubo e guardado no escuro a -20°C. Durante as reações evitar sempre a luz, p/ ex: cobrindo tubos e aquecedores com papel alumínio.

### 2. Degradação do RNA – volume: volume

- 1,5 ul 0,5M EDTA
- 1,5ul 1M NaOH

***Incubar a 70°C / 20 min***

- Neutralizar a reação: Add 1,5 ul 1M HCl.
- Incubar no escuro enquanto prepara a coluna

***Purificar em coluna Autoseq G50 (item 3) ou Precipitação (item 4).  
Recomendamos em G50.***

**3. Purificação do cDNA em coluna Autoseq G50 (Amersham Pharmacia, cat # 27-5340-01):**

- Ressuspender a resina vortexando.
- Quebrar a ponta da coluna e desatarrachar a tampa com meia volta (~ 40°).
- Spin a 2000g / 1min com a coluna adaptada em tubo eppendorf 2 mL.
- Add cada reação de cDNA (+/- 34,5 ul) sscDNA no centro da resina de cada coluna, adicionando gotinha por gotinha com cuidado. Use uma coluna por reação de cDNA.
- Transferir coluna para tubo eppendorf 2,0 mL.
- Spin a 2000g / 1min com a coluna adaptada em tubo de 2,0 mL.
- Guardar eluato a 4°C ao abrigo da luz ou - 20°C overnight.
- Os cDNAs poderão ser utilizados para a hibridização em estação de hibridização.

**4. Pré-hibridização das lâminas:**

- Incubar lâmina em tubo Falcon 50ml ou caixa de lâmina contendo 5X SSC; 0,2% SDS; 1% BSA, 5X solução Denhardt's (Filtrada em membrana 0,22 ou 0,45 um).
- ***Dica para preparo do tampão:*** Inicialmente dissolver o SSC, a BSA e o Denhardt's em H<sub>2</sub>Odd aquecida e então adicionar o SDS e completar o volume com H<sub>2</sub>Odd.

***42°C / ≥ 6hs ou overnight***

- Lavar mergulhando 10X em H<sub>2</sub>Odd nova e centrifugar a 1000 rpm / 1 min/(tubo Falcon 50 ml).
- Utilizar a lâmina imediatamente, pois a eficiência de hibridização diminui se os slides secarem por mais de uma hora.

**5. Hibridização**

***Tampão de hibridização 2X:*** 10 X SSC; 0,2% SDS. Filtrar (filtro 0,22 ou 0,45 µm), aliquotar e guardar em temp. ambiente. Aquecer a 50° a 70°C antes do uso até ressuspender todo o precipitado.

**5.1. Estação de hibridização:**

- Misture as duas reações de sscDNA Cy3 e Cy5 purificadas.
- Secar no *speed vacuum* protegido contra a luz. (Opção: guardar a sonda a - 20°C overnight).
- Add 6,4 ul de H<sub>2</sub>O/Milli-Q
- Add 2 ul polyA DNA (2 ug /ul, Amersham-Pharmacia, cat # 27-7988-01).
- Add 2 ul Cot1 DNA (2 ug/ul, Invitrogen BRL, cat# 15279-011).
- Add 2 ul de BSA 10 mg/mL.
- Mexer com o dedo e deixar hidratando por alguns minutos
- Add 9,5 ul de solução Denhardt's 50X.
- Add 2ul de Esperma de Salmão 5ug/ul.
- Add 23,75 ul de Formamida deionizada

- Add 47,5 ul de tampão de hibridização previamente aquecido a 70°C e devidamente solubilizado.
- Spin – vortex - spin
- Mix com cuidado e incubar a 45°C até o momento de uso.
- Somente após preparada a sonda, iniciar as lavagens das lâminas que já deverão estar pré-hibridizando.
- Montar a lâmina já pré-hibridizada na hyb station. Aquecer a 70°C / 2 min. Manter nesta temperatura até adicionar o cDNA.
- Aquecer o cDNA mix a 95°C / 5 min. Spin breve e manter aquecido a 70°C.
- Adicionar o cDNA mix na câmara de hibridização e incubar a 42°C por 24 hs com agitação.
- Desmontar o cassette de hibridização da hyb station. Retirar a lâmina com ajuda de pinça e imergir imediatamente em SSC 2X.

#### **6. Lavagem das lâminas**

- Pré-aquecer os tampões a 37°C antes do uso.
- Lavar as lâminas sob agitação e acondicionadas em estante própria. Para um máximo de 4 lâminas utilizar em torno de 400ml de cada tampão.
- 1X em 2X SSC / 10 min.
- 2X em 0,1X SSC; 0,1% SDS / 10 min.
- 2X em 0,1 X SSC / 10 min.
- Spin a 1000 rpm / 2 min / RT em tubo Falcon 50ml.
- Scannear as lâminas

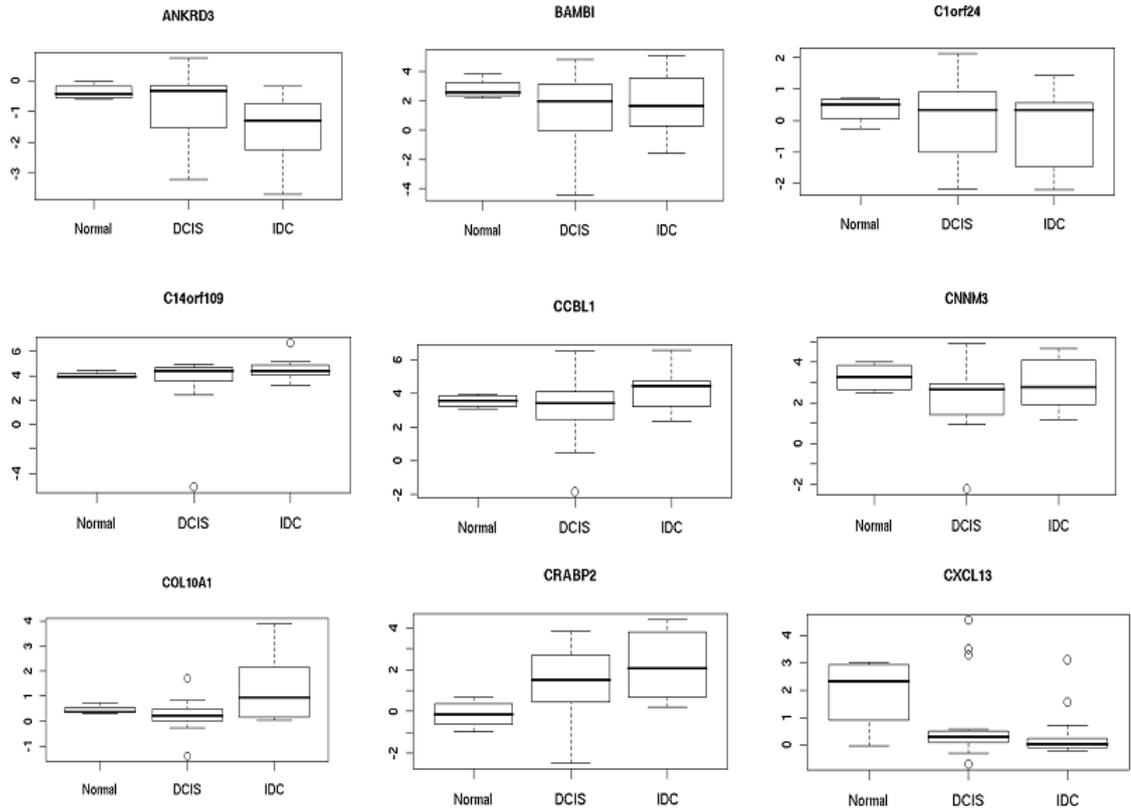
**Anexo 6 - Amostras Utilizadas No Estudo.**

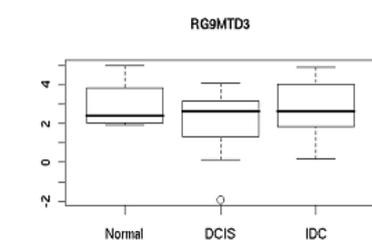
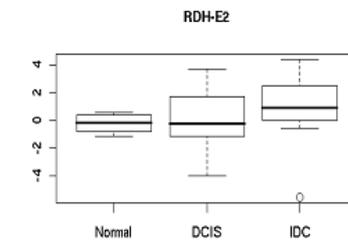
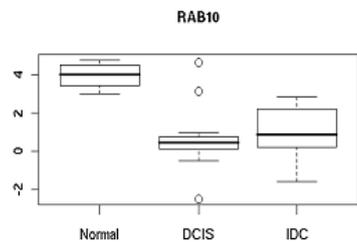
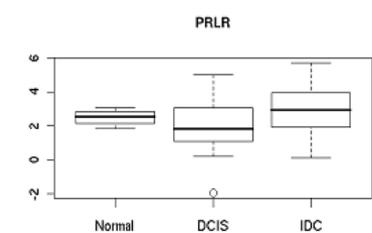
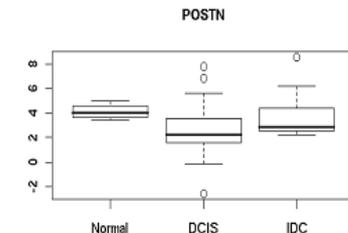
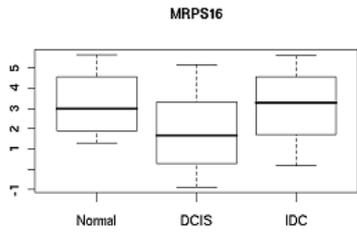
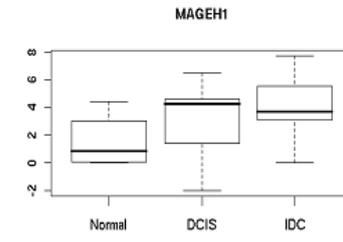
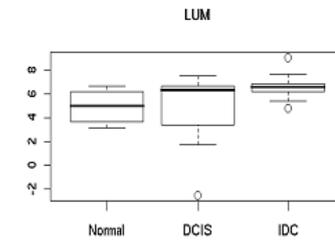
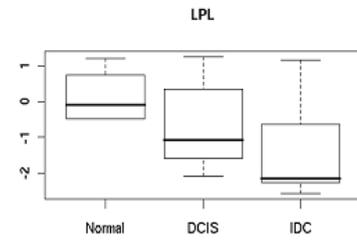
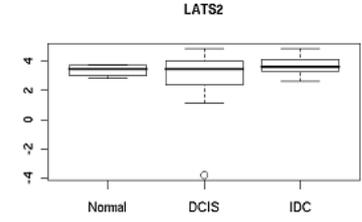
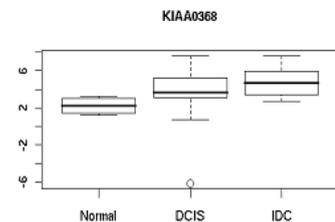
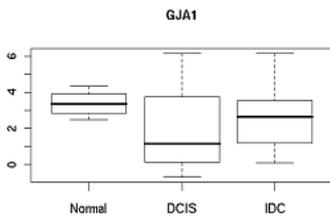
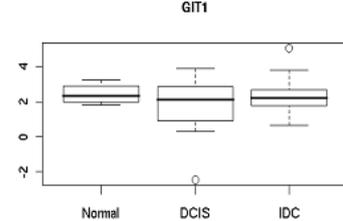
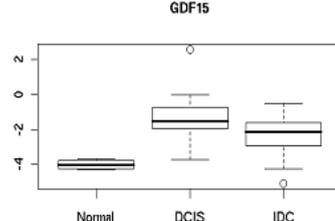
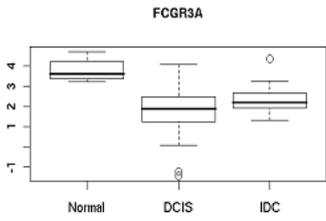
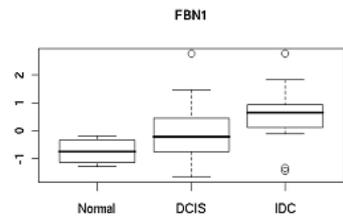
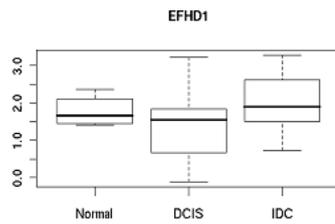
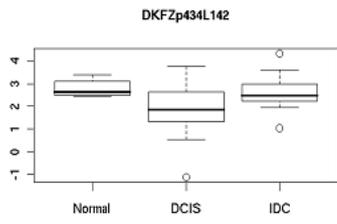
Plataforma 4.8K		Plataforma LV-01	
AMOSTRAS			
<u>PAREADAS</u>	<u>PROGRESSÃO</u>	<u>PAREADAS</u>	<u>PROGRESSÃO</u>
BP-LCM-1DCIS	<u>NORMAL</u>	BP-LCM-1DCIS	<u>NORMAL</u>
BP-LCM-1IDC	BP-LCM-MN14	BP-LCM-1IDC	BP-LCM-MN14
BP-LCM-2DCIS	BP-LCM-MN16	BP-LCM-2DCIS	BP-LCM-MN16
BP-LCM-2IDC	BP-LCM-MN18	BP-LCM-2IDC	BP-LCM-MN18
BP-LCM-3DCIS	BP-LCM-MN19	BP-LCM-3DCIS	BP-LCM-MN19
BP-LCM-3IDC	<u>DCIS Puro</u>	BP-LCM-3IDC	BP-LCM-MN23
BP-LCM-13DCIS	BP-LCM-43DCIS	BP-LCM-13DCIS	<u>DCIS Puro</u>
BP-LCM-13IDC	BP-LCM-44DCIS	BP-LCM-13IDC	BP-LCM-43DCIS-0
BP-LCM-15DCIS	BP-LCM-46DCIS	BP-LCM-25DCIS	BP-LCM-44DCIS-0
BP-LCM-15IDC	BP-LCM-48DCIS	BP-LCM-25IDC	BP-LCM-46DCIS-0
BP-LCM-24DCIS	BP-LCM-49DCIS	BP-LCM-69DCIS	BP-LCM-49DCIS-0
BP-LCM-24IDC	<u>DCIS/IDC</u>	BP-LCM-69IDC	<u>DCIS/IDC</u>
BP-LCM-25DCIS	P-LCM-2DCIS	BP-LCM-75DCIS	BP-LCM-45DCIS
BP-LCM-25IDC	BP-LCM-13DCIS	BP-LCM-75IDC	BP-LCM-69DCIS
BP-LCM-33DCIS	BP-LCM-25DCIS	BP-LCM-80DCIS	BP-LCM-75DCIS
BP-LCM-33IDC	BP-LCM-33DCIS	BP-LCM-80IDC	BP-LCM-82DCIS
BP-LCM-50DCIS	BP-LCM-66DCIS	BP-LCM-81DCIS	BP-LCM-83DCIS
BP-LCM-50IDC	BP-LCM-69DCIS	BP-LCM-81IDC	BP-LCM-85DCIS
BP-LCM-51DCIS	BP-LCM-75DCIS	BP-LCM-83DCIS	BP-LCM-86DCIS
BP-LCM-51IDC	BP-LCM-85DCIS	BP-LCM-83IDC	BP-LCM-87DCIS
BP-LCM-66DCIS	BP-LCM-86DCIS	<b>Pareadas - 10 casos</b>	BP-LCM-88DCIS
BP-LCM-66IDC	BP-LCM-87DCIS		BP-LCM-3NO-DCIS
BP-LCM-69DCIS	<u>IDC</u>		BP-LCM-8NO-DCIS
BP-LCM-69IDC	BP-LCM-1IDC		BP-LCM-15N1-DCIS
BP-LCM-75DCIS	BP-LCM-3IDC		<u>IDC</u>
BP-LCM-75IDC	BP-LCM-24IDC		BP-LCM-1IDC
BP-LCM-80DCIS	BP-LCM-50IDC		BP-LCM-2IDC
BP-LCM-80IDC	BP-LCM-51IDC		BP-LCM-3IDC
BP-LCM-81DCIS	BP-LCM-53IDC		BP-LCM-13IDC
BP-LCM-81IDC	BP-LCM-56IDC		BP-LCM-25IDC
BP-LCM-83DCIS	BP-LCM-80IDC		BP-LCM-53IDC
BP-LCM-83IDC	BP-LCM-81IDC		BP-LCM-56IDC
<b>Pareadas - 16 casos</b>	BP-LCM-83IDC		BP-LCM-63IDC
	<b>Normal - 4 casos</b>		BP-LCM-2NO-IDC
	<b>DCIS puro - 5 casos</b>		BP-LCM-80IDC
	<b>DCIS/IDC - 10 casos</b>		BP-LCM-81IDC
	<b>IDC - 10 casos</b>		<b>Normal - 5 casos</b>
			<b>DCIS puro - 4 casos</b>
			<b>DCIS/IDC - 12 casos</b>
			<b>IDC - 11 casos</b>

**Anexo 7 - Lista dos 33 genes diferencialmente expressos entre as amostras pareadas (DCIS x IDC) através do teste T de *Student* pareado ( $p < 0,01$  e Fold < |1,5| )**

Gene	Fold	p valor Pareado	pFDR	Descrição
LUM	-2.85	1,56E+04	0.35	lumican
KIAA0368	-2.34	4,97E+04	0.47	KIAA0368
SLC37A1	-2.26	3,73E+02	0.18	solute carrier family 37 (glycerol-3-phosphate transporter), member 1
RDH-E2	-2.21	2,80E+01	0.13	retinal short chain dehydrogenase reductase
POSTN	-2.12	1,92E+04	0.35	periostin, osteoblast specific factor
MRPS16	-2.08	2,12E+04	0.36	mitochondrial ribosomal protein S16
PRLR	-2.07	4,96E+04	0.47	prolactin receptor
STK3	-2.06	2,08E+02	0.15	serine/threonine kinase 3 (STE20 homolog, yeast)
COL10A1	-1.96	1,35E+04	0.35	collagen, type X, alpha 1(Schmid metaphyseal chondrodysplasia)
CCBL1	-1.95	9,67E+02	0.20	cysteine conjugate-beta lyase; cytoplasmic (glutamine transaminase K, kynurenine aminotransferase)
BAMBI	-1.925	1,90E+04	0.35	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)
C14orf109	-1.86	4,78E+04	0.47	chromosome 14 open reading frame 109
GJA1	-1.81	7,37E+03	0.31	gap junction protein, alpha 1, 43kDa (connexin 43)
MAGEH1	-1.74	1,83E+04	0.35	APR-1 protein
LATS2	-1.71	2,42E+04	0.38	LATS, large tumor suppressor, homolog 2 (Drosophila)
TMEFF2	-1.68	2,72E+04	0.39	transmembrane protein with EGF-like and two follistatin-like domains 2
CNNM3	-1.64	1,11E+04	0.34	cyclin M3
RG9MTD3	-1.63	1,92E+04	0.35	RNA (guanine-9-) methyltransferase domain containing 3
VMP1	-1.62	2,27E+04	0.37	likely ortholog of rat vacuole membrane protein 1
RAB10	-1.62	5,11E+03	0.28	RAB10, member RAS oncogene family
GIT1	-1.61	4,57E+04	0.47	G protein-coupled receptor kinase-interactor 1
TSTA3	-1.59	1,01E+04	0.33	tissue specific transplantation antigen P35B
CRABP2	-1.54	1,60E+04	0.35	cellular retinoic acid binding protein 2
FBN1	-1.53	2,21E+04	0.36	fibrillin 1 (Marfan syndrome)
FCGR3A	-1.53	4,88E+04	0.47	Fc fragment of IgG, low affinity IIIa, receptor for (CD16)
DKFZp434L142	-1.51	1,71E+04	0.35	hypothetical protein DKFZp434L142
EFHD1	-1.51	1,41E+03	0.20	EF hand domain containing 1
SLC35A4	-1.50	2,87E+04	0.40	solute carrier family 35, member A4
LPL	1.51	6,30E+03	0.30	lipoprotein lipase
ANKRD3	1.52	1,47E+03	0.20	ankyrin repeat domain 3
CXCL13	1.53	8,00E+02	0.20	chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)
C1orf24	1.60	1,97E+02	0.15	chromosome 1 open reading frame 24
GDF15	1.77	4,04E+03	0.27	growth differentiation factor 15

**Anexo 8 - Box Plot dos 33 genes com expressão alterada entre as amostras pareadas (Plataforma 4.8k)**





**Anexo 9** - Valores de diferença de expressão para cada par de amostra analisado através de qRT-PCR.

Gene	Microarray	qR-PCR - Amostras Pareadas									
		Diferença de Expressão (Fold) Pfaffl									
		1	2	3	4	5	6	7	8	9	10
<i>LUM</i>	-2,85	-2,8	1,3	1,0	-1,0	-104,9	8,2	-2,0	-6,6	6,3	-108,0
<i>RDH-E2</i>	-2,21	--	-1,8	-7,2	-1,2	1,1	-7,7	7,7	-14,0	-2,2	-5,9
<i>CRABP2</i>	-1,54	-1,1	-1,2	1,0	3,4	-2,1	-1,3	-1,9	-2,9	1,9	4,5
<i>KIAA0368</i>	-2,34	1,3	1,5	-1,1	-2,2	1,1	-2,6	1,1	-1,5	1,2	2,8
<i>CXCL13</i>	1,53	--	--	1,5	2,7	--	--	1,4	25,3	1,1	12,2
<i>FBN1</i>	-1,53	-1,9	1,3	2,7	-3,1	-11,1	5,7	-1,4	-1,9	8,8	-73,0
<i>POSTN</i>	-2,12	-2,5	2,0	1,5	-6,3	-124,0	14,1	-2,2	-3,4	5,9	-60,4
<i>GDF15</i>	1,77	-1,1	7,5	4,4	3,2	7,4	--	-1,8	1,0	-16,9	--

**Anexo 10** - Lista dos 90 genes diferencialmente expressos entre DCIS puro, DCIS/IDC e IDC ( $p < 0,01$ ).

Gene	DCIS/IDC x DCIS puro	IDC x DCIS puro	IDC x DCIS/IDC	Aliases	Descrição
ADCK5	-1,28	-1,29	-1,01	FLJ35454, MGC126708	aarF domain containing kinase 5
ADCY2	1,62	1,05	-1,54	FLJ16822, FLJ45092, HBAC2, KIAA1060, MGC133314	adenylate cyclase 2 (brain)
ADFP	-2,79	-2,26	1,24	ADRP, MGC10598	adipose differentiation-related protein
ALMS1	-3,80	-2,79	1,36	ALSS, DKFZp686A118, DKFZp686D1828, KIAA0328	Alstrom syndrome 1
AZGP1	-2,94	-1,52	1,93	ZA2G, ZAG	AZGP1- alpha-2-glycoprotein 1
B4GALT5	-2,77	-2,23	1,24	RP5-1063B2.1, B4Gal-T5, BETA4-GALT-IV, MGC138470, beta4Gal-T5, beta4GalT-V, gt-V	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5
BTRC	-1,70	-1,36	1,25	RP11-529I10.2, BETA-TRCP, FBW1A, FBXW1A, FWD1, MGC4643, bTrCP, bTrCP1, betaTrCP	beta-transducin repeat containing
C10orf26	-4,06	-2,81	1,44	RP11-753C18.10, FLJ20154, FLJ20367, MGC125648, OPAL1	chromosome 10 open reading frame 26
C12orf22	-2,32	-2,24	1,04	#N/A	
C13orf23	2,41	1,52	-1,59	RP11-50D16.2, FLJ12661, FLJ23780, bA50D16.2	chromosome 13 open reading frame 23
C13orf24	-2,49	-2,07	1,20	KIAA1008, PIBF1, RP11-505F3.1	chromosome 13 open reading frame 24
C16orf5	-5,45	-3,14	1,74	CDIP, I1	chromosome 16 open reading frame 5
C17orf63	-2,26	-1,63	1,38	FLJ10700	chromosome 17 open reading frame 63
C1orf21	-2,25	-1,56	1,44	RP4-768P8.1, PIG13	chromosome 1 open reading frame 21
C1orf66	-5,32	-5,96	-1,12	CGI-41	chromosome 1 open reading frame 66
C20orf42	-4,26	-2,70	1,58	RP5-1056H1.1, DTGCU2, FLJ20116, FLJ23423, KIND1, URP1	chromosome 20 open reading frame 42
CDC45L	-2,90	-2,01	1,44	CDC45, CDC45L2, PORC-PI-1	CDC45 cell division cycle 45-like ( <i>S. cerevisiae</i> )
CHRNB1	-5,11	-3,05	1,67	ACHRB, CHRNB, CMS1D, CMS2A, SCCMS	cholinergic receptor, nicotinic, beta 1 (muscle)
CHST10	-5,07	-4,68	1,08	HNK-1ST, HNK1ST, MGC17148	carbohydrate sulfotransferase 10
CLIPR-59	-1,98	-2,85	-1,44	CLIPR59, DKFZp586N1922, FLJ33413	CLIP-170-related protein
COL17A1	-2,52	-3,53	-1,40	RP11-16H23.2, BA16H23.2, BP180, BPAG2, KIAA0204, LAD-1	collagen, type XVII, alpha 1
CPNE3	-2,38	-1,76	1,35	CPN3, KIAA0636, PRO1071	copine III
CTTNBP2NL	-4,84	-3,27	1,48	DKFZp547A023, FLJ13278	CTTNBP2 N-terminal like
DDX50	1,85	-1,01	-1,87	GU2, GUB, MGC3199, RH-II/GuB	DEAD (Asp-Glu-Ala-Asp) box polypeptide 50

Gene	DCIS/IDC x DCIS puro	IDC x DCIS puro	IDC x DCIS/IDC	Aliases	Descrição
DENND3	-3,51	-1,85	1,89	KIAA0870	KIAA0870 protein
DHRS12	-2,45	-2,27	1,08	FLJ13639	dehydrogenase/reductase (SDR family) member 12
DHX35	-3,07	-2,38	1,29	C20orf15, DDX35, FLJ22759, KAIA0875	DEAH (Asp-Glu-Ala-His) box polypeptide 35
DOCK4	-1,77	-1,70	1,04	WUGSC:H_GS034D21.1, FLJ34238, KIAA0716, MGC134911, MGC134912	dedicator of cytokinesis 4
EPHA1	-2,15	-1,74	1,24	EPH, EPHT, EPHT1	EPH receptor A1
EPOR	-2,99	-2,64	1,13	MGC138358	erythropoietin receptor
ERC1	-3,86	-2,41	1,60	Cast2, ELKS, KIAA1081, MGC12974, RAB6IP2	ELKS/RAB6-interacting/CAST family member 1
FAM40B	-2,77	-2,26	1,22	homolog of yeast FAR11 protein 2; hypothetical protein LOC57464	family with sequence similarity 40, member B
FCN1	-3,79	-2,51	1,51	FCNM	ficolin (collagen/fibrinogen domain containing) 1
FUBP3	-1,59	-1,73	-1,09	FBP3	far upstream element (FUSE) binding protein 3
GBGT1	-4,06	-2,77	1,46	A3GALNT, FS, MGC44848, RP11-326L24.6, UNQ2513	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1
GIPC1	-2,74	-1,98	1,38	C19orf3, GIPC, GLUT1CBP, Hs.6454, IIP-1, MGC15889, MGC3774, NIP, RGS19IP1, SEMCAP, SYNECTIIN, TIP-2	GIPC PDZ domain containing family, member 1
GOSR2	-3,97	-2,70	1,47	Bos1, GS27, MEMBRIN	golgi SNAP receptor complex member 2
HECTD1	-2,33	-1,51	1,54	FLJ38315, KIAA1131	HECT domain containing 1
HK3	-4,64	-3,28	1,41	HKIII, HXK3	hexokinase 3 (white cell)
INOC1	-2,11	-1,97	1,07	INO80	INO80 complex homolog 1 (S. cerevisiae)
INPP5A	1,38	-1,53	-2,11	RP11-288G11.1, 5PTASE, DKFZp434A1721, MGC116947, MGC116949	inositol polyphosphate-5-phosphatase
IRF8	-4,33	-3,36	1,29	H-ICSBP, ICSBP, ICSBP1, IRF-8	interferon regulatory factor 8
ITPKC	-3,26	-1,55	2,11		inositol 1,4,5-trisphosphate 3-kinase C
KIAA0664	-3,45	-2,54	1,36	CLU1, DKFZp686L05242, FLJ23672	KIAA0664 protein
KIAA0748	-2,38	-2,42	-1,02	#N/A	KIAA0748
KIAA1211	2,30	2,49	1,09	hypothetical protein LOC57482	KIAA1211 protein
LPXN	-3,20	-2,58	1,24	LDPL	leupaxin
LRCH2	-2,33	-1,60	1,45	RP6-204F4.4, KIAA1495, MGC150671, dA204F4.4	leucine-rich repeats and calponin homology (CH) domain containing 2
LSM4	-2,85	-2,17	1,32	YER112W	LSM4 homolog, U6 small nuclear RNA associated (S. cerevisiae)

Gene	DCIS/DC x DCIS puro	IDC x DCIS puro	IDC x DCIS/DC	Aliases	Descrição
<i>LTBP3</i>	-2,74	-2,41	1,14	<i>DKFZP586M2123, FLJ33431, FLJ39893, LTBP-3, LTBP2, pp6425</i>	<i>latent transforming growth factor beta binding protein 3</i>
<i>MAPRE2</i>	-2,62	-1,65	1,59	<i>EB1, EB2, RP1</i>	<i>microtubule-associated protein, RP/EB family, member 2</i>
<i>MARK3</i>	-4,91	-3,24	1,51	<i>CTAK1, KP78, PAR1A</i>	<i>MAP/microtubule affinity-regulating kinase 3</i>
<i>MBD3</i>	-4,58	-3,14	1,46		<i>methyl-CpG binding domain protein 3</i>
<i>MED10</i>	2,15	1,60	-1,34	<i>L6, MGC5309, NUT2, TRG20</i>	<i>mediator of RNA polymerase II transcription, subunit 10 homolog</i>
<i>MGC21874</i>	-4,70	-2,72	1,73	<i>ADA2(beta), ADA2B</i>	<i>transcriptional adaptor 2 (ADA2 homolog, yeast)-beta</i>
<i>MRPS17</i>	-2,23	-2,23	-1,00	<i>HSPC011, MRP-S17, RPMS17</i>	<i>mitochondrial ribosomal protein S17</i>
<i>NCDN</i>	-5,19	-3,05	1,70	<i>KIAA0607</i>	<i>neurochondrin</i>
<i>NEDD9</i>	-6,02	-2,89	2,08	<i>CAS-L, CASL, HEF1, dJ49G10.2, dJ76112.1</i>	<i>neural precursor cell expressed, developmentally down-regulated 9</i>
<i>OSBPL7</i>	-4,91	-3,00	1,64	<i>MGC71150, ORP7</i>	<i>oxysterol binding protein-like 7</i>
<i>PCDH10</i>	-3,01	-2,50	1,21	<i>DKFZP761O2023, KIAA1400, MGC133344, OL-PCDH, PCDH19</i>	<i>protocadherin 10</i>
<i>PCLKC</i>	-2,75	-2,41	1,14	<i>FLJ20124, FLJ2038</i>	<i>protocadherin LKC</i>
<i>PEPD</i>	-3,65	-2,73	1,34	<i>MGC10905, PROLIDASE</i>	<i>peptidase D</i>
<i>PHTF1</i>	2,42	2,02	-1,20	<i>PHTF</i>	<i>putative homeodomain transcription factor 1</i>
<i>PITPNM2</i>	-4,28	-3,37	1,27	<i>KIAA1457, NIR3, RDGB2, RDGBA2</i>	<i>phosphatidylinositol transfer protein, membrane-associated 2</i>
<i>PLA2G5</i>	-2,17	-1,59	1,36	<i>DKFZp686C2294, GV-PLA2, MGC46205, PLA2-10, hVPLA(2)</i>	<i>phospholipase A2, group V</i>
<i>PMPCA</i>	-4,20	-3,61	1,16	<i>Alpha-MPP, INPP5E, KIAA0123, MGC104197</i>	<i>peptidase (mitochondrial processing) alpha</i>
<i>POLD3</i>	-2,69	-1,87	1,44	<i>KIAA0039, MGC119642, MGC119643, P66, P68</i>	<i>polymerase (DNA-directed), delta 3, accessory subunit</i>
<i>RAB3B</i>	1,01	1,53	1,52		<i>RAB3B, member RAS oncogene family</i>
<i>RABEPK</i>	-4,94	-5,60	-1,14	<i>RAB9P40, DKFZp686P1077, bA65N13.1, p40</i>	<i>Rab9 effector protein with kelch motifs</i>
<i>REC8L1</i>	-3,37	-3,66	-1,09	<i>HR21spB, MGC950, REC8, Rec8p</i>	<i>REC8-like 1 (yeast)</i>

Gene	DCIS/IDC x DCIS puro	IDC x DCIS puro	IDC x DCIS/IDC	Aliases	Descrição
RGMB	-2,28	-2,08	1,10	DRAGON, FLJ90406, MGC86970	RGM domain family, member B
RPA1	1,57	-1,14	-1,78	HSSB, REPA1, RF-A, RP-A, RPA70	replication protein A1, 70kDa
RPL3	-4,02	-3,84	1,05	MGC104284, TARBP-B	ribosomal protein L3
SAMD4A	-4,11	-2,67	1,54	DKFZp434H0350, KIAA1053, SAMD4, SMG, SMGA, Smaug, Smaug1	sterile alpha motif domain containing 4
SHC1	-2,54	-1,41	1,80	FLJ26504, SHC, SHCA, p52SHC, p66, p66SHC	SHC (Src homology 2 domain containing) transforming protein 1
SLC6A20	-3,54	-3,61	-1,02	SIT1, XT3, Xtrp3	solute carrier family 6 (proline IMINO transporter), member 20
SLC9A5	-2,32	-2,44	-1,05	NHE5	solute carrier family 9 (sodium/hydrogen exchanger)
SNRP70	-1,83	-1,28	1,44	RNPU1Z, RPU1, U170K, U1AP, U1RNP	small nuclear ribonucleoprotein 70kDa polypeptide (RNP antigen)
SNX6	1,49	-1,28	-1,90	MGC3157, MSTP010, TFAF2	sorting nexin 6
SOX13	-3,58	-2,47	1,45	RP11-74C13.1, ICA12, MGC117216, SRY-box 13, Sox-13	SRY (sex determining region Y)-box 13
STX11	-3,18	-2,32	1,37	FHL4, HLH4, HPLH4	syntaxin 11
TBL3	-1,63	-1,60	1,02	SAZD	transducin (beta)-like 3
TEAD2	-1,81	-1,42	1,27	ETF, TEF-4, TEF4	TEA domain family member 2
TRAP1	-3,50	-2,82	1,24	HSP75, HSP90L	TNF receptor-associated protein 1
TXNDC11	-6,00	-2,82	2,12	EFP1; LOC51061	thioredoxin domain containing 11
TXNL2	-4,24	-3,29	1,29	PICOT	thioredoxin-like 2
UTP20	-3,16	-2,18	1,45	DRIM	UTP20, small subunit (SSU) processome component, homolog (yeast). Other designation - down-regulated in metastasis
VPS29	-2,92	-4,56	-1,56	DC15, DC7, DKFZp564F0223, FLJ20492, PEP11	vacuolar protein sorting 29 (yeast)
YIPF5	-1,20	1,23	1,47	DKFZp313L2216, FinGER5, SB140, SMAP-5, SMAP5, YIP1A	Yip1 domain family, member 5
	-2,01	-2,47	-1,23		(NUT2, <i>S. cerevisiae</i> )

**Anexo 11** - Lista dos 147 genes supostamente envolvidos nos estágios iniciais do processo de progressão do DCIS na plataforma 4.8k ( $p < 0,01$  e diferença de expressão (Fold)  $> |2|$ ).

Gene	Fold (DCIS/DC x DCIS puro)	seq_acc	Aliases	Descrição
FN3K	-6,30	BE159100	-	fructosamine 3 kinase
NEDD9	-6,02	BF985269	CAS- L CASL HEF1 dJ49G10.2 d J761I2.1	neural recursor cell expressed, developmentally down-regulated 9
TXNDC11	-6,00	BE827235	EFP1	thioredoxin domain containing 11
C16orf5	-5,45	BE164925	CDIP 1	chromosome 16 open reading frame 5
C1orf66	-5,32	BE696370	CGI-41	chromosome 1 open reading frame 66
NCDN	-5,19	BF881774	KIAA0607	neurochondrin
CHRNB1	-5,11	BE926694	AChRB CHRNB CMS1D C MS2A SCCMS	cholinergic receptor, nicotinic, beta 1 (muscle)
CHST10	-5,07	AW606362	HNK- 1S HNK1S MGC17148	carbohydrate sulfotransferase 10
RABEPK	-4,94	sem_annotacao		
MARK3	-4,91	BG990012	CTAK1 KP78 PAR1A	MAP/microtubule affinity-regulating kinase 3
OSBPL7	-4,91	BF373541	MGC71150 ORP7	oxysterol binding protein-like 7
CTTNBP2NL	-4,84	BF911807	DKFZp547A023 FLJ13278	CTTNBP2 N-terminal like
MGC21874	-4,70	AW837140	ADA2(beta) ADA2B	transcriptional adaptor 2 (ADA2 homolog, yeast)-beta
MBD3	-4,58	BG009490	-	methyl-CpG binding domain protein 3
IRF8	-4,33	BF892591	H- ICSBP ICSBP ICSBP1 IRF- 8	interferon regulatory factor 8
PITPNM2	-4,28	BF881789	KIAA1457 NIR3 RDGB2 R DGBA2	phosphatidylinositol transfer protein, membrane-associated 2
C20orf42	-4,26	BF811913	DTGCU2 FLJ20116 FLJ23 423 KIND1 URP1	chromosome 20 open reading frame 42
TXNL2	-4,24	BQ317588	FLJ11864 PICOT	thioredoxin-like 2
PMPCA	-4,20	BE168869	Alpha- MPP INPP5E KIAA0123 M GC104197	peptidase (mitochondrial processing) alpha
SAMD4A	-4,11	AW889289	DKFZp434H0350 KIAA105 3 SAMD4 SMG SMGA Sm aug Smaug1	sterile alpha motif domain containing 4A
GBGT1	-4,06	BF985126	A3GALNT FS MGC44848  RP11-326L24.6 UNQ2513	globoside alpha-1,3-N- acetylgalactosaminyltransferase 1
C10orf26	-4,06	BF805138	FLJ20154 FLJ20367 MGC1 25648 OPAL1	chromosome 10 open reading frame 26
TARDBP	-4,03	AW939125	TDP-43	TAR DNA binding protein
RPL3	-4,02	BE155891	MGC104284 TARBP-B	ribosomal protein L3
GOSR2	-3,97	sem_annotacao		
ERC1	-3,86	BF858030	Cast2 ELKS KIAA1081 MG C12974 RAB6IP2	ELKS/RAB6-interacting/CAST family member 1
ALMS1	-3,80	BE935985	ALSS DKFZp686A118 DKF Zp686D1828 KIAA0328	Alstrom syndrome 1
FCN1	-3,79	BF897902	FCNM	ficolin (collagen/fibrinogen domain containing) 1

Gene	Fold (DCIS/IDC x DCIS puro)	seq_acc	Aliases	Descrição
PEPD	-3,65	BQ372768	MGC10905 PROLIDASE	peptidase D
RARRES3	-3,64	BE926700	HRASLS4 MGC8906 RIG1 TIG3	retinoic acid receptor responder (tazarotene induced) 3
UBXD1	-3,63	AW793772	DKFZp667D109 UBXDC2	UBX domain containing 1
C20orf23	-3,60	BF371959	KISC20ORF SNX23	chromosome 20 open reading frame 23
SOX13	-3,58	BF912572	ICA12 MGC117216 SRY-box 13 Sox-13	SRY (sex determining region Y)-box 13
SLC6A20	-3,54	AW816486	MGC161475 SIT1 XT3 Xtrp3	solute carrier family 6 (proline IMINO transporter), member 20
RBM9	-3,52	AW383995	Fox-2 HRNBP2 RTA dJ106I20.3 fxh	RNA binding motif protein 9
DENND3	-3,51	sem_annotacao		
TRAP1	-3,50	BQ312213	HSP75 HSP90L	TNF receptor-associated protein 1
KIAA0664	-3,45	AW604853	CLU1 DKFZp686L05242 DKFZp686M09233 FLJ23672	KIAA0664
HSPA6	-3,39	AW797111	-	heat shock 70kDa protein 6 (HSP70B')
REC8	-3,37	AW579882	HR21spB MGC950 REC8L1 Rec8p	REC8 homolog (yeast)
IFT57	-3,30	BE156247	ESRRBL1 FLJ10147 HIPPI MHSA4R2	intraflagellar transport 57 homolog (Chlamydomonas)
FGFBP1	-3,28	sem_annotacao		
ITGB2	-3,27	BF897511	CD18 LAD LCAMB LFA-1 MAC-1 MF17 MF17	integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)
ITPKC	-3,26	CV394766	-	inositol 1,4,5-trisphosphate 3-kinase C
LPXN	-3,20	BE711695	LDPL	leupaxin
STX11	-3,18	BF892936	FHL4 HLH4 HPLH4	syntaxin 11
UTP20	-3,16	BE926620	DRIM	UTP20, small subunit (SSU) processome component, homolog (yeast)
RHOU	-3,12	sem_annotacao		
DHX35	-3,07	BF858113	C20orf15 DDX35 FLJ22759 KAIA0875	DEAH (Asp-Glu-Ala-His) box polypeptide 35
C3AR1	-3,04	BF900259	AZ3B C3AR HNFAG09	complement component 3a receptor 1
PCDH10	-3,01	BI039957	DKFZP761O2023 KIAA1400 MGC133344 OL-PCDH PCDH19	protocadherin 10
UFD1L	-3,01	AW175974	UFD1	ubiquitin fusion degradation 1 like (yeast)
P4HB	-3,01	BE710887	DSI ERBA2L GIT PDI PDIA1 PHDB PO4DB PO4HB PROHB	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide
EPOR	-2,99	BE931886	MGC138358	erythropoietin receptor
SYN1	-2,97	BE702745	SYN1a SYN1b SYNI	synapsin I
DGCR2	-2,96	sem_annotacao		
AZGP1	-2,94	sem_annotacao		alpha-2-glycoprotein 1, zinc
IMMT	-2,92	sem_annotacao		inner membrane protein
PPTC7	-2,92	BF959926	DKFZp686M07120 MGC133072 TA-PP2C	PTC7 protein phosphatase homolog (S. cerevisiae)

Gene	Fold (DCIS/IDC x DCIS puro)	seq_acc	Aliases	Descrição
ARHGAP9	-2,92	BF892905	10C FLJ16525 MGC1295  RGL1	Rho GTPase activating protein 9
CDC45L	-2,90	BE087469	CDC45 CDC45L2 PORC- PI-1	CDC45 cell division cycle 45-like ( <i>S. cerevisiae</i> )
LSM4	-2,85	sem_annotacao		
HLTF	-2,84	BE768509	HIP116 HIP116A HLTF1 R NF80 SMARCA3 SNF2L3  ZBU1	helicase-like transcription factor
CAMP	-2,84	BQ334335	CAP18 FALL- 39 FALL39 HSD26 LL37	cathelicidin antimicrobial peptide
ZBTB5	-2,82	BF848481	-	zinc finger and BTB domain containing 5
ADFP	-2,79	BF086797	ADRP MGC10598	adipose differentiation-related protein
FAM40B	-2,77	BF881173	-	family with sequence similarity 40, member B
B4GALT5	-2,77	sem_annotacao		
PCLKC	-2,75	AW752023	FLJ20124 FLJ20383 MGC1 63154	protocadherin LKC
LTBP3	-2,74	BE931661	DKFZP586M2123 FLJ3343 1 FLJ39893 FLJ42533 FLJ 44138 LTBP- 3 LTBP2 pp6425	latent transforming growth factor beta binding protein 3
GIPC1	-2,74	BG010288	C19orf3 GIPC GLUT1CBP  Hs.6454 IIP- 1 MGC15889 MGC3774 N  P RGS19IP1 SEMCAP SY NECTIN TIP-2	GIPC PDZ domain containing family, member 1
KIF1A	-2,71	BF923058	ATSV C2orf20 DKFZp686l 2094 FLJ30229 HUNC- 104 MGC133285 MGC133 286 UNC104	kinesin family member 1A
FCGR3A	-2,70	AW805776	CD16 CD16a FCG3 FCGR 3 GFR3	Fc fragment of IgG, low affinity IIIa, receptor (CD16a)
POLD3	-2,69	sem_annotacao	KIAA0039 MGC119642 MG C119643 P66 P68	polymerase (DNA-directed)
MAPRE2	-2,62	BF884651	EB1 EB2 RP1	microtubule-associated protein, RP/EB family, member 2
BCHE	-2,62	BG877797	CHE1 E1	butyrylcholinesterase
NGDN	-2,59	AW878300	C14orf120 DKFZP564O09 2 LCP5 NGD pd-2	neuroguidin, EIF4E binding protein
ARFIP1	-2,55	AW062376	HSU52521 MGC117369	ADP-ribosylation factor interacting protein 1 (arfaptin 1)
KDR	-2,55	BF956971	CD309 FLK1 VEGFR VEG FR2	kinase insert domain receptor (a type III receptor tyrosine kinase)
SHC1	-2,54	AW606299	FLJ26504 SHC SHCA p52 SHC p66 p66SHC	SHC (Src homology 2 domain containing) transforming protein 1
GPR77	-2,52	sem_annotacao		
COL17A1	-2,52	BE703135	BA16H23.2 BP180 BPAG2  KIAA0204 LAD-1	collagen, type XVII, alpha 1
CLINT1	-2,50	BE768344	CLINT ENTH EPN4 EPNR  EPSINR KIAA0171	clathrin interactor 1
NUP50	-2,50	BF154492	MGC39961 NPAP60 NPAP 60L	nucleoporin 50kDa
C13orf24	-2,49	BE165549	KIAA1008 PIBF1 RP11- 505F3.1	chromosome 13 open reading frame 24
PPP2R3A	-2,48	BE711719	PPP2R3 PR130 PR72	protein phosphatase 2 (formerly 2A), regulatory subunit B", alpha
AOF2	-2,46	sem_annotacao		

Gene	Fold (DCIS/IDC x DCIS puro)	seq_acc	Aliases	Descrição
DHRS12	-2,45	BE163454	FLJ13639	dehydrogenase/reductase (SDR family) member 12
PIAS2	-2,43	BE708369	MGC102682 MIZ1 PIASX-ALPHA PIASX-BETA SIZ2 ZMIZ4 miz	protein inhibitor of activated STAT, 2
KBTBD10	-2,42	BE156689	SARCOSIN	kelch repeat and BTB (POZ) domain containing 10
ZFP36L1	-2,40	BE767943	BRF1 Berg36 ERF-1 ERF1 RNF162B TIS11B cMG1	zinc finger protein 36, C3H type-like 1
KIAA0748	-2,38	BF918997	HSPC257	KIAA0748
CPNE3	-2,38	BE087056	CPN3 KIAA0636 PRO1071	copine III
KCTD15	-2,35	sem_annotacao		
LRCH2	-2,33	BF856789	KIAA1495 MGC150671 dA204F4.4	leucine-rich repeats and calponin homology (CH) domain containing 2
FAM130A1	-2,32	BE156108	C12orf2 C12orf22 FLJ25576 TAIP-12	family with sequence similarity 130, member A1
SLC9A5	-2,32	BF742127	NHE5	solute carrier family 9 (sodium/hydrogen exchanger), member 5
KPNA6	-2,32	sem_annotacao		
RGMB	-2,28	AW392276	DRAGON FLJ90406 MGC86970	RGM domain family, member B
VPS54	-2,25	sem_annotacao		
POMGNT1	-2,25	AW373641	DKFZp761B182 FLJ20277 GnT1.2 MEB MGAT1.2	protein O-linked mannose beta 1,2-N-acetylglucosaminyltransferase
SPOCK2	-2,24	BQ345409	testican-2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2
RNF43	-2,23	BF819033	DKFZp781H02126 DKFZp781H0392 FLJ20315 MGC125630 RNF124 URCC	ring finger protein 43
ETNK2	-2,23	BE937936	EKI2 FLJ10761 HMFT1716	ethanolamine kinase 2
MRPS17	-2,23	BF874151	HSPC011 MRP-S17 RPMS17	mitochondrial ribosomal protein S17
CTSZ	-2,21	AW392304	CTSX	cathepsin Z
GPR98	-2,21	BQ339434	DKFZp761P0710 FEB4 KIAA0686 MASS1 USH2B USH2C VLGR1 VLGR1b	G protein-coupled receptor 98
SV2B	-2,21	BE700692	HsT19680 KIAA0735	synaptic vesicle glycoprotein 2B
ARHGAP19	-2,20	BE708546	DKFZp313K217 MGC138804 MGC138805 MGC14258	Rho GTPase activating protein 19
ANAPC13	-2,19	BE698000	APC13 DKFZP566D193 DKFZP566D193m SWM1	anaphase promoting complex subunit 13
CA3	-2,18	AW606293	CAIII Car3	carbonic anhydrase III, muscle specific
TBC1D9B	-2,14	BE156362	FLJ41880 KIAA0676 MGC157757	TBC1 domain family, member 9B (with GRAM domain)
INOX1	-2,11	BF897788	INO80	INO80 complex homolog 1 (S. cerevisiae)
BOP1	-2,10	BF799598	KIAA0124	block of proliferation 1
CAV1	-2,08	BF854340	CAV MSTP085 VIP21	caveolin 1, caveolae protein, 22kDa
BIN2	-2,08	BG005941	BRAP-1	bridging integrator 2
EYA2	-2,08	AW392633	EAB1 MGC10614	eyes absent homolog 2 (Drosophila)
ATF2	-2,07	BF369898	CRE-BP1 CREB2 HB16 MGC11558 TREB7	activating transcription factor 2

Gene	Fold (DCIS/DC x DCIS puro)	seq_acc	Aliases	Descrição
CLTCL1	-2,07	BF893495	CHC22 CLH22 CLTCL CLTD FLJ36032	clathrin, heavy chain-like 1
CPT1A	-2,06	BF892782	CPT1 CPT1-L L-CPT1	carnitine palmitoyltransferase 1A (liver)
PLEKHC1	-2,05	BE767038	FLJ34213 FLJ44462 KIND2 MIG2 UNC112 mig-2	pleckstrin homology domain containing, family C (with FERM domain) member 1
CIRBP	-2,04	BI011039	CIRP	cold inducible RNA binding protein
PPM1H	-2,02	BE768751	ARHCL1 FLJ13253 KIAA1157	protein phosphatase 1H (PP2C domain containing)
LYRM2	-2,01	BF841114	DJ12208.2 FLJ21130 MGC104253	LYR motif containing 2
RSL1D1	-2,00	BE702729	CSIG DKFZP564M182 L12 MGC138433 MGC142259 PBK1	ribosomal L1 domain containing 1
STK25	-2,00	BF087374	DKFZp686J1430 SOK1 YSK1	serine/threonine kinase 25 (STE20 homolog, yeast)
LOX	2,01	sem_annotacao		
C7orf24	2,12	BQ319951	FLJ11717 MGC3077	chromosome 7 open reading frame 24
MED10	2,15	AW934825	L6 MGC5309 NUT2 TRG20	mediator of RNA polymerase II transcription, subunit 10 homolog (NUT2, <i>S. cerevisiae</i> )
P4HA1	2,20	sem_annotacao	4-PH alpha-1 P4HA	procollagen-proline
KIAA1211	2,30	BF798551	DKFZp564H1363	KIAA1211 protein
HN1	2,30	AW363392	ARM2 HN1A	hematological and neurological expressed 1
IGSF6	2,32	sem_annotacao		immunoglobulin superfamily, member 6
SKIL	2,34	BQ346041	SNO SnoA SnoN	SKI-like oncogene
NCOA4	2,34	AW378456	ARA70 DKFZp762E1112 ELE1 PTC3 RFG	nuclear receptor coactivator 4
NOX4	2,35	AW391502	KOX KOX-1 RENOX	NADPH oxidase 4
CCT5	2,40	BE148625	CCT-epsilon CTE KIAA0098 TCP-1-epsilon	chaperonin containing TCP1, subunit 5 (epsilon)
C13orf23	2,41	BF096005	FLJ12661 FLJ23780 bA50D16.2	chromosome 13 open reading frame 23
PHTF1	2,42	BG988073	PHTF	putative homeodomain transcription factor 1
MARCH8	2,45	AW834812	MARCH-VIII MIR RNF178 c-MIR	membrane-associated ring finger (C3HC4) 8
CXCL9	2,70	BF379536	CMK Humig MIG SCYB9 crg-10	chemokine (C-X-C motif) ligand 9
RUNDC1	2,76	BF090114	DKFZp761H0421	RUN domain containing 1
PTBP2	2,84	BE166421	FLJ34897 MIBP PTB PTBLP brPTB nPTB nPTB5 nPTB6 nPTB7 nPTB8	polypyrimidine tract binding protein 2
RAD51AP1	2,96	sem_annotacao	PIR51	RAD51 associated protein 1
CORO1C	3,05	BQ359710	HCRNN4 coronin-3	coronin, actin binding protein, 1C
SULF1	3,92	BE934588	FLJ30905 FLJ38022 FLJ41750 HSULF-1 KIAA1077 SULF-1	sulfatase 1
NMU	4,12	BE715907	-	neuromedin U

**Anexo 12** - Lista dos 30 genes diferencialmente expressos entre as amostras pareadas ( $p < 0,05$ ) pertencentes as vias de sinalização WNT, PI3K e processo EMT.

Gene	Anotação	Fold	Pvalor
<i>MYC</i>	34129	-1,635	0,048
<i>WNT8A</i>	40098	-1,432	0,034
<i>TCF8</i>	MR1-RT0039-003-g12	-1,363	0,026
<i>ALS2CR7</i>	34010	-1,141	0,007
<i>NOX1</i>	IL3-CT0219-030-C12	1,112	0,034
<i>CLDN1</i>	CTC_RHFADT_030805_210_D12_NM_0	1,121	0,050
<i>PLAUR</i>	RC0-AN0041-031-e05	1,142	0,031
<i>PIK3R5</i>	PM3-EN0004-011-d03	1,168	0,040
<i>LRP3</i>	PM3-FN0058-004-a03	1,196	0,050
<i>DYRK2</i>	QV2-LT0038-108-c06	1,198	0,039
<i>PAX9</i>	MR0-HT0241-011-g06	1,218	0,005
<i>PIK3C2B</i>	MR1-SN0035-001-h12	1,218	0,019
<i>PLCB1</i>	QV3-BN0047-123-f02	1,243	0,010
<i>EYA1</i>	MR3-FN0004-006-d10	1,246	0,011
<i>HIPK2</i>	MR0-NT0039-001-c11	1,261	0,020
<i>DYRK4</i>	MR2-HN0036-007-f07	1,286	0,007
<i>TP53</i>	RC1-BT0314-016-g07	1,295	0,046
<i>ITGA2</i>	MR0-HT0407-017-f01	1,309	0,007
<i>TLE2</i>	MR3-FN0209-005-f12	1,311	0,015
<i>ITPR2</i>	CM1-CT0605-652-d02	1,312	0,011
<i>FOSL1</i>	MR2-CI0128-009-b10	1,321	0,040
<i>DVL2</i>	MR3-HB0013-002-d06	1,345	0,011
<i>EIF2AK2</i>	MR3-EN0087-012-b11	1,346	0,014
<i>CDH6</i>	RC2-BN0243-015-d03	1,348	0,020
<i>PIP5K2C</i>	MR0-HT0078-003-b09	1,361	0,034
<i>WT1</i>	PM3-BT0343-001-b10	1,413	0,043
<i>TBL1X</i>	MR2-OT0079-006-d06	1,432	0,017
<i>RIPK4</i>	QV4-UM0094-152-g09	1,572	0,020
<i>CITED1</i>	QV4-HT0538-128-c07	1,607	0,002
<i>SOX9</i>	QV1-CI0173-460-d06	1,613	0,043

**Anexo 13** - Listas geradas através do teste não paramétrico de *Wilcoxon* realizado para a combinação dois a dois dos tipos histológicos com nível de significância  $\alpha < 0,05$  e diferença de expressão (*Fold*)  $> 2$  |.

Normal x DCIS puro (26 genes)	FOLD	Normal x DCIS/IDC (205 genes)	FOLD	DCIS puro x DCIS/IDC (100 genes)	FOLD
MMP11	-2,668	ESR1	-8,039	GDAP1	-3,689
LRRC14	-2,247	SYT13	-7,630	CLIC6	-3,516
JUP	-2,229	TG	-3,718	MYB	-3,160
SLCO1C1	-2,131	THBS2	-3,703	RCD-8	-2,823
ACTG2	2,054	MB	-3,578	MED28	-2,763
EGR2	2,102	GDAP1	-3,564	RBM6	-2,615
JUN	2,142	HIPK2	-3,348	TGFB3	-2,417
XM_944191	2,170	UCP2	-3,314	PECAM1	-2,378
MED28	2,204	YPEL2	-3,284	EIF4G2	-2,368
C18orf1	2,266	FZD4	-3,249	MMP10	-2,300
PLCG2	2,352	MMP11	-3,201	MMP1	-2,236
HOXA4	2,486	RBM6	-3,053	WASPIP	-2,213
MARK3	2,487	GLUL	-2,965	PDIK1L	-2,190
PPP2R1A	2,506	MYB	-2,960	MPO	-2,189
RPL10A	2,586	PARP1	-2,841	FRAT2	-2,109
ABCA5	2,610	CDKN2D	-2,833	FOSL2	-2,093
TNFRSF6	2,779	GSDML	-2,833	FZD7	-2,090
CALD1	2,907	FAIM	-2,810	APAF1	-2,043
EIF5B	2,995	APH1B	-2,751	LRP6	2,032
PIK3R1	3,393	RCD-8	-2,726	HADHA	2,036
WIF1	3,806	ALDH1A3	-2,699	MAP3K12	2,042
EGR1	3,836	GALNAC4S-6ST	-2,651	PCK1	2,071
SAPS3	4,046	CCDC6	-2,579	SULT1A1	2,100
SNAI2	4,088	SLC1A2	-2,569	ALK	2,102
FZD7	5,712	TIMM44	-2,558	CTGF	2,107
CRYAB	7,417	PIK3R3	-2,550	CHRNA7	2,146
		TAPBP	-2,477	PTPN5	2,147
		EIF4G2	-2,473	ITPKC	2,149
		SPEN	-2,425	PPP2R2A	2,186
		PSENE1	-2,413	PCMT1	2,206
		COL1A2	-2,400	NID2	2,217
		AI925925	-2,392	HOXA4	2,226
		TSPAN1	-2,354	KIAA0423	2,234
		LOC653559	-2,344	DLEC1	2,241
		DA489876	-2,262	TRIP13	2,246
		L3MBTL2	-2,228	SNAP25	2,248
		BLNK	-2,209	RARB	2,248
		MMP10	-2,184	MAP1LC3A	2,263

	<b>Normal x DCIS/IDC (205 genes)</b>	<b>FOLD</b>	<b>DCIS puro x DCIS/IDC (100 genes)</b>	<b>FOLD</b>
	<i>BG167469</i>	-2,160	<i>ADAM15</i>	2,268
	<i>SDC1</i>	-2,095	<i>SEMA6C</i>	2,288
	<i>MGC4562</i>	-2,089	<i>PTCH</i>	2,294
	<i>DC-UbP</i>	-2,086	<i>MAP2</i>	2,302
	<i>FZD2</i>	-2,075	<i>BRCA2</i>	2,320
	<i>PIK3CD</i>	-2,043	<i>SALL1</i>	2,323
	<i>TBL1X</i>	-2,041	<i>CYP2C9</i>	2,327
	<i>SSH3</i>	-2,021	<i>EPHB6</i>	2,348
	<i>NFATC4</i>	-2,018	<i>PRKACG</i>	2,388
	<i>ARHGEF16</i>	-2,017	<i>PRL</i>	2,417
	<i>PCDH11X</i>	-2,012	<i>HOXA7</i>	2,420
	<i>TMEM16F</i>	-2,010	<i>CYP1A1</i>	2,431
	<i>CTNNBIP1</i>	-2,008	<i>EDN1</i>	2,479
	<i>PAX2</i>	-2,005	<i>PPARD</i>	2,494
	<i>DIO3</i>	2,004	<i>NR0B2</i>	2,496
	<i>SALL1</i>	2,008	<i>COL5A1</i>	2,518
	<i>HOXB5</i>	2,013	<i>SIM1</i>	2,525
	<i>DKK3</i>	2,016	<i>DLL1</i>	2,608
	<i>PPP3CC</i>	2,022	<i>MYOG</i>	2,622
	<i>RBP3</i>	2,024	<i>INPP1</i>	2,624
	<i>MLANA</i>	2,044	<i>WNT10B</i>	2,629
	<i>CAMK4</i>	2,063	<i>CR2</i>	2,633
	<i>HOXA7</i>	2,068	<i>LRAT</i>	2,655
	<i>RIPK4</i>	2,080	<i>KIAA1922</i>	2,663
	<i>PAX6</i>	2,086	<i>NGFB</i>	2,696
	<i>CYP26A1</i>	2,087	<i>B3GAT1</i>	2,719
	<i>SP8</i>	2,100	<i>L1CAM</i>	2,722
	<i>MITF</i>	2,107	<i>NOG</i>	2,732
	<i>LHX1</i>	2,124	<i>PAX4</i>	2,757
	<i>COL5A1</i>	2,126	<i>ALDH1A2</i>	2,775
	<i>BCDO2</i>	2,139	<i>FGF8</i>	2,785
	<i>CTSC</i>	2,141	<i>NPY</i>	2,831
	<i>ESR2</i>	2,142	<i>BDNF</i>	2,832
	<i>ABCA5</i>	2,147	<i>HOXB5</i>	2,848
	<i>INPP1</i>	2,157	<i>CALCA</i>	2,866
	<i>RORC</i>	2,163	<i>WNT1</i>	2,957
	<i>WNT7A</i>	2,167	<i>MBP</i>	2,976
	<i>PAX4</i>	2,172	<i>CAMK4</i>	2,978
	<i>DKK4</i>	2,181	<i>CTSC</i>	3,026
	<i>FZD9</i>	2,183	<i>PPP2R2B</i>	3,027
	<i>DLEC1</i>	2,192	<i>NCAM2</i>	3,036
	<i>CYP1A1</i>	2,195	<i>PCDHA8</i>	3,131
	<i>CALD1</i>	2,196	<i>HR</i>	3,154
	<i>NRGN</i>	2,200	<i>PAX6</i>	3,171
	<i>EDN1</i>	2,204	<i>RORB</i>	3,196

	<b>Normal x DCIS/IDC (205 genes)</b>	<b>FOLD</b>	<b>DCIS puro x DCIS/IDC (100 genes)</b>	<b>FOLD</b>
	<i>PHLDA1</i>	2,212	<i>ADH4</i>	3,215
	<i>DLL1</i>	2,223	<i>FCGR1A</i>	3,315
	<i>OMG</i>	2,235	<i>WISP2</i>	3,327
	<i>SALL3</i>	2,246	<i>NRG2</i>	3,343
	<i>MGC35048</i>	2,247	<i>TNFRSF8</i>	3,472
	<i>MAP2</i>	2,260	<i>CD79A</i>	3,490
	<i>PCDHA8</i>	2,265	<i>RELN</i>	3,505
	<i>WNT10B</i>	2,268	<i>GRB10</i>	3,511
	<i>L1CAM</i>	2,277	<i>DSCR1L2</i>	3,611
	<i>FZD1</i>	2,290	<i>CSEN</i>	3,619
	<i>COL2A1</i>	2,294	<i>MAPK8</i>	3,626
	<i>CNR1</i>	2,314	<i>HOXA1</i>	3,949
	<i>CR2</i>	2,318	<i>HTR2A</i>	4,096
	<i>WNT1</i>	2,320	<i>MTA1</i>	4,432
	<i>FCGR1A</i>	2,322	<i>OPN1LW</i>	5,275
	<i>RTN2</i>	2,341	<i>FANCD2</i>	5,536
	<i>DYRK3</i>	2,342	<i>APOD</i>	6,696
	<i>BRCA2</i>	2,347		
	<i>AKT1</i>	2,355		
	<i>NR0B2</i>	2,355		
	<i>TNFRSF6</i>	2,357		
	<i>TLE4</i>	2,362		
	<i>ALDH1A2</i>	2,375		
	<i>SPATA2</i>	2,384		
	<i>DNER</i>	2,388		
	<i>NGFB</i>	2,397		
	<i>TRHR</i>	2,411		
	<i>MBP</i>	2,427		
	<i>FOS</i>	2,438		
	<i>PI3</i>	2,442		
	<i>CD14</i>	2,450		
	<i>LOC647524</i>	2,468		
	<i>CD1A</i>	2,484		
	<i>ALDH1A1(3)</i>	2,487		
	<i>LAMA3</i>	2,488		
	<i>NRG2</i>	2,489		
	<i>CRABP1</i>	2,495		
	<i>RORB</i>	2,496		
	<i>RELN</i>	2,525		
	<i>DLL4</i>	2,560		
	<i>FGF8</i>	2,567		
	<i>CYP26B1</i>	2,575		
	<i>CHRNA7</i>	2,577		
	<i>MAP3K12</i>	2,581		
	<i>ABCB1</i>	2,603		

	<b>Normal x DCIS/IDC (205 genes)</b>	<b>FOLD</b>	
	<i>TTR</i>	2,608	
	<i>PCDH12</i>	2,640	
	<i>NISCH</i>	2,649	
	<i>RORA</i>	2,665	
	<i>PRKACG</i>	2,671	
	<i>RPS5</i>	2,681	
	<i>PIK3R1</i>	2,685	
	<i>RBP4</i>	2,689	
	<i>MYOZ1</i>	2,693	
	<i>RBP2</i>	2,693	
	<i>ACTG2</i>	2,712	
	<i>FZD7</i>	2,733	
	<i>NEUROD2</i>	2,754	
	<i>FZD10</i>	2,765	
	<i>SNAI1</i>	2,773	
	<i>CD79A</i>	2,773	
	<i>DSCR1</i>	2,775	
	<i>LRAT</i>	2,784	
	<i>SOST</i>	2,806	
	<i>TCF7L2</i>	2,818	
	<i>CALCA</i>	2,849	
	<i>IHH</i>	2,904	
	<i>EIF5B</i>	2,980	
	<i>LEP</i>	2,988	
	<i>JUN</i>	3,012	
	<i>ADH4</i>	3,016	
	<i>PTCH</i>	3,055	
	<i>NCAM2</i>	3,059	
	<i>FTH1</i>	3,065	
	<i>CSEN</i>	3,076	
	<i>LEF1</i>	3,084	
	<i>MYH11</i>	3,085	
	<i>MUC16</i>	3,109	
	<i>TNFRSF8</i>	3,147	
	<i>NPY</i>	3,168	
	<i>PRKCA</i>	3,179	
	<i>MYOD1</i>	3,193	
	<i>DOCK9</i>	3,211	
	<i>WISP2</i>	3,246	
	<i>FZD8</i>	3,248	
	<i>SNAI2</i>	3,261	
	<i>CSNK2A2</i>	3,300	
	<i>DIO2</i>	3,403	
	<i>RLBP1</i>	3,497	
	<i>NPYY2</i>	3,502	

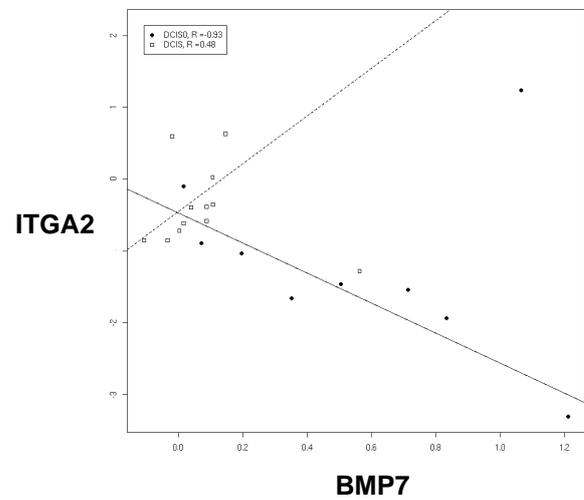
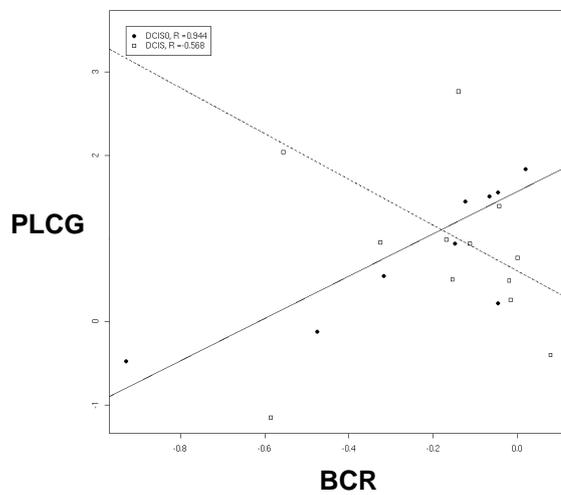
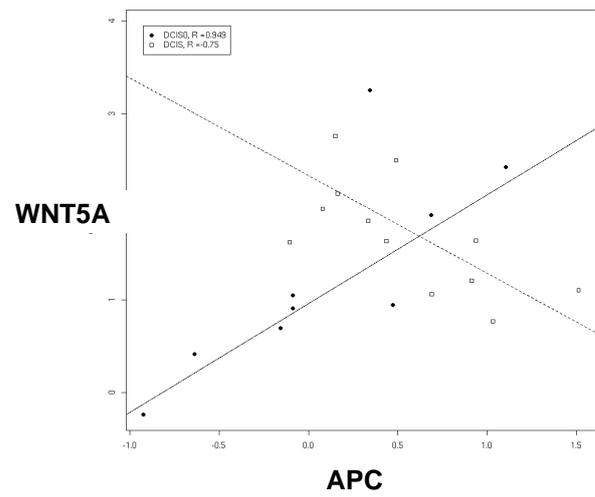
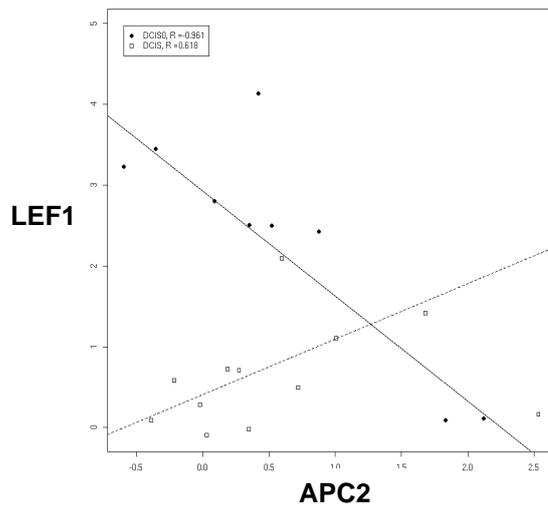
	<b>Normal x DCIS/IDC (205 genes)</b>	<b>FOLD</b>	
	<i>WIF1</i>	3,537	
	<i>DNTT</i>	3,552	
	<i>HOXB4</i>	3,633	
	<i>GRB10</i>	3,688	
	<i>MAP1B</i>	3,703	
	<i>MAPK8</i>	3,712	
	<i>DSCR1L2</i>	3,843	
	<i>ADAM23</i>	3,868	
	<i>PPP2R1A</i>	3,923	
	<i>S100B</i>	3,927	
	<i>OPN1LW</i>	4,219	
	<i>CDX2</i>	4,224	
	<i>ADAM22</i>	4,305	
	<i>NCOA7</i>	4,367	
	<i>MTA1</i>	4,387	
	<i>HOXA1</i>	4,424	
	<i>KRT15</i>	4,601	
	<i>EGR1</i>	4,779	
	<i>DKK1</i>	4,822	
	<i>KIT</i>	4,897	
	<i>ITPKC</i>	4,950	
	<i>MME</i>	4,962	
	<i>WNT3</i>	4,998	
	<i>MAP3K7</i>	5,379	
	<i>HOXA4</i>	5,533	
	<i>FANCD2</i>	5,796	
	<i>DDR2</i>	5,955	
	<i>HTR2A</i>	6,842	
	<i>OXTR</i>	7,068	
	<i>NUDT6</i>	9,380	
	<i>FOSB</i>	10,237	
	<i>CRYAB</i>	16,190	

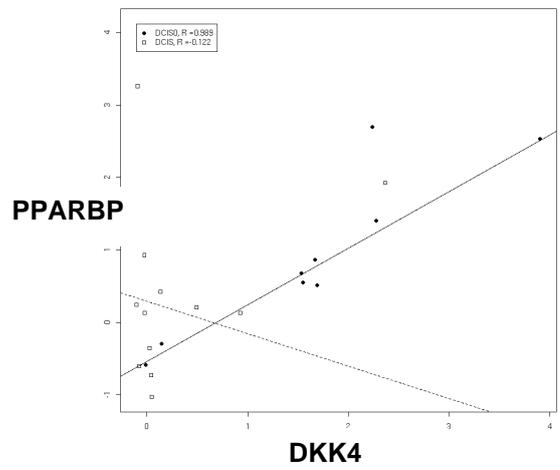
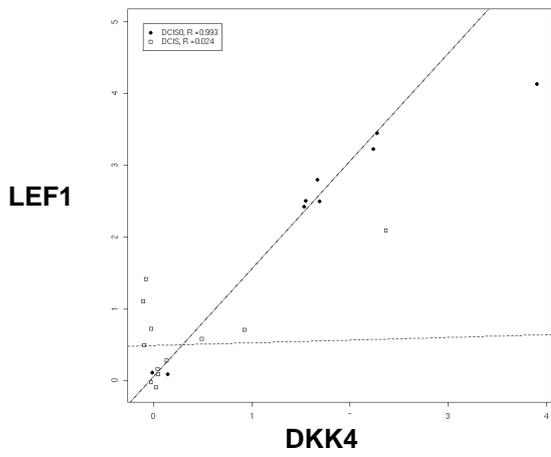
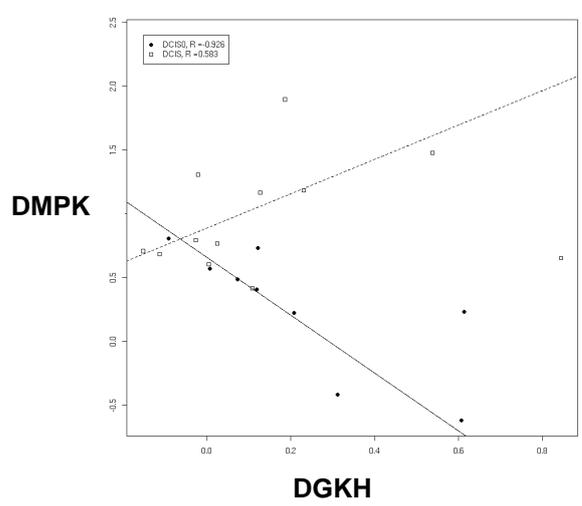
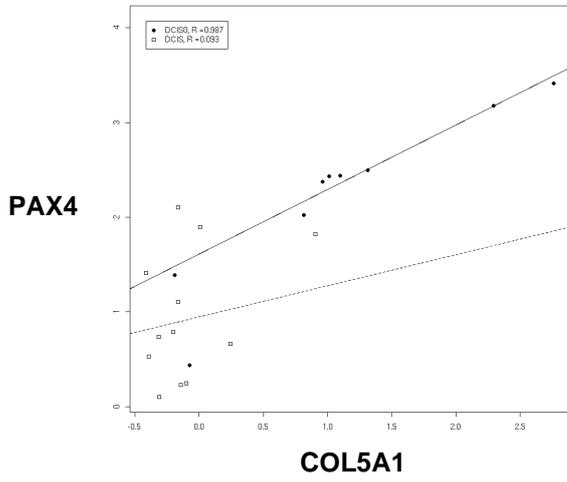
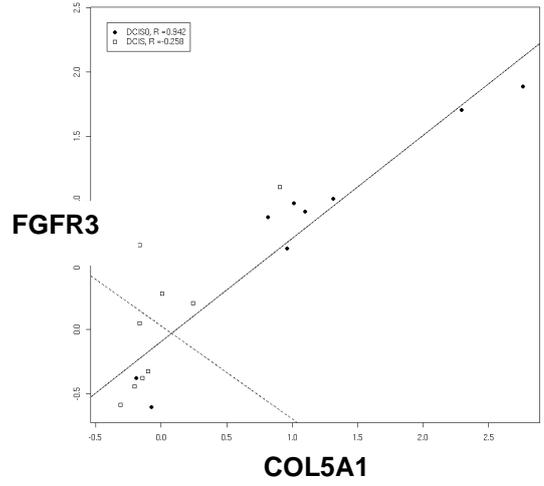
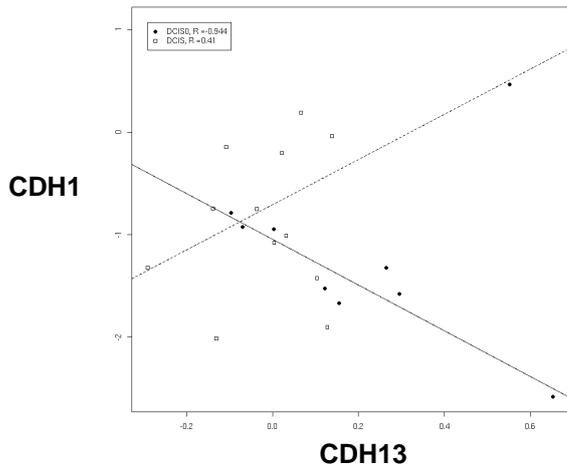
**Anexo 14** - Lista dos 57 genes diferencialmente expressos entre as amostras progressão ( $p < 0,05$  e  $Fold > 2$ ) pertencentes as vias de sinalização WNT, PI3K e processo EMT.

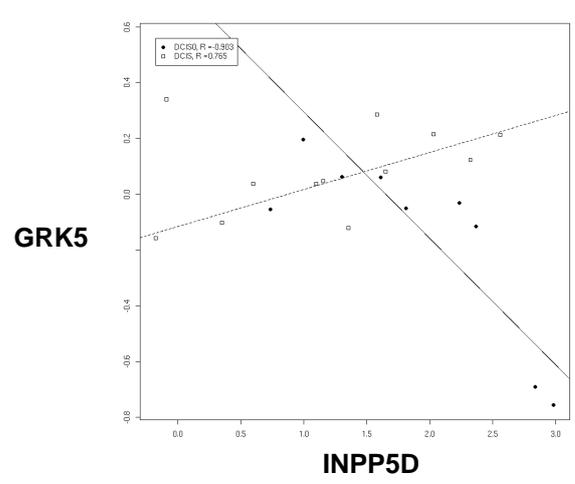
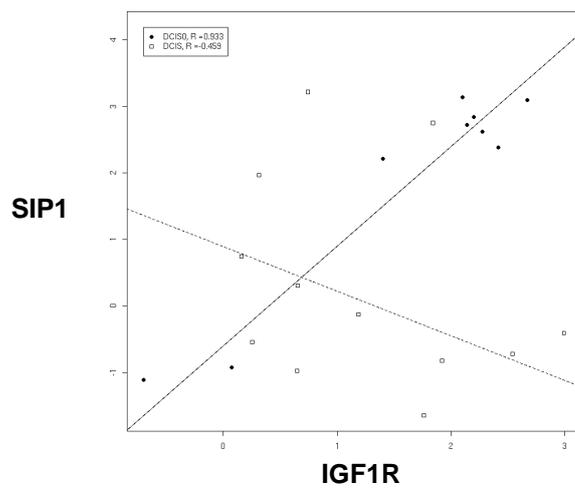
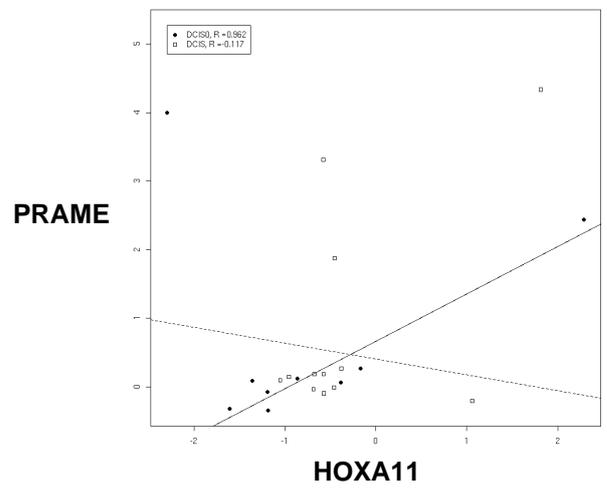
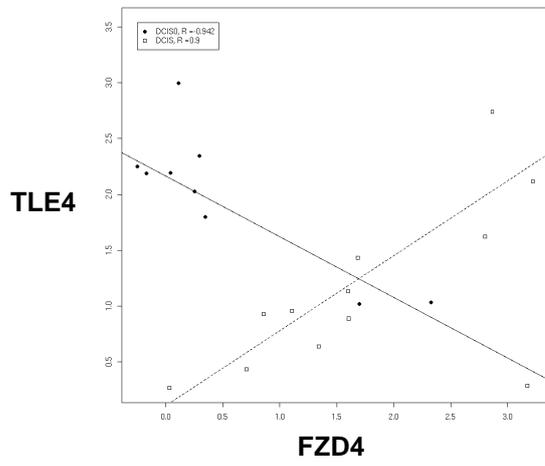
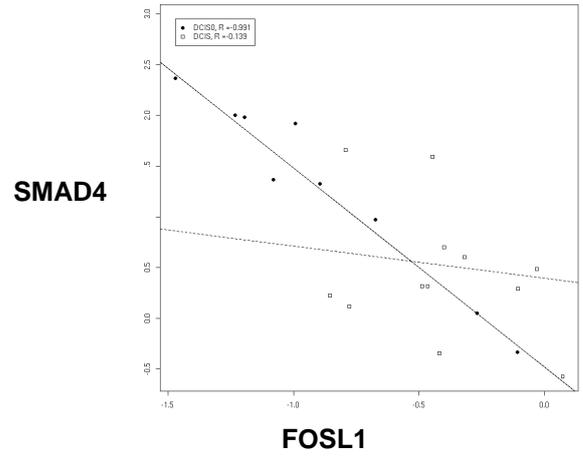
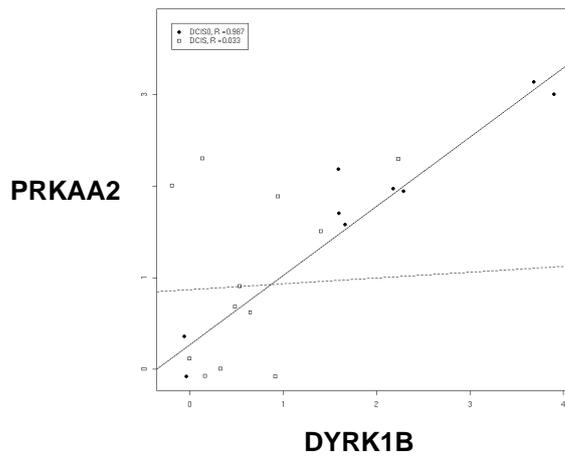
<b>Nome do Gene</b>	<b>Fold</b>	<b>Nome do Gene</b>	<b>Fold</b>
<i>RCD-8</i>	-2,726	<i>GDAP1</i>	-3,689
<i>HOXB5</i>	2,013	<i>MYB</i>	-3,160
<i>CAMK4</i>	2,063	<i>RBM6</i>	-2,615
<i>CTSC</i>	2,141	<i>EIF4G2</i>	-2,368
<i>WNT10B</i>	2,268	<i>MMP10</i>	-2,300
<i>CR2</i>	2,318	<i>DLEC1</i>	2,241
<i>BRCA2</i>	2,347	<i>MAP2</i>	2,302
<i>NGFB</i>	2,397	<i>SALL1</i>	2,323
<i>RORB</i>	2,496	<i>HOXA7</i>	2,420
<i>FGF8</i>	2,567	<i>CYP1A1</i>	2,431
<i>CHRNA7</i>	2,577	<i>EDN1</i>	2,479
<i>MAP3K12</i>	2,581	<i>NR0B2</i>	2,496
<i>PRKACG</i>	2,671	<i>COL5A1</i>	2,518
<i>LRAT</i>	2,784	<i>DLL1</i>	2,608
<i>CALCA</i>	2,849	<i>INPP1</i>	2,624
<i>PTCH</i>	3,055	<i>WNT10B</i>	2,629
<i>CSEN</i>	3,076	<i>L1CAM</i>	2,722
<i>NPY</i>	3,168	<i>PAX4</i>	2,757
<i>WISP2</i>	3,246	<i>ALDH1A2</i>	2,775
<i>MAPK8</i>	3,712	<i>MBP</i>	2,976
<i>ITPKC</i>	4,950	<i>NCAM2</i>	3,036
<i>RCD-8</i>	-2,726	<i>GDAP1</i>	-3,689
<i>PCDHA8</i>	3,131	<i>TNFRSF8</i>	3,472
<i>PAX6</i>	3,171	<i>CD79A</i>	3,490
<i>ADH4</i>	3,215	<i>RELN</i>	3,505
<i>FCGR1A</i>	3,315	<i>GRB10</i>	3,511
<i>NRG2</i>	3,343	<i>DSCR1L2</i>	3,611
<i>HOXA1</i>	3,949	<i>MTA1</i>	4,432
<i>HTR2A</i>	4,096	<i>OPN1LW</i>	5,275
<i>FANCD2</i>	5,536		

**Anexo 15 - Gráficos *scatter-plot*'s para aqueles pares de genes com coeficientes de correlação superiores a 0,9.**

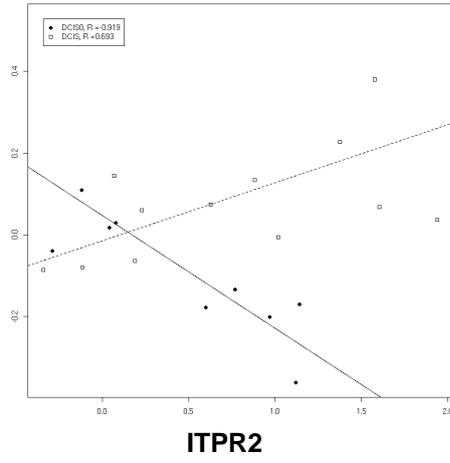
- Círculos preenchidos representam o grupo não-neoplásico + DCIS puro
- Quadrados representam o grupo DCIS/IDC



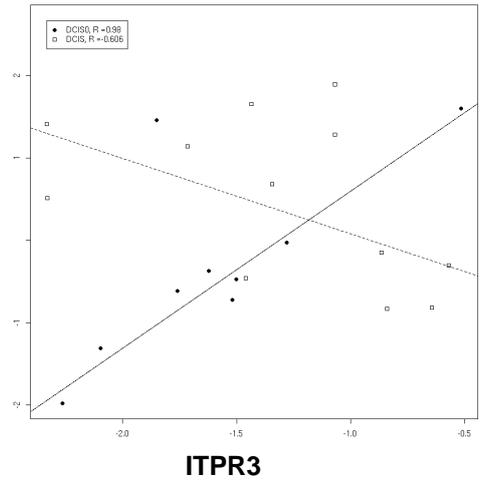




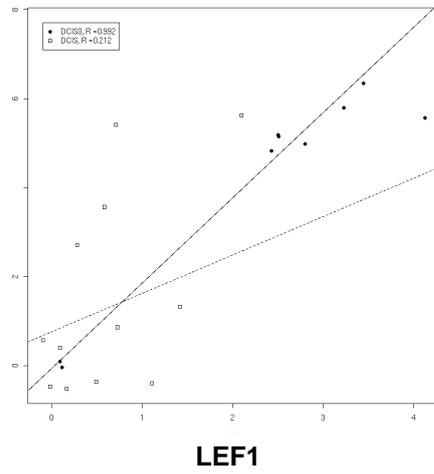
**ITPKA**



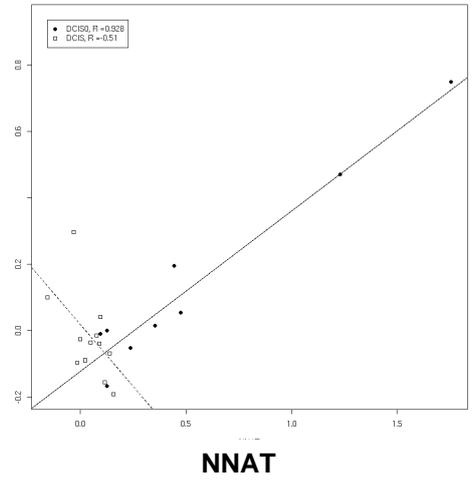
**TTK**



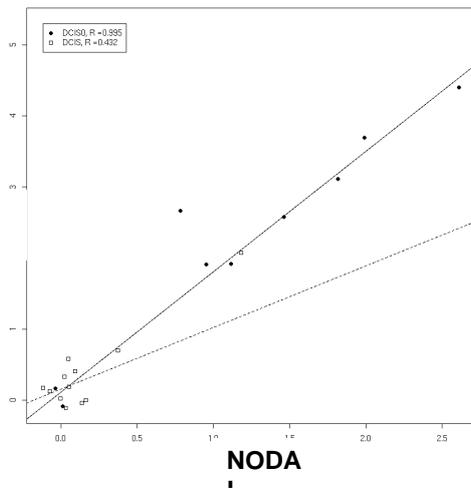
**WNT7B**



**TCF3**



**FZD8**



**CTNNA  
1**

