

**AVALIAÇÃO DA EXPRESSÃO DE GENES
CODIFICADORES E NÃO CODIFICADORES EM
CÉLULAS EPITELIAIS DO CARCINOMA DUCTAL
INVASIVO DE MAMA DE PACIENTES COM E SEM
LINFONODO ACOMETIDO**

MABEL GIGLIOLA PINILLA-FERNÁNDEZ

**Tese apresentada à Fundação Antônio Prudente
para a obtenção do título de Doutor em Ciências
Área de concentração: Oncologia**

Orientadora: Dra. Dirce Maria Carraro

**São Paulo
2015**

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Pinilla-Fernández, Mabel Gigliola

Avaliação da expressão de genes codificadores e não codificadores em células epiteliais do carcinoma ductal invasivo de mama de pacientes com e sem linfonodo acometido / Mabel Gigliola Pinilla-Fernández - São Paulo 2015. 98p.

Tese (Doutorado) -Fundação Antônio Prudente.

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

Orientadora: Dirce Maria Carraro.

Descritores: 1. CARCINOMA DUCTAL DE MAMA. 2. PROGNÓSTICO. 3. ANÁLISE DO PADRÃO DA EXPRESSÃO GÊNICA. 4. ANÁLISE EM MICROSSÉRIES.

AGRADECIMENTOS

Agradeço a todas as pessoas que contribuíram de alguma forma para realização deste trabalho.

À minha orientadora, Dra. Dirce Maria Carraro, pelas sugestões e críticas construtivas, que contribuíram muito para o desenvolvimento deste trabalho. Além disso, pela sua qualidade humana, apoio constante, paciência e incentivo que permitiram finalizar este trabalho.

Ao Dr. Alex Fiorini de Carvalho, que me acompanhou nos primeiros anos do desenvolvimento deste trabalho, contribuindo com seu conhecimento, alento e bom humor.

Aos colegas do laboratório de Genômica e Biologia Molecular (LGBM), os quais contribuíram no desenvolvimento do meu doutorado, Dra. Adriana Miti, Dr. Alex de Carvalho, Ana Paula Mitsue, Ms.c. Bianca Lisboa, Dra. Bruna Barros, Dra. Carolina Sens, Ms.c. Cristiane Gonçalves, Ms.c. Daiana Marin, Dra. Elisa Napolitano, Ms.c. Eloisa Olivieri, Dr. Felipe Carneiro, Dra. Giovana Torrezan, Louise Mota, Dra. Márcia Figueiredo, Ms.c. Rafael Canfield, Dr. Roberto Ferreira e Dra. Tatiana Ricca. Especialmente as minhas amigas do coração Dra. Danielle Braggio (Dani) e Dra. Vanina Elias (Vani) pela amizade, por sempre me ouvirem, me ajudarem e se importarem.

A todos os amigos do CIPE. Aos funcionários do A.C. Camargo Cancer Center, em especial ao pessoal do Biobanco, aos meninos do arquivo, SAME e ao Severino.

Ao pessoal da biblioteca, especialmente à Suely. À pós-graduação da fundação, especialmente à Ana Kuninari.

Ao Instituto de Química, Departamento de Bioquímica da Universidade de São Paulo, em nome do Dr. Sergio Verjovski, pela facilitação da lamina de *microarray* 244k e ao Dr. Yuri Moreira por sua ajuda com as análises de transcritos senso anti-senso.

À Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

À *Universidad de Concepción*, em especial ao Decano da Faculdade de Medicina Dr.

Raul González Ramos e ao Chefe de Carrera de Tecnologia Médica; T.M. Mónica Troncoso (2011-2012) e T.M. Fernando Rocha (2013-2015) por me apoiarem e estimularem a desenvolver o programa de doutorado, permitindo me ausentar de minhas obrigações docentes na universidade.

A minha colega Ms.c. Andrea Dibarrart Vera pelo seu apoio permanente, assumindo minha carga acadêmica na *Universidad de Concepción*.

A todos os amigos e familiares, os quais, mesmo à distância, estiveram muito presentes ao longo do doutorado, em especial as minhas amigas Carla Alarcón e Denise Chliplaski pelo suporte emocional e por acreditarem sempre na minha capacidade.

Ao “Fofó”, meu cachorro, que mesmo ele não sendo humano e provavelmente não entendendo, tem me acompanhado nos últimos três anos, proporcionando seu amor incondicional e me permitindo achar forças para querer continuar.

Agradeço de coração a minha família, aos meus pais Antonio e Ruth, e meus irmãos Pablo, Claudio e Dennis os quais sempre me apoiaram e tentaram ficar próximos mesmo separados por muitos quilômetros.

RESUMO

Pinilla-Fernández MG. **Avaliação da expressão de genes codificadores e não codificadores em células epiteliais do carcinoma ductal invasivo de mama de pacientes com e sem linfonodo acometido.** São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente].

O câncer de mama é o mais comum entre as mulheres, representando uma elevada taxa de mortalidade no mundo. O carcinoma ductal invasivo (CDI) é um grupo heterogêneo de tumores com diferentes desordens em nível molecular, ainda assim foram estabelecidos cinco subgrupos (Luminal A, Luminal B, HER-enriquecido, Basal e normal-like) os quais apresentam distintos padrões de expressão gênica com diferentes implicações em prognóstico. Clinicamente, o fator prognóstico mais importante é o acometimento linfonodal, devido a sua influência na sobrevida das pacientes. O maior conhecimento do genoma humano tem permitido identificar grandes transcritos que não codificam proteínas e são originados de sequências intrônicas ou intergênicas do genoma (lncRNAs) e diferenças de expressão desse tipo de transcrito em tecidos específicos e em alguns tumores tem sido reportada, sugerindo um papel regulatório importante, portanto, estudos nessa área seriam promissores para o melhor entendimento do câncer. Nesse contexto, o presente estudo tem como objetivo avaliar a expressão de transcritos codificadores e não-codificadores de proteínas nas células epiteliais tumorais de carcinoma ductal invasivo de mama com ou sem linfonodos acometidos.

A determinação da expressão gênica foi avaliada pela hibridização em uma lâmina de *microarray* customizada com sondas para transcritos codificadores de proteínas e transcritos longos não codificadores. Hibridizamos o RNA amplificado, proveniente de células epiteliais capturadas a laser, de 103 amostras de carcinoma ductal invasivo de mama, com e sem linfonodos acometidos, representativas dos subgrupos moleculares, caracterizados por marcadores imunoistoquímicos. A avaliação dos transcritos codificadores foi usada com dois propósitos. O primeiro foi a validação do ensaio *microarray* pela replicação do conjunto de genes PAM50. Nossos resultados mostraram um perfil de expressão muito similar aos achados por Parker et al. (2009), agrupando às amostras segundo os subtipos moleculares “intrínsecos” e validando os nossos experimentos de *microarray*. O segundo propósito foi a

identificação de perfis moleculares associados com acometimento linfonodal. Para esse propósito, conseguimos identificar 79 genes candidatos a estarem associados com comprometimento linfonodal, uma vez que apresentaram diferenças de expressão entre as amostras com e sem linfonodos acometidos. Esses candidatos foram submetidos a uma análise seriada utilizando o *software* STEM, o qual permite uma avaliação dinâmica da expressão dos genes em grupos representativos da invasão linfonodal de forma progressiva (grupos de amostras de pacientes sem acometimento linfonodal, amostras de pacientes com 1–3 linfonodos acometidos e pacientes com 4 ou mais linfonodos acometidos). Essa análise é baseada em 10 perfis de expressão, 5 que aumentam e 5 que diminuem a expressão ao longo da progressão do acometimento linfonodal, representado por grupos de amostras característicos de cada estágio. Desses, 3 perfis apresentaram significância estatística, um deles com expressão diminuída e dois com expressão aumentada. Para validação técnica dos genes pertencentes a esses perfis, foram selecionados 40 genes para o ensaio RT-qPCR 29 pertencentes aos perfis que apresentaram significância estatística e 11 aos que não apresentaram. O ensaio de RT-qPCR mostrou que sete genes confirmaram as diferenças de expressão entre as amostras LN0 e LN \geq 4, quatro com aumento da expressão (POU2AF1, ZDHHC20, CX3CL1 e FOXQ1), e genes com diminuição da expressão ao longo do processo de acometimento linfonodal (CDK5R2, NMX1 e SPTAN1). Desses sete genes, quatro estavam nos perfis do STEM com significância estatística. O conjunto de genes validados foi submetido a uma análise de interação gênica, obtendo uma única rede, composta por todos sete genes validados, na qual, interessantemente, predominaram as funções biológicas relacionadas ao movimento e à proliferação celular. Adicionalmente, avaliação dos transcritos não codificadores foi utilizada para identificar pares de transcritos senso-anti-senso com expressão invertida, sugestivo de regulação transcricional do transcrito codificante pelo não-codificante. Nossa análise identificou 15 pares de transcritos codificadores com seu respectivo transcrito anti-senso não codificante, com expressão invertida. Dois pares de transcritos apresentaram correlação de *Pearson* \geq 0.7, as sequências não codificadores estavam no mesmo *locus* do gene LCOR e RFWD2, sugerindo que esses genes podem estar sendo regulados por esses transcritos anti-sensos. No entanto, o papel desses genes na carcinogênese de mama e dos anti-sensos na sua regulação necessitam de análises adicionais.

SUMMARY

Pinilla-Fernández MG. [**Evaluating the expression of coding and non-coding genes in epithelial cells of invasive ductal breast cancer in patients with and without lymph node involved**]. São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente].

Breast cancer is the most common cancer among women with high mortality rate worldwide. Although Invasive ductal carcinoma (IDC) is a heterogeneous group of tumors with different disorders at the molecular level, it has been established five subgroups (Luminal A, Luminal B, HER-enriched, basal and normal-like) that have distinct patterns of gene expression with different implications for prognosis. Clinically the most important prognostic factor is lymph node involvement due to its influence on survival of patients. Greater knowledge of the human genome has allowed to identify large transcripts that do not code for proteins and are sourced from intronic and intergenic sequences of the genome (lncRNAs), analyzes of this type of transcript has shown expression differences in specific tissues and some tumors, showing that studies in this field would be promising for a better understanding of cancer. In this context, this study aims to evaluate the expression of coding and non-coding transcripts of proteins in tumor epithelial cells of breast invasive ductal carcinoma with or without lymph nodes metastases. The determination of gene expression was assessed by hybridization on a customized microarray platform containing probes for coding and non-coding protein transcripts. We employed for hybridization the amplified RNA derived from epithelial cells captured by microdissection laser of 103 samples of invasive ductal breast carcinoma, with or without affected lymph nodes, representative of molecular subgroups, characterized by immunohistochemistry markers. The evaluation of coding transcripts was used for two purposes. The first was the validation of microarray by the replication of the PAM50 gene set. Our results showed a very similar expression profile to the findings by Parker et al. (2009), gathering the samples according to "intrinsic" molecular subtypes and validating our microarray

experiments. The second purpose was to identify molecular profiles associated with lymph node involvement. For this purpose, we were able to identify 79 candidate genes to be associated with lymph node involvement, since they showed differential expression between samples with and without affected lymph nodes. These candidates were subjected to a serial analysis using STEM software which allows a dynamic assessment of gene expression in representative lymph node invasion groups in a progressive manner (patient samples without lymph node involvement, patient samples with 1-3 affected lymph nodes and patients with four or more affected lymph nodes). This analysis is based on 10 expression profiles, 5 up-regulated and 5 down-regulated along the progression of lymph node involvement, represented by a group of samples with characteristic of each stage. From them, 3 profiles were statistically significant, one with decreased expression and two with increased expression. For technical validation of genes belonging to these profiles, 40 genes were selected for RT-qPCR assay 29 belonging to the profiles that showed statistical significance and the 11 that did not. The RT-qPCR assay showed that seven genes confirmed the expression differences between the LN0 and LN \geq 4 samples, four with increased expression (POU2AF1, ZDHHC20, CX3CL1 and FOXQ1) and three genes with decreased expression during the process of lymph node involvement (CDK5R2, NMX1 and SPTAN1). From these seven genes, four were in STEM profiles with statistical significance. The set of validated genes were subjected to a gene interaction analysis, obtaining a single network, comprising all validated genes, with enrichment in biological functions related to cell proliferation and movement. Additionally, evaluation of non-coding transcripts was used to identify pairs of sense-antisense transcripts with inverted expression, suggestive of transcriptional regulation of the coding transcript coding by the non-coding. Our analysis identified 15 pairs of coding transcripts with their respective noncoding antisense transcript with inverted expression. Two pairs of transcripts showed Pearson correlation ≥ 0.7 , non-coding sequences were in the same locus of LCOR and RFWD2 gene, suggesting that these genes may be regulated by these antisense transcripts. However, the role of these genes in breast carcinogenesis and in its antisense regulation requires additional analyzes.

LISTA DE FIGURAS

Figura 1	Distribuição proporcional por sexo dos tipos de câncer com maior incidência no Brasil.....	2
Figura 2	Corte transversal de um ducto da mama com câncer invasivo.....	3
Figura 3	Representação esquemática da metodologia do projeto.....	23
Figura 4	Representação esquemática do procedimento microdissecção a laser....	27
Figura 5	Representação de uma mostra microdissecada a laser.....	28
Figura 6	Perfil analítico e RNA total por eletroforese capilar.....	29
Figura 7	Esquema das etapas da amplificação e marcação do RNA teste e RNA referência para ensaio <i>microarray</i> . Representa-se as etapas de sínteses e purificação dos cDNAs e cRNAs.....	33
Figura 8	Desenho do microarranjo de oligonucleotídeos intron-exon sonda 60-MER customizada.....	36
Figura 9	Lâmina 244K hibridada com RNA da amostra (Cy3 verde) teste e RNA referência (Cy5 vermelho).....	37
Figura 10	Esquematização da localização dos genes na placa PCRarray customizada.....	41
Figura 11	Distribuição das amostras, por principal componente de análises, durante os diferentes filtros.....	55

Figura 12	Identificação dos subtipos “Intrínsecos” moleculares no câncer de mama.....	58
Figura 13	Cluster hierárquico das amostras carcinoma ductal invasivo com e sem linfonodos acometidos.....	61
Figura 14	Distribuição dos 79 genes diferentemente expressos entre os grupos com e sem linfonodos acometidos nos 10 perfis de expressão gênica gerados com <i>software</i> STEM.....	62
Figura 15	Genes com diferenças de expressão em amostras de pacientes sem linfonodos acometidos e com mais de 3 linfonodos acometidos.....	65
Figura 16	Genes validados com diferenças de expressão na primeira etapa de invasão linfonodal.....	66
Figura 17	Genes validados com diferenças de expressão em etapas tardias de invasão linfonodal.....	66
Figura 18	Rede de interação molecular dos genes validados.....	67
Figura 19	Gráficos de dispersão dos transcritos codificantes para proteína (Eixo Y) e não codificantes (Eixo X) identificando os subgrupos moleculares e a invasão linfonodal.....	73

LISTA DE TABELAS E QUADROS

Tabela 1	Classificação baseada em marcadores imunoistoquímicos.....	25
Tabela 2	Idade por subtipos.....	48
Tabela 3	Subtipos de câncer de mama invasivo segundo <i>status</i> linfonodal.....	49
Tabela 4	Quantidade de cRNA amplificado e incorporação de fluoróforos.....	51
Tabela 5	Correlação da expressão de genes e transcritos não codificadores no mesmo <i>locus</i>	71
Quadro 1	Descrição do carcinoma ductal de mama em relação ao estágio, TNM e sobrevida.....	7
Quadro 2	Características específicas de cada fluoróforo utilizado.....	36
Quadro 3	Dados clínicos e anatomopatológicos.....	46
Quadro 4	Disfunção ou processos celulares associado aos genes validados.....	68
Quadro 5	Nodos de interconexões incluídos genes validados.....	69

LISTA DE ABREVIATURAS

AC	Linfonodos Acometidos
ANOVA	<i>Analysis Of Variance</i>
cDNA	DNA complementar
CDI	Carcinoma Ductal Invasivo
CDIS	Carcinoma Ductal <i>in situ</i>
CK5	Citoqueratina 5
cRNA	RNA complementar amplificado
CTP	Citosina Trifosfato
Cy3	Cianina 3
Cy5	Cianina 5
EGFR	<i>Epidermal Growth Factor Receptor</i>
ER	Estrogen Receptor
EST	<i>Expressed Sequence Tag</i>
FE	<i>Feature Extraction</i>
ID	Identificação das amostras
IM	Índice Mitótico
INCA	Instituto Nacional de Câncer
LN	Linfonodo
ncRNA	<i>Noncoding RNA</i>
nm	Nanômetros
nt	Nucleotídeos
PIN	<i>Partially Intronic Noncoding</i>
PR	<i>Progesteron Receptor</i>
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
Ress	Resseção
RIN	<i>RNA Integrity Number</i>
rRNA	Ribossomal RNA
SAM	<i>Significance Analisys for Microarray</i>
SBR	Scarff-Bloom-Richardson
TIN	<i>Totally Intronic Noncoding</i>

ÍNDICE

1	INTRODUÇÃO	1
1.1	Carcinoma Mamário	2
1.2	Estadiamento do Câncer de Mama	5
1.3	Marcadores Moleculares do Câncer de Mama	7
1.4	Perfil Molecular do Câncer da Mama.....	11
1.5	Potencial Prognóstico dos Perfis de Expressão	13
1.6	Expressão Gênica Associada à Progressão do Câncer	15
1.6.1	Estudo Independente dos Componentes Morfológicos do Câncer	17
1.7	Câncer e <i>Non Coding RNA</i>	18
2	JUSTIFICATIVA	20
3	OBJETIVOS	21
3.1	Objetivo Geral.....	21
3.2	Objetivos Específicos.....	21
4	MATERIAL E MÉTODOS	23
4.1	Amostras de Tecido	24
4.1.1	Classificação das Amostras.....	25
4.2	Microdissecção a Laser das Células Epiteliais de CDI.....	26
4.3	Extração e Amplificação do RNA das Células Microdissecadas a Laser.....	28
4.3.1	Avaliação da Qualidade e Integridade do RNA total.....	29
4.3.2	Amplificação do RNA com Cauda de Poli A (RNA poliA)	30
4.3.3	Hibridação em Plataforma de <i>Microarray</i>	34
4.3.4	Descrição do <i>Microarray</i> Utilizado	35
4.3.5	Captura de Sinal e Pré-análise das Lâminas de <i>Microarray</i>	36
4.4	Análise dos Resultados de CDNA <i>Microarray</i>	38
4.5	Validação dos Transcritos Identificados como Diferencialmente Expressos	40
4.6	Amostras para Validação POR RT-PCR Quantitativo	42

4.7	Análises dos Ensaaios RT-PCR Quantitativa	42
4.8	Análise Biológica	44
5	RESULTADOS.....	45
5.1	Dados Clínicos - Patológicos dos Pacientes analisados no Estudo.....	45
5.2	Análise de Expressão Gênica.....	50
5.2.1	Perfil de Expressão Gênica dos Transcritos Codificadores	55
5.3	Avaliação do Perfil de Expressão Gênica em Células Epiteliais de Carcinoma Ductal Invasivo de Pacientes com e sem Linfonodos Acometidos	59
5.4	Validação da Expressão Gênica em Amostras Provenientes de Pacientes em Três Estados de Invasão Linfonodal Diferentes, POR RT-QPCR.....	63
5.4.1	Avaliação por RT-qPCR	64
5.5	Redes de Interação Molecular e Vias Biológicas e Integrando essas Redes	66
5.6	Identificação de Pares de Transcritos Codificadores e não Codificadores com Nível de Expressão Invertida	70
6	DISCUSSÃO.....	74
7	CONCLUSÃO	84
8	REFERÊNCIAS BIBLIOGRÁFICAS.....	85

ANEXOS

Anexo 1	Carta de aprovação do Comitê de Ética em Pesquisa da Fundação Antônio Prudente-CEP
Anexo 2	Classificação Clínica Tumores de Mama
Anexo 3	Formulário de Coleta de Dados
Anexo 4	Protocolo de Coloração de Tecido Congelado
Anexo 5	Extração e Purificação do RNA
Anexo 6	Protocolo RNA 6000
Anexo 7	Protocolo Purificação RNA
Anexo 8	Amplificação do mRNA Microdissecado

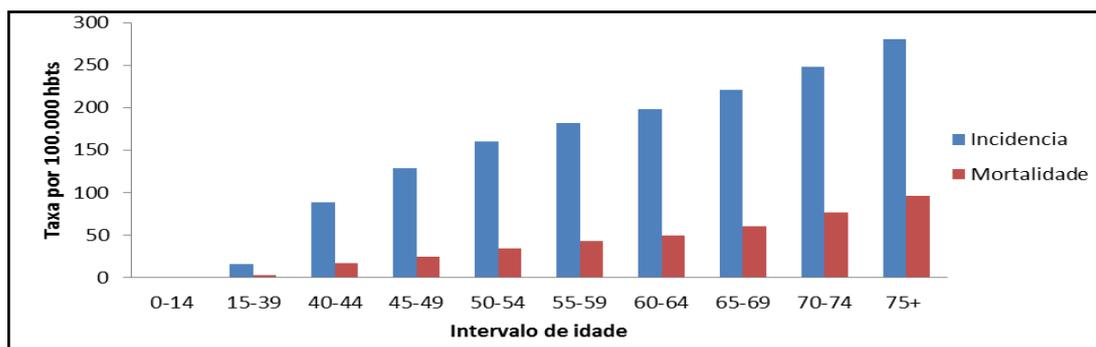
- Anexo 9** Dados Clínicos e Anatomopatológicos
- Anexo 10** Clusterização hierárquica de amostras baseadas nos 79 genes diferencialmente expressos entre os grupos com diferenças no *status* linfonodal
- Anexo 11** Avaliação de presença de DNA genômico nas amostras representativas de acometimento linfonodal
- Anexo 12** Interconexões de triângulo entre as moléculas e os genes validados na rede de interação molecular

1 INTRODUÇÃO

O câncer da mama é o tipo de câncer com maior prevalência nas mulheres no Mundo. Cerca de 1,4 milhões de casos novos dessa neoplasia foram esperados para o ano de 2008, correspondendo a 23% de todos os tipos de câncer. Nos Estados Unidos, cerca de uma em cada oito mulheres (pouco menos de 12%) irá desenvolver câncer de mama invasivo ao longo de sua vida. Na América do Sul, observa-se uma taxa de incidência de 44,3 por 100.000 habitantes, com taxa de mortalidade de 13,2 por 100.000 habitantes (JEMAL et al. 2011).

Para 2014, as estimativas para o Brasil foram de 57.120 casos novos de câncer da mama, com um risco estimado de 56,09 casos a cada 100 mil mulheres (Ministério de Saúde 2014). Sem considerar os tumores da pele não melanoma, o câncer de mama é o mais frequente nas mulheres das regiões Sudeste e Sul (71/100 mil), Centro-Oeste (51/100 mil) e Nordeste (37/100 mil). Na região Norte, é o segundo tumor mais incidente (21/100 mil) (Ministério de Saúde 2014).

As taxas de incidência do câncer de mama no Brasil revelam que a idade continua sendo o principal fator de risco, notando um aumento significativo dos 40 anos em diante (Figura 1).



Fonte: Ministério da Saúde (2014)

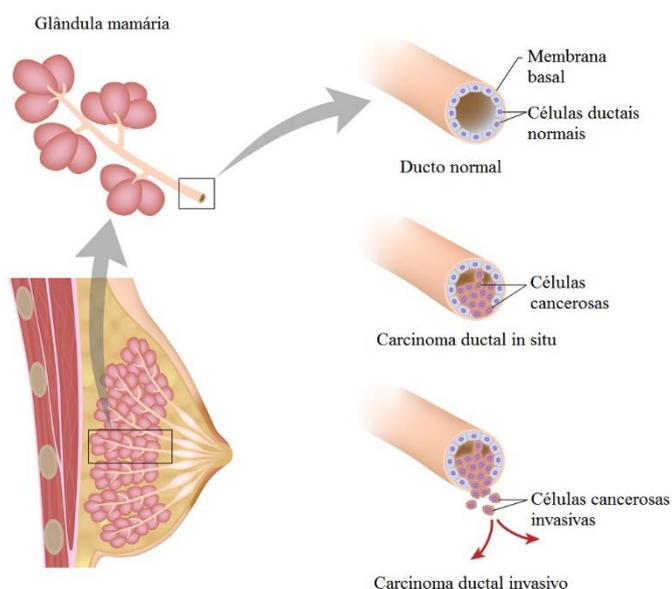
Figura 1 - Taxas de Incidência e Mortalidade por Idade, no Brasil. Pode-se observar um aumento da taxa de incidência e mortalidade a medida que aumenta a idade das mulheres.

Apesar de ser considerado um câncer relativamente de bom prognóstico se diagnosticado e tratado precocemente, as taxas de mortalidade por câncer da mama continuam elevadas no Brasil, muito provavelmente porque a doença ainda é diagnosticada em estádios avançados. A sobrevida média após cinco anos tem apresentado um discreto aumento na população de países desenvolvidos, atualmente é de 85%, entretanto nos países em desenvolvimento, como no Brasil, a sobrevida continua com valores entre 50 e 60% (Ministério de Saúde 2014).

1.1 CARCINOMA MAMÁRIO

O carcinoma mamário é uma doença complexa e heterogênea. Definido como o crescimento anormal e descontrolado das células epiteliais no tecido mamário. Quando a proliferação anormal permanece dentro dos ductos e/ou lóbulos é conhecido como carcinoma *in situ* (não-invasivo), ou seja, as células neoplásicas não ultrapassam a membrana basal. Se extrapolam a membrana basal o tumor é denominado invasivo.

O carcinoma ductal invasivo (CDI) é o câncer de mama mais comumente observado, não tem características histológicas especiais e corresponde a, aproximadamente, 90% dos carcinomas da mama. O CDI surge a partir de células epiteliais, cuboidais de revestimento ductal, neoplásicas que podem perder a capacidade de adesão e iniciar o processo de invasão, ultrapassando a membrana basal e a camada de células mioepiteliais (Figura 2) invadindo as estruturas preexistentes. Podem atingir o estroma, os vasos sanguíneos e linfáticos, chegando inclusive a produzir metástase, primeiro nos linfonodos axilares próximos e depois podem se espalhar por órgãos distantes (FIDLER 2003).



Fonte: <http://www.caperay.com/>

Figura 2 - Carcinoma ductal mamário. São observadas as etapas de invasão das células do carcinoma ductal.

O câncer de mama apresenta uma progressão clínica diferente em cada paciente, o que reflete nas diferenças observadas em relação ao crescimento tumoral, capacidade de invasão e potencial metastático. Os fatores prognósticos são

parâmetros possíveis de serem mensurados no momento do diagnóstico, que servem como preditores da sobrevida do paciente. No câncer de mama, são consideradas as características demográficas, características do tumor (comprometimento linfonodal, tamanho tumoral, tipo histológico) e marcadores biológicos como o padrão de expressão dos receptores hormonais (estrógeno e progesterona), bem como a amplificação e/ou aumento de expressão da proteína HER-2.

O *status* linfonodal axilar é o fator prognóstico mais importante do CDI da mama. O estudo do acometimento é feito por ressecção cirúrgica e posterior análise anatomopatológico do linfonodo, o que permite estabelecer o comprometimento metastático dos linfonodos axilares (CARTER et al. 1988). Estudos têm descrito que seis linfonodos são o mínimo necessário para avaliar adequadamente as metástases linfonodais (KATZ et al. 2008), mas geralmente são avaliados no mínimo 10 LNs (SCHIFFMAN et al. 2011).

O número de linfonodos ressecados é importante, porque tem-se demonstrado que existe uma associação entre quantidade de linfonodos positivos e o aumento da mortalidade (SCHIFFMAN et al. 2011). Doentes com ≥ 4 linfonodos axilares positivos têm um prognóstico pior em comparação com aqueles com 3 linfonodos positivos (CARTER et al. 1989). No Brasil, é observado que pacientes sem acometimento linfonodal apresentam uma elevada sobrevida livre de doença em 10 anos de 99% a 92%, mas desde o acometimento do primeiro linfonodo a taxa de sobrevida em 10 anos diminui para 82% a 65%, sendo que um maior número de linfonodos acometidos (≥ 4 linfonodos acometidos) a sobrevida cai ainda mais, chegando a 45%, e em pacientes com metástase a distância a sobrevida em 10 anos diminui para 14% (Ministério da Saúde 2014).

O tamanho do tumor é um parâmetro que mostra excelente correlação com a sobrevida global e risco de recorrência. Quanto maior a lesão, pior o prognóstico, desde que todas as outras variáveis permaneçam comparáveis. Pacientes com tumores < 1,0 cm têm uma sobrevida geral próxima de 99% comparado com 89% para tumores entre 1,0 e 3,0 cm, e 86% para tumores entre 3 e 5 cm (CARTER et al. 1988). Em pacientes com tumores menores do que 1,0 cm, observa-se que 10 a 20% apresentam comprometimento de linfonodos axilares, e em torno ao 90% dos pacientes sem linfonodos comprometidos têm uma sobrevida livre da doença em 10 anos (FITZGIBBONS et al. 2000; NAROD 2012).

O Colégio Americano de Patologistas recomenda a medida do tumor em pelo menos 2 dimensões, sempre considerando a maior dimensão do tumor invasor para a determinação do estágio. O tumor é medido na macroscopia e também na microscopia, se houver discrepância entre as duas medidas, a medida microscópica do componente invasor é aquela a ser usada para o estadiamento (FITZGIBBONS et al. 2000). Nos tumores com componente invasor e *in situ*, a medida do componente invasor determinará o estágio tumoral.

1.2 ESTADIAMENTO DO CÂNCER DE MAMA

O câncer de mama é classificado e estratificado pelas características histológicas e clínicas.

O sistema de estadiamento histológico mais utilizado para o câncer de mama é a classificação de *Scarff-Bloom-Richardson (SBR)*, o qual foi modificado pelo grupo de Nottingham (ELSTON e ELLIS 1991). O grau de diferenciação SBR é

avaliado de acordo com a tendência para formar glândulas, a presença de núcleos atípicos e o índice mitótico. A soma dos pontos dos três componentes determina os graus: 1 (bem diferenciado) 2 (moderadamente diferenciado) e 3 (fracamente diferenciado). Pacientes com tumores bem diferenciados mostram sobrevivência aumentada em relação aos tumores fracamente diferenciados.

Clinicamente, os tumores sólidos, como no caso do câncer de mama, são avaliados em três parâmetros que permitem descrever a extensão anatômica da doença (Ministério da Saúde 2011). Esses parâmetros são: extensão do tumor primário (T), presença ou a ausência de metástase em linfonodos regionais (N) e presença ou a ausência de metástase a distância (M). A partir das informações obtidas pela classificação do tumor nas categorias T, N e M, o tumor de mama é agrupado em quatro estádios (Quadro 1) (Ministério da Saúde 2011). A descrição extensa de cada categoria do TNM pode ser vista no “Anexo 2”.

Quadro 1 - Descrição do carcinoma ductal de mama em relação ao estágio, TNM e sobrevida

Estádio	TNM	Descrição	Sobrevida em 5 anos
0	Tis N0 M0	Carcinoma <i>in situ</i> . Sem metástase nos linfonodos regionais, sem metástases a distância	99%
I	T1 N0 M0	Tumor menor ou igual a 2 centímetros, sem tumor nos gânglios linfáticos regionais, sem metástases distantes.	92%
II	T0 N1 M0 T1 N1 M0 T2 N0 M0	Não há evidência de tumor primário, metástases nos gânglios, sem metástases a distância. Tumor menor ou igual a 2 centímetros, metástases nos gânglios, sem metástases distantes. Tumor é entre 2 e 5 centímetros, nenhum tumor é linfonodos regionais, sem metástases à distância	82%
IIB	T2 N1 M0 T3 N0 M0	Tumor é entre 2 e 5 centímetros, metástases nos linfonodos, sem metástases a distância. Tumor maior de 5 centímetros, sem linfonodos comprometidos, sem metástases à distância.	65%
IIIA	T0 N2 M0 T1 N2 M0 T2 N2 M0 T3, N1 N2 M0	Não há evidência de tumor primário, metástases nos linfonodos ipsilaterais fixos, sem metástases a distância. Tumor menor ou igual a 2 centímetros, metástases nos linfonodos ipsilaterais fixos, sem metástases distantes. Tumor é entre 2 e 5 centímetros, metástases nos linfonodos ipsilaterais fixos, sem metástases a distância. Tumor é mais de 5 centímetros, metástases nos linfonodos ipsilaterais fixos, sem metástases a distância.	47%
IIIB	T4 NX M0 TX N3 M0	Tumor estende à parede torácica, qualquer envolvimento nodal, sem metástases a distância. Qualquer envolvimento do tumor primário, metástases para nódulos mamários internos ipsilaterais, sem metástases a distância.	44%
IV	TX NX M1	Envolvimento do tumor primário, qualquer envolvimento nodal, metástases a distância.	14%

Nomenclatura: T: Tamanho; N: Linfonodos comprometidos; M: Metástases; X: Qualquer.

1.3 MARCADORES MOLECULARES DO CÂNCER DE MAMA

O câncer de mama é uma doença de comportamento diverso nas pacientes, portanto não poder ser utilizado um só método de tratamento a todos os pacientes. Geralmente as pacientes com câncer de mama são submetidas à cirurgia para remover o tumor e alguns dos linfonodos axilares. A cirurgia pode ser radical, total ou parcial. Com o objetivo de aumentar as chances de cura dos pacientes com a

doença, tem sido incorporadas terapias adjuvantes como: radioterapia, quimioterapia, hormonioterapia e tratamentos com anticorpos monoclonais.

O prognóstico do câncer de mama depende da extensão da doença (estadiamento). Quando a doença é diagnosticada no início, o tratamento tem maior potencial curativo. Quando há evidências de metástases (doença a distância), o tratamento tem por objetivos principais prolongar a sobrevida e melhorar a qualidade de vida.

Vários marcadores moleculares têm sido utilizados na classificação de rotina do carcinoma mamários. A atividade mitótica, por exemplo, pode ser avaliada por marcadores de proliferação celular, que permitem uma medida mais acurada da quantidade de células em proliferação, assim também a possibilidade de investigação prognóstica das neoplasias humanas de maneira simplificada, rápida e pouco dispendiosa (CHEANG et al. 2009). O Ki67 é um anticorpo monoclonal que reconhece um antígeno que está associado ao núcleo celular, sendo expresso em todas as fases do ciclo celular, exceto em G0. A expressão antigênica aumenta com a progressão do ciclo celular, alcançando um pico nas fases G2 e M. Alguns trabalhos têm mostrado a superioridade do Ki-67 como marcador de proliferação celular, por não sofrer tantas influências de fatores internos e externos (VERONESE et al. 1993). Além disso, sua expressão nuclear num período definido do ciclo celular pode representar uma vantagem em relação ao seu emprego como marcador biológico. Há estudos que sugerem um risco aumentado de recorrência da doença em pacientes com níveis elevados de Ki67 (DE AZAMBUJA et al. 2007; CHEANG et al. 2009; INWALD et al. 2013).

O *status* dos hormônios estrógeno e progesterona são fatores prognósticos bem estabelecidos e importantes na escolha do tratamento para câncer de mama. O estrógeno atua no crescimento e desenvolvimento normal das glândulas mamárias e também contribui na carcinogênese da mama (HENDERSON et al. 1988). Além disso, esse hormônio atua na proliferação celular mediada por receptores intracelulares (ER – do inglês, *estrogen receptor*), e, quando é ativado, atua como fator de transcrição, regulando a expressão de vários genes. Em geral, tumores ER negativos estão associados a uma menor sobrevida quando comparados com ER positivo, sendo o *status* de ER utilizado como marcador prognóstico e também na predição de resposta à hormonioterapia, sendo que pacientes ER negativo raramente respondem à terapia endócrina, reforçando a associação entre o nível de expressão de ER e a resposta à terapia (BEZWODA et al. 1991).

Os estrógenos também induzem a síntese de proteínas específicas, tal como receptor de progesterona. Portanto, a determinação da expressão do receptor de progesterona é um outro fator importante na determinação do prognóstico da doença e também na predição de resposta a hormonioterapia (PAYNE et al. 2008).

O Receptor do Fator de Crescimento Epidérmico (EGFR) é uma glicoproteína transmembrana, receptor de tirosina quinase, presente em pequenos níveis nas células epiteliais da mama normal. Trinta e cinco a 60 por cento dos cânceres de mama super-expressam o EGFR, o qual parece ter um efeito estimulador sobre o crescimento de alguns carcinomas de mama. Existe uma relação negativa entre o EGFR e o estado dos receptores hormonais, sendo a super-expressão de EGFR associada com a ausência do ER ou do PR e baixa resposta ao tamoxifeno (STACKER et al. 2002; ECCLES 2011). Sua maior positividade apresenta

associação positiva com pior grau tumoral e índices de proliferação aumentados. Pode ser relacionado à super-expressão de p53 anormal e de fatores angiogênicos (ECCLES 2011).

O oncogene HER-2/neu (c-erbB-2), membro da família do receptor de fator de crescimento epidérmico, localizado no 17q21, codifica uma glicoproteína transmembrana, p185HER2. A proteína HER-2/neu é expressa em níveis baixos nas células epiteliais e mioepiteliais do tecido mamário normal, encontrando-se amplificado em aproximadamente 25% dos casos de câncer de mama. A super-expressão está relacionada ao maior grau histológico, aumento da agressividade tumoral, redução da sobrevida e aumento da recorrência e mortalidade.

Diversos estudos da literatura indicam o c-erbB-2 como fator prognóstico independente e preditivo na resposta à quimioterapia adjuvante baseada no uso da adriamicina. Além disso, com a introdução do anticorpo monoclonal humanizado Herceptin®, como nova modalidade terapêutica em câncer de mama, o *status* do c-erbB-2 tornou-se muito importante na avaliação terapêutica global dessa doença (PAYNE et al. 2008).

Apesar do emprego de marcadores moleculares na rotina clínica, muitas pacientes com mesmo diagnóstico e prognóstico clínico podem apresentar resultados muito diferentes tanto em termos de resposta ao tratamento como recidiva e surgimento de metástase. Isso ocorre porque a classificação atual baseada nos marcadores utilizados é limitada e classifica pacientes com doenças molecularmente heterogêneas em uma única entidade (NAKHLIS e MORROW 2003), levando pacientes com perfis moleculares distintos a terem a mesma classificação por falta de outros indicadores (SØRLIE et al. 2001; SOTIRIOU et al. 2003).

Assim, é de muita importância na pesquisa contra o câncer de mama a identificação de novos marcadores moleculares tumorais que possibilitem o desenvolvimento de métodos diagnósticos de detecção precoce e com capacidade de prognosticar o desenvolvimento da doença.

1.4 PERFIL MOLECULAR DO CÂNCER DA MAMA

Devido à heterogeneidade do câncer de mama, tumores com a mesma classificação clínica e histopatológica podem apresentar diferentes evoluções clínicas no que diz respeito à recidiva, progressão e desenvolvimento de metástases. Os pesquisadores têm tentado compreender melhor a patologia analisando o componente molecular do câncer, os genes potencialmente envolvidos no controle do crescimento, morte e diferenciação das células cancerosas, com a intenção de identificar um padrão de expressão gênica (PEROU et al. 2000; MA et al. 2003).

Estudos de análise do perfil de expressão gênica usando *microarrays* foram aplicados com sucesso para conseguir classificar em quatro os subtipos moleculares do câncer de mama. A classificação intrínseca do câncer de mama os divide em: tumores com expressão de genes relacionados às células epiteliais luminais, tumores com expressão do receptor do fator de crescimento epidérmico humano-HER-2 (ERBB2), tumores com expressão gênica similar às células do epitélio basal e àqueles que expressavam um padrão semelhante ao tecido mamário normal (PEROU et al. 2000; SØRLIE et al. 2001).

O subtipo “Luminal” apresenta expressão dos receptores hormonais, estrogênio e progesterona, assim também apresenta a expressão da proteína GATA3,

da proteína 1 unida a *X-box*, do fator nuclear 3α do hepatócito e LIV-1 regulada por estrogênio. O subtipo Luminal pode se subdividir em “A” e “B”.

Luminal A: além de apresentar a expressão característica dos subtipos luminais, apresenta elevada expressão de *c-kit*, fator de crescimento hepatócito, *insulin-like growth factor-binding protein-3*, *ATF-3*, e componentes do fator transcricional AP-1, tais como: c-fos, c-jun, fosB, e jun-D e baixa expressão de genes relacionados ao crescimento celular como; topoisomerase II alfa, mitotic kinesin-like protein-1, *PCNA*, *CDC2*, *BUB1* e *MAD2L1* (SØRLIE et al. 2001; SOTIRIOU et al. 2003).

Luminal B: apresenta moderada expressão dos genes específicos do subtipo luminal, incluindo o grupo do receptor de estrogênio, além disso, nesse subgrupo é observada amplificação do *HER2* e padrão de proliferação, incluindo genes como: *CCNB1*, *MKI7*, e *MYBL2* (SØRLIE et al. 2001; CHEANG et al. 2009).

HER-2: esse grupo se caracteriza por apresentar elevada expressão de vários genes contidos na região 17q11 e 17q12-q22, incluindo *HER2/ERBB2* e *GRB7*. Além da elevada expressão de genes envolvidos na via RAS, tais como: *Ras-related GTPases*, *RALB*, e *RAB6A* (SORLIE et al. 2003; HU et al. 2006).

Basaloide: caracterizado pela elevada expressão de citoceratina 5 e 17, laminina e ácidos graxos unidos à proteína 7 junto com a ausência ou mínima expressão do receptor de estrogênio e *HER2*. Além da elevada expressão de *frizzled-related protein 1* e oncogene *KIT*, *MYC* e *SFRP1* (regulado pela via de sinalização *Wnt*) e baixa expressão da fibronectina 1 e mucina 1 (SØRLIE et al. 2001; SOTIRIOU et al. 2003).

Normal-like: esse grupo se caracteriza por expressar vários genes conhecidos por serem expressos por tecido adiposo e outros tipos celulares não epiteliais.

Além da classificação intrínseca do câncer de mama, a metodologia *microarray* tem sido usada para vários outros propósitos, os quais incluem a pesquisa de marcadores de resposta a quimioterapias (FOLGUEIRA et al. 2009), marcadores de progressão do carcinoma (CASTRO et al. 2008; MA et al. 2003 2009; ROZENCHAN et al. 2009). Na clínica, o perfil de expressão gênica tem contribuído de forma significativa para a avaliação prognóstica em relação aos fatores preditivos clínico-patológicos habituais.

1.5 POTENCIAL PROGNÓSTICO DOS PERFIS DE EXPRESSÃO

Uma das maiores contribuições dos *microarrays* na pesquisa do câncer de mama foi a identificação de padrões de expressão com potencial preditivo na evolução dos pacientes, identificando diferenças na sobrevida global associado aos subgrupos moleculares intrínsecos, assim os subtipos triplo negativos (basaloide) e *HER-2* foram relacionados a um pior prognóstico e o subtipo Luminal a um melhor prognóstico. (SOTIRIOU et al. 2003; THAKKAR e MEHTA 2011).

Estudos têm permitido determinar o perfil genético das neoplasias com mau prognóstico, cuja assinatura incluiu genes reguladores do ciclo celular, invasão, metástases, angiogênese e sinais de transdução que se apresentaram significativamente aumentadas nos subtipos de pior prognóstico (VAN'T VEER et al. 2002).

Dos estudos moleculares do câncer de mama, tem sido proposto vários conjuntos de genes que permitiriam classificar e prognosticar a doença, mas só dois deles estão disponíveis para uso clínico: o *MammaPrint* e *Oncotype*.

MammaPrint (Agendia, Huntington Beach, CA) é um ensaio *microarray* de DNA composto por 70 genes. O uso desse conjunto de genes pode ajudar a prever o risco de recidiva e metástases do câncer nos próximos 10 anos (VAN'T VEER et al. 2002). O teste é indicado para pacientes com câncer de mama com idade inferior a 61 anos, com estágio I ou II, tumor <5 centímetros e sem comprometimento linfonodal. Como resultado do teste, as pacientes são classificadas em dois grupos de risco: maior risco de metástase de 10 anos (50% de probabilidade) ou baixo risco (probabilidade de 10-15%), a fim de determinar a necessidade e conveniência de realizar tratamento com quimioterapia.

Oncotype DX (Genomic Health Inc) é um ensaio *RT-PCR*, composto por 21 genes, destinado a prever o risco de apresentar recidiva (PAIK et al. 2006). É dirigido a pacientes diagnosticadas em estádios I ou II, positivas para receptor de estrogênio e sem linfonodos acometidos. Como resultado do teste, para assim poder aconselhar melhor sobre a conveniência de realizar quimioterapia complementar à hormonioterapia, as pacientes são classificadas pelo risco de apresentar recidiva em três grupos (alto, intermédio e baixo).

1.6 EXPRESSÃO GÊNICA ASSOCIADA À PROGRESSÃO DO CÂNCER

No tecido mamário, uma vez que as células tumorais adquirem a capacidade de invasão, o tumor ganha o potencial de disseminação aos órgãos distantes. O modelo clássico de progressão do carcinoma ductal sugere uma progressão contínua, mas não obrigatória, através de uma série de etapas. O processo se inicia em um tecido normal, progredindo para um epitélio plano atípico, seguido por uma hiperplasia ductal atípica, avançando para um carcinoma ductal *in situ* e culminando em um CDI, que pode desencadear em metástases (WELLINGS e JENSEN 1973).

As modificações morfológicas estão acompanhadas por alterações moleculares, genéticas e epigenéticas, observando-se alterações do padrão de expressão nos diversos estados da progressão do câncer (KREEGER 2010; MICHOR et al. 2011).

Na fase pré-invasiva do carcinoma ductal, foi mostrado que as células epiteliais apresentam mudanças de expressão gênica que precedem às manifestações morfológicas, evidenciado que as células epiteliais tumorais do carcinoma ductal *in situ* puro sofrem maiores alterações no padrão de expressão gênica que o componente *in situ* de uma lesão invasiva, sendo que o último pouco diverge das células do componente invasivo (CASTRO et al. 2008). Alguns genes como *ANAPC13*, com expressão diminuída em eventos iniciais do carcinoma ductal mamário, têm sido sugeridos como potenciais biomarcadores para invasão do carcinoma ductal *in situ* (SENS-ABUÁZAR et al. 2012).

O microambiente parece desempenhar um papel importante na invasão das células tumorais ao tecido adjacente e tem sido estudado por vários grupos (MA et al. 2009; ROZENCHAN et al. 2009; KNUDSEN et al. 2012). Geralmente, o estroma no câncer apresenta uma abundância de células inflamatórias e fibroblastos ativadas, possivelmente seja o resultado de mudanças epigenéticas na fase pré-cancerosa (TLSTY 2001). Estudos têm demonstrado que existem diferenças de expressão entre as células mioepiteliais do tecido normal em comparação às células do carcinoma ductal *in situ*. As células mioepiteliais do CDIS expressam diminuição dos níveis de laminina, tenascina, trombospondina, citoqueratinas, tropomiosina e receptor da oxitocina, e elevada expressão de proteases (catepsinas, MMP-2, e PRSS11), inibidores de protease (trombospondina 2 SERP- ING1, cystatinC, e TIMP3) e colágenos (KHAMIS et al. 2012). Por meio dessas interações recíprocas que o tumor pode criar um microambiente favorável para a proliferação, neovascularização e a estimulação da produção de proteases capazes de degradar tecidos adjacentes, aumentando a probabilidade da invasão do tumor (BERGAMASCHI et al. 2008).

Em mama, uma vez que as células tumorais adquirem a capacidade de degradar a membrana basal, o tumor invasivo adquire potencial de disseminação. O primeiro local é o dos linfonodos regionais. Esse processo envolve alterações genéticas, desregulação de múltiplos genes e dos mecanismos regulatórios da transcrição (TALMADGE e FIDLER 2010).

Embora estudos genômicos e transcriptômicos venham sendo desenvolvidos para avaliar invasão linfonodal, é importante salientar que ainda não é possível distinguir uma única assinatura característica das pacientes com e sem linfonodos acometidos, ainda que alguns estudos estejam permitindo identificar genes

envolvidos no acometimento linfonodal (ABBA et al. 2007; SMEETS et al. 2011; BISWAS et al. 2014).

1.6.1 Estudo independente dos componentes morfológicos do câncer

Os tumores são formados por uma mistura complexa e heterogênea de tecidos compostos por vários tipos celulares. Durante a progressão da doença, acontecem mudanças que alteram a quantidade, morfologia e expressão gênica das células (SIMPSON et al. 2005; ANDERSON et al. 2006; CREIGHTON et al. 2010). No CDI de mama, as células epiteliais são predominantes, no entanto, estão presentes vários outros tipos de células, incluindo células mioepiteliais e células do sistema imune.

Muitas abordagens têm sido usadas para analisar às populações celulares de forma independente. Uma das metodologias que permite isolar populações celulares específicas é a microdissecção por captura laser (LCM), descrita pela primeira vez em 1996 (EMMERT-BUCK et al. 1996). Atualmente, essa metodologia está bem estabelecida como uma ferramenta poderosa para o isolamento de células de interesse sob controle morfológico por meio de identificação microscópica das células de interesse (ESPINA et al. 2006). Com essa metodologia é possível obter RNA íntegro, mas em baixas quantidades (WANG et al. 2006; BEVILACQUA et al. 2010), que precisa ser amplificado para possibilitar sua utilização em diversas abordagens moleculares. Vários protocolos de amplificação de RNA mensageiro estão disponíveis e vários estudos demonstram que esses procedimentos não enviesam o perfil de expressão quando avaliados por *microarray* (GOMES et al. 2003; WANG et al. 2006), por RT_PCR quantitativo (FELDMAN et al. 2002;

XIANG et al. 2003; FERREIRA et al. 2010) e por metodologias mais recentes como sequenciamento de RNA (RNAseq) (FERREIRA et al. 2015). Essa metodologia tem permitido estudar vários aspectos no contexto do câncer de mama.

1.7 CÂNCER E *NON CODING RNA*

Atualmente, os perfis de expressão do câncer de mama estão baseados em assinaturas de expressão que focalizam na análise de genes que codificam para proteínas. Nos últimos anos, tem sido demonstrado que o genoma humano contém mais do que somente genes. No projeto ENCODE foi mostrado que 90% do material genético é transcrito a RNA, do qual 1.2% é traduzido a proteína. O material genético transcrito, mas não traduzido, é conhecido como RNA não-codificante (ncRNA, do inglês *non coding RNA*) (ENCODE et al. 2007).

Esses ncRNAs funcionam diretamente como RNAs estruturais, catalíticos ou reguladores (HUNG e CHANG 2010; LEE 2012; MATSUI et al. 2013). De acordo com seus tamanhos, os ncRNAs podem ser classificados como: pequenos ncRNAs (menores a 200pb, por exemplo, miRNAs, siRNAs e piRNAs) e longos ncRNAs (lncRNAs) (maiores a 200pb, por exemplo, lincRNAs, macroRNAs) (YANG et al. 2014).

Muitos *loci* têm transcrição das duas fitas do DNA, uma fita com transcrição não codificante e a outra fita com um transcrito que codifica proteína. Os transcritos com orientação oposta no mesmo *locus* são conhecidos como pares senso-anti-senso (SAS) (LAPIDOT et al. 2006). A função dos pares de SAS ainda não foi

identificada, embora tem se proposto uma atividade regulatória da expressão gênica (PELECHANO e STEINMETZ 2013).

Foi observado que os lncRNAs desempenham um papel regulador na biologia do câncer (GIBB et al. 2011; YANG et al. 2014). Os lncRNAs estão desregulados em diferentes tipos de câncer e os níveis de expressão de certos lncRNAs estão associados aos diferentes estádios do câncer (MARUYAMA e SUZUKI 2012; YANG et al. 2014). Na progressão do câncer, tem-se identificado correlação entre o grau de diferenciação do câncer de próstata e a expressão do ncRNAs intrônicos (REIS et al. 2004; NAKAYA et al. 2007). Assim, também foram relatadas diferenças de expressão entre tumor primário e metástases no câncer pancreático (TAHIRA et al. 2011) e no câncer de mama (SØRENSEN et al. 2015).

2 JUSTIFICATIVA

A análise de expressão gênica tem se mostrado uma excelente ferramenta para caracterizar vários aspectos de câncer de mama, permitindo uma classificação molecular da doença e a análise de outros aspectos relacionados à heterogeneidade da doença.

Em relação à progressão da doença, o estudo do acometimento linfonodal seria de grande interesse tanto para contribuir com a compreensão da progressão do câncer de mama como para identificar biomarcadores potenciais que auxiliassem na discriminação de pacientes com maior potencial de progressão

Diversos estudos mostraram que os ncRNAs têm um importante papel na regulação da expressão gênica, inclusive em câncer. Além disso, os estudos de expressão gênica feitos anteriormente para caracterizar o câncer de mama foram realizados empregando-se somente microarranjos de sequências de genes codificadores para proteínas já conhecidas. Assim, neste estudo, propomos analisar os padrões de expressão gênica do CDI utilizando a técnica de *microarray* de DNA, uma lâmina customizada que contém as sequências de genes codificadores para proteínas já conhecidas e também incluindo transcritos não-codificadores intragênicos.

Assim, devido à grande necessidade de serem identificados genes envolvidos na progressão do câncer de mama, estudos que avaliem de forma pormenorizada às etapas importantes da progressão poderiam contribuir para a compreensão desse complexo processo, além de identificar potenciais biomarcadores.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Avaliar a expressão de transcritos codificadores para proteína e não codificadores nas células epiteliais tumorais de carcinoma ductal invasivo de mama com ou sem linfonodos acometidos.

3.2 OBJETIVOS ESPECÍFICOS

- Determinar o perfil de expressão gênica diferencial das células epiteliais tumorais de carcinoma ductal invasivo de mama em amostras microdissecadas a *laser* utilizando uma lâmina de *array* contendo 244.000 elementos codificadores e não codificadores.
- Avaliar o perfil de expressão gênica em subgrupos tumorais levando em conta o *status* do receptor de estrógeno e/ou progesterona, de amplificação ou expressão do oncogene *HER-2* e da presença ou ausência de acometimento dos linfonodos.
- Validar transcritos representantes de genes codificadores de proteína que foram diferencialmente expressos entre os subgrupos com e sem linfonodos acometidos.

- Buscar pares de transcritos codificadores e não codificadores com nível de expressão invertida nos diferentes grupos para identificar potenciais pares de transcritos coregulados.

4 MATERIAL E MÉTODOS

O estudo foi realizado em várias etapas sucessivas, desde a seleção das células epiteliais neoplásicas do CDI com microdissecção a laser, extração e purificação do RNA, transcrição e amplificação do material genético, marcação com cianinas para a posterior hibridização na lamina de *microarray* customizada com 244 sondas de transcritos codificadores e não codificadores, varredura da lamina, várias análises de expressão até, finalmente, a validação dos genes selecionados. Pode-se observar o fluxo do trabalho na “Figura 3”.

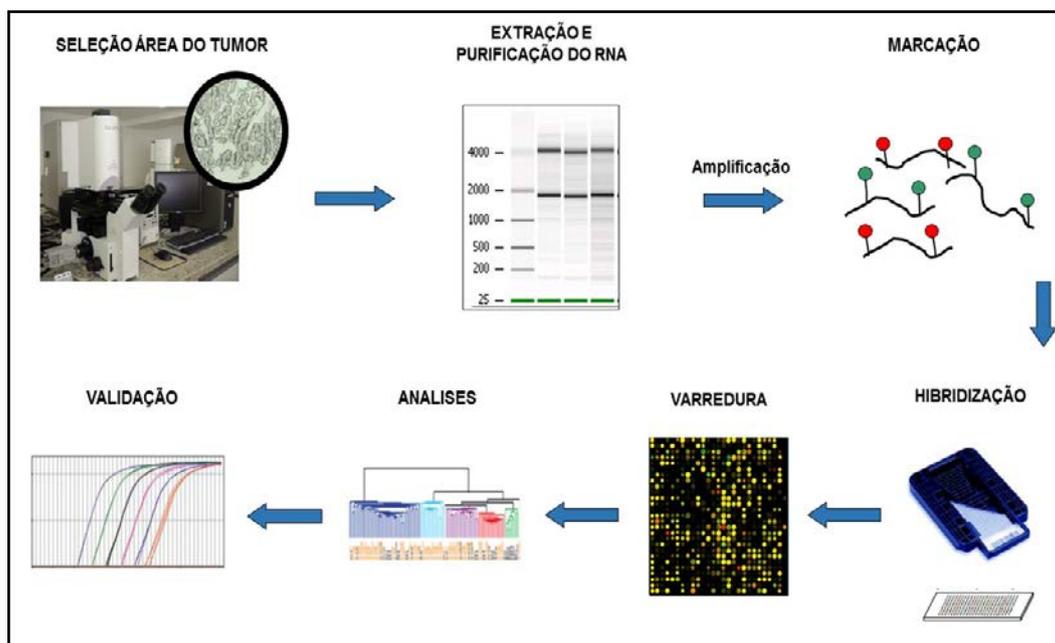


Figura 3 - Representação esquemática da metodologia do projeto.

4.1 AMOSTRAS DE TECIDO

Foram utilizadas 103 amostras de carcinoma ductal invasivo de mama do banco de tecidos do A.C.Camargo Cancer Center, São Paulo, coletadas entre 2002 e 2010, com consentimento livre e esclarecido assinado pelas pacientes para coleta e armazenamento do material no Biobanco/A.C.Camargo Cancer Center. As amostras foram examinadas pelo médico patologista Dr. Hugo Campos, a fim de certificar a presença de tecido tumoral e ausência de necrose e de infiltrado inflamatório intenso. A preparação da amostra para a avaliação do patologista foi realizada com o auxílio do pessoal do laboratório de macromoléculas.

Todas as amostras incluídas no projeto foram previamente submetidas aos ensaios imunoistoquímicos de rotina para receptores hormonais (estrógeno e progesterona) e do receptor C-erbB2-2. Os ensaios para C-erbB2-2 que apresentaram resultado inconclusivo (2+) foram avaliados pela técnica de FISH (*fluorescence in situ hybridization*), no Departamento de Anatomia Patológica do A.C.Camargo Cancer Center.

Todas as amostras de IDC incluídas neste estudo estavam livres de qualquer tipo de tratamento neoadjuvante antes da cirurgia (RT e/ou QT). No “Anexo 3” pode ser encontrado o instrumento de coleta dos dados clínicos e anatomopatológicos retirados dos registros do A.C.Camargo Cancer Center.

4.1.1 Classificação das amostras

As amostras foram divididas, baseadas na expressão IHQ dos receptores hormonais (receptor de estrógeno (ER) e progesterona (PR), do C-erbB2-2 (ou HER-2) e Ki-67, além disso foram considerados, quando existiam, os dados de FISH dirigida ao gene erbB2-2. Os grupos foram denominados Luminal A, Luminal B, Luminal B', HER-2, Triplo negativo (PEROU et al. 2000; SOTIRIOU et al. 2003; CHEANG et al. 2009), como é mostrado na “Tabela 1”.

Tabela 1 - Classificação baseada em marcadores imunoistoquímicos.

	ER	PR	C-erbB2-2	Ki-67
Luminal A	+	+	-	<15%
Luminal B	+	+/-	+	-
Luminal B'	+	+/-	-	≥15%
HER-2	-	-	+	+/-
Triplo Negativo	-	-	-	+/-

Simultaneamente, as amostras foram classificadas segundo o comprometimento linfonodal, baseado no procedimento clínico (LESTER et al. 2009), sendo categorizadas em três diferentes grupos:

Ln0: sem evidências histológicas ou nem IHQ de metástases nos linfonodos;

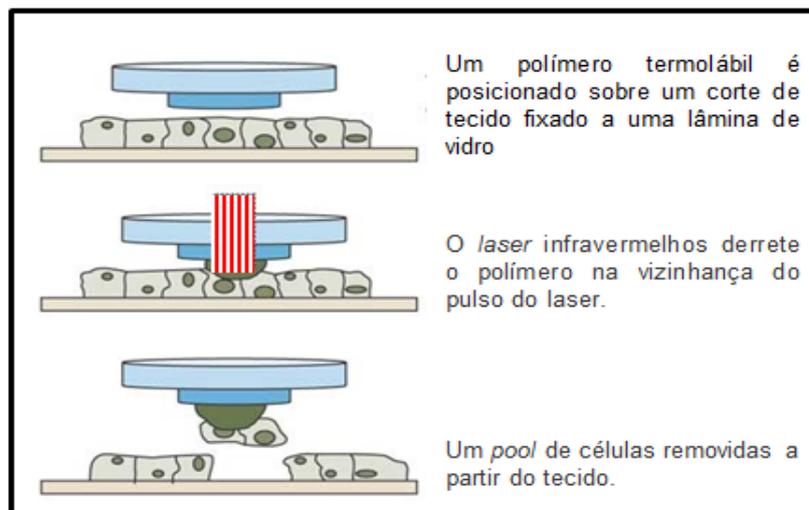
Ln1-3: metástases em 1-3 linfonodos axilares;

Ln≥4: metástases em 4 ou mais linfonodos axilares.

4.2 MICRODISSECÇÃO A LASER DAS CÉLULAS EPITELIAIS DE CDI

Com a intenção de estudar exclusivamente o componente epitelial do CDI, utilizamos a metodologia de microdissecção a *laser* com o equipamento *Pixcell II LCM* (*Arcturus Engineering, Mountain View, CA, USA*), o qual pertence ao Departamento de Anatomia Patológica do A.C. Camargo Cancer Center.

Esse 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 (ESPINA et al. 2006). A lâmina com o corte histológico, previamente corado, é fixada ao *chariot* por um mecanismo de sucção a vácuo. O feixe de raio *laser* incide sobre uma base de captura, ou *CapSure HS™* *Arcturus®*, revestido por um filme termoplástico. O *CapSure HS* fica posicionado entre o feixe de *laser* e o corte histológico. Com o disparo do feixe de *laser*, o filme termoplástico expande e a célula escolhida adere. Quando o pulso cessa, o filme retrai levando consigo a célula aderida à superfície do *CapSure* (Figura 4). Dessa forma, repete-se a operação quantas vezes forem necessárias. A quantidade de células necessárias varia de acordo com o tipo de tecido e com o estudo que será realizado.



Fonte: Modificado de ESPINA et al. (2006).

Figura 4 - Representação esquemática do procedimento microdissecção a laser.

Cada amostra foi cortada utilizando um criostato graduado para obter cortes de 10 μ m; esses cortes foram depositados em uma lâmina carregada eletrostaticamente e armazenados em freezer -80°C até o momento da microdissecção. As amostras foram mantidas congeladas com ajuda de gelo seco e descongeladas 30 segundos antes da coloração para início do processo de microdissecção; em seguida, foram coradas uma a uma com Hematoxilina (protocolo em extenso no “Anexo 4”) e imediatamente microdissecadas a *laser*, capturando exclusivamente às células epiteliais neoplásicas (ver “Figura 5”) num *CapSure HSTM Arcturus®*, usando feixe de *laser* com diâmetro de 7,5 μ m e um tempo de captura máximo de 10 minutos.

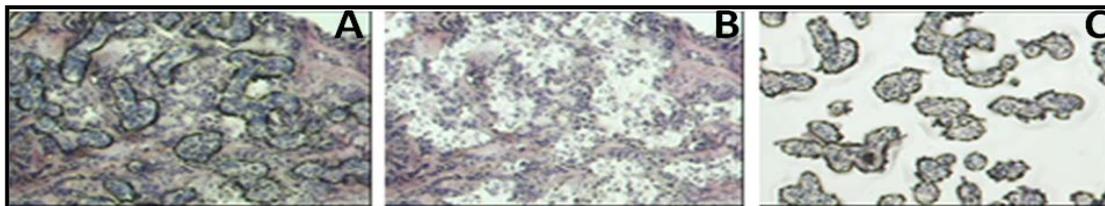


Figura 5 - Representação de uma amostra microdissecada a *laser*. As fotografias representam uma típica microdissecção a *laser* de células epiteliais, nas quais é possível observar a captura das células desejadas. A) Tecido antes da microdissecção; B) Tecido após a microdissecção; C) *CapSure* com as células capturadas.

4.3 EXTRAÇÃO E AMPLIFICAÇÃO DO RNA DAS CÉLULAS MICRODISSECADAS A LASER

Imediatamente após a captura das células epiteliais no *CapSure HST*TM utilizando o *kit Pico Pure RNA isolation*TM *Arcturus*®, o RNA total foi extraído segundo o protocolo do fabricante.

Foi acoplado o anel de extração ao *CapSure* e adicionado o *extraction Buffer* (XB), incubando a 42°C, extraíndo o RNA, o qual foi imediatamente congelado e mantido em gelo seco por um tempo máximo de 2 horas. O RNA foi purificado segundo o protocolo do fabricante, tratado com DNase I para remoção do DNA genômico contaminante e eluído em 15µL de *Elution Buffer*. Uma alíquota de 1,5µL foi separada para avaliação da qualidade do RNA extraído e o restante (13,5µL) foi concentrado até 4,15µL mediante o uso de uma centrifuga a vácuo (*Centrivap Labconco*). O protocolo está descrito em detalhes no “Anexo 5”.

4.3.1 Avaliação da Qualidade e Integridade do RNA Total

A qualidade do RNA total foi analisada utilizando o equipamento *Bioanalyser 2100 (Agilent Technologies)*, o qual realiza uma eletroforese capilar de alta tensão. O *software* do equipamento calcula a proporção entre as bandas ribossomais 18S e 28S para amostras eucarióticas. É construído um eletroferograma e calculado um número de integridade do RNA (*RIN-RNA Integrity Number*) que varia de 0 a 10. Um exemplo da qualidade do RNA total, extraído das amostras microdissecadas a laser, está representado na “Figura 6”. Devido à baixa quantidade de RNA obtido de células capturas mediante microdissecção a *laser*, foi utilizado o *RNA 6000 Pico LabChip Kit*, que permite uso de amostras com quantidade extremamente pequenas de RNA (de 200pg a 5ng). O protocolo pode ser visualizado no “Anexo 6”.

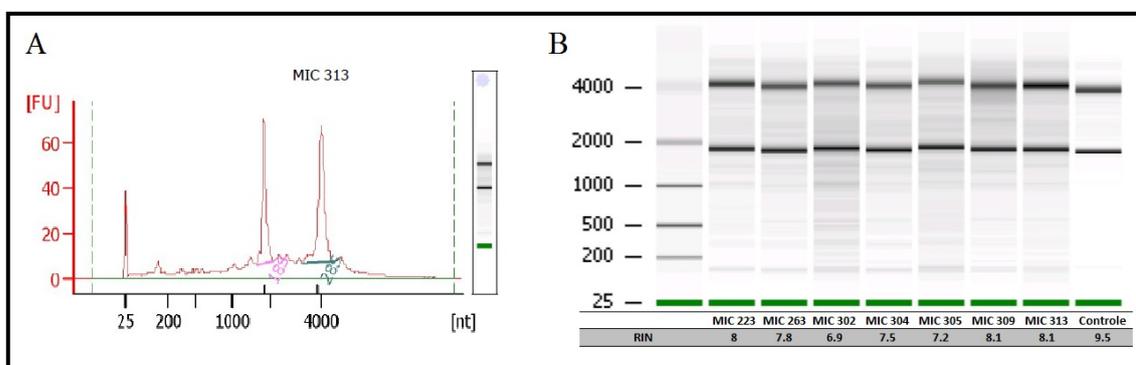


Figura 6 - Perfil analítico de RNA total por eletroforese Capilar (Bioanalyser 2100, Agilent Technologies). O painel apresenta o eletroferograma obtido a partir de 1µl de RNA total, tratado com DNasa, não quantificado. A: Gráfico do eletroferograma mostrando a fluorescência em relação ao tamanho dos nucleotídeos, observasse 3 picos, o primeiro em 25nt é o marcador, o pico em 1800nt corresponde a subunidade 18S e no 4000nt a subunidade 28S do ribossomo. B: Gel do eletroferograma das amostras, mostrando as subunidades 18S e 28S como bandas de eletroforeses, e as sombras o “*semear*” os picos de degradação do RNA.

4.3.2 Amplificação do RNA com cauda de poli A (RNA poliA)

Para este trabalho, nós adaptamos o protocolo de amplificação de RNA poliA utilizado pelo grupo e descrito em CASTRO et al. (2008) com a finalidade de possibilitar a utilização de RNA amplificado a partir de amostras microdissecadas a laser para hibridização na plataforma *Agilent*. A amplificação e marcação com fluoróforos foi realizada utilizando o kit de amplificação *Quick-Amp Labeling kit* (*Agilent Technologies*) com algumas modificações (Figura 7) descritas a seguir.

A partir de 7-12ng de RNA total em um volume de 4,15µL, foram adicionados 0,6µL de iniciador Oligo-dT ligado ao promotor de T7 e 1µL de RNA “*Spike*”, sendo *Spike A* para RNA das amostras teste marcadas com Cy3 e *Spike B* para o RNA referência marcado com Cy5. O RNA *Spike* é um RNA proveniente da transcrição de uma única sequência de 55-mer de adenovírus humano tipo 6, gene E1A, região13S, que funcionam como controle da amplificação e marcação, uma vez que o oligoarranjo construído pela *Agilent* possui sondas específicas para esses RNAs.

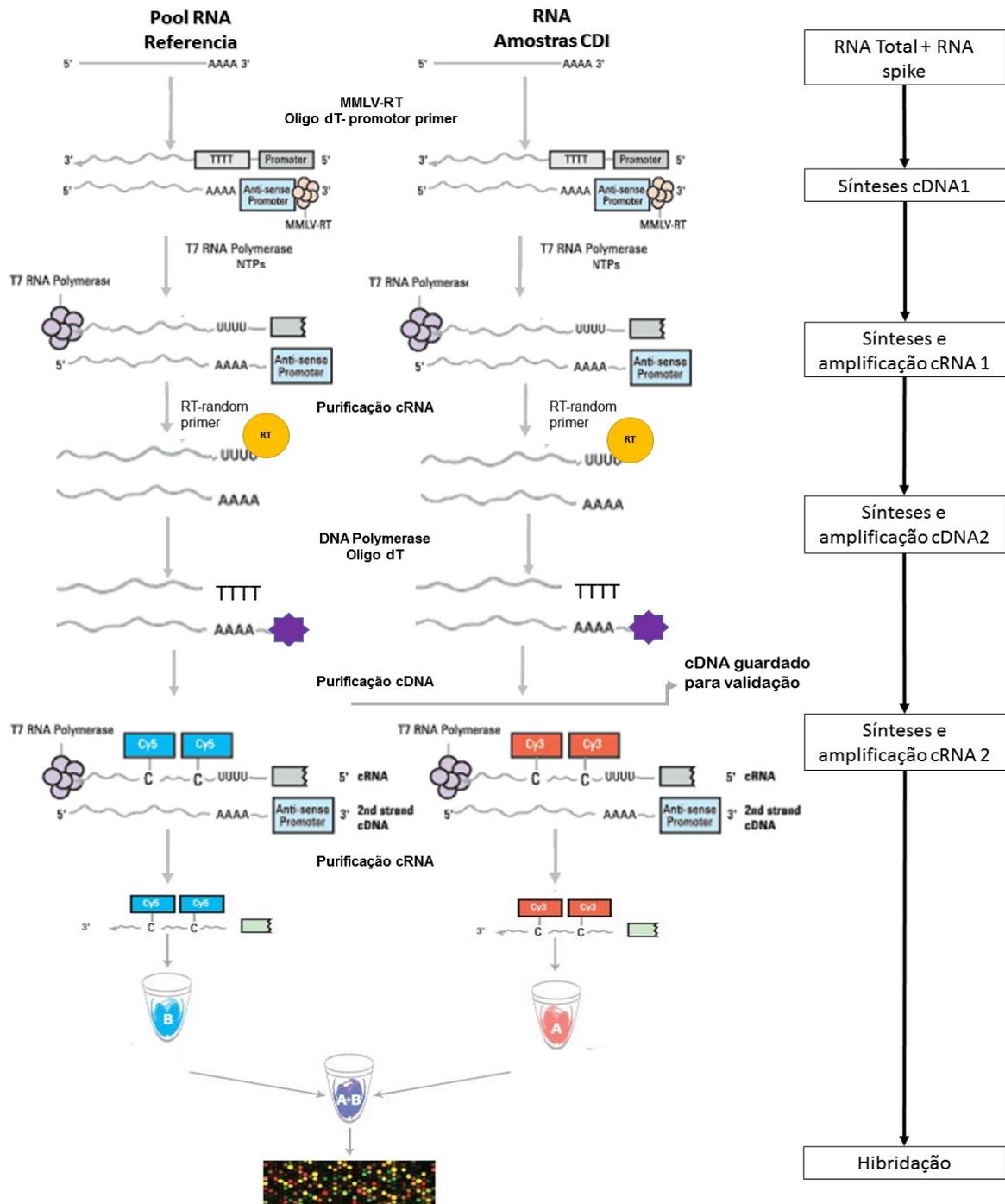
Os tubos foram incubados a 65°C por 10 min. para desnaturar o RNA, seguido por um rápido esfriamento em gelo por 10 min.; em seguida, foram adicionados 4,25µL de *master mix* para transcrição reversa, cDNA (2µL de 5X *First Strand Buffer* previamente aquecido por 5 min. a 80°C 1µL 0,1M DTT, 0,5µL 10mM dNTP mix, 0,5µL MMLV-RT, 0,25µL RNase OUT) incubando no termociclador por 2 horas a 40°C; posteriormente, foram adicionados 30µL do tampão de transcrição *Transcription master mix* (4,85µL H₂O DEPC 10µL 4X *transcription buffer*, 3µL 0,1M DTT, 8µL 25mM NTP mix (Promega), 3,2µL PEG previamente aquecido a 40°C por 1min, 0,25µL RNase OUT, 0,3µL *Pyrophosphatase inorganic*, 0,4µL T7

RNA polimerase) incubando por 3 horas a 40°C para sínteses do RNA complementar (cRNA) que permitisse o segundo ciclo de amplificação. Assim, o cRNA foi purificado empregando-se o *RNeasy* mini kit (Qiagen), seguindo o protocolo do fabricante. O protocolo de purificação de RNA pode ser visualizado no “Anexo 7”.

O cRNA gerado durante o primeiro ciclo de amplificação (cRNA1) foi concentrado na centrífuga a vácuo a 6,7µl e foi empregado no segundo ciclo de amplificação. O segundo ciclo de amplificação foi feito segundo protocolo otimizado em nosso grupo de pesquisa (CASTRO et al. 2008). A primeira fita de cDNA foi feita mediante adição de 2µg de iniciador *Random Hexameros (dN6)* e incubação a 70°C por 10 min., com rápido esfriamento em gelo por 10 min; em seguida, foram adicionados 12,3µl de *master mix* para transcrição reversa cDNA (4µl ImProm II 5X buffer, 4,8µl MgCl₂ 25mM 2µl dNTPs 10mM, 0,5µl Rnase OUT 1µl ImProm-II transcriptase reversa), incubando 2 horas a 42°C. Logo após foi feita a síntese da segunda fita de cDNA, adicionando 80µl de *master mix* (64,4µl H₂O-DEPC 10µl *Advantage PCR buffer*, 0,7µl iniciador Oligo -dT (Agilent) 2µl dNTPs 10mM, 0,72µl RNase H (2U/µl Gibco BRLCat# 18021-071) 2µl *Advantage Polymerase* (Clontech Cat# 8417-1). O cDNA foi purificado por fenol clorofórmio, o 20% do cDNA obtido foi guardado a -20°C para validação das amostras, o 80% do cDNA obtido foi utilizado segundo ciclo de amplificação. O segundo ciclo de amplificação utiliza a enzima T7 RNA polimerase, que simultaneamente amplifica o material alvo, gerando o cRNA e incorpora o CTP marcado com cianina 3 (Cy3) ou cianina 5 (Cy5) (protocolo em extenso no “Anexo 8”). As amostras testes foram marcadas com Cy3 e o RNA referência foi marcado com Cy5 para serem utilizados na hibridização competitiva.

O rendimento esperado da etapa de amplificação do protocolo disponibilizado pela *Agilent* é da ordem de 100 vezes quando comparado à concentração de RNA total inicial, no entanto, com a incorporação da etapa adicional de amplificação o rendimento chega a ordem de 1000 vezes.

Seguindo as recomendações do kit de amplificação *Quick-Amp Labeling kit* (*Agilent Technologies*), foram consideradas amostras aptas para a hibridação aqueles que tiveram uma massa de RNA amplificado $\geq 0.9\mu\text{g}$ e uma atividade específica das Cy3 ou Cy5 $> 8.0\text{pmol}/\mu\text{g}$ de cRNA amplificado. Um esquema explicando o processo está apresentado na “Figura 7”.



Fonte: Modificado do protocolo Two-Color Microarray-Based

Figura 7 - Esquema das etapas da amplificação e marcação do RNA teste e RNA referência para ensaio microarray. Representa-se as etapas de sínteses e purificação dos cDNAs e cRNAs.

4.3.3 Hibridação em plataforma de microarray

Neste estudo foi utilizada a hibridação competitiva. Nesse caso, as hibridações são realizadas utilizando uma amostra referência, comum a todas as hibridações, marcada com um corante fluorescente distinto da amostra teste, que se refere às amostras de interesse do estudo. As amostras teste e referência são misturadas e hibridadas em uma mesma lâmina. Como a hibridação é competitiva, os valores de fluorescência obtidos revelam níveis relativos de expressão de cada transcrito na amostra comparada com a amostra referência. Dessa forma, o uso da amostra referência possibilita comparações de experimentos independentes (NOVORADOVSKAYA et al. 2004).

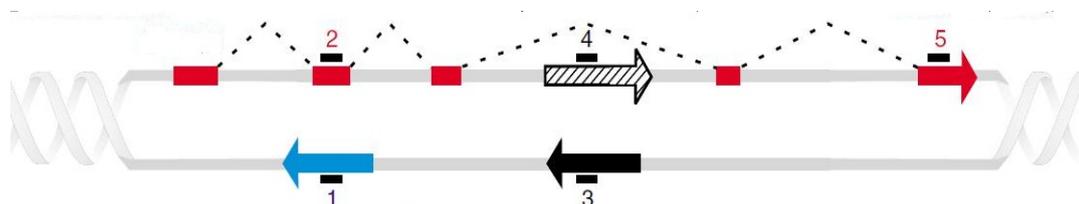
A hibridação dos *microarrays* foi realizada com 3µg de cRNA da amostra teste marcada com Cy3 e mesma quantidade do RNA referência marcados com Cy5, por um período mínimo de 18 horas e máximo 22 horas, a 65°C em forno rotatório a 10 RPM, utilizando o tampão de hibridação *Hi-RPM* fornecido pela *Agilent*. Após a hibridação, as lâminas foram lavadas utilizando tampões fornecidos pelo fabricante (*Agilent, USA*) para eliminação de alvos não específicos e eliminação de *background*. Nesse processo, as lâminas são retiradas do suporte e lavadas por 1 min. em tampão de lavagem #1 à temperatura ambiente, seguido pelo tampão #2 por 1 min a 37°C. A seguir, as lâminas passaram pelo processo de secagem que consiste em lavagens em acetone nitrila por 10s e na solução de estabilização por 30s.

4.3.4 Descrição do microarray utilizado

A fim de identificar RNAs intrônicos não-codificadores (ncRNAs) e RNAs codificadores de proteína expressos em tecido tumoral ou normal humano, foram utilizados *microarrays* contendo 244 mil sondas de oligonucleotídeos 60-mer, em um desenho essencialmente semelhante ao do *combined intron-exon oligoarray* 44k (NAKAYA et al. 2007), desenvolvido pelo grupo do Dr. Sergio Verjovski Almeida e do Dr. Eduardo Reis - Laboratório de expressão Gênica de Eucariotos – Instituto de Química da Universidade de São Paulo (MOREIRA 2010).

A lâmina 244K contém 44.074 sondas para um ou mais *éxons* de genes codificadores de proteínas humanas conhecidas [(sondas idênticas às da lâmina comercial *Whole Human Genome Oligo Microarray 44K (Agilent Technologies)*] (Figura 8, sonda 5). Além disso, a lâmina 244K possui sondas para todos os RNAs não-codificadores identificados em 75% de todos os genes humanos conhecidos (NAKAYA et al. 2007). Especificamente, a lâmina 244K possui 158.270 sondas para as fitas senso (Figura 8, sonda 4) e anti-senso (Figura 8, sonda 3) de todos os transcritos totalmente intrônicos não-codificadores (TINs) humanos conhecidos (para os quais existem duas ou mais ESTs no GenBank) (LOURO et al. 2007). Possui ainda 28.878 sondas para transcritos parcialmente intrônicos não-codificadores (PINs) conhecidos (com duas ou mais ESTs), tanto na fita anti-senso (Figura 8, sonda 1) quanto na fita senso correspondentes aos *éxons* dos genes codificadores (Figura 8, sonda 2), nos quais o ncRNA anti-senso PIN se sobrepõe (LOURO et al. 2007). Completando o total de 244 mil sondas, estão também presentes nessa plataforma dezenas de sondas de controles positivos e negativos recomendados pela

Agilent, incluindo sondas complementares a um *Spike* RNA que é adicionado à amostra de RNA para controle positivo da marcação e hibridação.



Fonte: NAKAYA et al. (2007).

Figura 8 - Desenho do microarranjo de oligonucleotídeos *intron-éxon*. Sonda 60-mer customizada pela empresa Agilent. Sonda 1: Sonda para transcritos parcialmente intrônicos antisense. Sonda 2: Sonda representando cada éxon que se sobrepõe, na fita oposta, a cada transcrito parcialmente intrônico. Sondas 3 e 4 pares de sonda reverso complementares para cada uma das duas possíveis fitas dos transcritos totalmente intrônicos. Sonda 5: Sonda para um éxon de um gene conhecido, codificador de proteínas.

4.3.5 Captura de sinal e pré-análise das lâminas de microarray

Os *arrays* foram digitalizados com o emprego do *Scanner Agilent Bundle modelo B* (Agilent, USA), o qual dispõe de 2 *lasers* com comprimento de ondas diferentes, que estimulam especificamente o Cy3 ou o Cy5 (Quadro 2).

Quadro 2 - Características específicas de cada fluoróforo utilizado.

Fluoróforo	Excitação (nm)	Emissão (nm)	Coloração
Cy3	550	570	Verde
Cy5	649	670	Vermelho

Quando as moléculas são excitadas, emitem um sinal proporcional ao número de moléculas hibridadas, que é captado pelo *scanner*. A quantificação desse sinal ocorre devido à conversão do comprimento de onda, emitido pelos fluoróforos

excitados, em energia elétrica, produzindo um sinal mensurável o que é convertido em uma imagem (Figura 8).

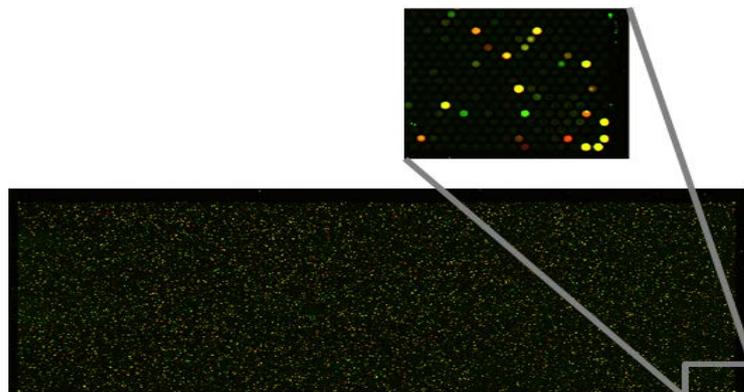


Figura 9 - Lâmina 244K hibridada com RNA da amostra (Cy3 verde) teste e RNA referência (Cy5 vermelho). A hibridação de ambos RNA em um mesmo *spot* geram a cor amarela, os *spots* nas esquinas da lâmina contem sondas dirigidas aos *spikes*, cuja hibridização é um controle positivo.

Os dados de intensidade de todos os *spots* digitalizados no *scanner* foram extraídos e pré-processados utilizando o programa *Agilent Feature Extraction* (v 10.7.1). Esse *Software* permite extrair os valores de intensidade de cada *spot*, obtendo os valores de fluorescência correspondente à expressão de cada transcrito na amostra-teste e na amostra-referência.

Além disso, o *Software FE* possui fatores de correção específicos para as lâminas da *Agilent*, como o fator de correção do “Efeito Domo”, produzido pela bolha de hibridização, característico do método de hibridização da *Agilent*. Durante a hibridização da lâmina, é deixada uma pequena bolha de ar na câmara que percorre a borda da lâmina durante a hibridização, enquanto a lâmina é girada no hibridizador giratório. Dessa forma, a periferia da lâmina passa por momentos sem solução (presença da bolha), entretanto o centro da lâmina está sempre em contato com a

solução; conseqüentemente, os *spots* localizados no centro da lâmina apresentam intensidade maior do que os localizados na periferia denominado de “Efeito Domo”.

O *software FE* também corrige as diferenças de sensibilidade de detecção dos dois fluoróforos (Cy3 é detectado pelo *scanner* com sensibilidade maior que Cy5, quando ambos estão em baixas intensidades). Foi feita também a subtração do *background* local (valor de intensidade ao redor de cada *spot*).

4.4 ANÁLISE DOS RESULTADOS DE CDNA MICROARRAY

Os *microarrays* foram analisados com o *software* genespring GX 12.1 (Agilent Technologies). Inicialmente, foram importados os dados de *Control Type*, *Probe Name*, *Signal* e *Feature*, obtidos com programa *Agilent Feature Extraction* (v 10.7.1). Os dados foram normalizados por transformação *log* e transformação de *baseline*.

Uma vez normalizados, os dados foram filtrados com a intenção de eliminar os *spots* com sinal baixo de *background* ou com saturação da hibridação. Consideramos saturação da hibridação os *spots* com emissão de sinal maior ou igual a 63.000 (sendo $65.550 = 16 \text{ BIT}$). Nesse caso, o excesso de material hibridado gera um sinal muito intenso que supera a capacidade do detector de medir a intensidade proporcional de fluorescência, invalidando a quantificação.

Foi realizado o controle de qualidade dos dados por Análise dos Componentes Principais (PCA). Essa análise permite comparar o perfil de expressão das amostras; assim amostras com a mesma condição experimental devem estar agrupadas. Desvios dessa presunção podem ser devido a amostras de má qualidade

no conjunto de amostras ou uma variação biológica verdadeira dentro das populações de estudo.

Em seguida, os dados foram filtrados para ficar só com sondas com evidência de expressão nas amostras analisadas. As sondas foram divididas em codificadores e não codificadores e foram realizadas análises de expressão procurando genes diferencialmente expressão nos subgrupos receptor hormonal e linfonodos acometidos. Foram considerados transcritos diferencialmente expressos aqueles que apresentaram *Fold Change* (FC) ≥ 1.8 e p-valor ≤ 0.01 .

Posteriormente, na análise para avaliar o perfil de expressão gênica em subgrupos com e sem acometimento linfonodal, o perfil de expressão dos genes avaliados foi verificado por meio do programa *Short Time-series Expression Miner* (STEM). Esse é um programa que permite agrupar, comparar e visualizar dados de expressão gênica serial, de experimentos de *microarray*.

No *software* STEM, todos os perfis de expressão devem começar em zero e ter, no mínimo, um intervalo de tempo. O resultado da análise permite evidenciar o comportamento dos perfis: estável, aumentar ou diminuir na análise serial, utilizando as diferenças de expressão dos genes entre os intervalos de tempo (ERNST e BAR-JOSEPH 2006). Em nossa análise, o valor definido como parâmetro para identificar perfil de expressão dos genes a serem avaliados pelo STEM foi fold-change $\geq |1,8|$. Em seguida, os valores de expressão gênica (\log_2) dos ensaios seriais para cada gene foram normalizados, de modo que Ln0 corresponde ao tempo 0 (Ln 0 = 0), e Ln 1-3 e Ln ≥ 4 são normalizados em concordância. Dependendo do número de intervalos inseridos ao programa e o número de perfis possíveis gerados, cada gene é atribuído a um perfil de expressão baseados no seu comportamento nos intervalos e no

coeficiente de correlação em relação ao perfil padrão do programa. O número de genes atribuído para cada perfil é então calculado. Além disso, o número esperado de genes por acaso a ser atribuído a um perfil é calculado permutando aleatoriamente a valores nos intervalos originais, renormaliza os valores de expressão e, em seguida, atribui genes nos perfis correspondentes (ERNST e BAR-JOSEPH 2006). O valor de significância estatística é calculado em relação ao número de genes atribuído a cada perfil contra o número esperado por acaso, indicando se o perfil apresenta maior ou menor quantidade de genes do esperado por acaso.

4.5 VALIDAÇÃO DOS TRANSCRITOS IDENTIFICADOS COMO DIFERENCIALMENTE EXPRESSOS

Para a validação dos dados de cDNA *microarray* provenientes da comparação pacientes com e sem linfonodos acometidos, realizamos um ensaio de RT-PCR quantitativa com sistema de detecção *Sybr Green*, no ensaio *PCR array*® de *Qiagen*, que consiste numa placa com *primers* desenhados e imobilizados pelo fabricante.

Para o ensaio *PCRarray* foi customizada uma placa de 384 poços (8 amostras/placa), foram escolhidos 40 genes para validação, 5 genes normalizadores internos e 3 genes controles sugeridos pelo ensaio (ver “Figura 10”). Para a escolha dos genes alvos avaliados por RT-PCR, foram privilegiados os genes nos perfis com diferenças estatisticamente significativas no *Análises STEM*, com tendência ao aumento ou diminuição da expressão nos primeiros sinais de acometimento linfonodal, com disponibilidade de partidores comerciais na plataforma *PCRarray* (Qiagen) e a função biológica dos genes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
2	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
3	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC
4	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
5	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
6	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC
7	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
8	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
9	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC
10	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
11	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
12	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC
13	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
14	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
15	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC
16	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
17	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
18	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC
19	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
20	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
21	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC
22	RGL3	FGFR3	PPT2	ZNF580	VILL	SNCB	SPTAN1	PHYHD1	ACSF2	H1F0	HLA-DQA1	CCDC88C	TBL1X	CDK5R2	ADRA1B	ADRA2A
23	IGF1R	B3GNT5	TUBB3	POU2AF1	IKZF2	RASSF4	ADAM8	DMD	CDC42EP4	GALNT3	JAKMIP1	MNX1	MAP6D1	EGR2	DGKH	CX3CL1
24	ZDHHHC20	THBS2	PDPN	FOXQ1	COL8A1	HOXB6	ZNF681	EDN2	18SrRNA	B2M	BCR	GUSB	RPLP0	HGDC	RTC	PPC

Figura 10 - Esquematização da localização dos genes na placa PCRarray Customizada. São representados os genes alvos do estudo em branco, os genes endógenos “normalizadores” em azul e os controles sugeridos pelo ensaio em rosa.

Os ensaios foram realizados utilizando 5ng de RNA amplificado (RNAa) em dois *rounds*, que foram convertidos em cDNA, conforme descrito anteriormente (ABUÁZAR 2012) e utilizando o protocolo do fabricante, no equipamento *Applied Biosystems 7900HT Real-Time PCR System (Life Technologies®)*.

4.6 AMOSTRAS PARA VALIDAÇÃO POR RT-PCR QUANTITATIVO

Com o objetivo de realizar a validação do experimento *microarray* utilizando as mesmas amostras, foi reservado durante o processo de amplificação 1/5 do cDNA primeiro *round* para ser novamente amplificado, mas sem a marcação necessária no *microarray*. Na segunda amplificação, o cRNA foi quantificado usando o equipamento *Nanodrop 1000*, posteriormente foi convertido a cDNA, estimando uma conversão 1:1, considerando a quantidade de cDNA necessário para os ensaios de PCR (5ng) e a quantidade de genes no estudo (55 genes). Foram consideradas aptas para estudo todas as amostras com cRNA > 600ng.

Finalmente, para selecionar as amostras participantes da validação, foi comprovada a integridade do material amplificado realizando a detecção do gene endógeno (*GAPDH*) por RT-qPCR (SYBR) e consideramos aptas para análises as amostras com ciclo de amplificação inferior a 30.

4.7 ANÁLISES DOS ENSAIOS RT-PCR QUANTITATIVA

Os resultados dos ensaios RT-PCR quantitativa foram analisados com o valor de expressão relativa, baseado na comparação entre o nível de expressão do gene alvo *versus* um gene controle, também chamado de gene referência, interno ou normalizador do inglês *housekeeping*.

$$\text{Expressão relativa: } 2^{-\text{ct}(\text{genes alvos})} / 2^{-\text{ct}(\text{housekeeping})}$$

O propósito de usar genes *housekeeping* é corrigir a variação não específica, como a diferença em quantidade e qualidade do RNA utilizado que pode afetar a eficiência da transcrição reversa e PCR (EISENBERG e LEVANON 2013), mas nem todos os genes podem servir como normalizadores. Para poder escolher os genes com maior estabilidade, utilizamos *software* geNorm, que, basicamente, realiza duas análises: a estabilidade dos genes e a conveniência de usar mais de um gene como normalizador.

Após calcular a expressão relativa de cada amostra por gene, foram calculadas as diferenças de expressão entre os grupos ou *Fold Change* (FC). Foi comparada a média do valor expressão relativa normalizada dos grupos, e consideramos como diferencialmente expressos aqueles genes com $FC \geq |1.5|$ e p-valor ≤ 0.05 .

Nas análises do ensaio *PCR-array* foi adicionada uma etapa prévia ao cálculo da expressão relativa, que consistiu em observar a expressão do controle interno HGDC (do inglês *Human Genomic DNA Control*). Foi considerada aceitável a aparição depois do ciclo 35 ou antes do ciclo 35 com T_m diferente a 75°C para considerar desprezável o nível de contaminação com DNA genômico (QIAGEN 2014).

4.8 ANÁLISE BIOLÓGICA

Os conjuntos de genes identificados com diferenças de expressão entre os grupos foram submetidos a uma análise de interpretação biológica, com a intenção de conhecer as funções biológicas e vias de sinalização nas quais participam os genes.

Utilizamos o *software Ingenuity Pathway Analysis® (IPA)*, que permite analisar dados de expressão provenientes de ensaios de *microarrays*, SNP, transcriptômica, proteômica e metabolômica, que gerem listas de genes ou proteínas, permitindo avaliar o enriquecimento de funções biológicas ou vias de sinalização e redes causais que integram relação de causa-efeito relatados anteriormente na literatura.

Neste estudo foi realizado um *core analysis*, utilizando parâmetros padrão e os dados de *gen symbol*, *Fold change* e p-valor do conjunto de genes identificados com diferenças de expressão.

5 RESULTADOS

5.1 DADOS CLÍNICOS E PATOLÓGICOS DOS PACIENTES ANALISADOS NO ESTUDO

As 103 pacientes que fizeram parte do estudo são uma representação ampla do carcinoma ductal invasivo de mama, considerando: idade, grau histológico (SBR), grau nuclear, índice mitótico, TNM, linfonodos acometidos/retirados e expressão proteica por dados da imunistoquímica dos receptores hormonais (ER e PR), CerbB2, Ck5, Ki-67, EGFR e p53 (informação resumida na “Quadro 3” e completa no “Anexo 9”). É importante ressaltar que nenhuma paciente recebeu tratamento neoadjuvante (radioterapia, quimioterapia ou hormonioterapia) prévio a ressecção cirúrgica da amostra tumoral analisada neste estudo.

Quadro 3 - Dados clínicos e anatomopatológicos.

Nomenclatura	Linfonodos Acometidos	Classificação	Grau Histológico (SBR)	TNM	Idade do Diagnóstico
MIC 010T	Ln ≥4	Luminal B	1	T4N2M0	55
MIC 021T	Ln 0	Triplo negativo	3	T1N0M0	43
MIC 038T	Ln 1-3	Luminal A	1	T2N1M0	67
MIC 040T	Ln 1-3	Luminal A	3	T4bN1M0	78
MIC 044T	Ln 0	Luminal A	2	T3N0M0	55
MIC 046T	Ln ≥4	HER2 +	3	T2N3M0	57
MIC 049T	Ln 0	Luminal B	2	T2N2M0	51
MIC 055T	Ln 0	Luminal B'	2	T1N1M0	58
MIC 057T	Ln 1-3	Luminal A	2	T1N1M0	51
MIC 063T	Ln 1-3	Luminal A	2	T4bN1M0	79
MIC 064T	Ln 1-3	Luminal B'	2	T2N0M0	46
MIC 069T	Ln ≥4	Luminal A	2	T2N3M0	58
MIC 073T	Ln ≥4	Luminal A	2	T2N2M0	72
MIC 081T	Ln ≥4	Luminal B	1	T1N2M0	56
MIC 085T	Ln ≥4	Luminal A	2	T1N2M0	64
MIC 090T	Ln 0	Triplo negativo	3	T1N0M0	48
MIC 120T	Ln ≥4	Luminal A	3	T2N3M0	41
MIC 137T	Ln 1-3	Triplo negativo	3	T1N1M0	44
MIC 141T	Ln 1-3	Triplo negativo	3	T2N0M0	60
MIC 143T	Ln 0	HER2 +	3	T1N0M0	75
MIC 144T	Ln ≥4	Triplo negativo	3	T2N3M0	62
MIC 145T	Ln 0	HER2 +	3	T2N0M0	78
MIC 146T	Ln ≥4	Triplo negativo			18
MIC 147T	Ln ≥4	HER2 +	2	T2N2M0	40
MIC 148T	Ln 0	HER2 +	3	T1N0M0	48
MIC 149T	Ln 1-3	HER2 +	2	T4bN1M0	32
MIC 173T	Ln 1-3	Triplo negativo	2	T1N0M0	38
MIC 175T	Ln 1-3	Triplo negativo	3	T2N1M1	41
MIC 176T	Ln 0	Luminal B	2	T1N0M0	50
MIC 177T	Ln 1-3	Triplo negativo	3	T2N1M0	43
MIC 179T	Ln 1-3	Triplo negativo	3	T2N1M0	44
MIC 180T	Ln 1-3	Triplo negativo	3	T2N1M0	64
MIC 183T	Ln 1-3	Triplo negativo	3	T1N1M0	40
MIC 184T	Ln 0	Triplo negativo	3	T2N0M0	41
MIC 185T	Ln 1-3	Triplo negativo	3	T2N1M0	52
MIC 186T	Ln ≥4	Luminal A	2	T4bN2M0	24
MIC 187T	Ln 0	Luminal A	3	T2N0M0	60
MIC 188T	Ln 1-3	HER2 +	3	T2N1micM0	48
MIC 189T	Ln ≥4	Triplo negativo	3	T2N2M0	53
MIC 190T	Ln ≥4	Luminal B'	3	T3N2M0	30
MIC 191T	Ln ≥4	Triplo negativo	2	T2N3M0	48

Nomenclatura	Linfonodos Acometidos	Classificação	Grau Histológico (SBR)	TNM	Idade do Diagnóstico
MIC 192T	Ln 0	Luminal B'	2	T1N1M0	60
MIC 194T	Ln 1-3	Triplo negativo	2	T2N1M0	69
MIC 195T	Ln 0	Luminal B	1	T1N0M0	56
MIC 197T	Ln 0	Luminal A	1	T2N0M0	83
MIC 199T	Ln 0	Triplo negativo	2	T2N0M0	41
MIC 201T	Ln 1-3	Triplo negativo	3	T3N1M0	40
MIC 202T	Ln 0	Luminal B	3	T2N0M0	38
MIC 204T	Ln 0	Luminal A	3	T1N0M0	72
MIC 207T	Ln 0	Luminal A	1	T1bN1M0	66
MIC 208T	Ln 1-3	HER2 +	3	T4N2M0	85
MIC 209T	Ln 0	Luminal B'	3	T2N0M0	42
MIC 211T	Ln 0	Luminal B	3	T2N0M0	42
MIC 212T	Ln \geq 4	Luminal B'	3	T2N3M0	56
MIC 213T	Ln 0	Luminal B'	1	T2N1M0	47
MIC 216T	Ln \geq 4	Luminal B'	2	T1N2M0	58
MIC 217T	Ln \geq 4	Luminal B'	2	T2N2M0	39
MIC 218T	Ln 0	Triplo negativo	3	T2N0M0	71
MIC 220T	Ln \geq 4	Luminal A	2	T1N3M0	80
MIC 221T	Ln \geq 4	Luminal B'	3	T4bN0M0	41
MIC 223T	Ln 1-3	HER2 +	1	T2N1M0	58
MIC 224T	Ln 0	Triplo negativo	3	T2N0M0	47
MIC 225T	Ln 0	Luminal B	2	T1N0M0	61
MIC 226T	Ln 0	Luminal B'	3	T2N0M0	29
MIC 228T	Ln \geq 4	Triplo negativo	3	T2N2M0	64
MIC 229T	Ln 1-3	HER2 +	3	T1N1M0	52
MIC 230T	Ln 0	Triplo negativo	3	T2N0M0	48
MIC 231T	Ln 0	Luminal A	3	T2N1M0	58
MIC 233T	Ln \geq 4	Luminal A	2	T1N2M0	49
MIC 234T	Ln 0	Triplo negativo	3	T2N0M0	56
MIC 235T	Ln \geq 4	Luminal B	3	T2N2M0	39
MIC 236T	Ln 1-3	HER2 +		T2N1M0	69
MIC 237T	Ln \geq 4	Luminal B'	2	T2N3M0	71
MIC 240T	Ln 0	Luminal A	2	T1N0M0	71
MIC 261T	Ln 0	HER2 +	3	T1N0M0	72
MIC 263T	Ln 1-3	Luminal B	2	T2N1M0	45
MIC 264T	Ln 0	Luminal B	2	T1N0M0	55
MIC 265T	Ln 0	Luminal B	2	T1N0M0	45
MIC 279T	Ln 1-3	Luminal B'	3	T2N1micM0	58
MIC 283T	Ln 0	Luminal B'	3	T1N0M0	71
MIC 284T	Ln 0	Luminal A	2	T1N0M0	60
MIC 286T	Ln 0	Luminal B'	3	T2N0M0	63
MIC 287T	Ln 0	Luminal B'	2	T1N0M0	54
MIC 290T	Ln 1-3	Luminal B	3	T1N1M0	40

Nomenclatura	Linfonodos Acometidos	Classificação	Grau Histológico (SBR)	TNM	Idade do Diagnóstico
MIC 291T	Ln \geq 4	HER2 +	3	T2N2M0	66
MIC 292T	Ln 0	HER2 +	3	T3N1M0	40
MIC 294T	Ln 1-3	Luminal B'	2	T2N1MX	75
MIC 295T	Ln 1-3	Luminal B'	3	T1N1micM0	74
MIC 296T	Ln 0	Luminal B'	3	T2N0M0	69
MIC 298T	Ln 1-3	Luminal B'	2	T1N2M0	47
MIC 300T	Ln 0	Luminal A	2	T1N0M0	47
MIC 301T	Ln 1-3	Luminal B'	1	T2N1M0	47
MIC 302T	Ln 0	Luminal A	2	T1N0M0	54
MIC 303T	Ln 0	Luminal B'	2	T2N0M0	88
MIC 304T	Ln 0	Luminal A	2	T2N0M0	64
MIC 305T	Ln 0	Luminal B'	2	T1N0M0	60
MIC 307T	Ln 1-3	Luminal B'	2	T1N1M0	42
MIC 308T	Ln 0	Luminal A	2	T2N0M0	55
MIC 309T	Ln 0	Luminal A	1	T2N0M0	39
MIC 312T	Ln 0	Luminal A	3	T1N0M0	39
MIC 313T	Ln 0	Luminal B'	2	T1N1micM0	58
MIC 314T	Ln 0	Luminal B'	2	T1N0M0	64
MIC 315T	Ln \geq 4	Luminal B	3	T3N3M0	44

A idade média apresentada pelo grupo de pacientes no momento do diagnóstico foi de 54,3 anos, com idade mínima de 18 e máxima de 88 anos. Ao observar as idades médias por categorias, constatamos que os subgrupos HR-HER-2- e HR+ HER-2 apresentam as menores idades, enquanto nos subgrupos HR+HER-2- e HR-HER-2 apresentam idade mais avançada (Tabela 2).

Tabela 2 - Idade por subtipos.

	Média de idade	Nº de pacientes	Desvio padrão
Subtipos	HR- HER-2-	24	12.5
	HR- HER-2+	14	15.7
	HR+ HER-2-	51	14.1
	HR+ HER-2+	14	7.4

As pacientes foram classificadas segundo os critérios definidos nos materiais e métodos (ver “Quadro 1”), nos subgrupos relacionados aos receptores hormonais denominados: Luminal A, Luminal B, Luminal B’, HER 2 e Triplo negativo. Assim, também foram considerados os dados de comprometimento linfonodal, obtendo uma população homogênea com número de pacientes representativos de cada subgrupo (Tabela 3). Três amostras luminais não puderam ser classificadas como luminal A ou B’, uma vez que não havia material disponível do bloco parafinado necessário para fazer IHQ do marcador Ki67.

Tabela 3 - Subtipos de câncer de mama invasivo segundo *status* linfonodal.

Classificação	Subtipos	Linfonodos Não	Linfonodos Acometidos		Total
		Acometidos	1-3 LN	≥4 LN	
HR-/HER-	Triplo Neg.	8	11	5	24
HR-/HER+	HER-2	5	6	3	14
HR+/HER+	Luminal B	8	2	4	14
	Luminal A	14	4	5	23
HR+/HER-	Luminal B’	12	7	6	25
	Indeterminado	1	0	2	3
Total		48	30	25	103

5.2 ANÁLISE DE EXPRESSÃO GÊNICA

Neste trabalho foram selecionadas unicamente as células epiteliais dos CDI, usando a metodologia LCM, com uma média de 1750 feixes (tiros) de *laser* por amostra, num tempo máximo de 10 minutos. Após esse procedimento, o RNA total foi extraído e avaliado quanto a sua pureza e integridade. Os valores RIN (*RNA integrity Number*) das amostras estão disponíveis na “Tabela 3”.

Depois da avaliação da integridade do RNA, todas as amostras foram submetidas a um ciclo de amplificação do RNA, através de uma adaptação do método de amplificação de RNA mensageiro (mRNA) usado anteriormente pelo grupo (CASTRO et al. 2008; FERREIRA et al. 2010). Cada amostra foi amplificada e marcada com Cy3-CTP e foi adicionado à reação de amplificação um controle positivo de RNA (“*spike A*”). Ao mesmo tempo, foi amplificado e marcado com Cy5-CTP um RNA referência (*pool* de RNA de várias culturas celulares) e adicionado o controle positivo de RNA (“*spike B*”). A qualidade do RNA total inicial e do RNA final obtido, assim como a incorporação das cianinas em todas, estão descritas na “Tabela 4”.

Tabela 4 - Quantidade de cRNA amplificado e incorporação de fluoróforos.

Nomenclatura	RIN	Massa RNA amplificado	Atividade específica ($\mu\text{mol dye}/\mu\text{gRNA}$)
MIC 010T	9.7	2.98	14.49
MIC 021T	8.4	6.58	4.07
MIC 038T	6.5	8	13.79
MIC 040T	5.7	10.83	5.67
MIC 044T	7.1	6.15	7.08
MIC 046T	5.2	20.95	26.92
MIC 049T	7.7	8.13	12.85
MIC 055T	N/D	3.48	4.31
MIC 057T	N/D	1.73	4.35
MIC 063T	5	8.24	4.14
MIC 064T	8.3	11.46	16.62
MIC 069T	N/A	9.87	5.05
MIC 073T	N/A	6.41	9.44
MIC 081T	N/A	2.78	5.75
MIC 085T	6.1	9.98	14.08
MIC 090T	6	13.23	4.81
MIC 120T	8.4	14.78	19.52
MIC 137T	N/A	6.63	4.07
MIC 141T	9.5	4.09	5.5
MIC 143T	7.2	19.03	11.24
MIC 144T	8.5	13.52	11.54
MIC 145T	5.2	2.9	8.28
MIC 146T	7	5.77	5.11
MIC 147T	N/A	11.6	5
MIC 148T	N/A	9.95	4.12
MIC 149T	6.1	5.51	6.8
MIC 173T	5,1	11,66	7,41
MIC 175T	9,2	9,34	3,94
MIC 176T	N/A	13,62	5,88
MIC 179T	5,4	5,48	12,26
MIC 180T	7,6	12,27	8,61
MIC 183T	8,1	4,93	7,55
MIC 184T	7,7	5,34	8,31
MIC 185T	6,7	3,75	3,68
MIC 186T	4,8	7,16	7,37

Nomenclatura	RIN	Massa RNA amplificado	Atividade específica (pmol dye/μgRNA)
MIC 187T	7,4	15,84	5,71
MIC 188T	6,7	13,66	5,15
MIC 189T	6,2	4,52	1,86
MIC 190T	N/A	4,53	2,38
MIC 191T	6,3	6,9	4,96
MIC 192T	5,1	3,11	6,18
MIC 194T	7,3	3,17	10,97
MIC 195T	6,2	5,86	9,01
MIC 197T	6	2,75	5,23
MIC 199T	8,1	11,59	5,18
MIC 201T	7,8	6,9	7,73
MIC 202T	5,4	24,09	6,53
MIC 204T	6,9	3,64	4,68
MIC 207T	4,4	5,1	10,10
MIC 208T	5,3	11,62	8,68
MIC 209T	7,4	4,3	8,65
MIC 211T	6,1	6,59	7,37
MIC 212T	5,7	21,37	6,29
MIC 213T	7,3	20,4	7,29
MIC 216T-A	6,7	3,75	7,04
MIC 217T	7,3	10,34	4,35
MIC 218T-A	6,7	4,48	10,44
MIC 220T	5,6	6,19	8,54
MIC 221T-A	6,1	6,34	9,7
MIC 223T	8	17,55	13,28
MIC 224T-A	6,0	5,49	9,84
MIC 225T	6,8	7,26	5,12
MIC 226T	6,9	4,81	13,71
MIC 228T	5	6,23	8,99
MIC 229T	7,4	12,97	14,25
MIC 230T	8,2	4,82	
MIC 231T	5,5	5,81	12,99
MIC 233T	7	12,6	3,72
MIC 234T	5,6	5,1	11,74
MIC 235T	5,8	8,02	6,66
MIC 236T	7,1	7,78	12,57
MIC 237T	6,5	6,65	9,03

Nomenclatura	RIN	Massa RNA amplificado	Atividade específica (pmol dye/ μ gRNA)
MIC 240T	5,4	9,64	11,05
MIC 261	6,9	3,65	7,72
MIC 263	7,8	15,35	23,56
MIC 264	5,2	4,57	11,16
MIC 265	5,4	2,96	9,73
MIC 279T	4,8	6,95	11,9
MIC 283T	5,6	3,17	4,355
MIC 284T	5,8	5,03	9,89
MIC 286T	4,9	8,78	12,644
MIC 287T	7,3	7,75	9,37
MIC 290T	4,8	7	9,94
MIC 291T	6,2	12,27	10,86
MIC 292T	6,7	12,58	7,15
MIC 294T	6,3	13,02	10,42
MIC 295T	6,5	7,01	8,47
MIC 296T	6,5	11,03	10,22
MIC 298T	5,8	6,33	7,21
MIC 300T	7,3	10,31	8,55
MIC 301T	7,5	10,71	8,07
MIC 302T	6,9	13,43	11,7
MIC 303T	5,2	6,33	5,59
MIC 304T	7,5	15,83	12,89
MIC 305T	7,2	20,48	14,12
MIC 307T	6,8	7,39	5,92
MIC 308T	5,9	8,68	6,91
MIC 309T	8,1	11,37	11,56
MIC 312T	6,4	4,86	5,06
MIC 313T	8,1	18,5	13,98
MIC 314T	5,8	5,86	6,28
MIC 315T	6,3	9,72	8,4

Os ensaios de *microarray* foram analisados mediante o *software GeneSpring GX 12.1 (Agilent Technologies)*, seguindo as recomendações do programa. Como primeira abordagem foi realizada uma análise dos componentes principais (PCA), que evidenciou que a amostra (MIC143) apresentava uma distribuição diferente

(separada) das outras amostras e foi observado, também, que tinha um padrão de expressão aberrante no controle negativo, portanto foi eliminada das análises.

Além disso, o gráfico de distribuição do PCA revelou a presença de dois grupos, que foram relacionados ao lote do kit de amplificação utilizado (Figura 11A). Para eliminar as possíveis diferenças de expressão entre os grupos amplificados com kit de lotes diferentes, foi realizado, de forma independente em cada grupo, um filtro por expressão, escolhendo as sondas com base nos valores de intensidade de sinal entre 20-100% em todas as amostras. Com esses parâmetros foi obtido um grupo com 69.766 sondas em comum (Figura 11B). Posteriormente, foi realizado o filtro por *flag*, selecionando as sondas que tivessem sido detectadas em, pelo menos, 65% das amostras, sendo que 26.130 sondas apresentaram tal característica. A seguir, foi analisado se ainda existiam sondas com diferenças de expressão entre os grupos amplificados com kit de lotes diferentes. Para isso, foi realizado um teste T (não pareado, com correção Bomferroni) e foram identificadas 2.915 sondas diferentemente expressas entre os grupos, as quais foram retiradas da análise com a finalidade de diminuir vieses nas análises posteriores. Finalmente, foi obtido um grupo único composto por 23.215 sondas (Figura 11C), as quais correspondem a 11.312 transcritos codificadores e 11.903 transcritos não codificadores (2.610 anti-sensos).

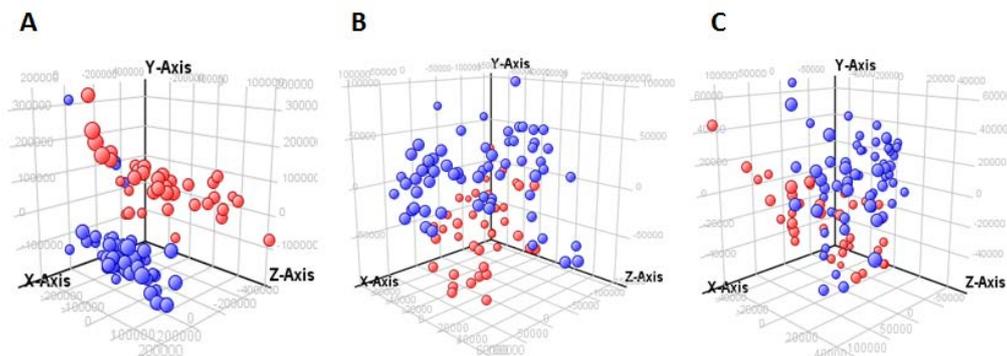


Figura 11 - Distribuição das amostras por Principal Componente de Análises, durante os diferentes filtros. A) representa a distribuição das amostras pela expressão de todas as sondas, B) representa a distribuição das amostras pela expressão das sondas após o filtro de intensidade, C) representa a distribuição das amostras pela expressão das sondas após a retirada das sondas com diferenças de expressão entre os grupos. As cores azul e vermelho representam as amostras amplificadas com os diferentes lotes de kit.

5.2.1 Perfil de Expressão Gênica dos Transcritos Codificadores

Para a identificação dos subtipos moleculares de câncer de mama comumente conhecidos como Luminal A, Luminal B, HER-2 enriquecido e basal *like*, nós utilizamos o conjunto de genes descrito por PARKER et al. (2009). Conhecido como PAM50, esse *set* de genes tem alta concordância na classificação com a grande quantidade de genes definidos previamente para subtipagem (PEROU et al. 2000; SORLIE et al. 2003; HU et al. 2006; PARKER et al. 2009), e recentemente tem sido comumente empregado com sucesso (ELLIS et al. 2011; CHIA et al. 2012).

Dos 50 genes que compõem o PAM 50, o *array* 244K contém sondas para 48 deles, com número de sondas variáveis por gene (total 70 sondas). No entanto, nem todas as sondas apresentaram sinal detectável em todas as amostras. Assim, o nosso *set* de genes foi reduzido a 40 genes (53 sondas), não podendo analisar os genes *BAG1*, *CDC20*, *CDC6*, *TYMS*, *MMP11*, *BLVRA*, *CDCA1*, *PTTG1*, *FGFR4* e *UBE2C*.

Em seguida, as amostras e os genes foram submetidos à clusterização hierárquica seguindo a recomendação de PARKER et al. (2009) (correlação de *Pearson*). O *cluster* gerado está formado por 3 grupos facilmente reconhecíveis (ver “Figura 12”), comparáveis com os subgrupos anteriormente relatados (PEROU et al. 2000).

O primeiro grupo está composto por 55 amostras, as quais, baseadas na caracterização imunohistoquímica, mostram-se 100% positivas para os receptores de estrógeno e progesterona, o que permite caracterizar as amostras como o subgrupo “Luminal”. Dentro do conjunto de 55 amostras estão presentes 46 amostras (85,2%) positivas para os receptores de estrógeno e progesterona e negativas para HER-2 e 8 amostras (14,8%) positivas para receptores de estrógeno, progesterona e HER-2. O segundo grupo foi associado ao subgrupo “*HER-2*” está formado por 22 amostras 13 (59%) delas foram positivas HER-2 e negativas para receptores de estrógeno e progesterona; outras 5 amostras (22,72%), foram positivas para HER-2, receptores de estrógeno e progesterona, assim também neste grupo foram incluídas 4 amostras sem positividade para HER-2, duas delas foram positivas para receptores de estrógeno e progesterona e as outras duas foram negativas para os três marcadores. O último grupo foi associado ao subgrupo “Triplo negativo” está formado por 19 amostras 18 (94,7%) delas são negativas para os três marcadores e uma amostra é positiva para *HER-2* e negativa para os receptores de estrógeno e progesterona.

Ao comparar os grupos originados com a classificação pela IHQ podemos evidenciar que grupo Basal ou triplo negativo foi o mais conservado, apresentando somente uma amostra que foi caracterizada pelos marcadores imunohistoquímicos como superexpressão de HER-2. Dentro do grupo HER-2 enriquecido foram

agrupadas as amostras previamente caracterizada como; Luminal B (HER-2) e 2 amostras caracterizadas como Triplo Negativo (amostras 228T e 179T).

Em relação à ausência de separação entre as amostras dos grupos Luminais A e B, acreditamos que possivelmente deve-se ao uso incompleto do painel PAM 50, faltando a informação do nível de expressão de 10 genes mencionados anteriormente, sendo que 3/10 genes são cruciais para poder diferenciar os subgrupos luminais (PRAT et al. 2012). Assim, de forma resumida, esta análise agrupou corretamente 87% das amostras luminais, 82% das amostras HER-2 e 81,8% das amostras triplo negativo, portanto confiáveis para proceder com a avaliação do perfil de expressão das pacientes com carcinoma ductal invasivo com e sem linfonodos acometidos e com a identificação dos pares de transcritos codificadores e não codificadores com nível de expressão invertida

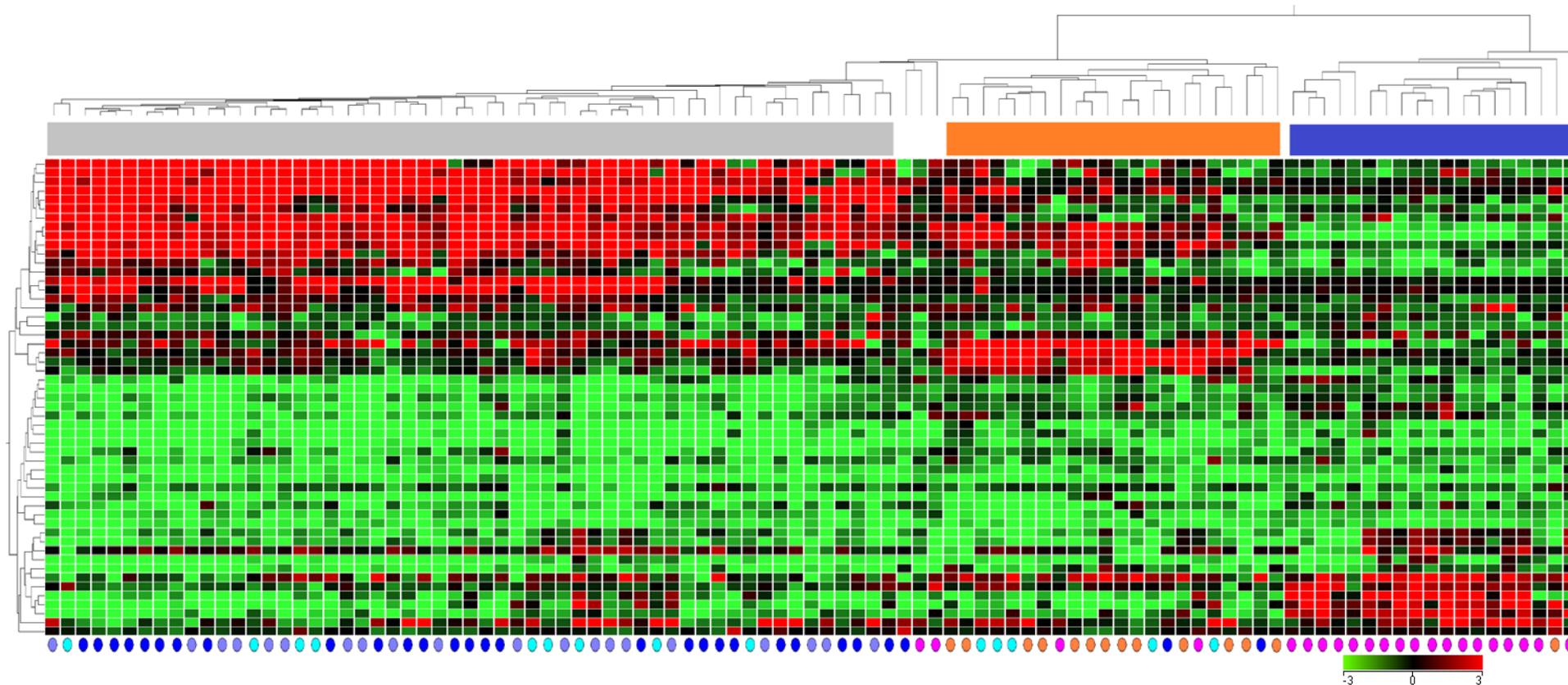


Figura 12 - Identificação dos subtipos “intrínsecos” moleculares no câncer de mama. Mediante Cluster hierárquico com correlação de Pearson. Cluster formado por 99 amostras de carcinoma ductal invasivo de mama e 40 genes dos que compõem o perfil de análise molecular PAM50. Identificam-se três agrupamentos que podem ser correlacionadas ao subgrupo Basal (azul), HER-2 enriquecido (laranja) e Luminal (Cinza). Na área inferior do cluster estão representadas as amostras previamente classificadas por IHQ denominadas: Luminal A (azul claro), Luminal B (azul escuro), Luminal B' (verde claro), HER-2 enriquecido (laranja) e Triplo negativo (verde escuro).

5.3 AVALIAÇÃO DO PERFIL DE EXPRESSÃO GÊNICA EM CÉLULAS EPITELIAIS DE CARCINOMA DUCTAL INVASIVO DE PACIENTES COM E SEM LINFONODOS ACOMETIDOS

Para identificar genes diferentemente expressos entre amostras com acometimento linfonodal ao diagnóstico, comparamos os grupos de amostras sem linfonodos acometidos (Ln0) com as amostras com 4 ou mais linfonodos acometidos (Ln \geq 4). A comparação foi realizada mediante um teste T (não pareado) e foram selecionados os genes com FC maior que 1.8 entre os grupos, resultando 79 genes, dos quais 36 estão mais expressos no grupo de amostras sem linfonodo acometido e 43 estão mais expressos no grupo de amostras com linfonodos acometidos. Na tentativa de avaliar se o padrão de expressão baseado nestes genes seria capaz de discriminar as amostras quanto ao *status* linfonodal utilizamos clusterização hierárquica baseada no nível de expressão destes 79 genes considerados diferentemente expressos (Figura 13). Não foi observado padrão de expressão associado com o *status* linfonodal das amostras do estudo. Pela observação das características de cada amostra, pode-se observar que o padrão de expressão baseado nestes 79 genes foi influenciado principalmente pela expressão de receptor hormonal. Assim, uma vez que a expressão dos receptores hormonais influencia grandemente o padrão de expressão, nosso próximo passo foi investigar o componente linfonodal dentro de cada grupo em separado – amostras ER e/ou PR positivas e amostras ER e PR negativas. Dessa forma, foi novamente realizada uma clusterização hierárquica, e mesmo nos grupos em separado não foi possível observar padrão de expressão associado com o *status* linfonodal (Anexo 10).

Em seguida, na tentativa de caracterizar melhor o perfil de expressão dos 79 genes diferencialmente expressos em relação ao status de acometimento linfonodal, foi avaliado o comportamento de expressão destes genes nos três grupos segundo a característica de acometimento linfonodal (sem linfonodo acometido (Ln0) 1-3 Linfonodos acometidos (Ln1-3) e ≥ 4 Linfonodos acometidos (Ln ≥ 4)). Para isso utilizamos o software *Short Time-series Expression Miner* (STEM) (ver materiais e métodos, pág. 33). Foram avaliados 10 perfis de expressão que apresentam tendência ao aumento ou diminuição de expressão de acordo com o maior número de linfonodos acometidos (FIGURA 14). Destes três deles (perfis 2 12 e 13) apresentaram significância indicando que o número de genes do perfil difere daquele esperado por chance, e conseqüentemente a modulação de expressão de acordo com o perfil pode ter um papel importante no processo de acometimento linfonodal.

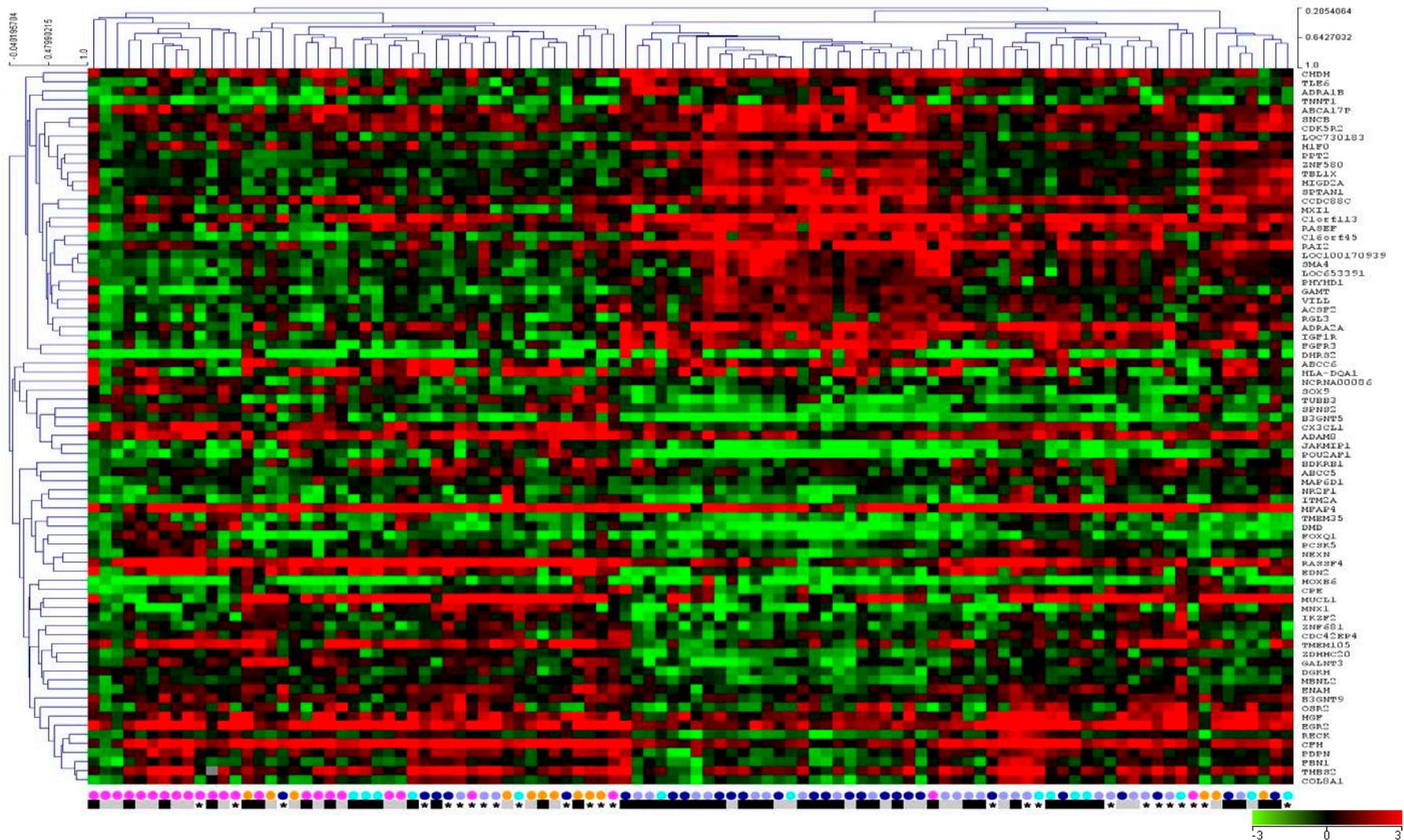


Figura 13 - Cluster Hierárquico das amostras Carcinoma Ductal Invasivo Com e sem linfonodos acometidos. Utilizando a métrica de correlação de Pearson. Cluster formado por 102 amostras de carcinoma ductal invasivo de mama e os 79 genes identificados como diferencialmente expressos. Não é possível observar padrão de expressão associado com o status linfonodal das amostras do estudo. Na área inferior do cluster está representado o status linfonodal das amostras, denominados: Não acometido ■ 1-3 Linfonodos acometidos ■ e ≥ 4 Linfonodos acometidos *. Também são representados os subgrupos Luminal A ●, Luminal B' ●, Luminal B ●, HER-2 enriquecido ● e Triplo negativo ●.

Genes que diminuem sua expressão em células epiteliais de pacientes com linfonodos acometidos					
Perfis STEM					
Genes	LOC100170939 ABCC6 RGL3 TNNT1 ABCA17P C16ORF45 FGFR3	ZNF580 PPT2 VILL SNCB SPTAN1 PHYHD1 LOC653391 LOC730183 ACSF2 H1FO HLA-DQA1	NCRNA00086 CDK5R2 CCDC88C TBL1X ADRA1B	GAMT ADRA2A SMA4 C1ORF113 IGF1R TLE6 RAI2 DHRS2	HIGD2A CHDH MXI1
Genes que aumentam sua expressão em células epiteliais de pacientes com linfonodos acometidos					
Perfis STEM					
Genes	ABCC5 B3GNT9 CFH ENAH FBN1 HGF MBNL2 MFAP4 MUCL1 NEXN NR2F1 OSR2 PCSK5 RECK TMEM35	B3GNT5 TUBB3 POU2AF1	IKZF2 RASSF4 ADAM8 DMD CDC42EP4 JAKMIP1 GALNT3 MNX1	MAP6D1 EGR2 DGKH CX3CL1 ZDHHC20 THBS2 PDPN FOXQ1 COL8A1 HOXB6 ZNF681 EDN2	ITM2A SPNS2 TMEM105

Figura 14 - Distribuição dos 79 genes diferentemente expressos entre os grupos com e sem linfonodos acometidos nos 10 perfis de expressão gênica gerados com software STEM. Na esquerda superior está apresentado o número identificador do perfil; na parte central o comportamento de expressão do perfil em relação aos 3 grupos (esquerdo LN0, no centro LN1-3 e na direita LN \geq 4). Os perfis coloridos representam aqueles com significância estatística. Perfis com a mesma cor, significa que tem correlação entre eles (com base em coeficientes de correlação \geq 0.7).

5.4 VALIDAÇÃO DA EXPRESSÃO GÊNICA EM AMOSTRAS PROVENIENTES DE PACIENTES EM TRÊS ESTADOS DE INVASÃO LINFONODAL DIFERENTES, POR RT-QPCR

Com a intenção de confirmar as diferenças de expressão encontradas nas análises de microarray, selecionamos 40 genes para avaliar sua expressão por RT-qPCR, neste ensaio foram usadas as mesmas amostras que no *microarray* (ver materiais e métodos, p. 27), selecionando amostras representativas de cada grupo de pacientes (sem linfonodo acometido, com 1-3 linfonodos acometidos e >3 linfonodos acometidos). Os genes selecionados para validação, foram aqueles que apresentaram mudança de expressão nos primeiros estádios da invasão do tumor, ou seja entre as amostras sem linfonodos acometidos e com 1-3 linfonodos acometidos, incluindo os genes presentes nos perfis, com significância estatística, da análises com o software STEM, escolhendo assim os seguintes genes: *ACSF2*, *ADAM8*, *ADRA1B*, *ADRA2A*, *B3GNT5*, *CCDC88C*, *CDC42EP4*, *CDK5R2*, *COL8A1*, *CX3CLI*, *DGKH*, *DMD*, *EDN2*, *EGR2*, *FGFR3*, *FOXQ1*, *GALNT3*, *HIF0*, *HLA-DQA1*, *HOXB6*, *IGF1R*, *IKZF2*, *JAKMIP1*, *MAP6D1*, *MNX1*, *PDPN*, *PHYHD1*, *POU2AF1*, *PPT2*, *RASSF4*, *RGL3*, *SNCB*, *SPTAN1*, *TBL1X*, *THBS2*, *TUBB3*, *VILL*, *ZDHHC20*, *ZNF580* e *ZNF681*.

5.4.1 Avaliação por RT-qPCR

Em todas as amostras foi avaliado um gene controle de DNA genômico (HGDC do inglês *Human Genomic DNA Control* do ensaio Qiagen PCR array) foi observado que algumas amostras apresentaram amplificação antes do ciclo 35 do gene nas amostras controle de DNA genômico, sugestivo de contaminação genômica. No entanto, todas as amostras com esta característica tiveram T_m (do inglês – *melting Temperature*) inespecífica para amplificação do fragmento genômico correspondente a este gene (análise recomendada pelo fabricante nestes casos), sendo, portanto, a amplificação observada não correspondente ao DNA genômico. Sendo assim, consideramos as amostras avaliáveis (Dados de cada amostra disponível no Anexo 11).

Cada um dos 40 genes escolhidos para validação foi analisado segundo a sua curva de dissociação específica, obtendo 30 genes com dados confiáveis para a avaliação de expressão diferencial. Assim, a expressão relativa dos 30 genes selecionados foi calculada como é descrito no materiais e métodos (Pág. 28), utilizando três genes como normalizadores (*18SrRNA*, *B2M*, *RPLP0*).

Foi avaliada a expressão diferencial entre os grupos mais extremos, o grupo sem linfonodo acometido (Ln_0) e o grupo com ≥ 4 Linfonodos acometidos ($Ln_{\geq 4}$). Foi mostrado que quatro genes tiveram diferenças de expressão (FC) concordantes com os dados observados no ensaio *microarray*, todos com aumento da expressão nas amostras com acometimento linfonodal (*POU2AF1*, *ZDHHC20*, *CX3CL1* e *FOXQ1*) (Figura 15).

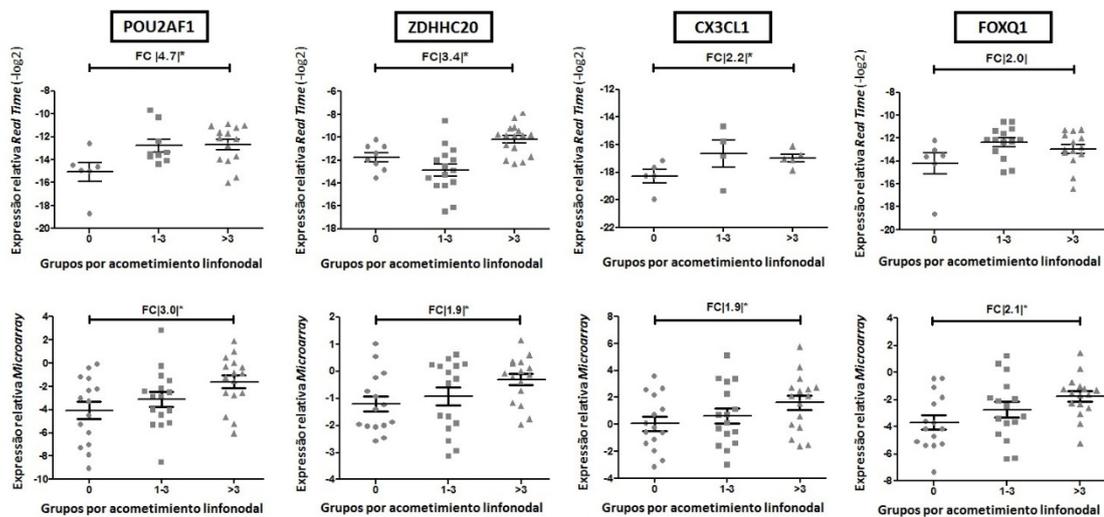


Figura 15 - Genes com diferenças de expressão em amostras de pacientes sem linfonodos acometidos e com mais de 3 linfonodos acometidos. Nos Gráficos é possível observar os dados de expressão relativa do ensaio real time (superior) e microarray (inferior). As diferenças de expressão estão representadas pelos valores de *Fold change*, símbolo * representa os dados estaticamente significantes.

Com a intenção de avaliar o comportamento da expressão gênica ao longo da progressão do acometimento linfonodal, foram avaliados todos os grupos de amostras representativas dos diferentes níveis de acometimento linfonodal. Foram identificadas diferenças de expressão concordantes com os dados de microarray, na primeira etapa do acometimento linfonodal, entre o grupo sem (Ln0) e o grupo com poucos linfonodos acometidos (Ln1-3) para os genes *POU2AF1*, *FOXQ1* e *SPTAN1* (Figura 16), e também para os genes *ZDHHC20*, *MNX1* e *CDK5R2* que apresentam diferenças de expressão em etapas mais tardias do acometimento linfonodal, entre Ln1-3 e Ln \geq 4 (Figura 17).

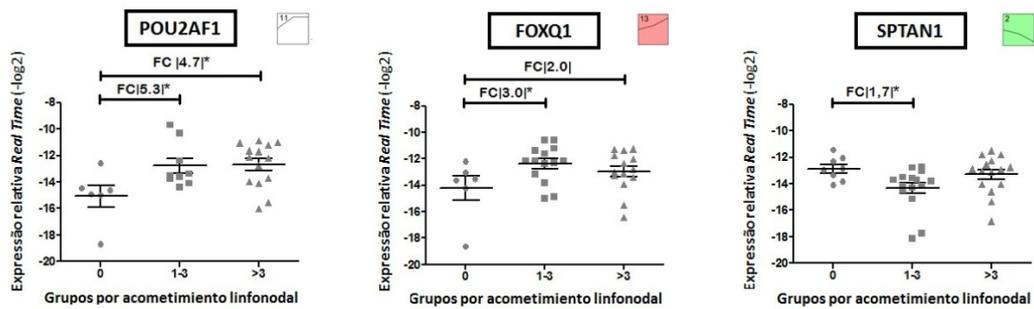


Figura 16 - Genes validados com diferenças de expressão na primeira etapa de invasão linfonodal. Nos Gráficos é possível observar os dados de expressão relativa do ensaio real time (superior) e microarray (inferior).

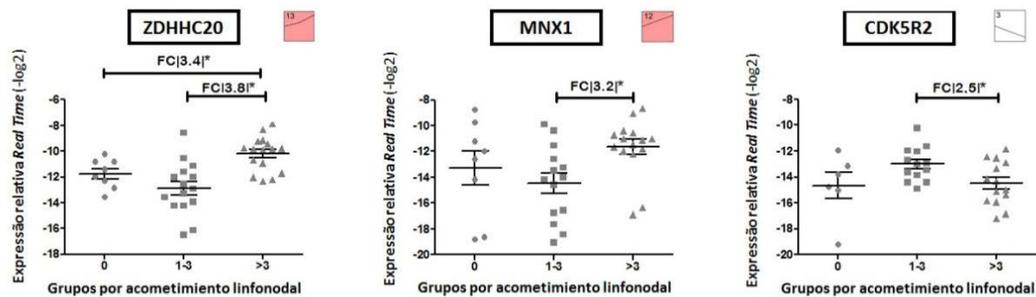


Figura 17 - Genes validados com diferenças de expressão em etapas tardias de invasão linfonodal. Nos Gráficos é possível observar os dados de expressão relativa do ensaio real time (superior) e microarray (inferior).

5.5 REDES DE INTERAÇÃO MOLECULAR E VIAS BIOLÓGICAS E INTEGRANDO ESSAS REDES

Com a intenção de conhecer melhor o conjunto de 7 genes validados das diferentes comparações e identificar possíveis relações entre eles, foi realizada uma análise de geração de redes e de vias, com uso do *software Ingenuity Pathway Analysis* (IPA). Esse software tem um algoritmo que inclui tanto a função biológica conhecida dos genes como dados recentemente publicados na literatura.

A análise gerou uma única rede composta por 20 genes que permite conectar os 7 genes validados juntamente com outros 13 genes necessários, formando uma rede enriquecida em processos de movimento e proliferação celular (Figura 18).

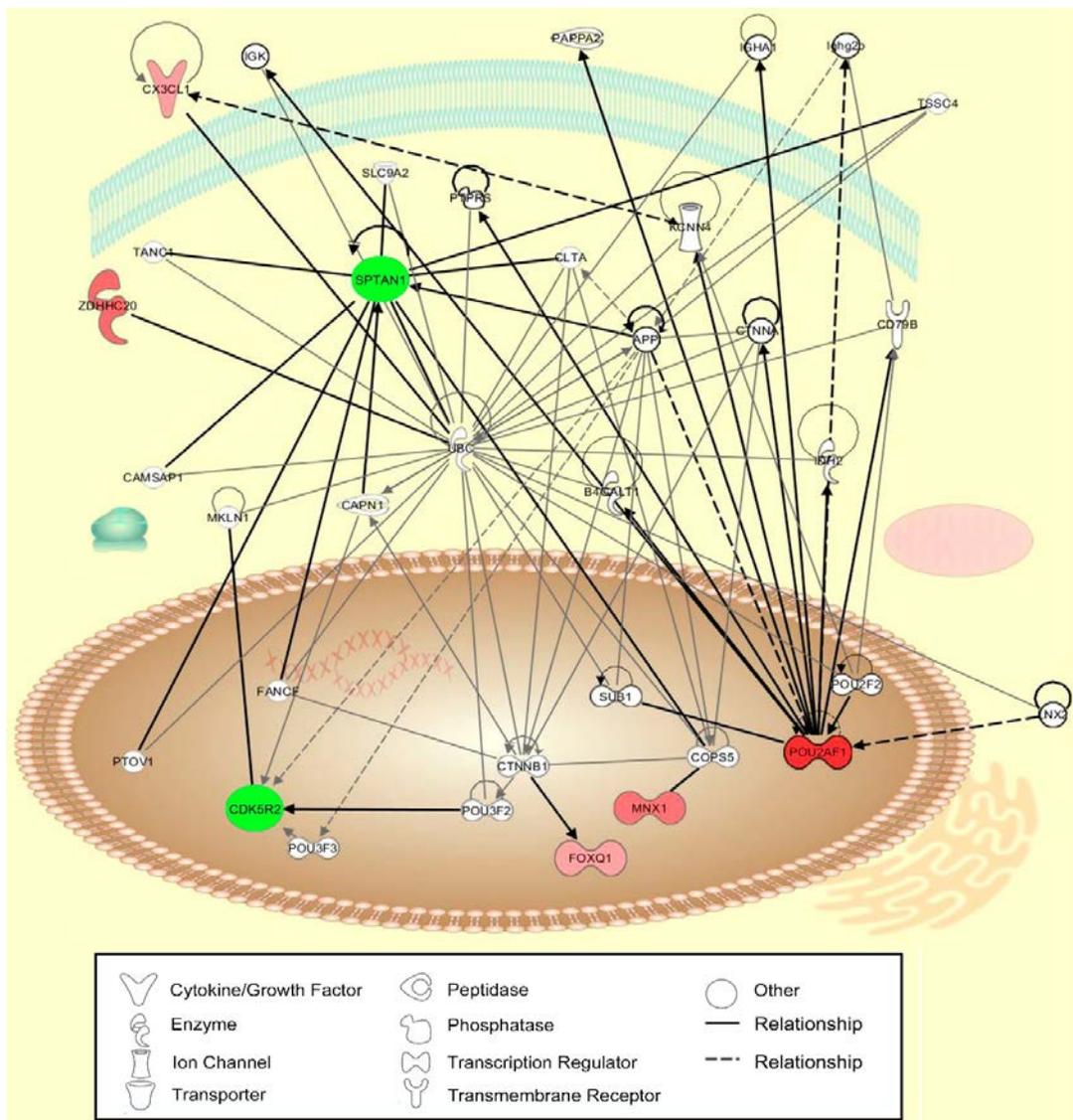


Figura 18 - Rede de interação molecular dos genes validados. A rede é representada por possíveis interações entre os genes. Os genes estão representados com sua função biológica e interconectados por linhas (que mostra a relação biológica entre genes). As cores presentes nos genes representam as diferenças de expressão validadas na comparação LN \geq 4/LN0, a cor vermelha indica genes com aumento da expressão e a cor verde com expressão diminuída em relação às pacientes com maior quantidade de linfonodos acometidos, as moléculas em branco são genes que o software inclui, por estar biologicamente ligados aos nossos genes validados.

O *software* mostrou que os genes que formaram a rede estão associados a processos de desenvolvimento e movimento celular. Uma análise mais acurada permitiu observar que os genes validados estão associados, principalmente, a processos de movimento e proliferação celular (ver “Quadro 4”).

Quadro 4 - Disfunção ou Processos Celulares Associado aos Genes Validados. São mostrados os genes validados em negrito e as setas indicam a expressão aumentada ou diminuída em relação ao acometimento linfonodal.

Disfunção ou Processo celular	Genes
Proliferação celular	↓ <i>CDK5R2</i> , ↑ <i>CX3CLI</i> , ↑ <i>FOXQ1</i> , ↑ <i>MNX1</i> , ↑ <i>POU2AF1</i> , ↓ <i>SPTANI</i> .
Quantidade de células	↑ <i>CX3CLI</i> , ↑ <i>MNX1</i> , ↑ <i>POU2AF1</i> .
Movimento celular	<i>APP</i> , <i>B4GALT1</i> , <i>CAPN1</i> , ↓ <i>CDK5R2</i> , <i>CTNNA1</i> , <i>CTNNB1</i> , ↑ <i>CX3CLI</i> , ↑ <i>FOXQ1</i> , <i>Ighg2b</i> , <i>IGK</i> , <i>KCNN4</i> , ↑ <i>MNX1</i> , ↑ <i>POU2AF1</i> , <i>POU3F2</i> , <i>POU3F3</i> .
Migração celular	↓ <i>CDK5R2</i> , ↑ <i>CX3CLI</i> , ↑ <i>FOXQ1</i> , ↑ <i>MNX1</i> , ↑ <i>POU2AF1</i> .
Movimento de linfócitos	↑ <i>CX3CLI</i> , ↑ <i>POU2AF1</i> .
Morfologia celular anormal	↓ <i>CDK5R2</i> , ↑ <i>MNX1</i> , ↑ <i>POU2AF1</i> .

Logo nós focalizamos nas interconexões que apresentavam os genes validados, mediante a contagem de triângulos de ligação entre os genes (Ingenuity Pathway Analysis 2005; BASE et al. 2014). Obtendo que 4 dos 7 genes validados apresentaram interconexão com pelo menos outras 2 moléculas, com formação de triângulo entre eles. Os genes que mostraram maior interconexão foram *POU2AF1*, *SPTANI*, *CDK5R2* e *CX3CLI*, no Quadro 5 mostra-se os genes com os quais estão interconectados (no Anexo 12 estão disponíveis as imagens que esclarecem as interações).

Quadro 5 - Nodos de interconexões incluídos genes validados.

Genes Validados	Moléculas interconectadas	Nº de triângulos por interconexões
<i>CX3CL1</i>	UBC-KCNN4	1
<i>CDK5R2</i>	POU3F3-APP CAPN1-APP	2
<i>POU2AF1</i>	POU2F2-KCNN4 POU2F2-CD79B CD79B-Ighg2b Ighg2b-APP APP-SUB1 APP-CTNNA1	6
<i>SPTAN1</i>	PTOV1-UBC TANC1-UBC TSSC-UBC TSSC-APP UBC-CLTA UBC-CAPN1 UBC-SLC9A2 UBC-FANCF CLTA-COP5	9

Das análises de interações foi evidenciado que três moléculas (APP, KCNN4 e UBC) interagem diretamente com pelo menos 2 dos genes validados.

5.6 IDENTIFICAÇÃO DE PARES DE TRANSCRITOS CODIFICADORES E NÃO CODIFICADORES COM NÍVEL DE EXPRESSÃO INVERTIDA

Para a identificação dos pares de transcritos Senso Antisenso (SAS) foi realizado uma primeira análise pelo Dr. Yuri Moreira mediante uso do software R. Foram separadas as sondas codificantes das sondas não codificantes anti-senso e feita uma análise para identificação dos transcritos diferencialmente expressos em cada grupo de sondas, separadamente (FDR do SAM de 10%). Em seguida, foi realizada uma comparação dos transcritos diferencialmente expressos com o objetivo de identificar transcritos antisenso e genes codificadores do mesmo *locus*. Foram encontrados 7 pares de transcritos senso antisenso, dois com correlação negativa (*LCOR* e *RFWD2*) e cinco com correlação positiva (*PAX8*, *TRPS1*, *TFF1*, *SCL39A6* e *MGC29506*).

Posteriormente em uma segunda análise utilizando o programa GeneSpring GX12.1 foi realizado um teste T (não pareado, com 1000 permutações, com correção Benjamini Hochberg FDR) nas amostras com receptor hormonal negativo e positivo tanto nos transcritos codificantes como nos não codificantes, de forma independente, com o objetivo de identificar transcritos diferencialmente expressos em cada grupo de sondas. Em seguida, foi realizada uma comparação dos transcritos diferencialmente expressos tanto no subgrupo receptor hormonal positivo e receptor hormonal negativo com o objetivo de identificar transcritos antisenso e genes codificadores do mesmo *locus*.

Foram identificados 53 pares de transcritos Senso-Anti-senso (SAS) com regulação inversa. Posteriormente foi realizado teste estatístico de correlação de Pearson e selecionados aqueles pares SAS com correlação absoluta maior que 0,5, originando 15 pares SAS, dos quais 2 já tinham sido identificados em análises anteriores (Tabela 5).

Tabela 5 - Correlação da expressão de genes e transcritos não codificadores no mesmo *locus*.

Gene Symbol	Sonda Gene	Sonda Transcrito No codificante	Correlação <i>Pearson</i> *
LCOR	A_23_P429670	CUST_1932_P1418160934	-0,74
RFWD2	A_23_P396194	CUST_62911_P1418160934	-0,73
AP3B1	A_23_P136635	CUST_32172_P1418160934	-0,69
PDXDC1	A_23_P147845	CUST_17876_P1418160934	-0,67
LRIG1	A_23_P109636	CUST_61569_P1418160934	-0,65
POGZ	A_23_P500734	CUST_10208_P1418160939	-0,60
CLEC16A	A_32_P194246	CUST_17861_P1418160934	-0,59
PDCD6	A_23_P218997	CUST_32018_P1418160934	-0,58
C11orf58	A_23_P47559	CUST_2256_P1418160934	-0,57
SAFB	A_23_P208446	CUST_62911_P1418160934	-0,57
IDE	A_24_P928038	CUST_39350_P1418160934	-0,57
AHCYL1	A_23_P368126	CUST_4401_P1418160939	-0,56
C3orf63	A_24_P371303	CUST_31330_P1418160934	-0,53
GALNT7	A_23_P144384	CUST_3363_P1418160939	-0,51
C18orf1	A_23_P207939	CUST_42595_P1418160934	-0,51

*p < 0,01

Os dados de expressão tanto do gene como do transcrito não codificadores no mesmo *locus*, foram representados num gráfico de dispersão, identificando (com cores) as características moleculares e de invasão linfonodal nas amostras, permitindo observar a distribuição das amostras caracterizadas nos transcritos codificadores e não codificadores dos genes anteriormente mencionados (ver Figura 19). Foi interessante perceber que tanto no *locus* do gene *LCOR* como do gene *RFWD2*, as amostras sem expressão do receptor hormonal (verde) mostravam uma tendência a maior expressão do transcrito Não Codificante e de modo inverso as amostras com expressão de receptor hormonal (azul) tiveram maior expressão do transcrito Codificante. Em relação a invasão linfonodal não foi observada qualquer associação desta característica com a expressão dos transcritos.

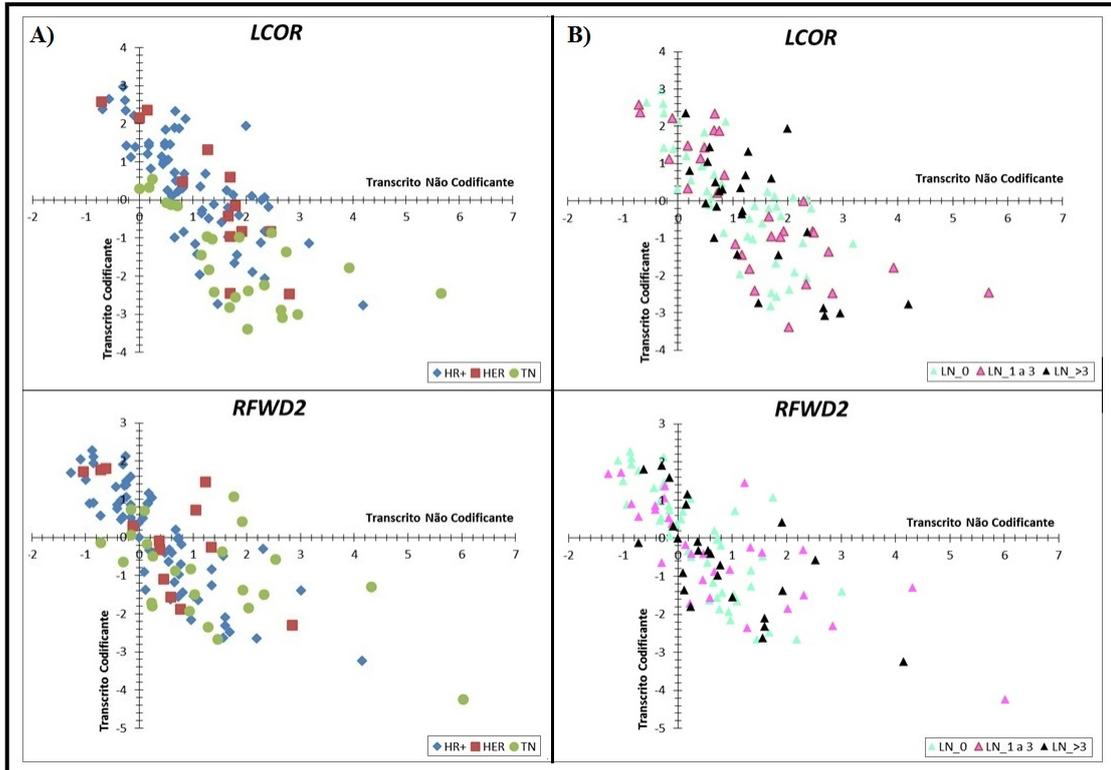


Figura 19 - Gráficos de dispersão dos transcritos codificadores para proteína (Eixo Y) e não codificadores (Eixo X) identificando os subgrupos moleculares e a invasão linfonodal. É possível observar a correlação entre os transcritos codificante para proteína e não codificante, localizados no mesmo *locus* dos genes *LCOR* e *RFWD2*, na Figura A) foram representados os subtipos moleculares; receptores hormonais (HR) (◆), HER-2 (■) e Triplo Negativo (●) mostrando que aparentemente existe uma maior expressão dos genes codificadores nas amostras que apresentam receptores hormonais (◆) e da sonda antisense nas amostras receptor hormonal negativo (●). Na Figura B) foi representada a característica de invasão linfonodal, da seguinte forma sem linfonodos acometidos (LN_0 ▲) de 1 a 3 linfonodos acometidos (LN_1 a 3 ▲) e maior que 3 linfonodos acometidos (LN_≥4 ▲).

6 DISCUSSÃO

O carcinoma ductal de mama consiste em um conjunto heterogêneo de tumores que apresenta diferentes alterações moleculares. As células tumorais podem adquirir características invasivas, como perda da adesão celular, aumento das capacidades migratórias e, finalmente, ultrapassagem da membrana basal. No câncer de mama, os fenótipos invasivos podem originar metástases, atingindo em primeira instância os gânglios linfáticos e, em seguida, os órgãos distantes. As mulheres que apresentam acometimento linfonodal têm pior prognóstico quando comparadas às mulheres com tumor restrito à mama, observando-se menor sobrevida livre de doença e aumento da taxa de mortalidade, sendo que a busca de alterações gênicas potencialmente envolvidas no processo progressão tumoral são de grande interesse da comunidade científica envolvidas em estudos do câncer de mama. Neste estudo foi utilizada a ferramenta de *cDNA microarray* com a intenção de avaliar transcritos codificadores e não codificadores expressos em células epiteliais de amostras tumorais de pacientes com câncer de mama primário, livres de tratamentos, para identificação de diferenças na expressão gênica que podem revelar informações moleculares importantes que auxiliem na compreensão deste processo.

A clusterização utilizando o perfil de expressão gênica do conjunto de genes PAM 50 (PARKER et al. 2009) permitiu agrupar as amostras em três grandes grupos: luminais, HER-2 e triplo negativo, observando-se um grande nível de concordância na classificação imunoistoquímica inicial e o perfil de expressão gênica. Especificamente foram agrupadas corretamente 87% das amostras luminais,

92,8% das amostras HER-2 e 81,8% das amostras triplo negativo. Resultados similares de concordância da classificação imunohistoquímica inicial e o perfil de expressão gênica foram descritos no trabalho de GUIU et al. (2012), sendo observado por ele concordância de 73%–100% nos subtipos Luminais, 41%–69% das amostras HER-2 e 80% das amostras triplo negativo, portanto a expressão gênica no conjunto geral é confiável para avaliar qualquer característica das pacientes com carcinoma ductal invasivo de mama.

Foram identificados 79 genes com diferenças de expressão entre as amostras dos pacientes sem acometimento linfonodal e com mais de 3 linfonodos acometidos, sendo que a clusterização hierárquica baseada no padrão de expressão desses genes não foi capaz de discriminar amostras dos 3 grupos representativos da progressão do acometimento linfonodal (LN0, LN1-3 e LN>3). No entanto, esse padrão se mostrou associado aos subgrupos de tumores, principalmente no que se refere aos receptores hormonais, sendo que o subgrupo de tumores TN foi totalmente discriminado dos demais. Esse resultado, utilizando todos os subtipos moleculares, é diferente aos de outros estudos (ABBA et al. 2007; SMEETS et al. 2011), por meio dos quais foi possível observar agrupamento por acometimento linfonodal utilizando somente amostras do subtipo Luminal A (HR+ e HER2-). No entanto, observamos que um dos genes identificados no trabalho de ABBA et al. (2007), o gene *SPTAN1*, foi também identificado e validado em nosso estudo, apresentando expressão diminuída em relação às pacientes com linfonodo acometido.

No nosso estudo, foram escolhidos 40 genes para validação por RT-qPCR com sistema *SYBR green*, utilizando cRNA 2º round de amplificação, e 10 dos 40 genes selecionados não puderam ser avaliados por apresentar amplificação

inespecífica. Dentre os 30 restantes, sete genes (*CDK5R2*, *CX3CLI*, *FOXQ1*, *MNX1*, *POU2AF1*, *SPTAN1* e *ZDHHC20*) apresentaram expressão concordante com os dados de cDNA *microarray*.

POU2AF1 (*POU class 2 associating factor 1*), um gene que mostrou expressão aumentada em amostras com maior número de linfonodos acometidos, está localizado no braço longo do cromossomo 11, numa região também chamada “supressora de tumor” (BACA-LÓPEZ et al. 2012). Esse gene codifica uma proteína membro da família POU, que tem capacidade de inibir a progressão do ciclo celular e modular a sobrevivência das células. A sigla POU é derivada dos três fatores de transcrição que compõem a família: fatores de transcrição específicos da pituitária de Pit-1, as proteínas com fator de transcrição em octâmero Oct-1 e Oct-2 (com sequência ATGCAAAT) e fator de transcrição neural Unc-86 de *Caenorhabditis elegans* (PHILLIPS e LUISI 2000). *POU2AF1* é também conhecido como *OBF-1* (*OCT-binding factor 1*) e codifica um coativador transcricional, que estimula a transcrição por meio da interação com as regiões promotoras dos genes, Oct-1 ou Oct-2 (SÁEZ et al. 2002; ZHAO et al. 2008). Na medula óssea, o OBF-1 promove a sobrevivência de células B de transição e a transcrição de um subconjunto de genes Ig (BORDON et al. 2008), sendo importante para a resposta a antígenos mediada por células B e a formação de centros germinais (SCHUBART et al. 1996). Foi observado expressão diferencial de *POU2AF1* em células de pacientes com leucemia linfocítica crônica em comparação com a expressão em células B normais (AUER et al. 2005). Tem sido detectado amplificação e super-expressão de *POU2AF1* em células de mieloma múltiplo, com reminiscência do crescimento celular (ZHAO et al. 2008). Foi mostrado que *AGR2* reprime a resposta imune humoral de *POU2AF1*,

postulando-se como um mecanismo de metástases do adenocarcinoma de pulmão (HUANG et al. 2014). Não há dados na literatura sobre o papel dessa proteína em câncer de mama.

CDK5R2 (*cyclin-dependent kinase 5, regulatory subunit 2 (p39)*) codifica para um ativador neuronal específico da ciclina 5 dependente de quinases (CDK5) (MURAVENKO et al. 2000). Nossos dados se mostraram menos expressos no grupo de amostras com mais de 3 linfonodos acometidos. Ciclina é uma família de serina/treonina quinases, CMGC (ciclina dependente de quinase, MAP quinase, glicogênio quinase sintase, quinases semelhantes a CDK) e desempenha um papel crucial na regulação da progressão celular, controlando vários eventos em diferentes fases do ciclo celular (MALUMBRES e BARBACID 2005). A CDK5R2 apresenta similaridade limitada as ciclinas, sugerindo se tratar de famílias de proteínas distintas das proteínas cinases dependentes de ciclinas. Não há dados na literatura sobre o papel dessa proteína em câncer de mama.

CX3CL1 (*chemokine (C-X3-C motif) ligand*) é uma citosina membro da família de quimosinas CX3C. Sintetiza uma proteína localizada na membrana celular que quando clivada dá origem à forma solúvel de CX3CL1, atuando como um potente quimioatraente de células T e monócitos, enquanto que a citosina ligada à membrana promove uma forte adesão de leucócitos a células endoteliais ativadas (DETERRE 2014). Ambas funções de quimiotaxia e adesão são mediadas pelo receptor CX3CR1 acoplado à proteína G, presente em células CD8 β , células T, células NK, células dendríticas e monócitos (IMAI et al. 1997). *CX3CL1* desempenha um papel crucial em diferentes tipos de câncer. Estudos anteriores em modelos animais mostraram efeitos antitumorais provocadas pelo sistema

imunológico (DEVAPATLA et al. 2012). Em linha com a sua função quimioatratante, a expressão de *CX3CLI* no câncer de mama tem sido correlacionada com o aumento da densidade de linfócitos infiltrantes no tumor (PARK et al. 2012; TSANG et al. 2013). Nossos resultados mostraram expressão aumentada de *CX3CLI* em amostras com maior número de linfonodos acometidos, sendo que os dados de expressão proteica observados por outros grupos são controversos. Por exemplo, TSANG et al. (2013), em concordância com nossos dados, reportaram maior expressão imunoistoquímica em tumores com linfonodos acometidos enquanto PARK et al. (2012), em discordância com nossos dados, mostraram maior expressão imunoistoquímica em tumores sem linfonodos acometidos. Os dados controversos podem ser, em parte, pela heterogeneidade do câncer de mama e os diferentes subgrupos que foram analisados nos diferentes estudos.

FOXQ1 (*forkhead box Q1*) codifica a proteína FOXQ1, que é membro da superfamília *forkhead box* (Fox) envolvida no crescimento, diferenciação celular e embriogênese (CHRISTENSEN et al. 2014). Essa proteína tem um domínio conservado FOX e um domínio extra FOX. O domínio FOX está envolvido na ligação de DNA, enquanto que as regiões extra FOX estão envolvidas na interação com os componentes de ativação e repressão da transcrição e complexos de reparo do DNA (KATOH et al. 2013). No nosso estudo, foi observado um aumento da expressão de *FOXQ1* entre amostras sem linfonodos acometidos contra os dois grupos de linfonodos acometidos (1 a 3 e maior que 3). Vários estudos têm associado a expressão de *FOXQ1* em câncer, especialmente com a transição epitélio-mesenquima (FEUERBORN et al. 2011; QIAO et al. 2011; ZHANG et al. 2011; ABBA et al. 2013; SEHRAWAT et al. 2013), processo que está intimamente

associado com propriedades pró-metastáticas, incluindo o aumento da motilidade e invasão (DRASIN et al. 2011). Foi observado aumento da expressão de *FOXQ1* em metástases de câncer de mama (QIAO et al. 2011; ROSS et al. 2015), juntamente com aumento da migração e invasão celular (ZHANG et al. 2011). Assim, também foi observado que no carcinoma colorretal existe aumento da expressão e correlação com a atividade da via Wnt em tumores primários (CHRISTENSEN et al. 2013). Em mama, a expressão desse genes foi associada a menor sobrevida, carcinoma de alto grau e imunofenótipo basal (QIAO et al. 2011).

SPTAN1 (*spectrin, alpha, non-erythrocytic 1*) Spectrins é uma família de proteínas do citoesqueleto filamentosos, as quais funcionam como proteínas estabilizadoras da membrana plasmática e organelas intracelulares. A proteína codificada tem sido envolvida na adesão e migração celular, comunicação intercelular, regulação do ciclo celular e reparação do DNA (SRIDHARAN et al. 2006; URSITTI et al. 2007; METRAL et al. 2009). A desregulação de *SPTAN1* afeta seriamente o comportamento celular e promove a progressão do tumor. Foi demonstrado aumento da expressão de *SPTAN1* em tumores de pulmão (SORMUNEN et al. 1994), cólon (YOUNES et al. 1989) e pele (TUOMINEN et al. 1996), sendo associado com agressividade local e comportamento metastático em carcinomas de tecidos moles (CUNHA et al. 2010). Além disso, *SPTAN1* foi vinculado à progressão do tumor de ovário (L'ESPÉRANCE et al. 2006). Nossos resultados mostraram expressão diminuída de *SPTAN1* em amostras com 1 a 3 de linfonodos acometidos, sendo, de certa forma, concordantes com o estudo de ABBA et al. (2007), por meio do qual se comparou a expressão usando a metodologia

SAGE das amostras de pacientes (HR+, HER-) com linfonodos acometidos *versus* sem acometimento linfonodal.

O gene *MNX1* (*motor neuron and pancreas homeobox 1*) está localizado no cromossomo 7q36.3 e pertence à família de genes *homeobox EHG*, que inclui também EN1, EN2, GBX1 e GBX2. *MNX1* é um gene de 12.801 pb, composto por 3 *éxons* e codifica para um fator de transcrição. Mutações no gene originam a síndrome Currarino, uma malformação congênita autossômica dominante, também conhecida como síndrome hereditário agenesia sacral (HSA; do inglês *Hereditary Sacral Agenesis*) (WILKENS et al. 2011). No nosso estudo, esse gene se mostrou menos expresso no grupo de amostras com mais de 3 linfonodos acometidos quando comparado com o grupo de amostras com 1 a 3 linfonodos acometidos. Não há dados na literatura sobre o papel dessa proteína em câncer de mama.

Por último, *ZDHHC20* (zinc finger, DHHC-type containing 20) é uma palmitoyl acyltransferase localizada no braço longo do cromossomo 13 (13q12.11) e se mostrou mais expresso nas amostras, com mais de 3 linfonodos acometidos quando comparadas com amostras dos dois outros grupos. Foi demonstrado que as proteínas palmitoiladas desempenham um importante papel na proliferação e nas vias de sinalização de sobrevivência das células cancerosas (DRAPER e SMITH 2010; KORYCKA et al. 2012). Além disso, foi observado que a supressão de *ZDHHC20* em linhagem de câncer de mama (MDA-MB231) aumenta a sobrevivência celular, sugerindo que palmitoilação da proteína DHHC20 regula aspectos importantes da progressão do câncer da mama (RUNKLE et al. 2014). Foi observado que a expressão de *ZDHHC20* está aumentada em vários tipos tumorais humanos, incluindo ovário, mama e próstata (DRAPER e SMITH 2010).

Os sete candidatos foram submetidos a uma análise biológica conjunta, utilizando o *software Ingenuity Pathway Analysis (IPA)*, por meio da qual a informação dos genes diferencialmente expressos são cruzados com conjuntos de genes que estão associadas a uma função biológica ou via particular com base nas relações de causa-efeito relatados na literatura (KRÄMER et al. 2014). Interessantemente, a análise permitiu obter uma única rede de interação gênica que incorpora todos os genes validados. A rede está associada, principalmente, a processos de movimento e proliferação celular, que está em concordância com o processo implicado ao longo do acometimento linfonodal. Portanto, a avaliação das redes de interação gênica permitiu incluir os genes identificados neste estudo em vias classicamente associadas ao fenômeno de progressão do câncer de mama, aumentando a credibilidade de que identificamos genes com grande potencial de estarem envolvidos nesse processo.

Os resultados deste estudo sugerem que a investigação das diferenças moleculares entre células epiteliais de carcinoma ductal invasivo de pacientes com e sem linfonodos acometidos pode resultar na identificação de novos genes envolvidos na progressão da invasão. Uma análise mais detalhada desse grupo de genes pode trazer informações importantes a respeito da biologia da progressão do CDI.

Em relação à análise dos transcritos codificadores e não codificadores no mesmo *locus*, foram identificado 15 pares de transcritos senso e anti-senso (SAS) com correlação inversa, devido à distribuição da expressão dos transcritos em relação aos receptores hormonais acreditamos que os transcritos mais interessantes são os localizados no *locus* dos genes *LCOR* e *RFWD2*.

O gene *LCOR* (co-repressor ligando dependente) é um coregulador transcricional com capacidade de ligar-se diretamente aos receptores nucleares (NR), através do motivo LxxLL (FERNANDES et al. 2003) reprimindo a transativação por hormônio via recrutamento de histonas desacetilases (HDACs) e proteínas de ligação C terminal (PALIJAN et al. 2009a). Dentre dos receptores nucleares tipo I, encontra-se os receptores de progesterona, estrogênio e andrógeno que na presença do hormônio recruta *LCOR* e modula a expressão gênica. Foi observado, em linhagem celular de câncer de mama deficientes de *LCOR*, aumento da transcrição do gene progesterona (PALIJAN et al. 2009b), assim como em câncer de próstata no qual foi observado que a repressão da expressão do receptor de andrógeno por *LCOR* leva a diminuição do crescimento do tumor (ASIM et al. 2011). Nossos resultados mostraram expressão invertida de transcritos codificante e não codificante no *locus* do Gene *LCOR*, sendo que não há dados na literatura sobre a expressão de transcrito não codificadores desse gene. Curiosamente nossos resultados mostraram uma aparente maior representatividade de amostras RH positivas nas amostras com maior expressão do gene e de TN nas amostras com menor expressão de *LCOR*.

O gene *RFWD2* (também conhecido como *COP1*) codifica uma ubiquitina ligase E3, membro da família de proteínas COP-DET-FUS. Participa do processo ubiquitinação e subsequente degradação de proteínas alvo no proteossomo. A função biológica conhecida até o momento indica que o gene está diretamente envolvido na ubiquitinação e degradação da oncoproteína c-Jun e a proteína supressora de tumor p53, suprimindo assim a transcrição e apoptose dependente de p53 (DORNAN et al. 2004b). O papel de *RFWD2* em câncer ainda não foi completamente estudado, não obstante, tem sido observado diferenças de expressão em vários tipos de câncer. Em

câncer gástrico os dados são controversos Li et al. (2012) relataram aumento da expressão de *RFWD2* no tumor e SAWADA et al. (2013) observaram diminuição da expressão. Foi sugerido que a diferença nos dados de expressão está influenciada pelo *status* de mutação da proteína supressora de tumor p53. Em câncer de mama os dados também estão pouco claros, os quais reportaram que 80% dos adenocarcinomas de mama (sem especificar o subtipo) têm expressão aumentada de *RFWD2* (DORNAN et al. 2004a). Em um estudo posterior com amostras TN OUYANG et al. (2015) reportaram diminuição da expressão em relação à mama normal. Nossos resultados mostraram expressão invertida de transcritos codificante e não codificante no *locus* do Gene *RFWD2*, sendo que não há dados na literatura sobre a expressão de transcrito não codificadores deste gene. Nossos dados de expressão gênica, de forma semelhante ao observado para o gene *LCOR*, mostram uma tendência de menor expressão de *RFWD2* com amostras negativas para RH concordando com os dados de OUYANG et al. (2015) que reportaram expressão diminuída do gene nesse grupo de tumores de mama.

Estudos mais detalhados dos transcritos senso e anti-senso são necessários para entender melhor o potencial processo de regulação da transcrição de genes pelos ncRNA longos.

7 CONCLUSÃO

- Foi avaliado o perfil de expressão gênica de células epiteliais em uma plataforma de transcritos codificadores e não codificadores de proteínas em uma coorte de amostras representativas dos diferentes subtipos de tumores de mama e com diferentes níveis de acometimento linfonodal gerando resultados robustos de expressão.
- Foi identificado um conjunto de 79 genes cujo padrão de expressão não foi capaz de discriminar as amostras das células epiteliais dos tumores em relação ao acometimento linfonodal nodal das pacientes, mas que mostrou uma tendência de associação com os subtipos de câncer de mama, principalmente em relação a expressão dos receptores hormonais.
- Sete genes foram validados por RT-qPCR - *CDK5R2*, *CX3CLI*, *FOXQ1*, *MNX1*, *POU2AF1*, *SPTAN1* e *ZDHHC20*, que participam de uma via gênica associada com processos de invasão e metástase, demonstrando grande potencial de realmente estarem associados ao processo de acometimento linfonodal, que é uma das etapas mais ameaçadoras do câncer de mama.
- Foi identificado dois pares de transcritos codificadores, *LCOR* e *RFWD2*, e não codificadores de proteínas (lncRNA), que podem estar intimamente relacionados e desempenhar um papel relevante na regulação da expressão gênica associada à tumorigênese.

8 REFERÊNCIAS BIBLIOGRÁFICAS

Abba MC, Sun H, Hawkins KA, et al. Breast cancer molecular signatures as determined by SAGE: correlation with lymph node status. **Mol Cancer Res** 2007; 5:881-90.

Abba M, Patil N, Rasheed K, et al. Unraveling the role of FOXQ1 in colorectal cancer metastasis. **Mol Cancer Res** 2013; 11:1017-28.

Abuázar CS. **Identificação e validação de marcadores moleculares para risco de progressão de carcinoma ductal de mama**. São Paulo; 2012. [Tese de Doutorado-Fundação Antonio Prudente].

Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. **Cell** 2006; 127:905-15.

Asim M, Hafeez BB, Siddiqui IA, et al. Ligand-dependent corepressor acts as a novel androgen receptor corepressor, inhibits prostate cancer growth, and is functionally inactivated by the Src protein kinase. **J Biol Chem** 2011; 286:37108-17.

Auer RL, Starczynski J, McElwaine S, et al. Identification of a potential role for POU2AF1 and BTG4 in the deletion of 11q23 in chronic lymphocytic leukemia. **Genes Chromosomes Cancer** 2005; 43:1-10.

Baca-López K, Mayorga M, Hidalgo-Miranda A, Gutiérrez-Nájera N, Hernández-Lemus E. The role of master regulators in the metabolic/transcriptional coupling in breast carcinomas. **PLoS One** 2012; 7: e42678.

Bergamaschi A, Tagliabue E, Sørli T, et al. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. **J Pathol** 2008; 214:357-67.

Bevilacqua C, Makhzami S, Helbling JC, Defrenaix P, Martin P. Maintaining RNA integrity in a homogeneous population of mammary epithelial cells isolated by laser capture microdissection. **BMC Cell Biol** 2010; 11:95.

Bezwoda WR, Esser JD, Dansey R, Kessel I, Lange M. The value of estrogen and progesterone receptor determinations in advanced breast cancer. Estrogen receptor level but not progesterone receptor level correlates with response to tamoxifen. **Cancer** 1991; 68:867-72.

Biswas S, Sengupta S, Roy Chowdhury S, et al. CXCL13-CXCR5 co-expression regulates epithelial to mesenchymal transition of breast cancer cells during lymph node metastasis. **Breast Cancer Res Treat** 2014; 143:265-76.

Bordon A, Bosco N, Du Roure C, et al. Enforced expression of the transcriptional coactivator OBF1 impairs B cell differentiation at the earliest stage of development. **PLoS One** 2008; 3: e4007.

Carter CL, Corle DK, Micozzi MS, Schatzkin A, Taylor PR. A prospective study of the development of breast cancer in 16,692 women with benign breast disease. **Am J Epidemiol** 1988; 128:467-77.

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.

Castro NP, Osório CA, Torres C, et al. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. **Breast Cancer Res** 2008; 10: R87.

Cheang MC, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. **J Natl Cancer Inst** 2009; 101:736-50.

Chia SK, Bramwell VH, Tu D, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. **Clin Cancer Res** 2012; 18:4465-72

Christensen J, Bentz S, Sengstag T, Shastri VP, Anderle P. FOXQ1, a novel target of the Wnt pathway and a new marker for activation of Wnt signaling in solid tumors. **PLoS One** 2013; 8:e60051.

Christensen J, Anderle P. **Atlas of genetics and cytogenetics in oncology and haematology** 2014; 19:11–4. FOXQ1 (forkhead box Q1). Available from: <URL:<http://AtlasGeneticsOncology.org/Genes/FOXQ1ID45906ch6p25.html>> [2015 nov 12]

Creighton CJ, Chang JC, Rosen JM. Epithelial-mesenchymal transition (EMT) in tumor-initiating cells and its clinical implications in breast cancer. **J Mammary Gland Biol Neoplasia** 2010; 15:253-60.

Cunha IW, Carvalho KC, Martins WK, et al. Identification of genes associated with local aggressiveness and metastatic behavior in soft tissue tumors. **Transl Oncol** 2010; 3:23-32.

de Azambuja E, Cardoso F, de Castro G Jr, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. **Br J Cancer** 2007; 96:1504-13.

Deterre P. **Atlas of genetics and cytogenetics in oncology and haematology** 2014; 18(4):260-3. CX3CL1 (chemokine (C-X3-C motif) ligand 1). Available from: <URL:<http://AtlasGeneticsOncology.org/Genes/CX3CL1ID46756ch16q13.html>> [2015 out 15]

Devapatla B, Sanders J, Samuelson DJ. Genetically determined inflammatory-response related cytokine and chemokine transcript profiles between mammary carcinoma resistant and susceptible rat strains. **Cytokine** 2012; 59:223-7.

Dornan D, Bheddah S, Newton K, et al. COP1, the negative regulator of p53, is overexpressed in breast and ovarian adenocarcinomas. **Cancer Res** 2004a; 64:7226-30.

Dornan D, Wertz I, Shimizu H, et al. The ubiquitin ligase COP1 is a critical negative regulator of p53. **Nature** 2004b; 429:86-92.

Draper JM, Smith CD. DHHC20: a human palmitoyl acyltransferase that causes cellular transformation. **Mol Membr Biol** 2010; 27:123-36.

Drasin DJ, Robin TP, Ford HL. Breast cancer epithelial-to-mesenchymal transition: examining the functional consequences of plasticity. **Breast Cancer Res** 2011; 13:226.

Eccles SA. The epidermal growth factor receptor/Erb-B/HER family in normal and malignant breast biology. **Int J Dev Biol** 2011; 55:685-96.

Eisenberg E, Levanon EY. Human housekeeping genes, revisited. **Trends Genet** 2013; 29:569-74.

Ellis MJ, Suman VJ, Hoog J, et al. Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype--ACOSOG Z1031. **J Clin Oncol** 2011; 29:2342-9.

Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. **Histopathology** 1991; 19:403-10.

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

ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature** 2007; 447:799-816.

Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. **BMC Bioinformatics** 2006; 7:191.

Espina V, Wulfkühle JD, Calvert VS, et al. Laser-capture microdissection. **Nat Protoc** 2006; 1:586-603.

Feldman AL, Costouros NG, Wang E, et al. Advantages of mRNA amplification for microarray analysis. **Biotechniques** 2002; 33:906-12, 914.

Fernandes I, Bastien Y, Wai T, et al. Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and independent mechanisms. **Mol Cell** 2003; 11:139-50.

Ferreira EN, Maschietto M, Silva SD, Brentani H, Carraro DM. Evaluation of quantitative rt-PCR using nonamplified and amplified RNA. **Diagn Mol Pathol** 2010; 19:45-53.

Ferreira EN, de Campos Molina G, Puga RD, et al. Linear mRNA amplification approach for RNAseq from limited amount of RNA. **Gene** 2015; 564:220-7.

Feuerborn A, Srivastava PK, Küffer S, et al. The Forkhead factor FoxQ1 influences epithelial differentiation. **J Cell Physiol** 2011; 226:710-9.

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

Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. **Arch Pathol Lab Med** 2000; 124:966-78.

Folgueira MA, Brentani H, Carraro DM, et al. Gene expression profile of residual breast cancer after doxorubicin and cyclophosphamide neoadjuvant chemotherapy. **Oncol Rep** 2009; 22:805-13.

Gibb EA, Becker-Santos DD, Enfield KS, et al. Aberrant expression of long noncoding RNAs in cervical intraepithelial neoplasia. **Int J Gynecol Cancer** 2012; 22:1557-63.

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.

Guiu, S.; Michiels, S.; André, F, et al Molecular subclasses of breast cancer: how do we define them? The impakt 2012 working group statement. **Annals Of Oncology : Official Journal Of The European Society For Medical Oncology / Esmo**, V. 23, N. 12, P. 2997–3006 2012

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.

Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. **BMC Genomics** 2006; 7:96.

Huang J, Wang L, Jiang M, Chen Q, Jiang Z, Feng H. AGR2-mediated lung adenocarcinoma metastasis novel mechanism network through repression with interferon coupling cytoskeleton to steroid metabolism-dependent humoral immune response. **Cell Immunol** 2014; B290:102-6.

Hung T, Chang HY. Long noncoding RNA in genome regulation: prospects and mechanisms. **RNA Biol** 2010; 7:582-5.

Imai T, Hieshima K, Haskell C, et al. Identification and molecular characterization of fractalkine receptor CX3CR1, which mediates both leukocyte migration and adhesion. **Cell** 1997; 91:521-30.

Ingenuity Pathway Analysis. **IPA Network Generation Algorithm**. 2005. Available from: <URL:<https://www.ingenuity.com/wp-content/themes/ingenuity-qiagen/pdf/ipa/IPA-netgen-algorithm-whitepaper.pdf>> [2015 Jan 15].

Inwald EC, Klinkhammer-Schalke M, Hofstädter F, et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. **Breast Cancer Res Treat** 2013; 139:539-52.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. **CA Cancer J Clin** 2011; 61:69-90. Erratum in: **CA Cancer J Clin** 2011; 61:134.

Katoh M, Igarashi M, Fukuda H, Nakagama H, Katoh M. Cancer genetics and genomics of human FOX family genes. **Cancer Lett** 2013; 328:198-206.

Katz A, Smith BL, Golshan M, et al. Nomogram for the prediction of having four or more involved nodes for sentinel lymph node-positive breast cancer. **J Clin Oncol** 2008; 26:2093-8.

Khamis ZI, Sahab ZJ, Sang QX. Active roles of tumor stroma in breast cancer metastasis. **Int J Breast Cancer** 2012; 2012:574025.

Knudsen ES, Ertel A, Davicioni E, Kline J, Schwartz GF, Witkiewicz AK. Progression of ductal carcinoma in situ to invasive breast cancer is associated with gene expression programs of EMT and myoepithelia. **Breast Cancer Res Treat** 2012; 133:1009-24.

Korycka J, Łach A, Heger E, et al. Human DHHC proteins: a spotlight on the hidden player of palmitoylation. **Eur J Cell Biol** 2012; 91:107-17.

Kreeger PK, Lauffenburger DA. Cancer systems biology: a network modeling perspective. **Carcinogenesis** 2010; 31:2-8.

Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. **Bioinformatics** 2014; 30:523-30.

L'Espérance S, Popa I, Bachvarova M, et al. Gene expression profiling of paired ovarian tumors obtained prior to and following adjuvant chemotherapy: molecular signatures of chemoresistant tumors. **Int J Oncol** 2006; 29:5-24.

Lapidot M, Pilpel Y. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. **EMBO Rep** 2006; 7:1216-22.

Lee JT. Epigenetic regulation by long noncoding RNAs. **Science** 2012; 338:1435-9.

Lester SC, Bose S, Chen YY, et al. Protocol for the examination of specimens from patients with invasive carcinoma of the breast. **Arch Pathol Lab Med** 2009; 133:1515-38.

Louro R, Nakaya HI, Amaral PP, et al. Androgen responsive intronic non-coding RNAs. **BMC Biol** 2007; 5:4.

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

Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. **Breast Cancer Res** 2009; 11:R7.

Malumbres M, Barbacid M. Mammalian cyclin-dependent kinases. **Trends Biochem Sci** 2005; 30:630-41.

Maruyama R, Suzuki H. Long noncoding RNA involvement in cancer. **BMB Rep** 2012; 45:604-11.

Matsui M, Prakash TP, Corey DR. Transcriptional silencing by single-stranded RNAs targeting a noncoding RNA that overlaps a gene promoter. **ACS Chem Biol** 2013; 8:122-6.

Metral S, Machnicka B, Bigot S, Colin Y, Dhermy D, Lecomte MC. AlphaII-spectrin is critical for cell adhesion and cell cycle. **J Biol Chem** 2009; 284:2409-18.

Michor F, Liphardt J, Ferrari M, Widom J. What does physics have to do with cancer? **Nat Rev Cancer** 2011; 11:657-70.

Ministério da Saúde. Instituto Nacional do Câncer. **Estimativa/2012 incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2011.

Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva. **Estimativa 2014: incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2014.

Moreira YB. **Identificação de perfis de expressão de rnas codificadores e não codificadores de proteína como preditores de recorrência de câncer de próstata**. São Paulo; 2010. [Tese de Doutorado-Universidade de São Paulo- Instituto de Química].

Muravenko OV, Gizatullin RZ, Protopopov AI, Kashuba VI, Zabarovsky ER, Zelenin AV. Assignment of CDK5R2 coding for the cyclin-dependent kinase 5, regulatory subunit 2 (NCK5AI protein) to human chromosome band 2q35 by fluorescent in situ hybridization. **Cytogenet Cell Genet** 2000; 89:160-1.

Nakaya HI, Amaral PP, Louro R, et al. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. **Genome Biol** 2007; 8:R43.

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

Narod SA. Tumour size predicts long-term survival among women with lymph node-positive breast cancer. **Curr Oncol** 2012; 19:249-53.

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

Ouyang M, Wang H, Ma J, et al. COP1, the negative regulator of ETV1, influences prognosis in triple-negative breast cancer. **BMC Cancer** 2015; 15:132.

Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. **J Clin Oncol** 2006; 24:3726-34.

Palijan A, Fernandes I, Verway M, et al. Ligand-dependent corepressor LCoR is an attenuator of progesterone-regulated gene expression. **J Biol Chem** 2009a; 284:30275-87.

Palijan A, Fernandes I, Bastien Y, et al. Function of histone deacetylase 6 as a cofactor of nuclear receptor coregulator LCoR. **J Biol Chem** 2009b; 284:30264-74.

Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. **J Clin Oncol** 2009; 27:1160-7.

Payne SJ, Bowen RL, Jones JL, Wells CA. Predictive markers in breast cancer--the present. **Histopathology** 2008; 52:82-90.

Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. **Nat Rev Genet** 2013; 14:880-93.

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

Phillips K, Luisi B. The virtuoso of versatility: POU proteins that flex to fit. **J Mol Biol** 2000; 302:1023-39.

Prat A, Parker JS, Fan C, Perou CM. PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer. **Breast Cancer Res Treat** 2012; 135:301-6.

QIAGEN. **RT 2 Profiler PCR Array Handbook**. 2014. Available from: <URL:<https://www.qiagen.com/br/resources/resourcedetail?id=6161ebc1-f60f-4487-8c9e-9ce0c5bc3070>> [2015 jan 12]

Qiao Y, Jiang X, Lee ST, Karuturi RK, Hooi SC, Yu Q. FOXQ1 regulates epithelial-mesenchymal transition in human cancers. **Cancer Res** 2011; 71:3076-86.

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

Ross JB, Huh D, Noble LB, Tavazoie SF. Identification of molecular determinants of primary and metastatic tumour re-initiation in breast cancer. **Nat Cell Biol** 2015; 17:651-64.

Rozenchan PB, Carraro DM, Brentani H, et al. Reciprocal changes in gene expression profiles of cocultured breast epithelial cells and primary fibroblasts. **Int J Cancer** 2009; 125:2767-77.

Runkle KB, Terkowski S, Stypulkowski E, Witze E. Identifying a function for DHHC20 in breast cancer. [abstract]. **Cancer Res** 2014; 74(Suppl):3298. [Presented at AACR Annual Meeting 2014; 2014 April 5-9; San Diego, CA]

Sawada G, Ueo H, Matsumura T, et al. Loss of COP1 expression determines poor prognosis in patients with gastric cancer. **Oncol Rep** 2013; 30:1971-5.

Schiffman SC, McMasters KM, Scoggins CR, Martin RC, Chagpar AB. Lymph node ratio: a proposed refinement of current axillary staging in breast cancer patients. **J Am Coll Surg** 2011; 213:45-52; discussion 52-3.

Schubart DB, Rolink A, Kosco-Vilbois MH, Botteri F, Matthias P. B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. **Nature** 1996; 383:538-42.

Sehrawat A, Kim SH, Vogt A, Singh SV. Suppression of FOXQ1 in benzyl isothiocyanate-mediated inhibition of epithelial-mesenchymal transition in human breast cancer cells. **Carcinogenesis** 2013; 34:864-73.

Sens-Abuázar C, Napolitano E, Ferreira E, et al. Down-regulation of ANAPC13 and CLTCL1: early events in the progression of preinvasive ductal carcinoma of the breast. **Transl Oncol** 2012; 5:113-23.

Simpson PT, Reis-Filho JS, Gale T, Lakhani SR. Molecular evolution of breast cancer. **J Pathol** 2005; 205:248-54.

Smeets A, Daemen A, Vanden Bempt I, et al. Prediction of lymph node involvement in breast cancer from primary tumor tissue using gene expression profiling and miRNAs. **Breast Cancer Res Treat** 2011; 129:767-76.

Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. **Proc Natl Acad Sci U S A** 2003; 100:8418-23.

Sormunen R, Pääkkö P, Palovuori R, Soini Y, Lehto VP. Fodrin and actin in the normal, metaplastic, and dysplastic respiratory epithelium and in lung carcinoma. **Am J Respir Cell Mol Biol** 1994; 11:75-84.

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.

Sridharan DM, McMahon LW, Lambert MW. alphaII-Spectrin interacts with five groups of functionally important proteins in the nucleus. **Cell Biol Int** 2006; 30:866-78.

Stacker SA, Achen MG, Jussila L, Baldwin ME, Alitalo K. Lymphangiogenesis and cancer metastasis. **Nat Rev Cancer** 2002; 2:573-83.

Sáez AI, Artiga MJ, Sánchez-Beato M, et al. Analysis of octamer-binding transcription factors Oct2 and Oct1 and their coactivator BOB.1/OBF.1 in lymphomas. **Mod Pathol** 2002; 15:211-20.

Sørensen KP, Thomassen M, Tan Q, et al. Long non-coding RNA expression profiles predict metastasis in lymph node-negative breast cancer independently of traditional prognostic markers. **Breast Cancer Res** 2015; 17:55.

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

Tahira AC, Kubrusly MS, Faria MF, et al. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. **Mol Cancer** 2011; 10:141.

Talmadge JE, Fidler IJ. AACR centennial series: the biology of cancer metastasis: historical perspective. **Cancer Res** 2010; 70:5649-69.

Thakkar JP, Mehta DG. A review of an unfavorable subset of breast cancer: estrogen receptor positive progesterone receptor negative. **Oncologist** 2011; 6:276-85.

Tlsty TD. Stromal cells can contribute oncogenic signals. **Semin Cancer Biol** 2001; 11:97-104.

Tsang JY, Ni YB, Chan SK, et al. CX3CL1 expression is associated with poor outcome in breast cancer patients. **Breast Cancer Res Treat** 2013; 140:495-504.

Tuominen H, Sormunen R, Kallioinen M. Non-erythroid spectrin (fodrin) in cutaneous tumours: diminished in cell membranes, increased in the cytoplasm. **Br J Dermatol** 1996; 135:576-80.

Ursitti JA, Petrich BG, Lee PC, et al. Role of an alternatively spliced form of alphaII-spectrin in localization of connexin 43 in cardiomyocytes and regulation by stress-activated protein kinase. **J Mol Cell Cardiol** 2007; 42:572-81.

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.

Veronese SM, Gambacorta M, Gottardi O, Scanzi F, Ferrari M, Lampertico P. Proliferation index as a prognostic marker in breast cancer. **Cancer** 1993; 71:3926-31.

Wang H, Owens JD, Shih JH, Li MC, Bonner RF, Mushinski JF. Histological staining methods preparatory to laser capture microdissection significantly affect the integrity of the cellular RNA. **BMC Genomics** 2006; 7:97.

Wellings SR, Jensen HM. On the origin and progression of ductal carcinoma in the human breast. **J Natl Cancer Inst** 1973; 50:1111-8.

Wilkens L, Jaggi R, Hammer C, Inderbitzin D, Giger O, von Neuhoff N. The homeobox gene HLXB9 is upregulated in a morphological subset of poorly differentiated hepatocellular carcinoma. **Virchows Arch** 2011; 458:697-708.

Xiang CC, Chen M, Kozhich OA, et al. Probe generation directly from small numbers of cells for DNA microarray studies. **Biotechniques** 2003; 34:386-8, 390, 392-3.

Yang G, Lu X, Yuan L. LncRNA: a link between RNA and cancer. **Biochim Biophys Acta** 2014; 1839:1097-109.

Younes M, Harris AS, Morrow JS. Fodrin as a differentiation marker. Redistributions in colonic neoplasia. **Am J Pathol** 1989; 135:1197-212.

Zhang H, Meng F, Liu G, et al. Forkhead transcription factor foxq1 promotes epithelial-mesenchymal transition and breast cancer metastasis. **Cancer Res** 2011; 71:1292-301.

Zhao C, Inoue J, Imoto I, et al. POU2AF1, an amplification target at 11q23, promotes growth of multiple myeloma cells by directly regulating expression of a B-cell maturation factor, TNFRSF17. **Oncogene** 2008; 27:63-75.

**Anexo 1 - Carta de aprovação do Comitê de Ética em Pesquisa da Fundação
Antônio Prudente-CEP**

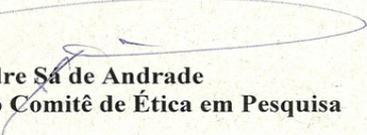
São Paulo, 14 de Julho de 2011.

À
Dra. Silvia Regina Rogatto

**Ref.: Projeto de Pesquisa nº. 1174/08 -B
“Determinação dos perfis de expressão do câncer de mama e próstata utilizando genes
codificadores e não codificadores de proteínas”.**

Os membros do Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antonio Prudente – Hospital do Câncer - A.C. Camargo/SP, em sua última reunião de 28/06/2011, **tomaram conhecimento e aprovaram** a solicitação de dispensa da apreciação do estudo intitulado “**Avaliação da expressão de genes codificadores e não codificadores em células epiteliais capturadas do carcinoma ductal invasivo de mama de pacientes com e sem linfonodo acometido**”, em razão deste estudo estar inserido em projeto supracitado, previamente aprovado por este comitê.

Atenciosamente,


Dr. Alexandre Sa de Andrade
2º Vice-Coordenador do Comitê de Ética em Pesquisa

Anexo 2 - Classificação Clínica Tumores de Mama

Os tumores de mama são classificados clinicamente pelo tamanho do tumor (T), acometimento linfonodal (N) e metástases a distância (M), a continuação são descritas cada característica segundo as possibilidades clínicas

T Exame físico ou diagnóstico por imagem (inclusive a mamografia)	TX	O tumor não pode ser avaliado.
	T0	Não há evidência do tumor primário.
	Tis	Carcinoma <i>in situ</i> e Paget – Doença de Paget do mamilo sem tumor de mama.
	T1	Tumor com 2 cm ou menos em sua maior dimensão. T1a – com mais de 0,1 cm até 0,5 cm em sua maior dimensão. T1b – com mais de 0,5 cm até 1 cm em sua maior dimensão. T1c – com mais de 1 cm, porém não mais de 2 cm em sua maior dimensão.
	T2	Tumor com mais de 2 cm até 5 cm em sua maior dimensão.
	T3	Tumor com mais de 5 cm em sua maior dimensão.
	T4	Tumor de qualquer tamanho com extensão direta à parede torácica ou à pele, somente como descritos em T4a e T4b . T4a – Extensão à parede torácica. T4b – edema (inclusive aspecto em casca de laranja) ou ulceração da pele da mama ou nódulos cutâneos satélites confinados à mesma mama. T4c – Ambos (T4a e T4c). T4d – Carcinoma inflamatório.

<p style="text-align: center;">N</p> <p>Exame físico ou diagnóstico por imagem (inclusive a mamografia)</p>	NX	Os linfonodos regionais não podem ser avaliados.
	N0	Ausência de metástase em linfonodos regionais.
	N1	Metástase em linfonodo(s) axilar(es), homolateral(ais), móvel(eis).
	N2	Metástase em linfonodo(s) axilar(es) homolateral(is) fixo(s) ou metástase clinicamente aparente em linfonodo(s) mamário(s) interno(s) homolateral(is), na ausência de evidência clínica de metástase em linfonodo(s) axilar(es). N2a – metástase em linfonodo(s) axilar(es) fixos uns aos outros ou a outras estruturas. N2b – metástase clinicamente aparente em linfonodo(s) mamário(s) interno(s), na ausência de evidência clínica de metástase em linfonodo(s) axilar(es).
	N3	Metástase em linfonodo(s) infraclavicular(es) homolateral(ais) com ou sem envolvimento de linfonodo(s) axilar(es); ou clinicamente aparente em linfonodo(s) mamário(s) interno(s) homolateral(is), na presença de evidência clínica de metástase em linfonodo(s) axilar(es); ou metástase em linfonodo(s) homolateral(is) com ou sem envolvimento de linfonodo(s) axilar(es) ou mamário(s) interno(s). N3a – Metástase em linfonodo(s) infraclavicular(es). N3b – Metástase em linfonodo(s) mamário(s) interno(s) e axilar(es). N3c – Metástase em linfonodo(s) supraclavicular(es).
<p style="text-align: center;">M</p> <p>Exame físico ou diagnóstico por imagem (inclusive a mamografia)</p>	MX	A presença de metástase a distância não pode ser avaliada.
	M0	Ausência de metástase a distância.
	M1	Metástase a distância.

Anexo 3 - Formulário de Coleta de Dados

Os dados foram coletados em formulário que contém 41 questões fechadas e uma coluna resumo para inserir o código da resposta. Posteriormente cada questão foi tabulada numa planhilla excel, para o analises.

MIC.....

Informação pessoal

Nome:	
Idade ao diagnostico:	RGH:

Informação clinica

Nº	Perguntas	Resumo
1.	Primeiro evento (1) DCIS (2) IDC (3) lesão benigna (99) ignorado (4) outros. Quais:	
2.	Mama primeiro evento: (1) Direita (2) Esquerda (3) bilateral (99) ignorado	
3.	Biópsia (0) Não (1) SIM (2) fora do Hospital (99) ignorado	
4.	AP da biópsia:	
5.	Data da biópsia:...../...../.....	
6.	Revisão de lâminas (0) NÃO (1) SIM (99) ignorado	
7.	Data da revisão:...../...../.....	
8.	AP da revisão:.....	
9.	Tipo histológico da biópsia ou na revisão da lamina 1er evento MD..(1) Intraductal (Ca <i>in situ</i>) (2) lobular <i>in situ</i> (3) Ductal invasivo (4) Ductal invasivo com componente intraductal predominante (5) lobular invasivo (6) mucinoso (colóide) (7) medular (8) papilífero (9) tubular (10) adenocístico (11) secretor (juvenil) (12) apócrino (13) metaplásico (99) Ignorado.	
10.	Grau Histológico (SBR) (0) In situ (1) I (2) II (3) III (88) Não se aplica (99) Ignorado	
11.	Grau Nuclear (1) Grau 1 (2) Grau 2 (3) Grau 3 (88) Não se aplica (99) Ignorado	
12.	Índice Mitótico.....(99) Ignorado	
13.	Infiltrado Inflamatório (0) não (1) Leve (2) Moderado ou discreta (3) Intenso (9) Sim (SE) (99) Ignorado	

14.	Positividade Marcadores IHQ	
	ER: (0) Não (1) Sim score.....% (99) ignorado	
	PR (0) Não (1) Sim score.....% (99) ignorado	
	HER2 (0) Não (1) Sim score.....(99) ignorado	
	P53 (0) Não (1) Sim score.....% (99) ignorado	
	CKT 5 (0) Não (1) Sim score.....% (99) ignorado	
	Ki67 (0) Não (1) Sim score.....% (99) ignorado	
	EGFR (0) Não (1) Sim score.....% (99) ignorado	
15.	Cirurgia...(0) NÃO (1) SIM	
16.	Tipo de cirurgia (1) mastectomia simples (2) mastectomia radical modificada (3) Ressecção segmentar (4) quadrantectomia (88) Não se aplica (99) ignorado	
17.	Mama da cirurgia (1) direita (2) esquerda (3) bilateral (88) não se aplica (99) ignorado	
18.	Data da cirurgia...../...../.....	
19.	AP da cirurgia.....	
20.	Tipo histológicos cirurgia (1) Intraductal (Ca <i>in situ</i>) (2) lobular <i>in situ</i> (3) Ductal invasivo (4) DI com componente intraductal predominante (5) lobular invasivo (6) mucinoso (colóide) (7) Medular (8) papilífero (9) tubular (10) adenocístico (11) secretor (juvenil) (12) Apócrino (13) Metaplásico (99) ignorado	
21.	Grau Histológico (SBR) (0) In situ (1) I (2) II (3) III (88) Não se aplica (99) ignorado	
22.	Grau Nuclear (1) Grau 1 (2) Grau 2 (3) Grau 3 (88) Não se aplica (99) ignorado	
23.	Índice Mitótico.....(99) ignorado	
24.	Infiltrado Inflamatório (0) não (1) Leve (2) Moderado ou discreto (3) Intenso (9) Sim (SE) (99) Ignorado	
25.	T.....N.....M.....	
26.	Linfonodos ressecados.. (0) NÃO (1) SIM (99) ignorado	
27.	Número de linfonodos positivos.....	

28.	Presença de Micrometástases.. (0)NÃO (1) SIM (99)ignorado	
29.	Presença de tumor residual (0) Não (1) SIM (99)ignorado	
30.	Positividade Marcadores IHQ	
	ER: (0) Não (1) Sim score.....% (99) ignorado	
	PR (0) Não (1) Sim score.....% (99) ignorado	
	HER2 (0) Não (1) Sim score.....(99) ignorado	
	P53 (0) Não (1) Sim score.....% (99) ignorado	
	CKT 5 (0) Não (1) Sim score.....% (99) ignorado	
	Ki67 (0) Não (1) Sim score.....% (99) ignorado	
	EGFR (0) Não (1) Sim score.....% (99) ignorado	
31.	Tratamento MD_1 (1)Hormonioterapia (2) Radioterapia (3) Quimioterapia	
32.	Data de inicio ao tratamento...../...../.....	
33.	Segundo evento Presença de segundo evento (0)NÃO (1) SIM (99)ignorado	
34.	Mama segundo evento: (1) Direita (2) Esquerda (3) bilateral (88) não se aplica (99)ignorado.	
35.	Data do segundo evento:	
36.	Tipo de segundo evento.. (1) DCIS local (2) IDC (3) metástase à distância (4) lesão benigna (88) não se aplica (99) ignorado	
37.	AP do segundo evento	
38.	Evolução Óbito.....(0)NÃO (1) SIM (88) não se aplica (99)ignorado	
39.	Causa de óbito (1) câncer (2) outra causa (99) ignorado (88) não se aplica	
40.	Data do óbito:	
41.	Data da última consulta:	

Anexo 4 - Protocolo de Coloração de Tecido Congelado

Arcturus® HistoGene™

1. Retirar as lamina a ser microdissecadas e manter em gelo seco.
2. Descongelar uma lamina, por poucos segundos. Em seguida tratar com os seguintes reagentes fornecidos pelo kit:
3. ETOH 75%, 30sec.
4. Staining / Hematoxilina – Histogene Part # 12241-05, 30 sec. Drenar o excesso.
5. H₂O, retirar o excesso.
6. H₂O, 30 sec.
7. ETOH 75%, 30 segundos
8. ETOH 95%, 30 segundos
9. ETOH 100% 2 minutos
10. Xilol 5 minutos
11. Deixar secar por 1 minuto.

Anexo 5 - Extração e Purificação do RNA

Arcturus® PicoPure® RNA Isolation Kit, Life™

Extração do RNA

1. Acoplar anel de extração ao CAP e adicionar 14µl do Extraction Buffer (XB)
2. Cobrir com tubo de 0,5ml (Applied Biosystems, cat# N801-611), devidamente rotulado.
3. Incubar a 42°C por 30 min.
4. Spin a 800 X g por 2 min.
 - **Manter congelado a -80°C. Processar o RNA o mais breve possível.**

Purificação do RNA

1. Adicione 250µl de buffer CB à coluna e incubar por 5min a temperatura ambiente.
2. Centrifugar a 16.000 X g por 1 minuto.
3. Adicionar 14µl ETOH 70% misturar com a pipeta e transferir para coluna (28ul aprox.).
4. Centrifugar a 100 X g por 2 minutos.
5. Centrifugar a 16.000 X g por 30 segundos.
6. Adicionar 100µl do tampão de lavagem (W1) e centrifugar a 8.000 X g por 1 minuto.

Tratamento com DNase: Misturar 2,5µl de DNase Qiagen (cat# 79254) e 17,5 ul do tampão de diluição de DNase. Adicionar à coluna 20µl da mistura à coluna e incubar a temperatura ambiente por 15 minutos.

7. Adicionar 40µl do tampão W1.
8. Centrifugar a 8.000 X g por 1 minuto.
9. Adicionar 100µl do tampão de lavagem (W2).
10. Centrifugar a 8.000 X g por 1 minuto.
11. Adicionar 100µl do tampão W2.
12. Centrifugar a 16.000 X g por 2 minutos.
13. Centrifugar novamente 16.000 X g por 1 minuto.
14. Trocar de tubo por o de 0,5ml.

15. Adicionar 15µl do Elution Buffer.
16. Incubar a Temperatura ambiente por 1 minuto.
17. Centrifugar a 1.000 X g por 1 minuto.
18. Centrifugar a 16.000 X g por 1 minuto.
19. Retirar alíquota de 1,5µl para analisar no Bioanalyser/Picochip e congelar a80C.
20. Concentrar o restante em speed-vac para volume de 4,15ul.

Anexo 6 - Protocolo RNA 6000

Bioanalyzer® Agilent Technology™

Preparação do gel (reagentes estocados na geladeira)

1. Colocar 550µL de RNA 6000 Pico Gel (matriz) num tubo e dar um spin rápido.
2. Centrifugar a 1.500 g por 10 minutos à temperatura ambiente.
3. Fazer alíquotas de 32,5µL 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 minutos.
2. Dar um vortex no corante durante 10 seg e em seguida dar um spin para depois adicionar 0,5µL do corante numa alíquota de 32,5µL do gel já preparado.
3. Vortexar bem a solução. Dar um spin no tubo a 13.000 x g durante 10 min à temperatura ambiente.

Transferência do mix gel-corante

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

Transferência da solução condicionante e marcador

1. Pipetar 9µL do reagente condicionante no poço / CS.
2. Pipetar 5µL do reagente marcador em todos os 11 poços de amostra e no poço do *ladder*.

Transferência da amostra padrão (ladder) e das amostras

1. Pipetar 1µL de amostra padrão (Ladder) no poço marcado.
2. Pipetar 1µL de amostra em cada um dos 11 poços de amostra. Pipetar 1µL de marcador em cada poço de amostra não utilizada.
3. Agitar o Chip, no vortex especial, durante 1 min a 2000 rpm.
4. Analisar o chip no Bioanalisador Agilent 2100. Usando Plataforma RNA total Pico Chip.

Lavagem da estação

1. Adicionar ~300µL de água Sigma, ao Chip transparente, observar que a água se espalhe em todos os poços
2. Colocar na estação do Bioanalisador por 1 min, permitindo que os sensores entrem em contatos com a água.
3. Deixar a tampa aberta do equipamento, para secar os sensores.

Anexo 7 - Protocolo Purificação RNA

RNeasy® Mini kit, Qiagen™

- 1.** Ajustar o volumem da amostra a 100µl, com água livre de RNase.
- 2.** Adicionar 350µl de tampão de lisis RLT, e misturar.
- 3.** Adicionar 250µl de Etanol 100% e misturar com a pipeta.
- 4.** Transferir a amostra (700µl) à coluna RNeasy Mini spin ajustada a um tudo de coleta (2ml).
- 5.** Centrifugar por 2 min a 100 x g. Seguido por uma segunda centrifugação por 30 s a 8000 x g.
- 6.** Retirar o tubo de coleta e trocar por um novo.
- 7.** Adicionar 500µl de tampão de lavagem RPE à coluna RNeasy Mini spin.
- 8.** Centrifugar por 30 s a 8000 x g.
- 9.** Retirar o tubo de coleta e trocar por um novo.
- 10.** Adicionar 500µl de tampão de lavagem RPE à coluna RNeasy Mini spin.
- 11.** Centrifugar por 30 s a 8000 x g.
- 12.** Retirar o tubo de coleta e trocar por um novo.
- 13.** Centrifugar por 1 min a 8000 x g.
- 14.** Retirar o tubo de coleta e trocar por um tubo de micro centrifuga (1,5ml).
- 15.** Adicionar, diretamente na membrana da coluna, 60µl a 80µl de H₂O livre de RNase. Centrifugar por 1min a 8000 x g.
- 16.** Manter em gelo

Anexo 8 - Amplificação do mRNA Microdissecado

Quick Amp Labeling kit ® GE™ e CASTRO et al. 2008

Preparo dos RNAs Spikes – Agilent cat# 5188-5279.

Os RNAs Spikes A e B fornecidos pela Agilent devem ser retirados do freezer -80°C e descongelados em gelo. Vortexar e spin breve.

Diluir seguindo a tabela abaixo. Manter os tubos sempre em gelo, vortexar a cada passo durante 10 sec e spin breve. A diluição na etapa 1 pode ser armazenada a -80°C por até 2 meses e a diluição na etapa 2 por 2 semanas.

Etapa #	Diluição	Volume RNA	Volume de tampão de diluição
1	1/20	1	19
2	1/40	4	156
3	1/4	10	30
4	1/10	10	90
5	1/10	10	90

*Diluição na etapa 1 e 2 são feitas em tampão, nas etapas 3,4 e 5 podem ser em água.

Síntese do cDNA, primeira e segunda fita. (Quick Amp Labeling kit ® GE™)

Misturar: 4,15µl do RNA total com 1µl do RNA SPIKE diluição etapa #5 e 0,6µl iniciador oligo-dT T7.

Incubar à 65°C por 10 minutos e resfriar rapidamente em gelo por 10 min. Preparar o master mix cDNA.

Reagentes	1 reação (µl)
5X buffer *	2
0,1 M DTT	1
10 mM dNTP mix	0,5
MMLV-RT	0,5
RNase Out	0,25
TOTAL	4,25

*Importante - Aquecer 5X buffer @ 80°C/ 4 min.

Adicionar 4,25µl de mix e incubar a 40°C por 2hs, seguido de 65°C por 15 minutos e resfriar em gelo.

Esta reação pode ser armazenada à -80°C.

1º CICLO DE AMPLIFICAÇÃO.

TRANSCRIÇÃO *IN VITRO*

(Quick Amp Labeling kit ® GE™)

Preparar o mix de transcrição:

Reagentes	1 reação (µl)
H ₂ O	4,85
4X Transcription buffer	10
0,1 M DTT	3
*NTP mix / 25mM	8
**PEG	3,2
RNase Out	0,25
Pyrophosphatase inorganic	0,3
T7 RNA pol	0,4
TOTAL	30

*** NTPs (U,A,C,G) 25mM “in house” . Não usar o NTP mix que do kit!!!!**

****Esquentar a 40°C por 5 min.**

Adicionar ao cDNA, derivado do procedimento anterior, 30µl de mix por tubo e incubar a 40°C por 3hs, no termociclador, em seguida resfriar em gelo.

Purificação do RNA RNeasy® Mini kit, Qiagen™ (Ver Anexo 6)

- Adicionando uma etapa de lavagem com **ETOH 80%**, depois das duas lavagens com tampão RPE.
- Recuperar o RNA por eluição com 80µl de água livre de RNase. Recomendado reeluir a coluna usando o primeiro eluato.
- Armazenar o cRNA imediatamente a -80°C.

SÍNTESE do cDNA

(CASTRO et al. 2008)

Primeira Fita do cDNA

1. Reduzir o volume do cRNA para 6,7µl através de centrifuga com vácuo.
2. Adicionar 1µl de iniciador Random Hexameros (2ug/µl). Misturar bem e centrifugação breve.
3. Incubar a 70°C por 10 min - resfriar imediatamente em gelo por 10 min.
4. Preparar o máster mix da síntese da Primeira fita de cDNA:

Reagentes	1 reação (µl)
5X buffer	4
MgCl ₂	4,8
10 mM dNTP mix	2
RNAsin	0,5
Improm II	1
Total	12.3

5. Distribuir 12,3µl por tubo.
6. Incubar, em termociclador, a 42^oC por 2hs.

Segunda Fita do cDNA

Preparar o máster *mix* da síntese da Segunda fita de cDNA

Reagentes	1 reação (µl)
H ₂ O DEPC	64,6
Advantage PCR buffer	10
Oligo dT do kit Agilent	0,7
10 mM dNTP mix	2
RNAse H (2U/ul Gibco BRL Cat# 18021-071)	0,7
Advantage Polymerease (Clontech Cat# 8417-1)	2
Total	80

Adicionar à primeira fita do cDNA, derivado do procedimento anterior, 80µl de *mix* por tubo e incubar no termociclador, nas seguintes temperaturas

37°C por 5min para digerir o mRNA,

94°C por 4min para denaturar

65°C por 6min para elinemento específico do iniciador

75°C por 30min para a extensão.

Deter a reação com 5µl da solução 1M NaOH/2mM EDTA.

Incubar a 65°C por 10min para inativar a enzima.

Purificação do cDNA dupla fita.

Adicionar ao produto da reação

1. 100µl **fenol pH 8.0/clorofórmio/álcool isoamílico** 25:24:1 (guardado 8°C)
2. Homoginizar por inversão por 20 s.
3. Centrifugar a 16.000 X g por 5 min. a temperatura ambiente.
4. Transpassar a fase acuosa (superior) a outro tubo que tem 100µl de **clorofórmio/álcool isoamílico** 24:1
5. Homoginizar, por inversão, por 20 segundos.
6. Centrifugar a 16.000 X g por 5 min. a temperatura ambiente.
7. Recuperar a fase aquosa e manter no gelo.

Precipitação do cDNA

1. Adicionar à fase aquosa:
 - 1,5µl de **acrilamida** (5mg/ml, Ambion, cat # 9520)
 - 0,5 vezes do volumem recuperado de **NH₄AC 7,5 M** (aprox. 75µl)
 - 2,5 vezes do volumem de **ETOH 100%** gelado (aprox. 375 ul)
2. Misturar gentilmente no vortex
3. Manter 1hr a -80°C ou toda a noite a -20°C
4. Centrifugar a 20.000 X g por 30 min a 4°C.
5. Lavar o *pellet* com 1,5ml de **ETOH 75%**
6. Centrifugar a 20.000 X g por 10 min a 4°C
7. Remover o sobrenadante e secar totalmente ao ar.
8. Ressuspender pellet em 15 ul **H₂O depc.** Incubar a 37°C por 20min

2º CICLO DE AMPLIFICAÇÃO

Transcrição in vitro/ marcação

1. Preparar o mix de transcrição:

Reagentes	1 reação (µl)
H2O	7,65
4X Transcription buffer	10
0,1 M DTT	3
NTP mix / kit Agilent	4
PEG 50%	3,2
RNAse Out	0,25
Pyrophosphatase	0,3
T7 RNA pol	0,4
* CIANINAS	1,2
Total	30

2. Distribuir 28,8µl da mix por tubo.
3. Adicionar 1,2µl do flouoroforo (**CY3-UTP (spike A) ou CY5-UTP (Spike B)**) por reação. Misturar com pipeta e incubar, no termociclador, a 40°C por 3hs, resfriar rapidamente em gelo.
4. Purificar o RNA, usando o **kit RNeasy mini kit da Qiagem (Anexo 6)**.
5. **Eluir** a coluna com **60 ul de água DEPC**. Recomendo reeluir a coluna usando o primeiro eluato.
6. **Quantificar o cRNA amplificado** e a concentração dos fluorocromos incorporados utilizando NanoDrop ND-1000 UV-VIS Spectrofotometro.
7. **Determine a atividade específica da sonda sintetizada**. Use a concentração do cRNA (ng/µL) e cyanine 3 or cyanine 5 (pmol/µL) para determinar a atividade específica:
$$(Conc\ do\ Cy3\ ou\ Cy5) / (Conc\ do\ cRNA) * 1000 = pmol\ Cy3 / \mu g\ cRNA$$
8. Armazenar o cRNA imediatamente @ -80°C.
9. Retirar a quantidade a ser fragmentada / hibridizada e armazenar o restante para hibridizações futuras.

Anexo 9 - Dados Clínicos e Anatomopatológicos

ID	Idade	Grau SBR	Grau Nuclear	Índice mitótico	TNM	LN Ac/Res	ER	Score	PR	Score	HER-2	Score	Ki67	Score	p53	EGFR	CK 5
MIC 010T	55	1	1	2	T4bN2M0	7/36	+	90	-	0	+	3	N/D	N/D	+	N/D	N/D
MIC 021T	43	3	2	25	T1cN0M0	0/20	-	0	-	0	-	1	N/D	N/D	N/D	N/D	N/D
MIC 038T	67	1	2	8	T2N1M0	1/19	+	95	+	100	-	0	+	5	-	N/D	N/D
MIC 040T	78	3	3	10	T4bN1M0	2/16	+	95	+	95	-	1	+	<20	-	N/D	N/D
MIC 044T	55	2	2	3	T3N0M0	0/17	+	100	+	30	-	1	+	<5	-	N/D	N/D
MIC 046T	57	3	3	18	T2N3M0	17/31	-	0	-	0	+	3	N/D	N/D	+	N/D	N/D
MIC 049T	51	2	3	4	T2N2M0	0/5	+	90	+	60	+	3	N/D	N/D	+	N/D	N/D
MIC 055T	58	2	2	9	T1cN1M0	0/17	+	100	+	N/D	-	1	+	40	-	N/D	N/D
MIC 057T	51	2	2	10	T1N1M0	1/14	+	40	+	60	-	1	+	10	N/D	N/D	N/D
MIC 063T	79	2	3	3	T4bN1M0	3/15	+	100	+	100	-	1	+	60	-	N/D	N/D
MIC 064T	46	2	3	3	T2N0M0	1/11	+	100	+	100	-	0	+	30	N/D	N/D	N/D
MIC 069T	58	2	2	1	T2N3Mx	10/29	+	70	+	40	-	0	+	10	-	N/D	N/D
MIC 073T	72	2	2	9	T2N2M0	7/17	+	100	+	100	-	1	+	20	-	N/D	N/D
MIC 081T	56	1	2	3	T1N2M0	8/25	+	90	+	15	+	3	N/D	N/D	N/D	N/D	N/D
MIC 085T	64	2	3	3	T1cN2M0	5/17	+	80	+	80	-	0	+	10	N/D	N/D	N/D
MIC 090T	48	3	3	23	T1cN0M0	0/0	-	0	-	0	-	0	N/D	N/D	N/D	N/D	N/D
MIC 120T	41	3	3	2	T2N3M0	17/19	+	30	+	50	-	1	+	40	-	N/D	N/D
MIC 137T	44	3	3	50	T1cN1M0	1/23	-	0	-	0	-	1	N/D	N/D	-	N/D	+
MIC 141T	60	3	3	76	T2N0M0	3/26	-	0	-	0	-	0	N/D	N/D	+	N/D	+
MIC 143T	75	3	3	3.7	T1cN0M0	0/17	-	0	-	0	-	2	N/D	N/D	+	N/D	+
MIC 144T	62	3	3	27	T2N3M0	34/46	-	0	-	0	-	1	N/D	N/D	+	N/D	-
MIC 145T	78	3	3	12	T2N0M0	0/32	-	0	-	0	+	2/FISH+	N/D	N/D	-	N/D	N/D
MIC 146T	18	3	3	50		17/40	-	0	-	0	-	1	N/D	N/D	+	N/D	+
MIC 147T	40	2	3	4	T2N2M0	4/10	-	0	-	0	+	3	N/D	N/D	+	N/D	N/D
MIC 148T	48	3	3	15	T1cN0M0	0/3	-	0	-	0	+	3	N/D	N/D	+	N/D	N/D

Cont/ Anexo 9

ID	Idade	Grau SBR	Grau Nuclear	Índice mitótico	TNM	LN Ac/Ress	ER	Score	PR	Score	HER-2	Score	Ki67	Score	p53	EGFR	CK 5
MIC 149T	32	2	3	3	T4bN1M0	2/21	-	0	-	0	+	3	N/D	N/D	+	N/D	N/D
MIC 173T	38	2	3	4	T1cN0M0	1/1	-	0	-	0	-	0	+	10	+	+	+
MIC 175T	41	3	3	22	T2N1M1	2/14	-	0	-	0	-	0	N/D	N/D	+	N/D	-
MIC 176T	50	2	3	3	T1N0M0	0/2	+	30	+	20	+	3	N/D	N/D	+	N/D	N/D
MIC 177T	43	3	3	16	T2N1MX	2/23	-	0	-	0	0	1	N/D	N/D	+	N/D	+
MIC 179T	44	3	3	62	T2N1	1/16	-	0	-	0	-	0	N/D	N/D	+	N/D	N/D
MIC 180T	64	3	3	10	T2N1M0	1/23	-	0	-	0	-	0	N/D	N/D	+	N/D	N/D
MIC 183T	40	3	3	20	T1cN1M0	1/24	-	0	-	0	-	0	0	0	+	N/D	-
MIC 184T	41	3	3	34	T2N0M0	0/1	-	0	-	0	-	0	N/D	N/D	+	N/D	+
MIC 185T	52	3	3	32	T2N1M0	1/34	-	0	-	0	-	0	N/D	N/D	+	N/D	+
MIC 186T	24	2	2	18	T4bN2M0	7/28	+	80	+	20	-	1	N/D	N/D	-	N/D	-
MIC 187T	60	3	3	32	T2N0M0	0/20	+	80	-	0	-	1	N/D	N/D	+	N/D	-
MIC 188T	48	3	3	23	T2N1micM0	1/2	-	0	-	0	+	3	N/D	N/D	+	N/D	N/D
MIC 189T	53	3	3	20	T2N2M0	4/43	-	0	-	0	-	0	N/D	N/D	+	N/D	N/D
MIC 190T	30	3	3	11	T3N2	30/36	-	5	-	5	-	1	+	40	+	N/D	-
MIC 191T	48	2	2	2	T2N3M0	14/62	-	0	-	0	-	0	N/D	N/D	+	N/D	N/D
MIC 192T	60	2	2	5	T1bN1M0	0/23	+	90	+	60	-	0	+	30	+	N/D	N/D
MIC 194T	69	2	3	9	T2N1M0	2/24	-	0	-	0	-	0	N/D	N/D	-	N/D	+
MIC 195T	56	1	2	11	T1bN0M0	0/8	+	80	+	70	+	3	N/D	N/D	N/D	N/D	N/D
MIC 197T	83	1	1	2	T2N0M0	0/26	+	80	+	30	-	0	+	10	+	N/D	N/D
MIC 199T	41	2	3	1	T2N0M0	0/19	-	0	+	10	-	0	+	5	-	N/D	N/D
MIC 201T	40	3	3	29	T3N1M0	2/19	-	0	-	0	-	0	N/D	N/D	-	N/D	+
MIC 202T	38	3	3	17	T2N0M0	0/27	+	40	-	0	+	3	N/D	N/D	+	N/D	N/D
MIC 204T	72	3	3	15	T1cN0M0	0/24	+	70	+	90	-	0	+	10	-	N/D	-
MIC 207T	66	1	2	3	T1bN1M0	0/18	+	80	+	70	-	0	+	10	N/D	N/D	N/D
MIC 208T	85	3	3	12	T4N2MX	3/16	-	0	-	0	+	3	N/D	N/D	-	N/D	N/D

Cont/ Anexo 9

ID	Idade	Grau SBR	Grau Nuclear	Índice mitótico	TNM	LN Ac/Ress	ER	Score	PR	Score	HER-2	Score	Ki67	Score	p53	EGFR	CK 5
MIC 209T	42	3	3	13	T2N0M0	0/21	+	70	+	10	-	0	+	50	N/D	N/D	N/D
MIC 211T	42	3	3	40	T2N0MX	0/2	+	50	+	60	+	3	N/D	N/D	N/D	N/D	N/D
MIC 212T	56	3	3	13	T2N3	29	+	70	+	40	-	0	+	30	-	N/D	-
MIC 213T	47	1	2	8	T2N1M0	10/21	+	30	+	40	-	1	+	30	N/D	N/D	N/D
MIC 216T	58	2	2	8	T1cN2MX	6/23	+	90	-	5	-	0	+	40	+	N/D	-
MIC 217T	39	2	3	5	T2N2M0	5/6	+	20	+	70	-	1	+	20	-	N/D	-
MIC 218T	71	3	3	26	T2N0M0	0/28	-	0	-	0	-	0	N/D	N/D	+	N/D	+
MIC 220T	80	2	2	5	T1N3M0	26/40	+	90	-	5	-	1	+	10	+	-	-
MIC 221T	41	3	3	24	T4bN0M0	5/41	+	10	+	70	-	0	+	50	-	N/D	-
MIC 223T	58	1	3	12	T2N1	1/18	-	0	-	0	+	3	+	10	+	-	-
MIC 224T	47	3	3	30	T2N0M0	0/2	-	0	-	0	-	0	+	20	+	N/D	+
MIC 225T	61	2	2	12	T1cN0M0	0/2	+	100	-	0	+	3	+	20	+	-	-
MIC 226T	29	3	3	14	T2N0M0	0/23	+	90	+	90	-	0	+	80	+	-	-
MIC 228T	64	3	3	23	T2N2M0	9/34	-	0	-	0	-	2/FISH-	N/D	N/D	+	N/D	-
MIC 229T	52	3	3	21	T1N1MX	1/15	-	0	-	0	+	2/FISH+	-	0	-	+	-
MIC 230T	48	3	3	7	T2N0M0	0/1	-	0	-	0	-	1	-	5	-	+	+
MIC 231T	58	3	3	25	T2N1M0	0/23	+	70	+	60	-	0	+	15	+	+	-
MIC 233T	49	2	2	5	T1cN2M0	8/22	+	90	+	90	-	0	+	10	N/D	N/D	-
MIC 234T	56	3	3	26	T2N0M0	0/36	-	0	-	0	-	1	+	80	+	-	+
MIC 235T	39	3	3	30	T2N2M0	6/20	-	5	+	30	+	2/FISH+	N/D	N/D	-	N/D	N/D
MIC 236T	69	N/D	N/D	99	T2N1M0	3/18	-	0	-	0	+	3	-	1	-	-	N/D
MIC 237T	71	2	2	6	T2N3M0	25/25	+	80	+	70	-	0	+	30	+	N/D	-
MIC 240T	71	2	3	7	T1cN0M0	0/25	+	95	+	90	-	0	-	5	-	-	-
MIC 261T	72	3	3	11	T1bN0M0	0/11	-	0	-	0	+	3	+	30	+	N/D	N/D
MIC 263T	45	2	2	15	T2N1M0	2/15	+	95	-	0	+	3	+	30	-	-	-
MIC 264T	55	2	3	5	T1bN0M0	1/0	+	95	+	80	+	2/FISH+	+	30	-	-	-

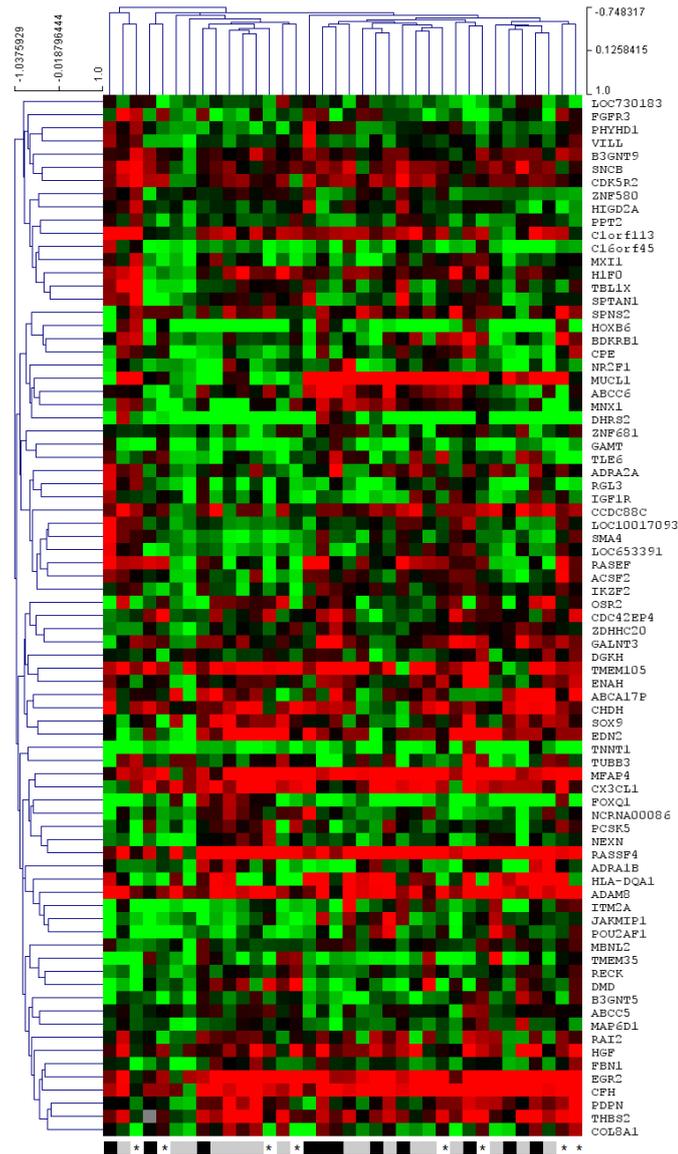
Cont/ Anexo 9

MIC 265T	45	2	3	6	T1cN0M0	1/0	+	80	-	5	+	2/FISH+	-	5	-	-	-
MIC 279T	58	3	3	23	T2N1micM0	1/1	+	95	+	95	-	2/FISH-	+	60	+	-	-
MIC 283T	71	3	3	10	T1N0M0	0/1	+	90	+	40	-	1	+	30	+	-	-
MIC 284T	60	2	3	8	T1cN0M0	0/18	+	100	+	10	-	2/FISH-	-	5	+	-	-
MIC 286T	63	3	3	11		0/24	+	90	+	90	-	1	+	40	+	-	N/D
MIC 287T	54	1	2	6		0/19	+	95	-	1	-	2/FSIH-	+	25	-	N/D	-
MIC 290T	40	3	3	20	T1cN1M0	2/20	+	30	+	50	+	3	+	40	-	-	+
MIC 291T	66	3	3	25	T2N2a M0	4/25	-	0	-	0	+	3	+	70	-	-	+
MIC 292T	40	3	3	22		0/16	-	0	-	0	+	3	+	60	-	-	-
MIC 294T	75	2	3	8	T2cN1MX	3/9	+	100	+	95	-	2/FSIH-	+	25	-	-	-
MIC 295T	74	3	3	17	T1cN1micM0	3/1	+	90	-	5	-	0	+	70	-	-	-
MIC 296T	69	3	3	18	T2N0M	0/6	+	100	+	100	-	0	+	20	-	-	-
MIC 298T	47	2	3	3	T1c N2 M0	2/3	+	70	-	5	-	1	+	20	-	-	-
MIC 300T	47	2	3	6	T1cN0M0	0/3	+	100	+	70	-	0	-	5	-	-	-
MIC 301T	47	1	2	1	T2N1M0	3/25	+	90	+	90	-	1	+	25	-	-	-
MIC 302T	54	2	3	3	T1c N0M0	0/1	+	95	+	90	-	1	-	1	-	-	-
MIC 303T	88	2	3	3	T2NoMx	0/9	+	80	+	70	-	1	+	20	+	-	-
MIC 304T	64	2	3	6	T2 N0 M0	0/6	+	60	-	0	-	0	+	10	-	-	-
MIC 305T	60	2	3	5	T1cN0M0	0/1	+	95	+	95	-	0	+	70	-	-	-
MIC 307T	42	2	3	1	T1N1M0	1/19	+	80	+	60	-	0	+	20	+	-	N/D
MIC 308T	55	2	2	5	T2N0M0	0/1	+	60	+	80	-	0	-	5	-	-	N/D
MIC 309T	39	1	2	4	T2N0M0	0/24	+	95	+	95	-	0	-	5	-	-	-
MIC 312T	39	3	2	3	T1cN0M0	0/1	+	80	+	80	-	0	-	5	-	-	-
MIC 313T	58	2	2	11	T1N1micM0	0/2	+	100	+	30	-	0	+	30	-	-	-
MIC 314T	64	2	3	3	T1cN0M0	0/2	+	90	+	90	-	0	+	20	-	-	-
MIC 315T	44	3	2	8	T3N3M0	40/44	+	60	+	80	+	3	+	30	-	-	-

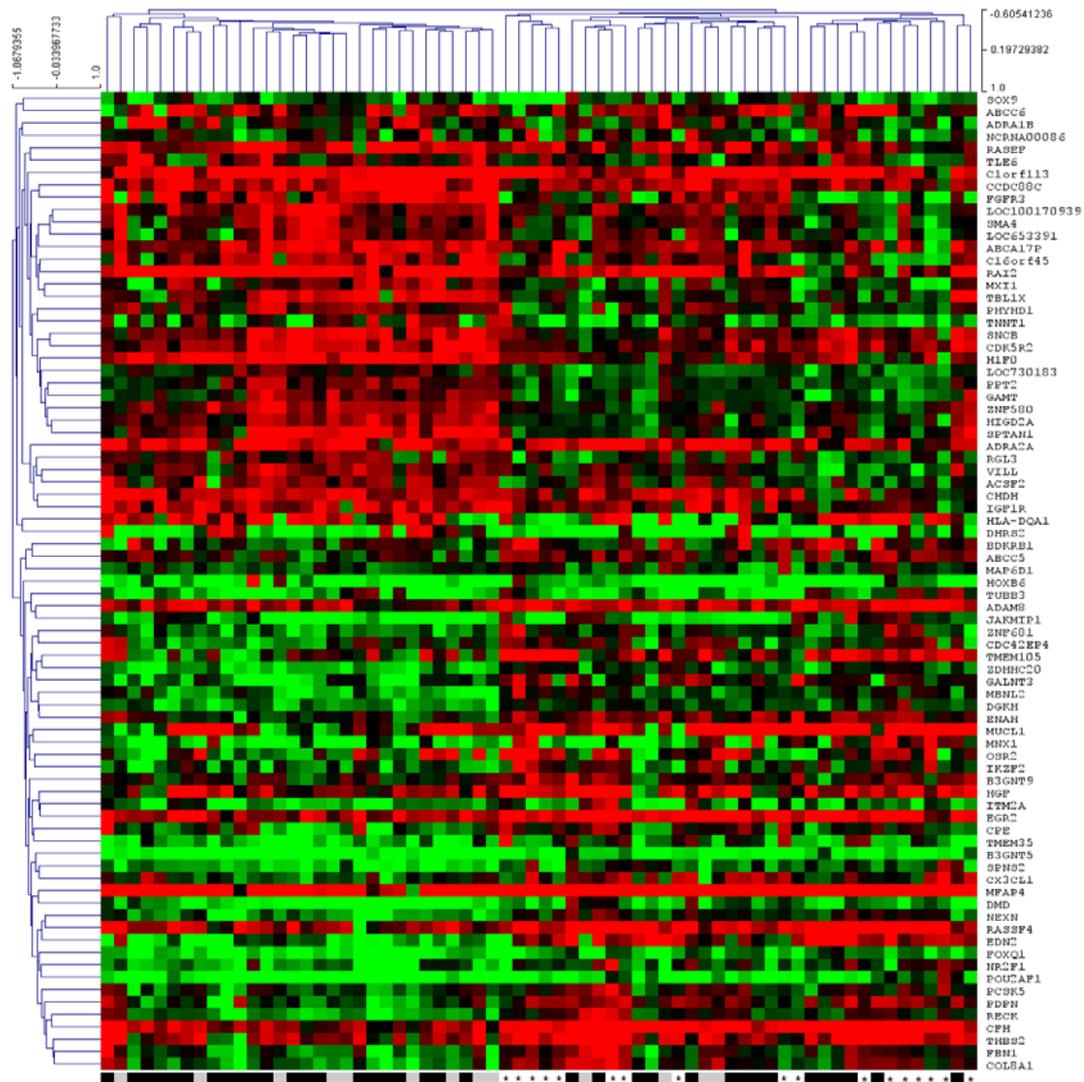
Legenda: Ac: Acometido ; LN: Linfonodo ; N/D: Não determinado ; Ress:Ressecados

Anexo 10 – Clusterização hierárquica de amostras baseadas nos 79 genes diferenciantemente expressos entre os grupos com diferenças no *status* linfonodal

Dois diferentes gráficos foram gerados com a intenção de avaliar separadamente os grupos de amostras luminais (ER e/ou PR positivo) e não luminais (ER e PR negativos)



Cluster Hierárquico de amostras de Carcinoma Ductal Invasivo não Luminais (ER e PR negativos) baseado em um grupo de 79 genes diferencialmente expressos em amostras com e sem linfonodos acometidos. Cluster utilizando a métrica de correlação de Pearson, formado por 36 amostras de carcinoma ductal invasivo de mama negativos para os receptores hormonais ER e PR. Na área inferior do cluster está representado o status linfonodal das amostras, denominados: Não acometido ■ 1-3 Linfonodos acometidos ■ e ≥4 Linfonodos acometidos *.



Cluster Hierárquico de amostras de Carcinoma Ductal Invasivo Luminais (ER e/ou PR positivos) baseado em um grupo de 79 genes diferencialmente expressos em amostras com e sem linfonodos acometidos. Cluster utilizando a métrica de correlação de Pearson, formado por 66 amostras de carcinoma ductal invasivo positivas para receptores hormonais. Na área inferior do cluster está representado o status linfonodal das amostras, denominados: Não acometido ■ 1-3 Linfonodos acometidos ■ e ≥ 4 Linfonodos acometidos *.

Anexo 11 - Avaliação de presença de DNA genômico nas amostras representativas de acometimento linfonodal

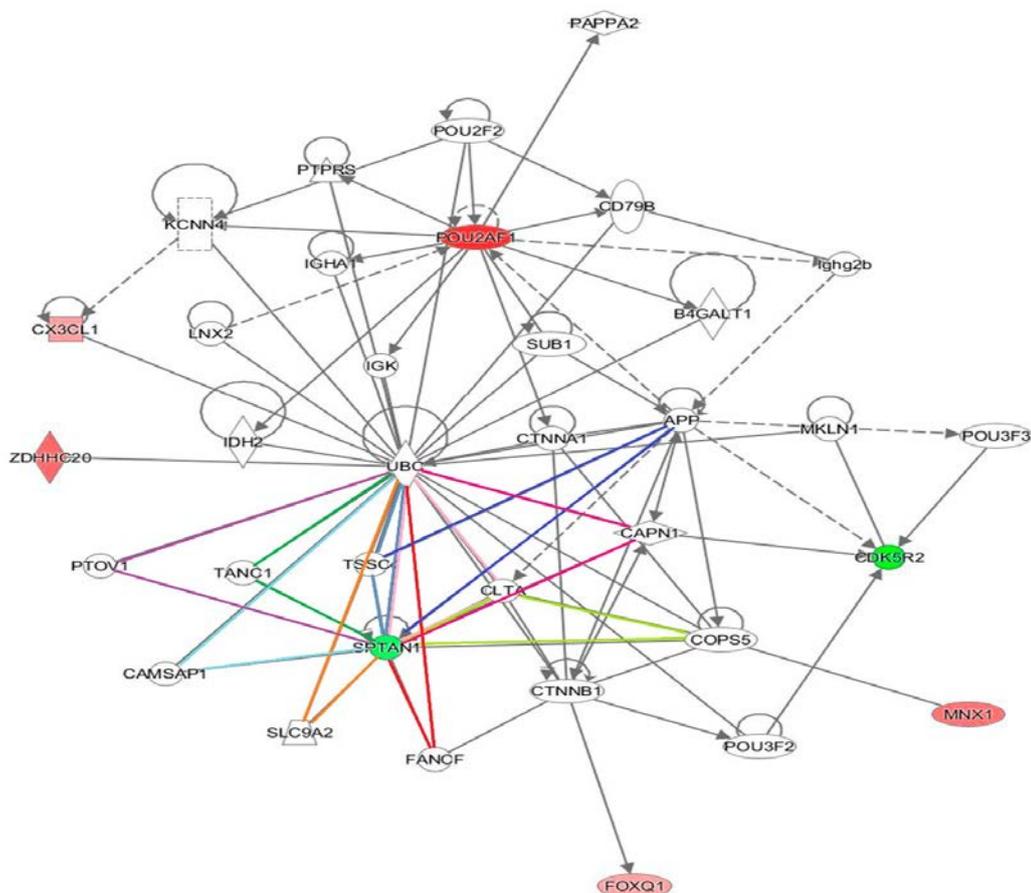
Previa avaliação da expressão diferencial entre os sub- grupos com e sem acometimento linfonodal, as amostras foram avaliadas enquanto a presença de DNA genômico residual, para isso foi incluído o gene HGDC dentro da placa de RT-qPCR (Qiagen), e foi considerado aceitável, segundo as indicações do fabricante, amplificação a partir do ciclo 35 ou à amplificação antes do ciclo 35 com temperatura de melting (tm) $\neq 79^{\circ}\text{C}$, os dados de cada uma das amostras estão disponíveis na tabela a seguir.

Amostras	Classificação	Linfonodos	Ct	Tm
MIC 143T	HR-HER+	Ln 0	37.0	77.6
MIC 145T	HR-HER+	Ln 0	34.1	72.2
MIC 049T	HR+HER+	Ln 0	32.8	71.1
MIC 304T	HR+HER-	Ln 0	32.1	78.3
MIC 292T	HR-HER+	Ln 0	32.6	78.2
MIC 199b	HR-HER-	Ln 0	32.9	80.4
MIC 055	HR+HER-	Ln 0	32.7	71
MIC 296T	HR+HER-	Ln 0	32.7	75.2
MIC 149T	HR-HER+	Ln 1-3	32.9	71.8
MIC 064T	HR+HER-	Ln 1-3	31.3	72.3
MIC 180T	HR-HER-	Ln 1-3	32.7	72.2
MIC 294T	HR+HER-	Ln 1-3	33.7	72
MIC 301T	HR+HER-	Ln 1-3	29.5	77.3
MIC 298T	HR+HER-	Ln 1-3	28.0	77.8
MIC 263	HR+HER+	Ln 1-3	30.2	71.7
MIC 236T	HR-HER+	Ln 1-3	30.6	78.5
MIC 201	HR-HER-	Ln 1-3	36.0	71.6
MIC 175T	HR-HER-	Ln 1-3	33.5	72.1
MIC 177T-A	HR-HER-	Ln 1-3	33.4	72
MIC 290	HR+HER+	Ln 1-3	32.3	78.1
MIC 038T	HR+HER-	Ln 1-3	34.8	77.5
MIC 040	HR+HER-	Ln 1-3	34.2	79.7

Amostras	Classificação	Linfonodos	Ct	Tm
MIC 183	HR-HER-	Ln 1-3	34.2	71.1
MIC 056	HR+HER-	Ln 1-3	36.4	71.4
MIC 120T	HR+HER-	Ln \geq 4	32.7	72.2
MIC 046T	HR-HER+	Ln \geq 4	33.6	72.7
MIC 146T	HR-HER-	Ln \geq 4	32.6	77.7
MIC 073T	HR+HER-	Ln \geq 4	33.0	71.9
MIC 144T	HR-HER-	Ln \geq 4	33.8	71.9
MIC 085T	HR+HER-	Ln \geq 4	32.2	71.8
MIC 315	HR+HER+	Ln \geq 4	36.0	71.6
MIC 186T	HR+HER-	Ln \geq 4	32.6	71.5
MIC 189T	HR-HER-	Ln \geq 4	34.4	72.3
MIC 147T	HR-HER+	Ln \geq 4	33.4	71.9
MIC 221T-A	HR+HER-	Ln \geq 4	40.0	84.2
MIC 212	HR+HER-	Ln \geq 4	33.9	78.5
MIC 220	HR+HER-	Ln \geq 4	36.1	71.6
MIC 235T	HR+HER+	Ln \geq 4	28.3	77.7
MIC 190T	HR+HER-	Ln \geq 4	36.3	71.3
MIC 228T	HR-HER-	Ln \geq 4	33.5	71.3

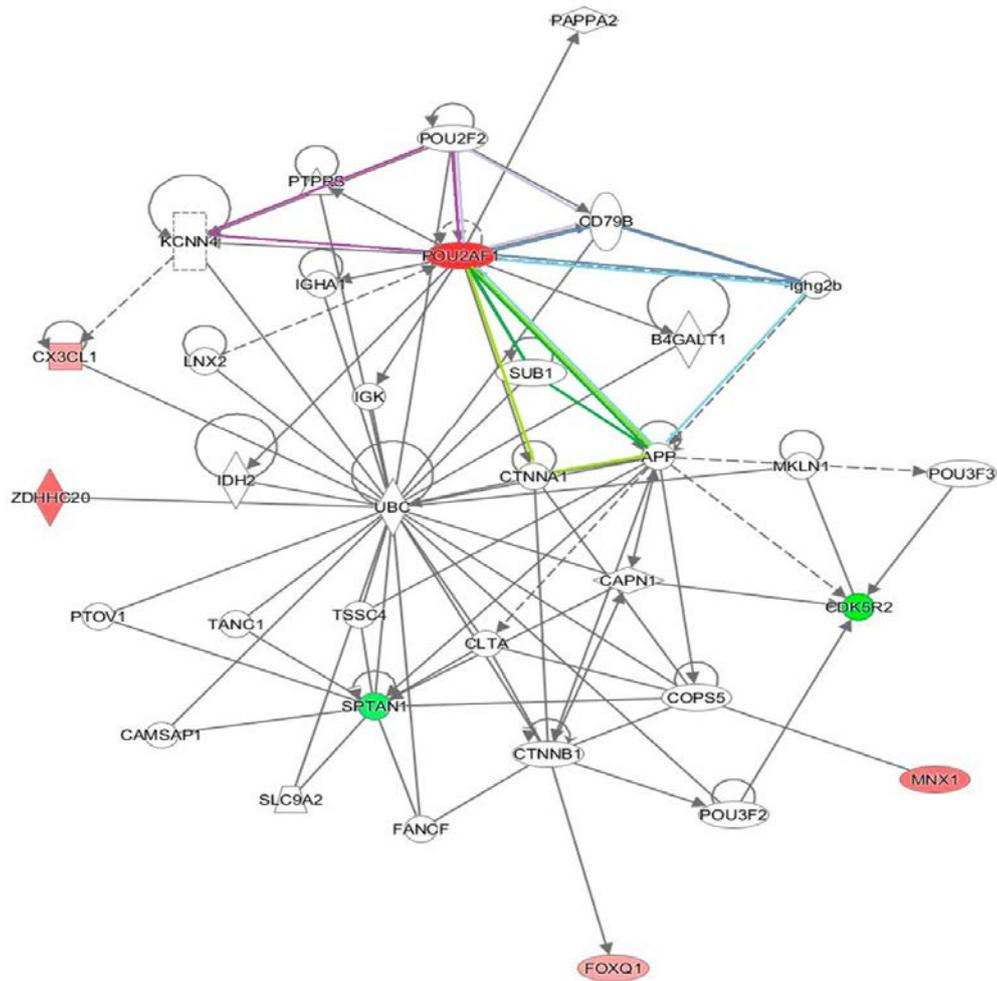
Anexo 12 - Interconexões de triângulo entre as moléculas e os genes validados na rede de interação molecular

A rede de interação molecular gerada no software IPA foi analisada segundo o número de interações entre as moléculas, considerando interessante aquelas nas que participam 3 moléculas. Nas imagens a continuação se destacam as interações em triângulo que incluem algum dos genes validados no estudo.

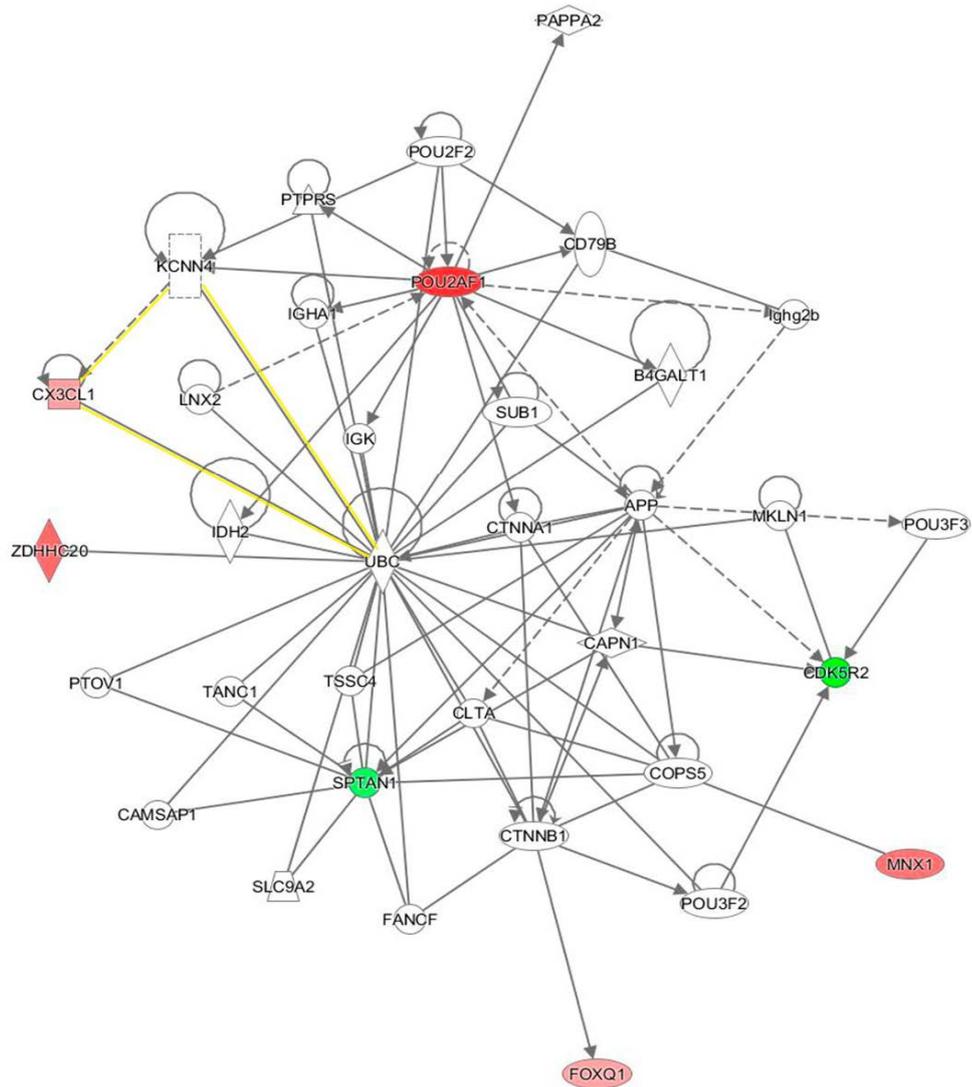


Na Imagem estão representadas as 10 interações em triângulo nas que participa o gene *SPTAN1*.

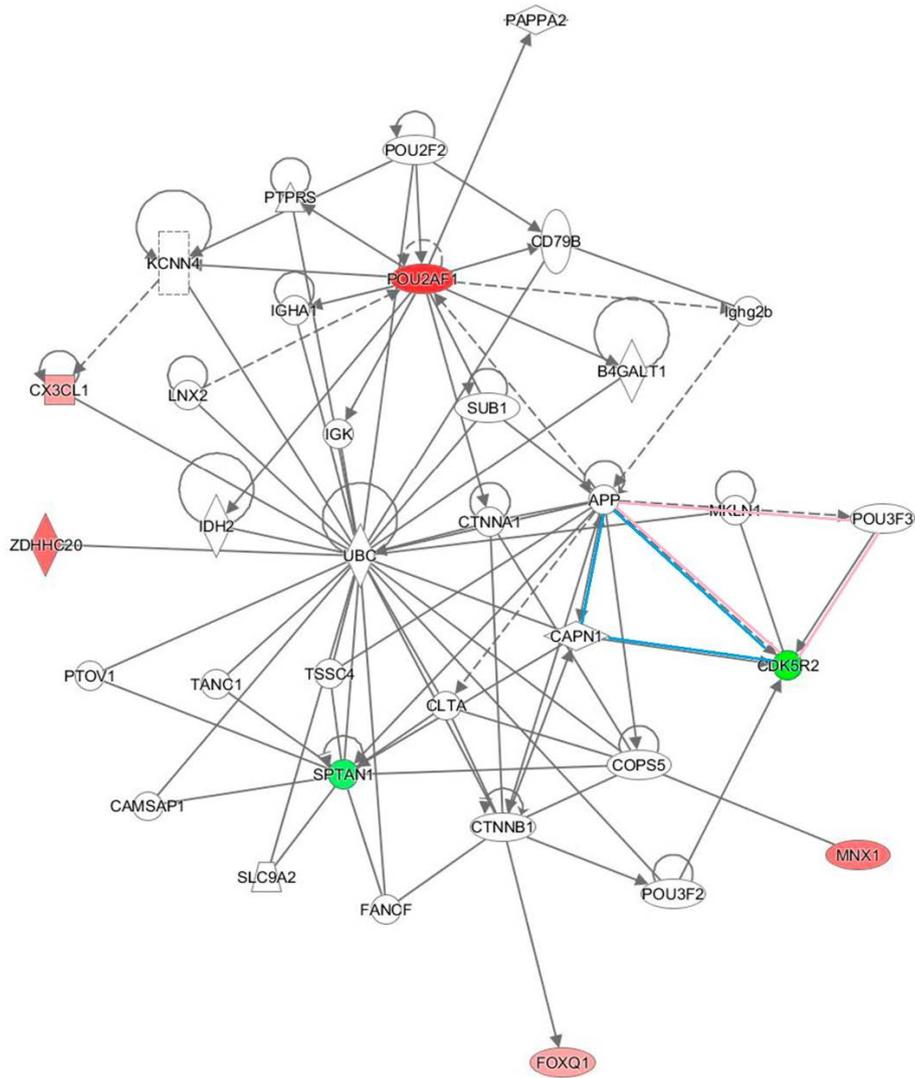
SPTAN1-CAPN1-UBC	SPTAN1- PTOV1-UBC
SPTAN1-CLTA-UBC	SPTAN1- TSSC-UBC
SPTAN1-FANCF-UBC	SPTAN1- TANC1-UBC
SPTAN1-SLC9A2-UBC	SPTAN1-COPS5-CLTA
SPTAN1-CAMSAP1-UBC	SPTAN1-APP-TSSC
POU2AF1-CD79B-POU2F2	POU2AF1- Ighg2b-CD79B
POU2AF1- POU2F2-KCNN4	POU2AF1-APP-SUB1
POU2AF1-CTNNA1-APP	POU2AF1-APP-Ighg2b



Na Imagem estão representadas as 6 interações em triângulo nas que participa o gene *POU2AF1*.



Na Imagem está representada a única interação em triângulo na que participa o gene CX3CL1 (CX3CL1-RCNN4-UBC).



Na Imagem estão representadas as 2 interações em triângulo nas que participa o gene *CDK5R2*

CDK5R2-POU3F3-APP	CDK5R2-APP-CAPN1
-------------------	------------------