

**ANÁLISE DO PERFIL DE METILAÇÃO EM TUMORES DE
MAMA UTILIZANDO *DIFFERENTIAL METHYLATION*
HYBRIDIZATION (DMH)**

ANNA CHRISTINA DE MATOS SALIM

**Tese de doutorado apresentada à Fundação Antônio
Prudente para a obtenção do Título de Doutor em
Ciências**

Área de concentração: Oncologia

Orientadora: Dra. Anamaria Aranha Camargo

Co-orientadora: Dra. Dirce Maria Carraro

SÃO PAULO

2009

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Salim, Anna Christina de Matos

Análise do perfil de metilação em tumores de mama utilizando differential methylation hybridization (DMH) / Anna Christina de Matos Salim – São Paulo, 2009.

110p.

Tese (doutorado) Fundação Antônio Prudente.

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

Orientador: Anamaria Aranha Camargo

Descritores: 1. CÂNCER DA MAMA. 2. METILAÇÃO DO DNA. 3. PROGNÓSTICO. 4. HIBRIDIZAÇÃO DE ÁCIDO NUCLÉICO.

EPÍGRAFE

“A vida é aquilo que acontece enquanto fazemos planos”

John Lennon

DEDICATÓRIA

A minha querida família, Benedito, Cleide, Sérgio, Ariana, Izabella e Giulia pelo amor infinito e incondicional, pelo carinho, compreensão, colinho, sorrisos e lágrimas.....

Ao meu amor Eduardo, por tudo o que vivemos juntos, pela lição de vida, pelo seu amor sem fim e acima de qualquer obstáculo.

Aos queridos agregados, Breno e Isabel, por tudo de bom que vocês trouxeram para a nossa família.

Muito obrigada por vocês fazerem parte da minha vida. AMO VOCÊS!!!!!!

AGRADECIMENTOS

À Dra. Anamaria A. Camargo, pela disponibilidade e paciência em me ensinar boa parte de tudo que aprendi em todos estes anos. Muito obrigada por todo o seu carinho e dedicação tanto no lado profissional quanto pessoal.

A Fabi e a Fel, e que fique bem claro que os nomes estão em ordem alfabética, pelo auxílio nos momentos de dúvidas, por compartilhar suas experiências, pela paciência nas horas difíceis, pelos conselhos pessoais..... Muito obrigada pela amizade e companheirismo.

À Bruna, Cibele, Daniel, Daniela, Érico, Paula, Raphael e Tamara pela amizade, convivência, conversas durante o trabalho e pelos momentos de descontração.

À Paula, Bruna e Renata, por trazerem tanta alegria e harmonia ao LMBG. Muito obrigada!

A todos os colegas do Laboratório de Biologia Molecular e Genômica, Ana Paula, Bruna, Cibele, Daniela, Daniel, Débora, Érico, Fabi, Fabrício, Fel, Júlia, Lílian Inoue, Maria Cristina, Paula, Raphael, Renata, Ricardo, Tamara e Valéria, pelo convívio diário, pelo ambiente de trabalho e por tudo o que aprendi com vocês.

À Adriana, Alex, Aline Pacífico, Lara, Enrique e Vlad pelo convívio e pelas conversas durante esses anos.

À Susan, Marcelo e Brina pela amizade, pelos momentos de alegria e pelos jogos dinâmicos.

Ao João pela amizade e companheirismo durante estes 11 anos de convívio.

Ao Dr. Alex Fiorini pela ajuda com os experimentos de *microarray*.

Ao Dr. Pedro, Dr. Jorge e Patrícia pela imensa ajuda no desenrolar deste projeto.

A Dra. Helena Brentani, Diogo, Luiz Paulo e César pela ajuda com análises deste projeto.

Ao Dr. Michael McClelland e ao Dr. Wang Yipeng, por terem cedido as lâminas de *microarray* utilizadas nesse projeto.

Ao Dr. Humberto Torloni, do Hospital A.C. Camargo, pela ajuda com as lâminas de tumores.

Ao Dr. Fernando Augusto Soares, à Dra. Cynthia Toledo e à Dra. Isabela Cunha médicos do Departamento de Anatomia Patológica do Hospital A.C. Camargo, pela disponibilização e análise histo-patológica das amostras de tumores de mama do Banco de Tumores.

À Dra. Mariângela Corrêa, ao Dr. Daniel Gimenes e ao Dr. Vladimir Cláudio, médicos do Departamento de Oncologia Clínica do Hospital A.C. Camargo, pela ajuda no levantamento dos dados clínico-patológicos das pacientes incluídas nos estudos com amostras de tumores de mama.

À Dra. Dirce Maria Carraro, do Laboratório de Análise de Expressão Gênica, do Hospital A.C. Camargo, pela amizade e colaboração neste trabalho.

Aos colegas do Instituto Ludwig de Pesquisa sobre o Câncer, pela boa convivência no ambiente de trabalho, pelo suporte técnico e administrativo.

À Profa. Dra. Luísa Villa, diretora do Instituto Ludwig de Pesquisa sobre o Câncer, por oferecer as condições necessárias para a realização deste trabalho e pelas críticas e sugestões nos seminários do Instituto.

À Ana Maria Kuninari e Luciana Pitombeira, da Secretaria de Pós-Graduação, pela atenção e dedicação aos alunos.

À Suely Francisco, da biblioteca da Fundação Antônio Prudente, pelo auxílio na revisão e organização desta dissertação. Aos demais funcionários da biblioteca, obrigada pela disponibilidade na aquisição de artigos científicos.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pela avaliação dos relatórios científicos periódicos, pelo suporte financeiro.

À minha segunda família, Leda e Eloá, pelos abraços, pelo carinho, pela atenção, pelo apoio em todos os momentos, por terem como uma filha amada!

As famílias Matos e Salim, pelo amor e amizade, pelos pensamentos positivos e pela torcida incondicional.

Aos meus avós, Alcindo e Lilia, Salim e Murita, pelo carinho e atenção que sempre dedicaram a mim. Muito obrigada pelo exemplo de vida e sabedoria.

Às minhas sobrinhas, Izabella e Giulia, por renovarem a vida e a alegria em nossa família.

Ao meus irmãos, Ari e Gus que, muito mais que irmãos, sempre foram amigos verdadeiros e maravilhosos. Obrigada pela participação de vocês em minha vida e, sobretudo, pelo amor!

Aos meu pais, Benes e Cleidoca, por sempre acreditarem em mim, pelo incentivo em todos os momentos, pelos ensinamentos ao longo da vida, pela amizade e, principalmente, pelo amor incondicional e pelo ensinamento do que é ser uma família!

Ao meu amor Eduardo, por ter olhado atrás da porta da geladeira e assim entrar para a minha vida, pela paciência, pela compreensão e pelos momentos inesquecíveis que foram tão importantes ao longo destes anos! Muito obrigada por tudo!!!

RESUMO

Salim ACM. **Análise do perfil de metilação em tumores de mama utilizando Differential Methylation Hybridization (DMH)**. São Paulo; 2009. [Tese de Doutorado-Fundação Antônio Prudente]

A presença de células tumorais nos linfonodos axilares constitui o fator prognóstico mais importante para pacientes com câncer de mama. No entanto, uma porcentagem das pacientes sem comprometimento de linfonodos morrem em decorrência do desenvolvimento de doença metastática. Uma das áreas mais promissoras na pesquisa sobre o câncer de mama é a identificação de marcadores moleculares capazes de prever a evolução da doença. Alterações no padrão de metilação do DNA podem ser utilizadas de forma eficiente como marcadores moleculares. A proposta deste projeto foi identificar uma assinatura de metilação capaz de discriminar dois grupos de pacientes com carcinoma ductal invasivo de mama linfonodo negativos sendo que um grupo não desenvolveu doença metastática, e o outro desenvolveu metástase à distância em um período mínimo de 5 anos. Para tanto utilizamos a metodologia de hibridização de fragmentos de DNA diferencialmente metilados desenvolvida por WANG et al. A etapa de identificação de fragmentos de DNA diferencialmente metilados entre os dois grupos de amostras foi realizada em 3 etapas utilizando diferentes critérios. Na primeira etapa, identificamos 234 fragmentos de DNA diferencialmente metilados entre os dois grupos de pacientes com $p < 0.01$ e $FDR < 0.05$. Na segunda etapa, o número de fragmentos de DNA diferencialmente metilados foi reduzido para 117 fragmentos com $z \text{ score} = 3.0$ e $p < 0.001$. Na terceira e última etapa, identificamos os 52 fragmentos de DNA com alterações no padrão de metilação mais frequentes em seus respectivos grupos. Todas as 3 assinaturas de metilação quando utilizadas para a clusterização hierárquica das amostras foram capazes de discriminar com 100% de acerto as pacientes que desenvolveram doença metastática das que não desenvolveram metástase à distância. Para a validação da assinatura de metilação contendo 117 fragmentos de DNA diferencialmente metilados utilizamos dados de metilação e dados de expressão

gênica disponíveis em bancos de dados públicos e comerciais. Em conjunto, nossos resultados demonstram que a estratégia utilizada neste trabalho foi eficiente na identificação de uma assinatura de metilação capaz de discriminar os pacientes que desenvolveram doença metastática dos pacientes que não desenvolveram metástase a distância e que 50 dos 117 genes correspondentes aos fragmentos de DNA diferencialmente metilados apresentam evidências de metilação e ou diminuição de expressão já descritas na literatura.

SUMMARY

Salim ACM. **[Global DNA methylation analysis in breast tumors using Differential Methylation Hybridization (DMH)]**. São Paulo; 2009. [Tese de Doutorado-Fundação Antônio Prudente]

The presence of tumors cells in the axillary lymph nodes is the most important prognosis factor for breast cancer patients. However, 30% of patients with negative lymph nodes die because of metastatic disease. One of the most promising areas in breast cancer research is the identification of molecular markers capable of accurately predicting the evolution of the disease. Alterations in DNA methylation patterns can be efficiently used as molecular marker. The proposal of this work was to identify a methylation signature capable to discriminate between two groups of lymph node negative breast invasive ductal carcinoma patients, one that did not develop metastatic disease and other that developed distant metastasis. For that, we used a differentially methylated DNA fragments hybridization strategy developed by WANG et al. The identification of differentially methylated DNA fragments between both groups of patients was done in 3 steps using different criterias. In the first step, we identified 234 differentially methylated DNA fragments between both groups of patients with $p < 0.001$ and $FDR < 0.05$. In the second step, the number of differentially methylated DNA fragments between both groups of patients was reduced to 117 fragments with $z \text{ score} = 3.0$ and $p < 0.001$. In the third and last step, we identified DNA fragments with the most frequent alterations in the metylation pattern in their respective groups. All of these three methylation signatures when used for hierarchical clustering of samples were able to discriminate with 100% accurance the patients that did and did not develop distant metastasis. For methylation signature validation we used methylation and gene expression data available in public and commercial data bases. Together, our results demonstrated that the strategy used in this work was efficient in identify a methylation signature capable to discriminate between patients that developed metastatic disease and patients that did not developed distant metastases and that 50 out of 117 genes corresponding to

differentially methylated DNA fragments present evidences of methylation and or diminished gene expression already described in literature.

LISTA DE FIGURAS

Figura 1 Representação esquemática de um corte de mama.	6
Figura 2 Representação esquemática da progressão tumoral.	7
Figura 3 A metilação do DNA em uma ilha de CpG localizada na região promotora de um gene e o silenciamento gênico.	19
Figura 4 Desenho esquemático das técnicas de hibridização de metilação diferencial.	27
Figura 5 Eletroforese representativa do DNA genômico extraído de amostras tumorais por fenol-clorofórmio em gel de agarose 0,8% corado com Gel <i>Red</i> .	35
Figura 6 Eletroforese representativa de uma reação de amplificação teste em gel de agarose 1% corado com Gel <i>Red</i> .	38
Figura 7 Eletroforese representativa da marcação do DNA <i>Cot1</i> humano com biotina em gel de agarose 1% corado com Gel <i>Red</i> .	40
Figura 8 Eletroforese representativa da reação de amplificação dos amplicons em gel de agarose 1% corado com Gel <i>Red</i> .	41
Figura 9 Interface gráfica disponibilizada pelo programa PubMeth.	48
Figura 10 Interface gráfica disponibilizada pelo programa Oncomine.	53
Figura 11 Análise da qualidade da hibridização de duas amostras tumorais independentes.	60
Figura 12 Normalização intra lâminas por <i>LOESS</i> .	62
Figura 13 Representação gráfica da normalização entre lâminas por <i>LOESS</i> .	63
Figura 14 Clusterização hierárquica das amostras com base no perfil de metilação dos 234 fragmentos de DNA diferencialmente metilados.	67
Figura 15 Clusterização hierárquica das amostras com base no perfil de metilação dos 117 fragmentos de DNA diferencialmente metilados.	68
Figura 16 Clusterização hierárquica das amostras com base no perfil de metilação dos 52 fragmentos de DNA diferencialmente metilados.	69
Figura 17 Clusterização hierárquica dos 90 fragmentos de DNA diferencialmente metilados utilizando as amostras do estudo de expressão de mama ¹⁰⁸ .	79

LISTA DE TABELAS

Tabela 1 Estimativa para o ano de 2008 das taxas brutas de incidência por 100.000 mulheres e do número de casos novos por câncer, em mulheres, segundo localização primária.....	3
Tabela 2 Estadiamento do câncer de mama.....	10
Tabela 3 Aspectos clínico-patológicos das pacientes com carcinoma ductal invasivo de mama do estudo.	33
Tabela 4 Resumo das análises de expressão diferencial.....	51
Tabela 5 Fontes externas integradas ao banco de dados do Oncomine.	52
Tabela 6 Frequência de metilação dos genes em diferentes tumores disponibilizada pelo PubMeth.....	72
Tabela 7 Lista de assinaturas de expressão gênica de mama do Oncomine com evidência de diminuição de expressão de genes da lista 117 genes.	75

LISTA DE ABREVIATURAS

µg	Micrograma
µJ	Microjoule
µL	Microlitro
µM	Micromolar
AC	Drogas quimioterápicas antracíclicas
AJCC	do inglês <i>The American Joint Committee on Cancer</i>
ATP	do inglês <i>adenosine 5'-triphosphate</i>
BLAST	do inglês <i>Basic Local Alignment Search Tool</i>
BLAT	do inglês <i>BLAST-Like Alignment Tool</i>
BSA	do inglês <i>Bovine Serum Albumin</i>
cm	Centímetro
cm²	Centímetro quadrado
CMF	Drogas quimioterápicas ciclofosfamida, metotrexato e 5-fluoracil
CpG	do inglês <i>Cytosine phosphate Guanine</i>
Cpm	Contagem por minuto
DMH	do inglês <i>Differential Methylation Hybridization</i>
DMSO	Dimetilsulfóxido
DNA	Ácido desossorribonucléico
DNMTs	DNA metiltransferases
dNTP	Deoxinucleotídeos fosfato
DO	Densidade óptica
EDTA	Ácido etilenodiaminotetracético
EMBL	<i>European Molecular Biology Laboratory</i>
HCl	Ácido clorídrico
HE	Coloração hematoxilina-eosina
INCA	Instituto Nacional do Câncer
IPTG	β-D-tiogalactopiranosídeo de isopropila
Kb	kilobase
LOESS	do inglês <i>Locally Weighted Scatterplot Smoothing</i>

LoTE	do inglês <i>low Tris-EDTA</i>
M	Molar
MBP	do inglês <i>Methylcytosine Binding Protein</i>
MgCl₂	Cloreto de Magnésio
ml	Mililitro
mM	Milimolar
NaCl	Cloreto de Sódio
NaOAc	Acetato de Sódio
NaOH	Hidróxido de Sódio
NEB	do inglês <i>New England Biolabs</i>
ng	Nanograma
nm	Nanômetro
nM	Nanomolar
OMS	Organização Mundial da Saúde
pb	Pares de base
PCR	do inglês <i>Polimerase Chain Reaction</i>
pH	Potencial hidrogeniônico
pT	Tamanho do tumor definido após medição da peça cirúrgica
RNA	do inglês <i>acid ribonucleic</i>
rpm	do inglês <i>rotation per minute</i>
RT-PCR	do inglês <i>Reverse Transcriptase Polimerase Chain Reaction</i>
SDS	Dodecil sulfato de sódio
SSC	do inglês <i>Saline Sodium Citrate</i>
TNM	do inglês <i>Tumor Node Metastasis</i>
Tris	Tris (hidroximetil) amino metano
UCSC	<i>University of California Santa Cruz</i>
UICC	<i>Union Internationale Contre le Cancer</i>
UV	Ultravioleta
X-gal	do inglês <i>5-bromo-4-chloro-3-indolyl-β-D-galactosidase</i> .

ÍNDICE

1	INTRODUÇÃO	2
1.1	O câncer de mama	2
1.1.1	Aspectos epidemiológicos, clínicos e histopatológicos do câncer de mama	2
1.1.2	Diagnóstico e tratamento do câncer de mama	8
1.1.3	Fatores prognósticos	11
1.1.4	Marcadores moleculares de prognóstico em câncer de mama	12
1.2	A metilação do DNA como mecanismo epigenético de regulação da transcrição gênica	16
1.3	Metilação do DNA e câncer	20
1.4	Metodologias em larga escala utilizadas na detecção de regiões diferencialmente metiladas	22
2	OBJETIVOS	29
2.1	Objetivo geral	29
2.2	Objetivos específicos	29
3	MATERIAL E MÉTODOS	31
3.1	Tumores primários de mama	31
3.2	Macrodissecção dos tumores primários de mama	34
3.3	Extração de DNA genômico	34
3.4	Hibridização de fragmentos de DNA diferencialmente metilados	36
3.4.1	<i>Microarray</i> enriquecido para seqüências promotoras	36
3.4.2	Amplificação dos fragmentos diferencialmente metilados e condições de hibridização	37
3.4.3	Digitalização das imagens e quantificação do sinal de hibridização	43
3.4.4	Análise da qualidade das hibridizações e normalização	44
3.5	Identificação de fragmentos de DNA diferencialmente metilados	45
3.6	Clusterização hierárquica	46

3.7	Validação dos fragmentos de DNA diferencialmente metilados utilizando o Pubmeth e o Oncomine	46
3.8	Validação da assinatura de metilação utilizando o dado de expressão do estudo desenvolvido por VAN'T VEER et al.	54
4	RESULTADOS	57
4.1	Caracterização do perfil de metilação de amostras tumorais de mama	57
4.1.1	Hibridização de fragmentos diferencialmente metilados	57
4.1.2	Análise de qualidade das hibridizações e normalização dos sinais de hibridização	58
4.2	Identificação de fragmentos de DNA diferencialmente metilados	64
4.3	Validação dos fragmentos de DNA diferencialmente metilados utilizando informações disponíveis em banco de dados públicos.	70
4.3.1	Validação dos fragmentos de DNA diferencialmente metilados utilizando dados de metilação diferencial do PubMeth	70
4.3.2	Validação dos fragmentos de DNA diferencialmente metilados utilizando dados de expressão gênica do Oncomine.	73
4.3.3	Validação da assinatura de metilação utilizando o dado de expressão do estudo desenvolvido por VAN'T VEER et al. ¹⁰⁸	76
5	DISCUSSÃO	81
6	CONCLUSÕES	96
7	REFERÊNCIA BIBLIOGRÁFICA	98

ANEXOS

Anexo 1 Assinatura de 234 genes diferencialmente metilados entre os dois grupos de amostras.

Anexo 2 Assinatura de 117 genes diferencialmente metilados entre os dois grupos de amostras.

Anexo 3 Assinatura de 52 genes diferencialmente metilados entre os dois grupos de amostras.

INTRODUÇÃO

1 INTRODUÇÃO

1.1 O CÂNCER DE MAMA

1.1.1 Aspectos epidemiológicos, clínicos e histopatológicos do câncer de mama

O câncer de mama é uma doença cada vez mais comum em todo o mundo e representa, nos países ocidentais, a segunda causa de morte entre as mulheres. Estima-se que o câncer de mama seja responsável por um quinto dos óbitos de mulheres na faixa etária entre 40-50 anos⁸⁵. Nos Estados Unidos, estima-se a identificação de 1.437.180 novos casos de câncer invasivo em 2008, sendo que 182.460 correspondem a tumores invasivos de mama o que representa 26% dos novos casos de tumores entre as mulheres. Além disto, são estimados aproximadamente 40.480 óbitos decorrentes de câncer de mama nos EUA, o correspondente a 15% de todos os óbitos por câncer e o segundo tipo de câncer que mais provoca mortes, atrás apenas do câncer de pulmão. Entre as mulheres americanas, a probabilidade de desenvolver câncer de mama em algum momento da vida é de 1 em cada 8 mulheres (13%) e o risco de óbito decorrente deste tipo de tumor é de 1 em cada 33 (3%) mulheres⁵⁵.

Segundo dados do Inca, dos 466.730 novos casos de câncer com previsão de serem diagnosticados em 2008 no Brasil, o câncer de mama será o segundo de maior incidência entre os indivíduos do sexo feminino, sendo responsável por 49.400 novos casos da doença, com um risco estimado de 51 casos a cada 100 mil mulheres⁷⁶ (Tabela 1). No Brasil, o câncer de mama é o câncer que mais causa mortes entre as

mulheres e é a terceira causa de morte entre os indivíduos do sexo feminino. Dados consolidados de 2005 registraram 10.208 óbitos de mulheres brasileiras decorrentes do câncer de mama⁷⁷.

Tabela 1 Estimativa para o ano de 2008 das taxas brutas de incidência por 100.000 mulheres e do número de casos novos por câncer, em mulheres, segundo localização primária.

Localização primária Neoplasia maligna	Estimativa de casos novos	Taxa bruta
Mama feminina	49.400	51
Colo do Útero	18.680	19
Cólon e reto	14.500	15
Traquéia, Brônquio e pulmão	9.460	10
Estômago	7.720	8
Leucemias	4.320	4.4
Cavidade oral	3.780	4
Pele melanoma	2.970	3
Esôfago	2.650	2.7
Outras localizações	62.270	64
Subtotal	175.750	180
Pele não melanoma	59.120	60
Total	234.870	241

Fonte: Modificado do Instituto Nacional do Câncer⁷⁷.

O câncer de mama é relativamente raro antes dos 35 anos de idade, mas acima desta faixa etária sua incidência cresce rápida e progressivamente alcançando seu pico na faixa de 65 a 70 anos. Nas últimas décadas, o crescimento contínuo na incidência do câncer de mama foi provocado pela melhora do nível sócio-econômico e conseqüentemente maior acessibilidade aos serviços de saúde, permitindo um diagnóstico mais eficiente da doença. Segundo a Organização Mundial da Saúde (OMS), nas décadas de 60 e 70 registrou-se um aumento de 10 vezes em suas taxas de incidência, ajustada por idade, nos registros de câncer de diversos continentes⁷⁷. Nos Estados Unidos, a partir da década de 80 foi observado um aumento de 40% na taxa de incidência do câncer de mama devido à implementação da mamografia como

exame de rotina no sistema de saúde, o que permitiu a detecção de doença localizada e tumores menores que 2 cm. No final da década de 90, houve diminuição na taxa de incidência devido à saturação na varredura através da mamografia. Além disto, observou-se também diminuição da taxa em mulheres americanas entre 50 e 69 anos devido à dramática redução no uso de terapia de reposição hormonal a partir de 2002⁵⁶. Por outro lado, houve uma redução de 2,3% ao ano na taxa de mortalidade em decorrência da doença devido ao avanço dos métodos de diagnóstico e a melhora efetiva dos tratamentos oferecidos²⁵.

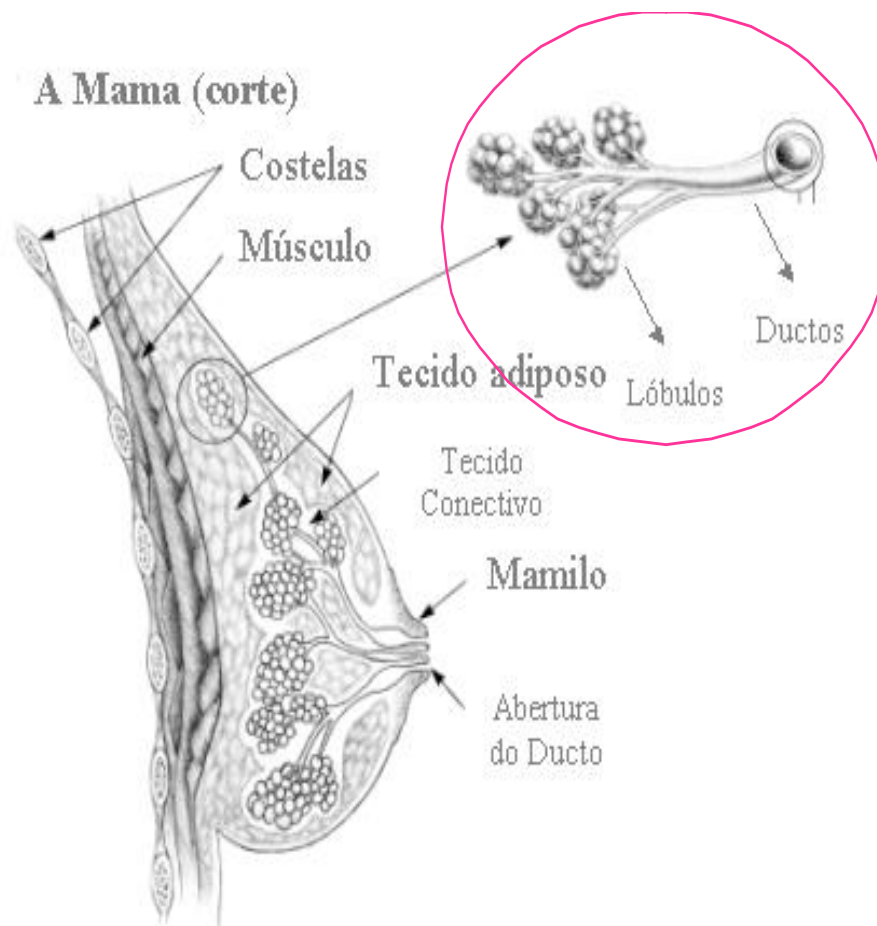
O diagnóstico precoce do câncer de mama é fundamental para a eficácia do tratamento e a conseqüente cura da doença. Quando o câncer de mama é detectado em estádios iniciais (I e II) a taxa de sobrevida livre de doença em 5 anos para pacientes com doença local e sem comprometimento de linfonodos axilares é superior a 95% e, no caso de comprometimento dos linfonodos axilares, diminui para 83%. Entretanto, quando a doença metastática é constatada ao diagnóstico a taxa de sobrevida em cinco anos diminui drasticamente, girando em torno de 26%⁵⁶.

A metástase representa a principal causa de morte para pacientes com câncer de mama. Aproximadamente 10 a 15% das pacientes com tumor de mama apresentam doença agressiva e desenvolvem metástase em até 3 anos após o diagnóstico inicial da doença. Entretanto, não é incomum o aparecimento de metástases em sítios distantes 10 ou mais anos após a identificação do tumor primário¹⁰¹. Os sítios mais comuns de formação de metástases são ossos, pulmão e fígado⁶⁶.

A etiologia do câncer de mama ainda não é totalmente conhecida sendo que os fatores de risco identificados explicam somente uma pequena parte dos casos.

Baseado em estudos epidemiológicos conduzidos em diferentes populações, os fatores de risco já bem estabelecidos incluem idade avançada, localização geográfica (risco aumentado nos EUA e países do nordeste europeu, risco diminuído na África e América do Sul), eventos reprodutivos (menarca precoce antes de 12 anos, menopausa tardia após os 55 anos, nuliparidade, gravidez após os 30 anos, maior tempo de amamentação), hormônios exógenos (uso de contraceptivo oral e terapia de reposição hormonal), estilo de vida (consumo de álcool, obesidade e sedentarismo), alta densidade da mama, história de doença de mama benigna, exposição à radiação ionizante quando crianças, altos níveis de IGF1 e prolactina, e fatores genéticos (mutação nos genes BRCA1 e BRCA2)²⁵.

O câncer de mama é uma doença bastante heterogênea, apresentando uma grande variabilidade clínica e histopatológica. A apresentação clínica do câncer de mama pode variar desde formas não invasivas até outras altamente metastáticas. O tipo histológico mais freqüente dentre os tumores malignos da mama é o carcinoma que tem origem a partir das células epiteliais que revestem os ductos e os lóbulos da glândula mamária, mais especificamente das células epiteliais luminais (Figura 1)^{45,81}.

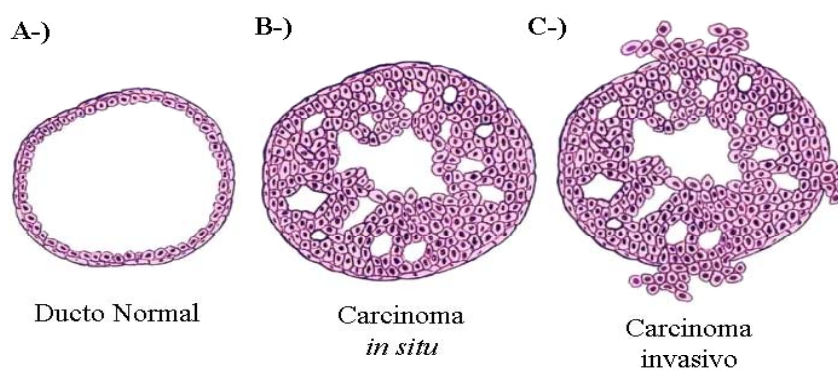


Legenda: Em destaque estão representados os lóbulos e ductos da glândula mamária.

Fonte: Modificado de www.cancervic.org.au/about-cancer/cancer_types/breast_cancer

Figura 1 Representação esquemática de um corte de mama.

Os carcinomas podem se apresentar na forma *in situ*, quando o tumor permanece confinado ao tecido mamário, ou invasiva, quando rompem a membrana basal e invadem tecidos adjacentes. Os carcinomas invasivos representam de 70 a 80% de todos os tumores de mama identificados, sendo que destes cerca de 80% correspondem a carcinomas ductais invasivos, 10% a carcinomas lobulares invasivos, sendo a agressividade destes dois tipos de carcinomas semelhante. O restante dos carcinomas correspondem a outros subtipos, como carcinomas tubular, mucinoso, medular e papilar que correspondem a menos de 10% de todos os casos. A doença de Paget é um tipo raro de carcinoma *in situ* que se inicia nos ductos dos mamilos e acomete a epiderme (Figura 2)¹⁰⁶.



Legenda: Ducto normal (A); Carcinoma ductal *in situ* (B) e Carcinoma ductal invasivo (C).

Fonte: Modificado de www.breastcancer.org/is_cancer_invasive.html

Figura 2 Representação esquemática da progressão tumoral.

1.1.2 Diagnóstico e tratamento do câncer de mama

Entre os principais métodos de detecção do câncer de mama estão o auto-exame, o exame clínico palpatório e a mamografia. A mamografia utiliza radiação ionizante de baixa potência, o que pode ser prejudicial ao paciente. Mesmo assim, seus benefícios superam os riscos. Atualmente, a ultra-sonografia e a ressonância magnética também têm sido empregadas como métodos de diagnóstico complementares à mamografia, indicados para pacientes mais jovens (abaixo de 35 anos) que apresentam tecido mamário denso ou que possuem história familiar da doença. A sensibilidade destes métodos varia entre 46% e 88%, e sua especificidade varia entre 82% e 99%, e são influenciadas por fatores como tamanho e localização da lesão, densidade do tecido mamário, qualidade dos recursos técnicos e habilidade e interpretação dos profissionais^{18,56}. Desta forma, estas ferramentas quando combinadas e somadas aos exames cito e histopatológicos da biópsia, conferem uma acuidade de até 95% no diagnóstico da doença⁶ e possibilitam a detecção de carcinomas menores que 1 cm, porém o custo é bastante elevado¹⁰³. Em vista disto, no Brasil, 50-70% dos casos são diagnosticados tardiamente, o que acaba dificultando enormemente o tratamento e a cura da doença⁴¹.

Uma vez confirmada a ocorrência do carcinoma de mama através de biópsia e avaliação anatomo-patológico, exames complementares são realizados para se determinar o estadiamento clínico da doença o qual, por sua vez, definirá a conduta terapêutica a ser seguida. Exames complementares incluem o raio-X de tórax, a cintilografia óssea e a tomografia computadorizada e são geralmente realizados para avaliar a presença de doença metastática¹¹¹.

O estadiamento do câncer de mama é realizado através do sistema de classificação TNM estabelecido em 1987 pelo *Union Internationale Contre le Cancer* (UICC) e pelo *The American Joint Committee on Cancer* (AJCC). Esse sistema se baseia no tamanho do tumor (T), no acometimento de linfonodos (N) e na presença de metástases à distância (M) (Tabela 2)^{98,103}.

Diversos tratamentos têm sido adotados no combate ao câncer de mama e se baseiam em terapias locais, sistêmicas e biológicas. O principal tratamento é a ressecção cirúrgica do tumor que pode ou não ser seguida de terapia adjuvante. A cirurgia conservadora com esvaziamento dos linfonodos e a mastectomia radical são as modalidades de ressecção mais amplamente utilizadas. Hoje em dia a terapia adjuvante é adotada em pacientes com tumores de diferentes tamanhos, incluindo os menores que 1 cm e que apresentam metástases nos linfonodos axilares¹¹¹. Três modalidades de terapia adjuvante são comumente utilizadas: i) quimioterapia (taxanos e antracíclicos), ii) radioterapia e iii) terapia hormonal (tamoxifeno e inibidores de aromatase). Individualmente, estes quimioterápicos produzem resposta efetiva em 20 a 80% dos pacientes com câncer de mama metastático. Porém, respostas completas são raras e menos de 20% dos pacientes as mantêm por mais de cinco anos. Recentemente, a terapia biológica utilizando o anticorpo monoclonal Trastuzumabe (Herceptina®), que reconhece o domínio extracelular da proteína ERBB2, produto do oncogene ERBB2, tem sido indicada para pacientes com tumores de mama que apresentam superexpressão do gene ERBB2, o que é verificado em aproximadamente 30% dos casos. Estes tumores geralmente são muito agressivos e estão associados a altas taxas de recidivas^{96,97}.

Tabela 2 Estadiamento do câncer de mama.

Estádio	Tamanho	Linfonodos	Metástase
0	Tis	N0	M0
I	T1	N0	M0
IIa	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
IIb	T2	N1	M0
	T3	N0	M0
	T4	N2	M0
IIIa	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1/N2	M0
IIIb	T4	Qualquer N	M0
	Qualquer T	N3	M0
IV	Qualquer T	Qualquer N	M1

Legenda: T0 – sem evidência de tumor primário, Tis – carcinoma *in situ*, T1 – tumor com 2 cm ou menos em sua maior dimensão, T2 – tumor com mais de 2 cm e até 5 cm em sua maior extensão, T3 – tumor com mais de 5 cm em sua maior extensão, T4 – tumor de qualquer tamanho com extensão direta à parede torácica ou à pele; N0 – ausência de metástase nos linfonodos regionais, N1 – metástase em 1 a 3 linfonodo (s) axilar (es) homolateral (is) móvel (is), N2 - metástase em 4 a 9 linfonodos axilares homolaterais fixos uns aos outros ou a outras estruturas, N3 - metástase em mais de 10 linfonodos da cadeia mamária interna homolateral incluindo metástase em linfonodos infraclaviculares e supraclaviculares ou nódulos linfáticos mamário; M0 – ausência de metástase à distância e M1- metástase à distância (incluindo metástase em linfonodos supraclaviculares).

1.1.3 Fatores prognósticos

A presença de metástases em linfonodos axilares e o tamanho do tumor são os dois fatores prognósticos mais importantes no câncer de mama e determinam o uso ou não de terapia adjuvante. O grau de comprometimento linfonodal constitui o mais importante fator prognóstico para doença metastática, devendo virtualmente toda mulher com comprometimento linfonodal axilar ser submetida a tratamento adjuvante.

No entanto, uma porcentagem das pacientes que não apresentam comprometimento linfonodal morrem em decorrência da doença¹⁰³. Essas pacientes teoricamente se beneficiariam da administração de terapia adjuvante, contudo não é possível, até o momento, identificar tais pacientes. Por outro lado, a adoção de terapia adjuvante em todas as pacientes submetidas à ressecção cirúrgica é inviável devido à alta toxicidade e ao elevado custo do tratamento^{107,111}. Estima-se também que cerca de 30 a 40% das pacientes com tumor de mama e que possuem linfonodos acometidos por metástases permaneceriam livres de recidiva da doença em 10 anos e não necessitariam receber terapia adjuvante⁴⁷.

Para pacientes que não apresentam comprometimento linfonodal (linfonodo negativos) é necessário recorrer a outros critérios para indicação do tratamento adjuvante. CARTER et al.¹⁶ em estudo realizado com 13.464 pacientes linfonodo negativos detectaram sobrevida global em 5 anos de 99% para tumores menores de 1 cm comparados a 89% para tumores de 1 a 3 cm e 86% para tumores entre 3 e 5 cm. Para pacientes linfonodo negativos o tamanho do tumor é o mais poderoso fator prognóstico e é utilizado rotineiramente em decisões terapêuticas devendo as pacientes com tumores maiores que 1 a 2 cm serem tratadas com terapia adjuvante.

Além destes, outros fatores patológicos e moleculares também são empregados no prognóstico, como o grau histológico, a presença de invasão linfática e/ou vascular, a expressão dos receptores hormonais de estrogênio e progesterona (RE e RP, respectivamente) e a superexpressão do oncogene ERBB2.

Assim sendo, uma das áreas mais promissoras na pesquisa sobre o câncer de mama é a identificação de marcadores moleculares que possam ser utilizados na avaliação do prognóstico das pacientes¹⁰⁰.

1.1.4 Marcadores moleculares de prognóstico em câncer de mama

Muitos estudos têm procurado identificar alterações no padrão de expressão de grupos de genes no tumor primário em estádios iniciais capazes de predizer a evolução da doença através da metodologia de cDNA *microarray*^{28,75,108,109,114}. Em um estudo caso-controle realizado por VAN'T VEER et al.¹⁰⁸ no *Netherlands Cancer Institute* uma assinatura gênica contendo 70 genes foi identificada utilizando tecido tumoral congelado de 78 pacientes jovens com câncer de mama em estágio inicial, sem comprometimento de linfonodos e que desenvolveram ou não metástase em um seguimento de pelo menos 5 anos. Neste estudo, os tumores das pacientes que desenvolveram metástase à distância rapidamente tinham um perfil de expressão distinto dos tumores de pacientes que permaneceram livres da doença. Este perfil de expressão gênica foi então aplicado a um segundo grupo de validação contendo 295 amostras de tumores coletados de pacientes jovens com ou sem comprometimento de linfonodos, incluindo 61 pacientes do primeiro estudo, no qual observaram resultados semelhantes¹⁰⁹. Esta assinatura gênica de pior, ou melhor, prognóstico, que se revelou significativamente associada à sobrevida livre de metástase em 5 e 10

anos destas pacientes, foi capaz de predizer com melhor acurácia as pacientes com maior chance de desenvolver metástase quando comparado as variáveis clínicas tais como, idade, *status* do linfonodo, tamanho do tumor, invasão vascular, expressão de receptor de estrógeno, além do tipo de cirurgia e tratamento.

Subseqüentemente, WANG et al.¹¹⁴ em 2005, utilizando uma plataforma de *array* diferente, avaliaram uma população de 286 pacientes com câncer de mama linfonodo negativas que não foram submetidas à terapia neo-adjuvante e adjuvante sistêmica. Neste estudo, primeiramente as pacientes foram agrupadas de acordo com o *status* de receptor de estrógeno e cada subgrupo foi analisado independentemente. Finalmente, foi identificada uma assinatura gênica composta por 76 genes (60 no grupo positivo e 16 no grupo negativo para estrógeno) capaz de predizer metástase à distância independente da idade do paciente, do tamanho e grau do tumor.

No entanto, apesar dos resultados destes estudos sugerirem que perfis de expressão gênica possam fornecer uma poderosa ferramenta prognóstica, os estudos têm falhado em identificar marcadores moleculares tumorais que possam ser reproduzidos em grupos de pacientes independentes e, portanto, sua aplicação clínica é incerta. Além disto, a inconsistência das assinaturas gênicas entre os diferentes estudos tem sido evidenciada em análises independentes subseqüentes que mostram, por exemplo, que a assinatura contendo os 70 genes, identificada por VAN'T VEER et al.¹⁰⁸, não era a única assinatura possível dentro do banco de dados do estudo e que foi fortemente influenciada pelo grupo de pacientes utilizados na seleção dos genes da assinatura^{28,75}. Estes autores observaram que muitas outras listas com valor igualmente preditivo poderiam ser produzidas pela mesma análise e que 3 principais propriedades dos dados poderiam explicar esta sensibilidade: i) muitos genes têm

correlação com a sobrevida, ii) as diferenças entre estas correlações são pequenas e iii) as correlações flutuam fortemente quando medidas em diferentes subgrupos de pacientes do estudo.

Apesar dessas limitações, vários testes multigênicos já estão sendo comercializados. Dois destes investigam a expressão de um painel específico de genes pela medida dos níveis de RNA em tumores de mama utilizando diferentes metodologias como o RT-PCR em tempo real e cDNA *microarray*. O *oncotype DX*TM (Genomic Health, Redwood City, Ca) é um ensaio que quantifica a expressão de 21 genes por RT-PCR em tempo real⁸². Este teste pretende predizer a probabilidade de desenvolvimento de metástases em mulheres de todas as idades, com câncer de mama em estádios iniciais, linfonodo negativas e RE positivas, tratadas com tamoxifeno. O outro teste, *Mammaprint*[®], é o primeiro ensaio multigênico comercial baseado na metodologia de *microarray* e utiliza o perfil de expressão de 70 genes desenvolvido por VAN'T VEER et al.¹⁰⁸ sendo comercializado pela Agendia (Amsterdam, the Netherlands). Este é um teste prognóstico para mulheres com idade igual ou inferior a 61 anos, com câncer de mama em estádios iniciais, linfonodo negativas e RE negativas ou positivas aprovado pelo *Food and Drug Administration* (FDA). Os dois testes possuem protocolos definidos para avaliar o conteúdo da amostra tumoral a ser analisada, para preparar o RNA das amostras e para normalizar o dado bruto da medida de expressão, a fim de produzir um número associado ao risco de desenvolver metástases. Como as decisões terapêuticas são baseadas em estimativas de risco, testes que melhoram tais estimativas têm grande possibilidade de alterar a evolução clínica de pacientes com câncer de mama no que se refere a evitar tratamentos adjuvantes desnecessários ou

oferecê-los às pacientes que se beneficiarão ao usá-los, diminuindo assim o risco de metástase.

Entretanto, ainda não há na literatura um estudo que apresente o uso destes testes como intervenção primária que direcione a conduta terapêutica através do resultado do teste e avalie a evolução do paciente levando em consideração morbidade, mortalidade e ou qualidade de vida. Embora algumas evidências suportadas pelos testes sejam provenientes de testes randomizados, o que os confere alguma confiabilidade, é fato que os estudos de validação necessitam lidar com populações para as quais as implicações dos vários grupos de risco em relação às decisões tomadas estejam claras. Para todos os testes, exceto o *oncotype DX*, tanto o estudo de desenvolvimento quanto o de validação foram realizados em populações heterogêneas, sem número de amostras suficiente para a estratificação de grupos homogêneos para guiar a conduta clínica. Além disto, as amostras usadas na validação são freqüentemente reutilizadas em diferentes estudos. Em vista disto, dois estudos clínicos, o *TAILORx* e o *MINDACT*, que estão em andamento poderão atribuir maiores evidências aos testes *oncotype DX*TM e *MammaPrint*, respectivamente, permitindo quase uma inferência direta sobre o impacto na evolução da doença.

Apesar do avanço na identificação de genes diferencialmente expressos a avaliação da expressão de transcritos específicos através de RT-PCR em tempo real e *microarray* apresenta vários desafios que podem comprometer a reprodutibilidade e a confiança das medidas obtidas sendo de difícil aplicação na prática clínica. Dada a complexidade dos procedimentos usados neste tipo de investigação, as fontes de variabilidade são múltiplas, do tipo de material biológico utilizado (RNA) até a

análise computacional usada para quantificar os níveis de expressão dos genes. O RNA é uma molécula muito instável e sua degradação pode ser causada por mudanças no pH assim como por ribonucleases presentes nas células que podem permanecer ativas caso o isolamento do RNA não seja realizado da maneira correta. Diferentemente do RNA, a molécula de DNA é bastante estável e de fácil manipulação. Sendo assim, alterações que ocorrem na molécula de DNA e que sejam específicas da célula tumoral como, por exemplo, alterações no padrão de metilação da molécula de DNA nas células tumorais, podem ser mais facilmente detectadas e seriam mais apropriadas para a implantação na prática clínica^{30,31,83,113}.

1.2 A METILAÇÃO DO DNA COMO MECANISMO EPIGENÉTICO DE REGULAÇÃO DA TRANSCRIÇÃO GÊNICA

A metilação de citosinas no carbono 5 dos dinucleotídeos CpG é uma característica de genomas eucarióticos e um dos principais mecanismos epigenéticos de regulação da expressão gênica⁴. A transferência do grupo metil da S-adenosilmetionina ao carbono 5 das citosinas localizadas 5' de guaninas é catalisada pelas DNA metiltransferases (DNMTs). Estima-se que em células somáticas humanas as citosinas metiladas correspondam a 1% do total de bases do genoma e que, portanto aproximadamente 70-80% dos dinucleotídeos CpG estejam metilados no genoma humano^{3,14,27,112}.

Os dinucleotídeos CpGs possuem uma baixa representação no genoma humano, e isto se deve à alta taxa de mutações associadas às metilcitosinas, que podem sofrer deaminação hidrolítica espontânea e serem convertidas em timinas

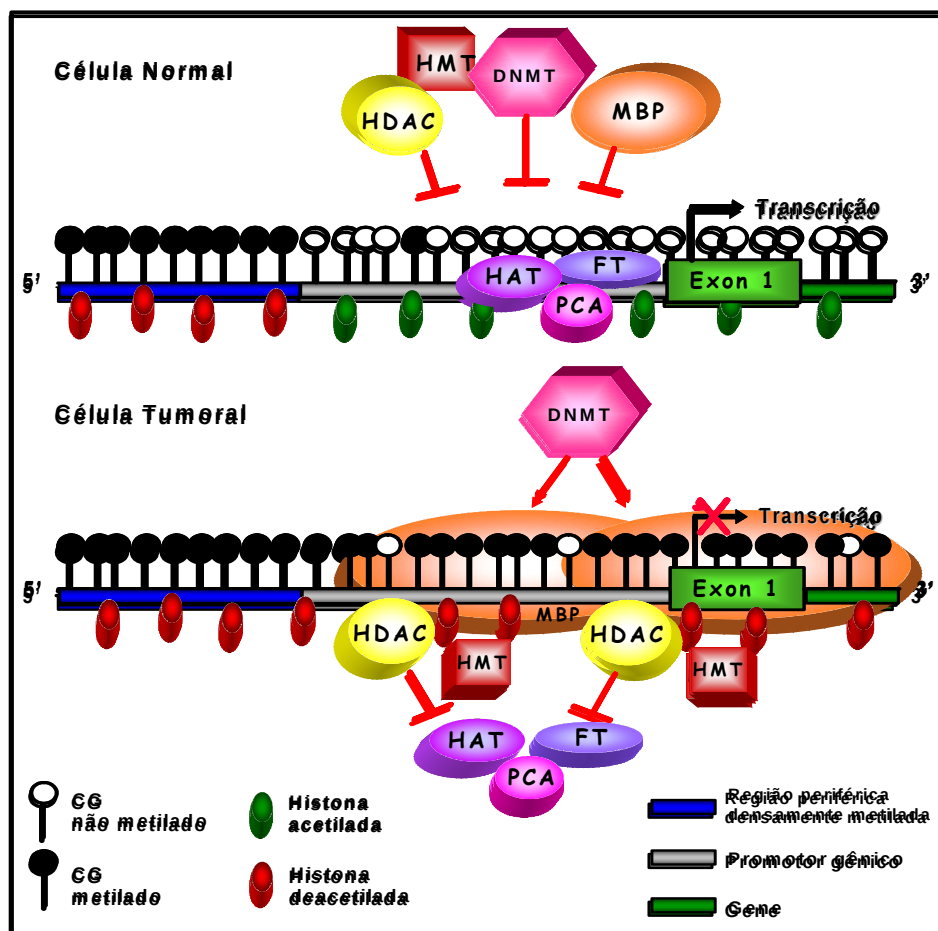
durante a replicação do DNA⁴⁵. Sua distribuição no genoma não é randômica, estando estas freqüentemente agrupadas em regiões genômicas com alta proporção de conteúdo G+C, chamadas “Ilhas de CpG”⁷². As ilhas de CpG foram originalmente definidas como seqüências de DNA com mais de 200 bp, com conteúdo de guaninas e citosinas igual ou superior a 50% e uma razão entre a freqüência observada e a freqüência esperada de dinucleotídeos CpG igual ou superior a 0,6³⁵. Em um estudo mais recente, TAKAI e JONES¹⁰⁵ revisaram as regras para a definição de ilhas de CpG a fim de excluir outras seqüências genômicas ricas em CG como repetições ALU. Com base em extensivas buscas na seqüência completa dos cromossomos 21 e 22, redefiniram as ilhas de CpG como seqüências com mais de 500 pb, com conteúdo de guaninas e citosinas igual ou superior a 55% e uma razão entre a freqüência observada e a freqüência esperada de dinucleotídeos CpG igual ou superior a 0,65.

A relação entre metilação e expressão gênica tornou-se aparente quando se observou que regiões do genoma acessíveis às nucleases e, portanto, ativas transcricionalmente, continham DNA hipometilado⁸⁸. Estudos posteriores de distribuição dos dinucleotídeos CpG demonstraram que ilhas de CpG são encontradas preferencialmente na região 5' dos genes humanos em comparação à região 3', sugerindo, desta forma, um papel importante na regulação da expressão gênica.

Diversos estudos demonstraram que a metilação em regiões promotoras está diretamente correlacionada à ausência ou diminuição da atividade transcricional⁸⁹. A metilação do DNA, juntamente com a modificação de histonas e o *imprinting*

genômico, são os principais mecanismos epigenéticos, capazes de alterar o padrão de expressão gênica sem afetar a sequência de nucleotídeos do DNA^{57,65}.

Diferentes mecanismos foram propostos para explicar o papel da metilação na diminuição da atividade transcricional (Figura 3). Em um primeiro modelo foi proposto que os resíduos de citosina metilados poderiam interferir diretamente na interação de fatores de transcrição aos seus sítios de ligação ao DNA, uma vez que o grupo metil se projeta no sulco maior da dupla fita de DNA⁹⁰. Além desta interferência direta, foi proposto que membros de uma família de proteínas que se ligam a resíduos metilados (MBPs - *Methylcytosine Binding Proteins*) como MeCP2, MBD1-MBD4, poderiam competir pela ligação com os próprios fatores de transcrição⁷². Finalmente, sabe-se que as MBPs podem recrutar complexos protéicos contendo co-repressores e histona deacetilases e, com base nessa interação, foi proposto que a ligação destes complexos ao DNA levaria a uma mudança na estrutura de cromatina impedindo a transcrição⁷⁹. Recentemente, outras proteínas que participam deste processo, tais como CTCF (do inglês *CCCTC binding factor*) e BORIS ou CTCFL (do inglês *CCCTC binding factor like*) tem sido identificadas^{11,63,67,68}. Nenhum dos modelos propostos é exclusivo e, muito provavelmente todos esses eventos estão atuando simultaneamente na repressão da expressão gênica pela metilação.



Legenda: Em células normais, as ilhas de CpGs nos promotores estão protegidas do espalhamento da metilação do DNA a partir de regiões periféricas densamente metiladas. Ao redor da região promotora estão dispostas histonas acetiladas e a cromatina está pouco compactada, o que permite que fatores de transcrição se liguem à fita de DNA e ocorra a transcrição gênica. Nas regiões vizinhas densamente metiladas estão presentes histonas deacetiladas e a cromatina está bastante compactada. Neste contexto, complexos com enzimas DNMTs não se aproximam do promotor. Em células tumorais, as DNMTs têm acesso à região promotora, juntamente com MBPs, e a metilação do DNA se espalha a partir das regiões periféricas em direção às ilhas de CpGs. As histonas tornam-se deacetiladas e a cromatina adquire conformações compactadas, impedindo a ligação de fatores de transcrição à fita de DNA. Deste modo, é estabelecido o silenciamento gênico. DNMTs, DNA metiltransferases; FT, fator de transcrição; HAT, histona acetil transferase; HDAC, histona deacetilase; HMT, histona metiltransferase; MBPs, proteínas de ligação a CpGs metilados; PCA, proteína co-ativadora.

Fonte: Modificado de JONES e BAYLIN⁵⁷.

Figura 3 A metilação do DNA em uma ilha de CpG localizada na região promotora de um gene e o silenciamento gênico.

1.3 METILAÇÃO DO DNA E CÂNCER

Alterações no padrão de metilação são freqüentemente encontradas em diversas patologias incluindo as neoplasias^{8,10,94}. Dois tipos de alterações no padrão de metilação são freqüentemente encontrados em tumores: a hipometilação global do genoma e a hipermetilação localizada em ilhas de CpG presentes na região promotora de genes relacionados ao câncer²⁹.

A perda da metilação do DNA em dinucleotídeos CpGs foi a primeira anormalidade epigenética identificada em células tumorais. FEINBERG e VOGELSTEIN³² compararam o padrão de digestão do DNA com enzimas de restrição sensíveis a metilação entre células normais e tumorais e descobriram que uma significativa proporção de CpGs metilados em tecidos normais estava desmetilada nas células tumorais. Em seguida, GAMA-SOSA et al.³⁴ demonstraram uma redução global de 10 a 20% no conteúdo de citosinas metiladas no genoma de vários tipos de tumores, tanto benignos quanto malignos, quando comparados aos tecidos normais correspondentes. Posteriormente, muitos estudos comprovaram a ocorrência da hipometilação global do genoma de células tumorais provenientes de diferentes neoplasias. Esta modificação do DNA tumoral é observada freqüentemente em diferentes tipos de seqüências repetitivas do genoma, incluindo regiões de heterocromatina do DNA, retrotransposons, retrovírus endógenos e seqüências satélites, bem como em seqüências únicas, incluindo algumas de controle da transcrição. A hipometilação do DNA induz a atividade transcricional dessas seqüências e favorece a instabilidade cromossômica nos tumores. A hipometilação também tem sido associada à expressão diferencial de oncogenes e antígenos

tumorais²⁶. O gene que codifica para a ornitina descarboxilase (ODC1) e os proto-oncogenes ERBA1 e BCL2, por exemplo, se mostraram hipometilados em leucemia linfocítica crônica²⁴. Da mesma forma, a ativação do gene MAGE1, que codifica um antígeno específico em melanomas, parece estar relacionada a desmetilação do promotor deste gene e o gene TCL1 é reativado através da demetilação do seu promotor em Linfoma de Burkitt e Leucemias Linfocíticas Crônicas de células B¹²³. A perda de metilação é um evento freqüentemente observado no início do desenvolvimento de tumores, e se correlaciona com a severidade da doença e com o potencial metastático em muitos tipos de tumores^{92,118}.

Além da hipometilação global do genoma, as células tumorais também são caracterizadas pela presença de hipermetilação em ilhas de CpGs localizadas nas regiões promotoras de genes relacionados ao câncer. A primeira associação entre hipermetilação e genes supressores de tumor foi feita com o primeiro gene supressor de tumor descoberto, o gene do retinoblastoma (RB1)⁴⁰. Alguns anos depois, foi descoberto que a inativação dos genes VHL, CDKN2A (também conhecido como P16 ou INK4A), CDH1 e do gene de reparo MLH1 também estavam associadas a hipermetilação de suas regiões promotoras^{38,39,48,60,73}. A partir daí, os relatos de inativação por hipermetilação de genes envolvidos com diversas funções celulares que estão alteradas nas células tumorais como regulação do ciclo-celular, reparo do DNA, remodelamento da cromatina, sinalização celular, transcrição e apoptose, aumentaram progressivamente na literatura. Estudos recentes sugerem que centenas de ilhas de CpG podem estar metiladas em células tumorais¹⁹. Cada tipo de tumor parece ter seu próprio padrão de genes específicos ligados ao câncer que são mais ou menos suscetíveis a metilação¹⁷. Os genes CDKN2B, CDH1 e CALCA, por

exemplo, são comumente metilados em AML (“*Acute Myeloid Leukaemia*”), mas raramente metilados em câncer de próstata¹⁰².

Alterações no padrão de metilação têm sido amplamente utilizadas na caracterização de novos genes envolvidos no processo de formação e progressão tumoral. Na literatura podemos encontrar evidências diretas ou indiretas do silenciamento por metilação de mais de 40 genes em tumores de mama¹¹⁹. Entre eles estão genes envolvidos com a aquisição das seis características essenciais para a sobrevivência das células tumorais, definidas por HANAHAN e WEINBERG⁴³: auto-suficiência em relação a fatores de crescimento (ER, PR, RASSF1A, SOCS1), potencial replicativo ilimitado (CDKN2A, CDKN1A, CCND2, RARB), insensibilidade a sinais inibitórios do crescimento (HIN1, TGFBR2, RFC), evasão da morte celular programada (CASP8, TP53, DAPK), angiogênese sustentada (TP73, MASPIN, THBS1), invasão tecidual e metástase (CDH1, TIMP3, APC, BCSG1, prostatina, TWIST).

1.4 METODOLOGIAS EM LARGA ESCALA UTILIZADAS NA DETECÇÃO DE REGIÕES DIFERENCIALMENTE METILADAS

Em 1999, HUANG et al. adaptaram a metodologia disponível para a confecção de cDNA *microarray* e desenvolveram uma técnica denominada *Differential Methylation Hybridization* (DMH), que permite uma análise em larga escala das ilhas de CpG diferencialmente metiladas entre diferentes amostras⁴⁹. O primeiro passo da técnica de DMH envolve a obtenção de fragmentos de DNA genômico correspondentes a ilhas de CpG que serão organizadas em suportes sólidos

(ex. membranas de nylon ou lâminas de vidro). Esses fragmentos devem, necessariamente, conter sítios de restrição para enzimas de restrição sensíveis a metilação, tais como a *Bst*UI (CGCG) ou *Hpa*II (CCGG). O segundo passo envolve a amplificação, através da técnica de PCR, de uma população de fragmentos metilados a partir de diferentes amostras que será utilizada como sonda contra membranas contendo as ilhas de CpG imobilizadas. Para a amplificação dos fragmentos metilados, o DNA genômico isolado a partir de tecidos normais e tumorais é separadamente digerido com a enzima *Mse*I, que cliva o DNA em fragmentos de aproximadamente < 200 bp. Como a enzima de restrição *Mse*I reconhece a sequência TTAA, as ilhas de CpG se mantêm, relativamente, intactas após a digestão. Sequências adaptadoras são, então, ligadas aos fragmentos *Mse*I e os elementos repetitivos presentes na sequência genômica são subtraídos. Após a subtração do DNA repetitivo, os fragmentos de DNA genômico ligados aos adaptadores são digeridos com uma enzima de restrição sensível a metilação (*Bst*UI ou *Hpa*II) e, posteriormente, amplificados utilizando iniciadores específicos para as sequências adaptadoras. Os produtos amplificados representam uma população de fragmentos genômicos metilados nas diferentes amostras uma vez que na ausência de metilação a amplificação por PCR não é possível devido à digestão do fragmento não metilado pela enzima de restrição. Na última etapa, os produtos amplificados são marcados (radioativamente ou com fluorófos) e utilizados como sonda contra o *array* de ilhas CpG no qual as diferenças entre a intensidade do sinal de hibridização das diferentes amostras refletem alterações no padrão de metilação de cada uma das ilhas de CpG presentes (Figura 4). Um ponto limitante da técnica de DMH consiste no isolamento de fragmentos genômicos correspondentes a ilhas de CpG que serão imobilizados

nos suportes sólidos. As seqüências isoladas a partir de bibliotecas genômicas enriquecidas para ilhas de CpG contém uma parcela significativa das seqüências correspondendo a elementos repetitivos (DNA ribossomal rico em CG ou elementos retrovirais da família ALU) e nem todas as seqüências possuem sítios de restrição para enzimas sensíveis a metilação. Além disto, outro ponto limitante para o sucesso da DMH é o estabelecimento de uma correlação direta entre as ilhas de CpG imobilizadas no suporte sólido e os genes correspondentes as mesmas, pois somente através do estabelecimento desta correlação é possível identificar genes diferencialmente metilados entre as amostras que possam então ser utilizados como marcadores.

Uma segunda metodologia de hibridização de fragmentos diferencialmente metilados foi desenvolvida por WANG et al.¹¹⁵ em 2004. Nesta metodologia, o DNA genômico é digerido com a enzima *HpaII* (sensível a metilação e com sítio de restrição encontrado com freqüência nas ilhas de CpG), ligado a adaptadores específicos para o sítio desta enzima e amplificado por PCR. Após a amplificação, os fragmentos são marcados com fluoróforos e hibridizados contra uma lâmina de vidro contendo 12800 fragmentos de DNA enriquecidos para seqüências promotoras de genes humanos (1000 bp *upstream* e 500 bp *downstream* do sítio de início da transcrição) que foram identificadas com base na montagem do genoma humano disponibilizada pela Universidade Santa Cruz, Califórnia (UCSC) através do programa BLAT (www.genome.ucsc.edu). Ao contrário do DMH, a utilização dessa metodologia permite a amplificação de fragmentos pequenos que possuem sítios de restrição próximos e não metilados no genoma. A presença de metilação nos sítios de restrição da enzima *HpaII* impede a digestão e gera fragmentos de maior tamanho

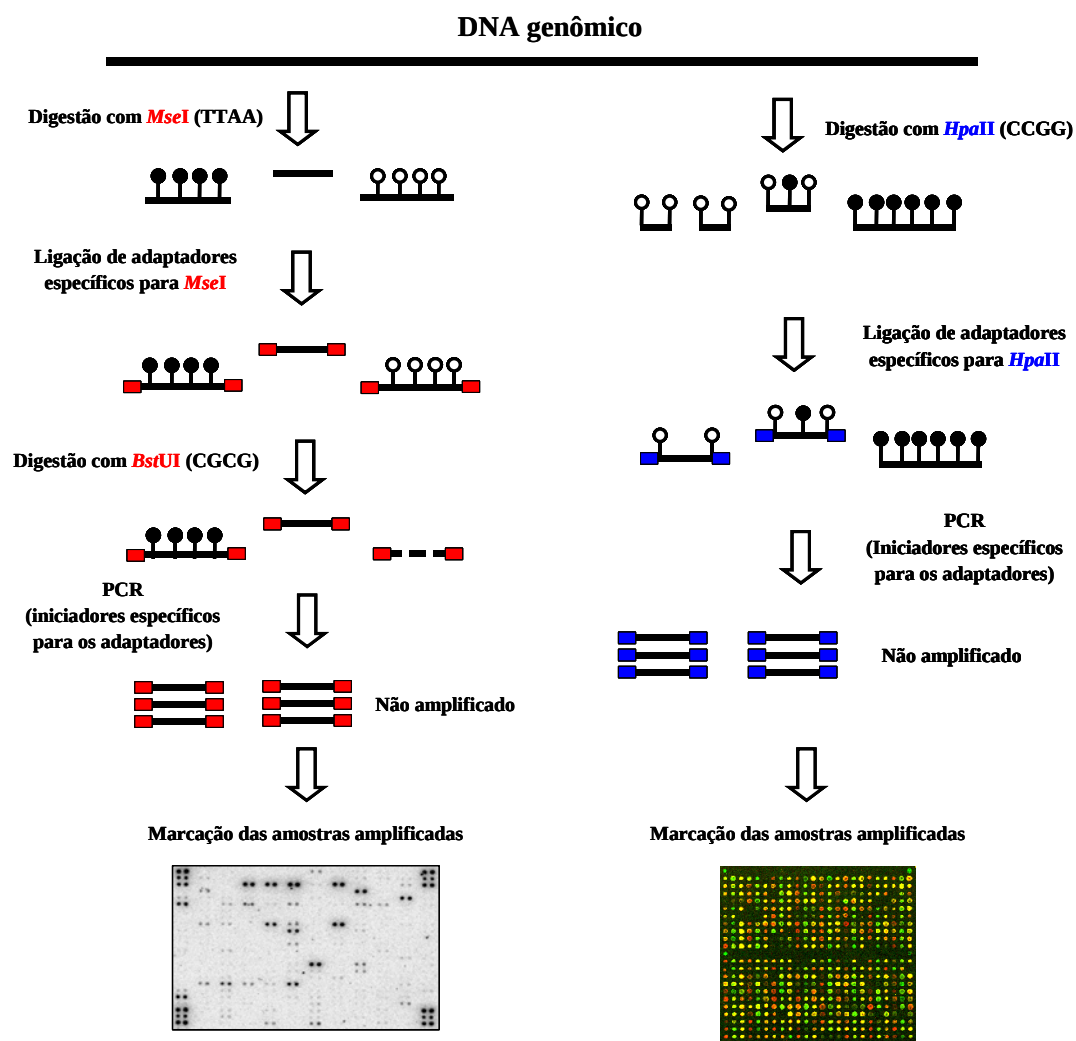
que são amplificados com menor eficiência, e portanto, representados em menor quantidade na sonda gerada (Figura 4). A eficiência de amplificação de um determinado fragmento e sua representatividade na sonda utilizada na hibridização deverão permanecer inalterada entre diferentes amostras de DNA a não ser que haja diferença nos níveis de metilação entre as mesmas. Essas diferenças podem então ser visualizadas através da intensidade do sinal de hibridização de cada uma das amostras.

Vários estudos têm utilizado a hibridização de metilação diferencial para identificar genes diferencialmente metilados em tumores. Em 2000, YAN et al.¹²¹ aplicaram a técnica de DMH para avaliar um grupo de amostras pareadas de tecido normal e tumoral de mama a fim de determinar se um padrão específico de alterações epigenéticas tinha correlação com parâmetros patológicos nas análises dos pacientes. As análises estatísticas do padrão de metilação diferencial das amostras revelaram que a hipermetilação das ilhas de CpG tinha correlação com o grau histológico dos tumores de mama analisados. Tumores que se apresentavam pouco diferenciados exibiam um maior número de ilhas de CpG hipermetiladas em relação aos tumores que se apresentavam moderadamente ou bem diferenciados ($p = 0,041$). Estes resultados sugerem, então, que os diferentes padrões de metilação observados em várias amostras clínicas podem refletir diferentes estádios ou tipos de câncer.

AHLUWALIA et al.¹, em 2001, também aplicaram a técnica de DMH para avaliar o padrão de metilação de ilhas de CpG em câncer de ovário. Estes autores observaram que algumas ilhas de CpG estavam comumente hipermetiladas em amostras de tumores de mesmo estágio. WEI et al.¹¹⁶, avaliaram o perfil de metilação de amostras de pacientes com carcinoma de ovário em estágio 3 e 4 e observaram

que o alto grau de metilação em um determinado grupo de amostras identificado na clusterização hierárquica estava relacionado à recorrência precoce da doença nestas pacientes após a quimioterapia. Em 2005, VAN DOORN et al.¹¹⁰ utilizando esta metodologia identificaram a hipermetilação de genes supressores de tumor incluindo BCL7A, PTPRG e P73 em linfoma cutâneo de células T. Em 2006, RAHMATPANAHI et al.⁸⁶ utilizaram esta metodologia para estudar diferentes subtipos de linfoma de célula B pequena (SBCL) e observaram padrões de metilação preferencial que têm ajudado a definir os diferentes subtipos de SBCL que não são reconhecidos pelos sistemas convencionais de classificação além de fornecer informações sobre a biologia destes tumores.

No entanto, até o presente momento, não há relato na literatura de estudos de metilação diferencial em larga escala que identificou padrões de metilação diferencial entre pacientes com câncer de mama sem comprometimento linfonodal que desenvolveram ou não metástase à distância.



Legenda: O esquema descreve os protocolos utilizados na preparação do DNA genômico que será utilizado para a detecção de diferenças de metilação proveniente da amplificação dos fragmentos digeridos entre as amostras. À esquerda metodologia desenvolvida por HUANG et al. e à direita WANG et al. Vale ressaltar que enquanto na metodologia desenvolvida por HUANG et al. os produtos amplificados são metilados, na metodologia desenvolvida por WANG et al. são amplificados os produtos não metilados.

● – CpG metilado, ○ – CpG não metilado, ■ – adaptador específico para *MseI*, ■ – adaptador específico para *HpaII*.

Figura 4 Desenho esquemático das técnicas de hibridização de metilação diferencial.

OBJETIVOS

2 OBJETIVOS

2.1 OBJETIVO GERAL

Identificar uma assinatura de metilação capaz de discriminar dois grupos de pacientes com carcinoma ductal invasivo de mama linfonodo negativas sendo um grupo que não desenvolveu doença metastática e outro que desenvolveu metástase em um período mínimo de 5 anos.

2.2 OBJETIVOS ESPECÍFICOS

- 1 Estabelecer o perfil de metilação e identificar fragmentos de DNA diferencialmente metilados entre amostras de pacientes com carcinoma ductal invasivo de mama linfonodo negativas que desenvolveram ou não metástase à distância utilizando a metodologia de hibridização de fragmentos diferencialmente metilados;
- 2 Avaliar se o perfil de metilação desses fragmentos de DNA diferencialmente metilados (assinatura de metilação) é capaz de discriminar os dois grupos de pacientes em estudo;
- 3 Comparar nossos dados de metilação diferencial com as assinaturas de expressão gênica capazes de prever o prognóstico de pacientes com câncer de mama linfonodo negativas.

MATERIAL E MÉTODOS

3 MATERIAL E MÉTODOS

3.1 TUMORES PRIMÁRIOS DE MAMA

Foram utilizadas neste estudo 21 amostras de carcinoma ductal invasivo de mama de pacientes linfonodo negativas armazenadas no Banco de Tumores do Centro de Tratamento Médico e Pesquisa do Hospital A.C. Camargo e do Instituto de Lucha Contra El Cáncer (SOLCA) – Cuenca – Equador. Os linfonodos foram removidos durante a cirurgia para a remoção do tumor primário, sendo analisado um número mínimo de 20 linfonodos por paciente. Foram consideradas linfonodo negativas as pacientes que não apresentaram evidências de macrometástases e micrometástases em cortes histológicos corados com hematoxilina-eosina (HE) e que também não apresentaram evidências de micrometástases em cortes histológicos analisados por imunoistoquímica com anticorpos específicos para marcadores epiteliais (citoceratina AE1/AE3). Todas as amostras do estudo foram obtidas mediante consentimento informado das pacientes e aprovação do comitê de ética de ambas as instituições.

Os critérios utilizados para inclusão de pacientes em nosso estudo foram os seguintes:

- i) Pacientes com carcinoma ductal invasivo;
- ii) Linfonodo negativo;
- iii) Idade ao diagnóstico igual ou inferior a 55 anos;
- iv) Tamanho do tumor igual ou menor que 5 cm (pT1 ou pT2);

- v) Sem uso de quimioterapia ou radioterapia neoadjuvante;
- vi) Ter sido submetido aos esquemas de quimioterapia adjuvante 8 ciclos Ciclofosfamida, Metotrexato e 5-Fluoracil (CMF) ou 4 ciclos Antracíclicos (AC);
- vii) Ter pelo menos cinco anos de seguimento;
- viii) Não ter história familiar de câncer de mama.

Seguindo os critérios acima descritos foram incluídas em nosso estudo 21 pacientes linfonodo negativas das quais 13 não desenvolveram metástase à distância e 08 desenvolveram metástase à distância em um período de seguimento mínimo de 5 anos (Tabela 3).

Dentre as 13 pacientes que não desenvolveram doença metastática, 5 pacientes apresentavam tumores pT1 (<2 cm) e 8 pacientes tumores pT2 (2 a 5 cm). Dentre as 8 pacientes que desenvolveram metástase à distância 2 pacientes tinham tumor pT1 e 6 pT2. Em relação ao tratamento quimioterápico adjuvante 12 das 13 pacientes sem doença metastática e 6 das pacientes que desenvolveram metástase foram submetidas ao tratamento sistêmico com CMF. Análises de imunoistoquímica com anticorpos específicos para receptores de estrogênio (RE) e progesterona (RP) evidenciaram que das 13 pacientes que não desenvolveram metástase à distância 6 foram consideradas RE e RP positivas, duas eram somente RE positivas e duas somente RP positivas e destas 8 receberam tamoxifeno. Dentre as pacientes que desenvolveram metástase à distância 2 eram RE e RP positivas e uma paciente somente RE positiva e destas 5 não receberam tamoxifeno. Para duas e três destas pacientes não foi possível avaliar o status de RE/RP e tamoxifeno, respectivamente.

Em relação aos sítios de metástase, uma paciente desenvolveu metástase óssea, uma paciente desenvolveu metástase no sistema nervoso central, uma paciente desenvolveu metástase pulmonar, uma paciente desenvolveu metástase óssea, hepática e outros (medula ou tiróide), uma paciente desenvolveu metástase óssea, pulmonar, hepática e outros e duas pacientes desenvolveram metástases óssea e pulmonar. Vale ressaltar que para o rastreamento da doença metastática foram realizados os exames de raio-X de tórax, cintilografia óssea e tomografia computadorizada.

Tabela 3 Aspectos clínico-patológicos das pacientes com carcinoma ductal invasivo de mama do estudo.

Pacientes	pT	RE	RP	QTAdj	RTAdj	Rec	Idade	Tamoxifeno
43	1	1	1	CMF	0	0	46	1
59	1	0	0	AC	1	0	49	0
200	2	0	1	CMF	1	0	39	0
333	1	1	1	CMF	1	0	50	1
494	2	1	1	CMF	1	0	39	1
496	2	0	0	CMF	1	0	50	0
756	2	0	0	CMF	1	0	48	0
759	2	0	0	CMF	1	0	32	0
773	1	1	0	CMF	1	0	45	1
834	2	1	0	CMF	1	0	51	1
1151	2	1	1	CMF	1	0	53	1
1587	2	1	1	CMF	1	0	48	1
1737	2	1	1	CMF	0	0	48	1
536	1	0	0	CMF	1	1	50	0
630	2	1	0	CMF	1	1	34	0
752	2	1	1	CMF	0	1	35	0
1938	2	0	0	CMF	1	1	45	0
2568	2	1	1	AC	1	1	42	0
04	2	0	0	AC	1	1	41	-
A	1	-	-	CMF	0	1	<55	-
C	2	-	-	CMF	1	1	<55	-

Legenda: pT: tamanho do tumor (1 - até 2 cm e 2 - entre 2 e 5 cm); Linfonodo: RE: receptor de estrógeno; RP: receptor de progesterona; QTAdj: quimioterapia adjuvante; RTAdj: radioterapia adjuvante; Rec: Recidiva à distância; Idade: idade da paciente ao diagnóstico. 0 – negativo e 1 – positivo.

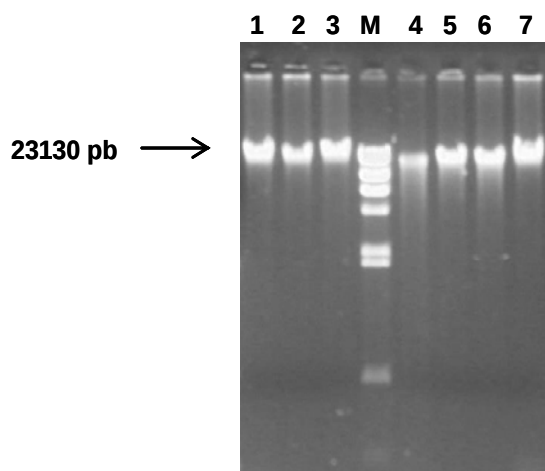
3.2 MACRODISSECÇÃO DOS TUMORES PRIMÁRIOS DE MAMA

As amostras selecionadas para este estudo foram macrodissecadas para enriquecer a amostra para células tumorais. Cortes histológicos dos tumores foram corados por hematoxilina-eosina (HE) e as áreas necróticas e de tecido normal foram demarcadas. Tais áreas foram eliminadas da peça com o auxílio de um bisturi e uma lupa. Foram utilizadas somente amostras com pelo menos 80% de células tumorais.

3.3 EXTRAÇÃO DE DNA GENÔMICO

O DNA genômico dos tumores primários de mama foi extraído através do método convencional de fenol clorofórmio. As amostras de tumores foram incubadas a 55°C durante a noite em 4 ml de tampão de extração TES (10 mM Tris-HCl / 50 mM EDTA / 0,5% SDS) contendo 100 µg/ml de Proteinase K. Após digestão completa do tecido adicionou-se 4 ml de fenol equilibrado (pH 8.0) à amostra e esta foi mantida sob agitação por 5 minutos seguida de centrifugação durante 15 minutos a 5.000xg, 4°C. Após a recuperação da fase aquosa, adicionou-se 4 ml de solução clorofórmio-álcool isoamílico (24:1), seguindo-se novamente a agitação da amostra por 5 minutos e sua centrifugação durante 15 minutos a 5.000xg, 4°C. A fase aquosa foi novamente separada seguida de nova extração com clorofórmio-álcool isoamílico. Ao final desta extração, uma fase aquosa transparente livre de proteínas foi coletada e transferida para tubo corex estéril. Em seguida, a precipitação do DNA foi realizada adicionando-se 1/10 do volume da amostra de 3M NaOAc pH 6.0 e 2.5 vezes o volume final de etanol absoluto gelado. As amostras foram incubadas a -

20°C durante a noite. No passo seguinte, as amostras foram centrifugadas durante 30 minutos a 22.000xg e lavadas com etanol 70%. Após a secagem do DNA à temperatura ambiente, o mesmo foi ressuspenso em 100-500 µl de tampão TE (10 mM Tris HCl, 1.0 mM EDTA pH 8.0) e armazenado a 4°C. A qualidade do DNA obtido foi avaliada em gel de agarose 0,8% (Figura 5). A quantificação do DNA obtido foi obtida em espectrofotômetro com comprimento de onda equivalente a 260 nm, considerando-se que 1DO₂₆₀ nm equivale a 50µg/ml de DNA. A relação entre as leituras realizadas a 260 e 280 nm foi utilizada como parâmetro para estimar o grau de contaminação do DNA por proteínas.



Legenda: M - Marcador de massa molecular *Lambda HindIII*; 1 a 7 - amostras de tumores de mama.

Figura 5 Eletroforese representativa do DNA genômico extraído de amostras tumorais por fenol-clorofórmio em gel de agarose 0,8% corado com Gel Red.

3.4 HIBRIDIZAÇÃO DE FRAGMENTOS DE DNA DIFERENCIALMENTE METILADOS

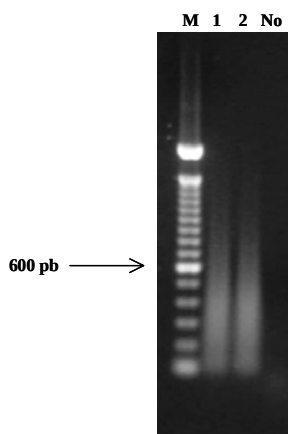
3.4.1 *Microarray* enriquecido para seqüências promotoras

O *microarray* enriquecido para seqüências promotoras utilizado nesse projeto foi gentilmente cedido pelo Dr. Michael McClelland do Sidney Kimmel Cancer Center, San Diego, Ca. Em uma breve descrição da construção do *microarray*, seqüências promotoras (1000 pb *upstream* e 500 pb *downstream* do sítio de início da transcrição) foram identificadas com base na montagem do genoma humano disponibilizada pela Universidade Santa Cruz, Califórnia (UCSC) através do programa BLAT (www.genome.ucsc.edu). Os oligonucleotídeos utilizados para a amplificação destas seqüências promotoras foram desenhados utilizando uma versão modificada do programa *Primer3* (www.broad.mit.edu/genome_software/other/primer3.html). Fragmentos de seqüências promotoras apresentando um tamanho médio de 1200 pb foram amplificados, purificados e imobilizados em *UltraGAPS-coated slides* (Corning, Inc., Corning, NY) na presença de 50% de DMSO. Nos experimentos realizados em nosso laboratório foi utilizado o *microarray* contendo triplicatas de 11586 seqüências promotoras, 752 seqüências não promotoras controle e 192 controles representados por seqüências não humanas imobilizadas em duas lâminas sendo 6400 seqüências em triplicata em cada uma das lâminas (Lâminas Set A e Set B). Muitos dos fragmentos de seqüências promotoras imobilizados no *microarray* estão relacionados a genes que possuem particular importância no câncer e genes sabidamente regulados por metilação em câncer.

3.4.2 Amplificação dos fragmentos diferencialmente metilados e condições de hibridização

Os produtos amplificados que foram utilizados como sonda na hibridização dos *microarrays* foram gerados como descrito por WANG et al.¹¹⁵. Brevemente, 0.5 µg de DNA genômico foram completamente digeridos com 15U da endonuclease sensível a metilação *HpaII* (New England Biolabs - NEB) em uma reação de 20 µL por 16 horas à 37°C. Como referência comum nos experimentos de hibridização foi utilizado o DNA genômico da linhagem celular normal de mama HB4a digerido com 15U da endonuclease *MspI* que reconhece o mesmo sítio de restrição da enzima *HpaII* porém, não é sensível a metilação em uma reação de 20 µL por 16 horas à 37°C. O produto da digestão foi então ligado a 20 pmol de cada um dos adaptadores Meth_1a1/Meth_1a2/Meth_4a1/Meth_4a2 específicos para o sítio de restrição da enzima *HpaII* na presença de 400U de T4 DNA ligase (NEB) e 1 mM de ATP (Amersham Biociences) à temperatura ambiente por 4 horas em um volume final de 20 µl. As seqüências dos oligonucleotídeos são Meth_1a1 (5'-AAGTATCGGAGGAGTCTTTGTTA-3'), Meth_1a2 (5'-Phosphate-CGTAACAAAGACTCCTCCGATACTT-Amine-3'), Meth_4a1 (5'-TCTCTTGAAGAGTAACTTGTTGG-3') e Meth_4a2 (5'-Phosphate-CGCCAACAAGTTACTCTTCAAGAGA-Amine-3'). Os iniciadores foram solubilizados em água em uma concentração final de 50 µM cada um. Os oligonucleotídeos Meth_1a1 e Meth_1a2 foram anelados por resfriamento gradual da mistura (20 µl de cada oligonucleotídeo, 1X NEB Buffer 2 (NEB) e 50% de LoTE pH 7.5, volume final de 50 µl), que foi de 100°C à 25°C. O mesmo procedimento foi realizado para os oligonucleotídeos Meth_4a1 e Meth_4a2. Para verificar a eficiência

da digestão com as endonucleases *HpaII* e *MspI* e a eficiência da reação de ligação dos adaptadores ao DNA digerido foi realizada uma PCR teste. Nessa reação, um microlitro do DNA digerido e ligado aos adaptadores, 0,4 μ M de cada um dos iniciadores Meth_1p (5'-GTATCGGAGGAGTCTTTGTTACGG-3') e Meth_4p (5'-CTTGAAGAGTAACTTGTTGGCGG-3'), 1 M de Betaína, 0,2 mM de dNTPs, 3 mM de MgCl₂, mais 1 U de *Taq* Polimerase (Invitrogen) em um volume final de 50 μ l foram submetidos a seguinte ciclagem: 95°C por 1 min; 95°C por 15 seg, 63°C por 15 seg e 72°C por 30 seg, por 24 ciclos; e 72°C por 7 min. Como podemos observar na figura 6, a reação foi avaliada em gel de agarose 1%. A presença de um rastro difuso com tamanho entre 0.2 – 1.2 Kb indica a amplificação bem sucedida dos fragmentos aos adaptadores.

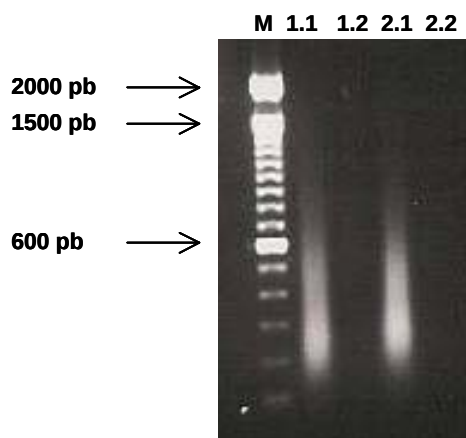


Legenda. M - Marcador de massa molecular 100 pb (Fermentas); 1 – amostra referência e 2 – amostra de tumor de mama e No – controle sem DNA da reação.

Figura 6 Eletroforese representativa de uma reação de amplificação teste em gel de agarose 1% corado com Gel *Red*.

As seqüências de DNA repetitivo foram subtraídas do DNA digerido com a endonuclease *HpaII* e ligado aos adaptadores usando o protocolo de hibridização subtrativa por *Cot1* humano (Invitrogen) descrito por CRAIG et al.²⁰ com algumas modificações. Dez microgramas de *Cot1* humano enriquecido de seqüências repetitivas foi marcado com biotina utilizando o *BiotinULS Labeling Kit* (Fermentas Life Sciences) por 30 minutos à 85°C. O DNA *Cot1* humano marcado foi purificado por precipitação com etanol e solubilizado em 100 µL do tampão de ligação TEN100 (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.5) dos quais 5 µL foram reservados para posterior controle de qualidade da marcação do *Cot1* humano com biotina por separação em gel de agarose (1*). Cem microlitros (1 mg) de partículas magnéticas com estreptavidina (F. Hoffmann-La Roche Ltd.) preparados de acordo com instruções do fabricante foram adicionados ao DNA *Cot1* humano marcado com biotina e incubado por 15 minutos à temperatura ambiente. O tubo contendo a mistura foi colocado no separador de partículas magnéticas (F. Hoffmann-La Roche Ltd.), o sobrenadante foi removido e deste 5 µL foram reservados para controle da qualidade de marcação por separação em gel de agarose (2*). Duzentos microlitros de TEN100 fresco foram adicionados e as partículas foram incubadas por mais 15 minutos à temperatura ambiente seguido da remoção do sobrenadante. As partículas foram então lavadas duas vezes com o tampão de lavagem TEN1000 (10 mM Tris-HCl, 1 mM EDTA, 1000 mM NaCl, pH 7.5). O controle da qualidade da marcação do *Cot1* humano com biotina foi feito por separação em gel de agarose 1% das alíquotas reservadas em 1* e 2*. Se $1 = 2$, a marcação foi ineficiente, se $1 > 2$, a marcação foi eficiente (Figura 7). O DNA digerido com *HpaII* e ligado aos adaptadores foi adicionado as partículas magnéticas e o volume ajustado para 100 µL

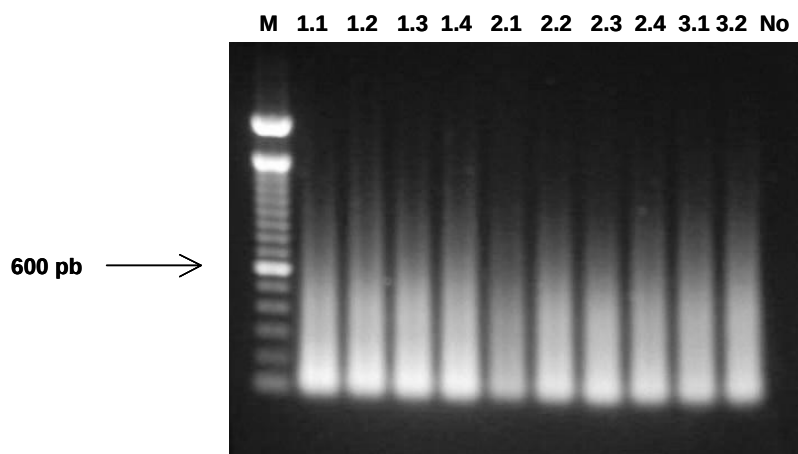
com 6x SSC e 0,1% SDS . A mistura, DNA genômico digerido e ligado, e DNA *Cot1* humano biotilado e ligado às partículas magnéticas com estreptavidina, foi desnaturada por 10 minutos e hibridizada durante 20 horas à 65°C em forno de hibridização com rotação. Após a hibridização, outros 100 µL de partículas magnéticas com estreptavidina (1 mg) foi preparado, adicionado a mistura hibridizada, complementada com 200 µL de tampão de ligação TEN100 e incubada por 30 minutos à temperatura ambiente. Os tubos foram colocados no separador de partículas magnéticas e o sobrenadante separado. Este sobrenadante foi incubado com outros 100 µL de partículas magnéticas com estreptavidina (1 mg) por mais 30 minutos à temperatura ambiente. Após a incubação o sobrenadante foi separado e purificado utilizando o *MinElute PCR Purification Kit* (Qiagen, Inc.) como recomenda o fabricante.



Legenda: Se 1 é igual a 2, a marcação foi ineficiente. Se 1 é maior que 2, a marcação foi eficiente. M - marcador de massa molecular 100 bp (Fermentas), 1.1 e 1.2 - alíquotas 1 e 2 da marcação de uma amostra e 2.1 e 2.2 - alíquotas 1 e 2 da marcação de uma amostra independente.

Figura 7 Eletroforese representativa da marcação do DNA *Cot1* humano com biotina em gel de agarose 1% corado com Gel Red.

Os fragmentos *HpaII* e *MspI* foram amplificados em cinco reações contendo 20 ng do DNA digerido, ligado aos adaptadores e subtraído de seqüências repetitivas, 0,4 μ M de cada iniciador Meth_1p e Meth_4p, 2M de Betaína e 2U de *Taq* Polimerase (Invitrogen Corporation) em 50 μ L de volume final cada. Uma amplificação de baixa ciclagem (95°C 1 min; 95°C 15 seg, 63°C 15 seg e 72°C 30 seg por 21X e 72°C 7 min) foi utilizada a fim de manter a representatividade dos diferentes fragmentos e evitar a saturação da reação de amplificação. Os produtos amplificados foram reunidos e então purificados utilizando o kit *MinElute PCR Purification Kit* (Qiagen, Inc.). A reação foi avaliada em gel de agarose 1%. A presença de um rastro difuso com tamanho entre 0.2 – 1.2 Kb indica a amplificação bem sucedida dos fragmentos aos adaptadores. (Figura 8).



Legenda. M – marcador de massa molecular 100 pb (Fermentas), 1.1 a 1.4 – amplificação de uma amostra tumoral e 2.1 a 2.4, amplificação de uma amostra tumoral independente, 3.1 e 3.2 – amplificação da amostra referência e No – controle sem DNA da reação de amplificação.

Figura 8 Eletroforese representativa da reação de amplificação dos amplicons em gel de agarose 1% corado com Gel Red.

Um micrograma do produto *HpaII* amplificado (amostra tumoral) foi marcado com 2 µL de cianina 5 (Cy5, Amersham Biociences) e o produto *MspI* amplificado (amostra referência) foi marcado com 2 µL de cianina 3 (Cy3, Amersham Biociences) utilizando o sistema *Read-To-Go DNA Labeling Beads* (Amersham Biociences) a 37°C por 90 minutos, no escuro. Após a reação de incorporação dos corantes, purificamos o produto da reação em coluna G50 (Amersham Biociences) centrifugando por 1 minuto a 1000g. O DNA marcado e purificado foi quantificado por espectrofotometria sob os comprimentos de onda específicos para cada fluoróforo (Cy3=550nm e Cy5=649). Os melhores resultados de hibridização são obtidos com o DNA apresentando uma frequência de incorporação (FOI) de pelo menos 20 nucleotídeos marcados por mil nucleotídeos. Um FOI menor que 20 diminui a sensibilidade do ensaio e um maior que 50 é indicativo de remoção incompleta dos nucleotídeos marcados não incorporados (Corning Protocol, Inc., Corning).

As lâminas de *microarray* foram pré-hibridizadas em 23,0 ml de solução contendo 5X SSC, 0,1% SDS, e 0,1 mg/ml de albumina bovina (BSA) por 3 horas à 42°C. Após a pré-hibridização das lâminas, foram feitas duas lavagens de 5 minutos cada em 0,1X SSC à temperatura ambiente, e então as lâminas foram enxaguadas em água MilliQ e centrifugadas por 2 minutos a 1600 rpm para secagem. Após esta etapa as lâminas foram colocadas na estação de hibridização (Genomic Solutions).

Os produtos amplificados e marcados com Cy3/Cy5 (referência e amostra) foram reunidos em um único tubo ao qual foram adicionados 10 µg de DNA *Cot-1* humano (Invitrogen). O volume desta solução foi reduzido para 8 µl e foram adicionados 5X SSC, 0,1% SDS, 25% de formamida para um volume final de 95 µl

que foi mantido sob agitação a 500rpm à 45°C até o momento da hibridização. Para a hibridização a solução de hibridização contendo os produtos marcados foi incubada a 95°C por 5 minutos para desnaturação dos fragmentos e mantida a 70°C até sua aplicação na lâmina. A hibridização foi realizada em três etapas: 70°C por 5 minutos, 55°C por 45 minutos e 42°C de 16 a 20 horas. Após a hibridização, as lâminas foram lavadas uma vez de 5 minutos em solução de lavagem I (2X SSC / 0,1% SDS) à 42°C, duas vezes de 5 minutos cada em solução de lavagem II (0,1X SSC / 0,1% SDS) à temperatura ambiente, quatro vezes de 1 minuto cada em solução de lavagem III (0,1X SSC) à temperatura ambiente e enxaguadas por 10 segundos em solução de lavagem IV (0,01X SSC) à temperatura ambiente (Corning Protocol, Inc., Corning). Após a etapa de lavagem as lâminas foram secas por centrifugação a 1600 rpm por 2 minutos.

3.4.3 Digitalização das imagens e quantificação do sinal de hibridização

Os sinais gerados pela hibridização das amostras foram capturados por *scanner* confocal a *laser* com canal duplo (*ScanArrayTM Express Microarray Acquisition System – Packard Bioscience Biochip Technologies*), usando uma resolução de 10 microns e sensibilidade de captação de 70% para ambos os canais. Para cada lâmina a leitura dos fluoróforos é feita separadamente gerando assim duas imagens independentes e então é feita a composição das imagens pelo próprio programa de captura. A quantificação da intensidade dos *spots* e do background médio local de todas as lâminas foi obtida com o *Quantarray Microarray Analysis Software* (Perkin Elmer, Inc.). Os valores de intensidade R e G referem-se à incorporação dos corantes Cy5 (*RED*) e Cy3 (*GREEN*).

3.4.4 Análise da qualidade das hibridizações e normalização

Após a captura de sinal pelo *scanner* e formação da imagem, os dados foram submetidos a algumas análises de critério de qualidade que são medidas em escala logarítmica dos valores de intensidade do sinal emitido por cada ponto do *array*. Essa análise foi realizada pelo laboratório de Bioinformática da Fundação Antônio Prudente – Hospital AC Camargo. Estas medidas podem ser visualizadas graficamente pelo *MA plot* e *Box plot*. Há também uma representação gráfica do dado bruto da emissão de sinal dos fluoróforos captado pelo *scanner* onde o Ch1 (*Green_Cy3*) representa o sinal da amostra referência e o Ch2 (*Red_Cy5*) representa o sinal da amostra teste e outra representação da distribuição deste sinal para toda a lâmina dividida em *sub-arrays*. Inicialmente, foram excluídos os pontos saturados, ou seja, aqueles que emitiram um sinal igual ou maior a 63.000 ($65.550 = 16 \text{ BIT}$). Após a subtração dos pontos saturados, foram também eliminados todos os pontos da lâmina que apresentaram emissão de fluorescência igual ou menor ao sinal emitido pelo ruído (*background*) através da subtração do sinal emitido por um ponto menos o valor do ruído no local do ponto. Após a subtração do ruído local os sinais quantificados foram transformados em escala logarítmica na base 2. Então foi feita a normalização de cada lâmina (intralâminas) seguida por normalização entre lâminas para minimizar as variações experimentais entre as hibridizações que afetam a medida do nível de metilação. A abordagem escolhida para a normalização dos nossos dados foi o *Locally Weighted Scatterplot Smoothing (LOESS)* desenhado para ajustar os dados por intensidade e localização espacial do *spot* na lâmina.

3.5 IDENTIFICAÇÃO DE FRAGMENTOS DE DNA DIFERENCIALMENTE METILADOS

Para a identificação de fragmentos de DNA diferencialmente metilados foram considerados apenas os sinais de hibridização dos fragmentos de DNA imobilizados na lâmina que correspondem a seqüências promotoras de genes humanos conhecidos uma vez que estão presentes na lâmina fragmentos controles correspondentes à região não promotora. Para cada um desses fragmentos foi calculada, separadamente para os dois grupos de pacientes envolvidos em nosso estudo, a média do sinal de hibridização das diferentes amostras. O teste *t* de *Student* moderado, corrigido pelo método de correção múltipla *False Discovery Rate* (FDR)¹², foi aplicado para determinar a significância da diferença observada entre a média do sinal de hibridização para cada um dos fragmentos nos dois grupos de amostras. Foram considerados fragmentos diferencialmente metilados aqueles cuja diferença na média das intensidades dos sinais de hibridização apresentaram um *p* valor menor que 0.01 e FDR menor que 0.05. Em seguida, para cada um dos fragmentos diferencialmente metilados identificados através dessa abordagem estatística foi calculado um *Z score* a partir da média de intensidade do sinal de hibridização e do desvio padrão desse sinal nos dois grupos de amostras. Foram considerados fragmentos diferencialmente metilados entre os dois grupos aqueles que apresentaram um *z score* = 3.0 e *p* < 0.001).

Para reduzir o número de fragmentos de DNA diferencialmente metilados entre os dois grupos de amostras, priorizando aqueles com alterações no padrão de metilação mais freqüentes em seus respectivos grupos, selecionamos os fragmentos

que estavam diferencialmente metilados em 75% das amostras de um dos grupos e em apenas 25% das amostras do outro grupo.

3.6 CLUSTERIZAÇÃO HIERÁRQUICA

A clusterização hierárquica foi realizada utilizando um algoritmo de clusterização baseado em distância euclidiana (*Euclidian average with support*) e a significância da classificação foi testada através do método de reamostragem *Bootstrap* (número de reamostragem = 1000)¹⁰⁴. O valor de probabilidade do *bootstrap* (BP) de um cluster é o valor real da frequência com que ele aparece nas replicatas do *bootstrap*.

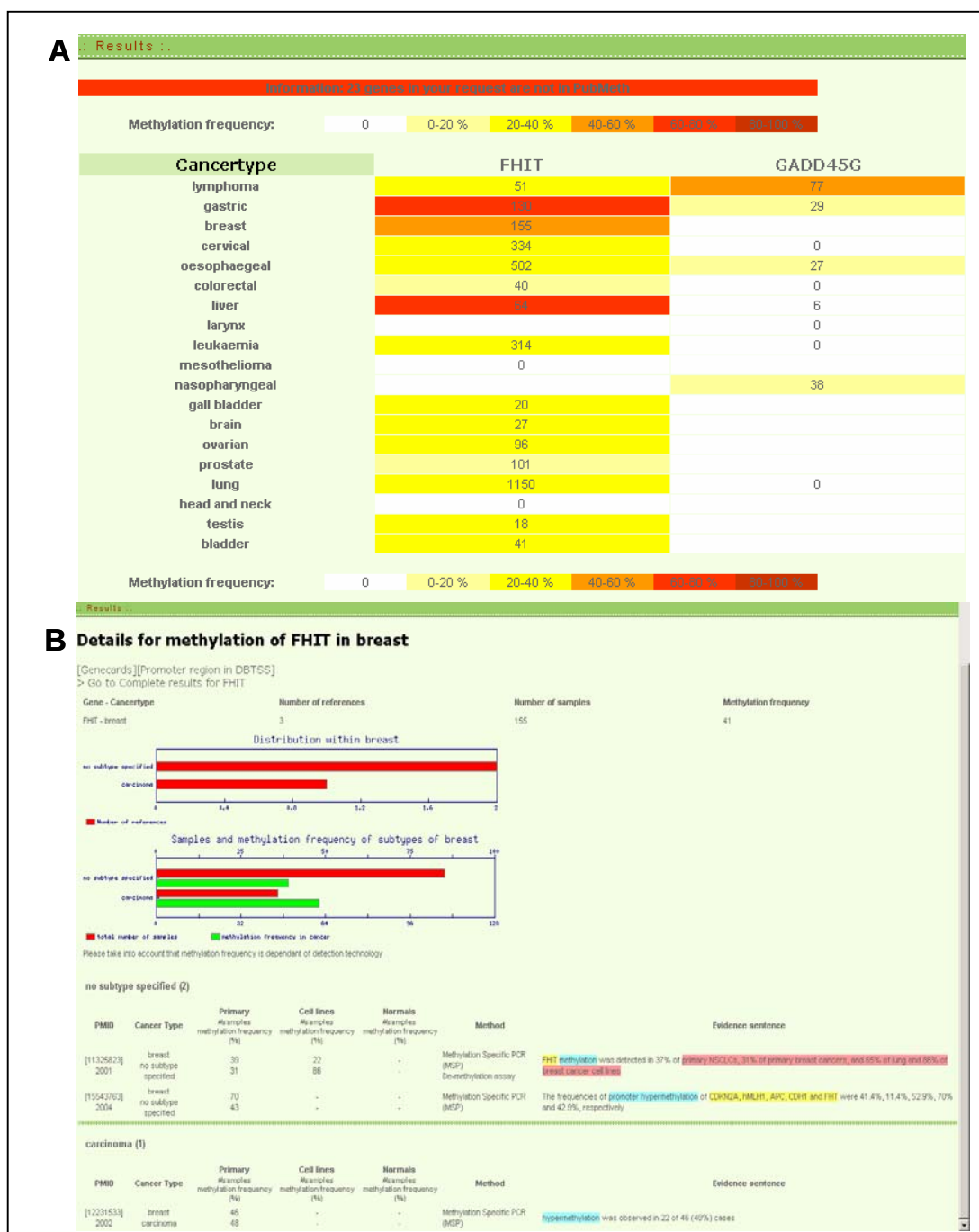
3.7 VALIDAÇÃO DOS FRAGMENTOS DE DNA DIFERENCIALMENTE METILADOS UTILIZANDO O PUBMETH E O ONCOMINE

Para validar nossos dados de metilação diferencial utilizamos informações disponíveis nos bancos de dados PubMeth⁸⁰ (www.pubmeth.org) e Oncomine⁹¹ (www.oncomine.org).

O PubMeth é um banco de dados de metilação público que inclui genes descritos como metilados em vários tipos de câncer. O PubMeth contém informações extraídas dos resumos disponíveis no Medline/PubMed (www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed) combinada à leitura e anotação manual dos resumos pré-selecionados. A pesquisa no PubMeth pode ser baseada na busca por genes específicos (em quais tipos de câncer esse gene se apresenta

metilado) ou por tipos de câncer (quais genes estão metilados em um determinado tipo ou subtipo de câncer). O resultado da análise é disponibilizado em uma tabela onde estão descritas as frequências de metilação encontradas em diferentes tipos tumorais para cada gene (Figura 9a). Além disso, uma interface gráfica fornece informações detalhadas sobre as amostras e a metodologia utilizada para a detecção do nível de metilação em cada estudo (Figura 9b).

Para a análise dos genes identificados como diferencialmente metilados no PubMeth, utilizamos o número de acesso no GenBank referente a cada um desses genes e também o nome do gene e seus aliases.



Legenda: A - Para cada gene pesquisado é possível obter informações sobre o número de amostras estudadas. As cores observadas variam de acordo com a frequência de metilação do gene em um determinado tumor (barra de frequência de metilação). B – É possível obter informações sobre o estudo e acessar o artigo de onde os dados de metilação foram extraídos.

Fonte: PubMeth⁸⁰

Figura 9 Interface gráfica disponibilizada pelo programa PubMeth.

O Oncomine é um banco de dados de expressão gênica em câncer disponibilizado pela Universidade de Michigan. Esse banco de dados contém informações referentes a 362 estudos de expressão gênica em câncer utilizando a técnica de *microarray* disponibilizados no *Gene Expression Omnibus* (GEO)⁷ e no *Stanford Microarray Database* (SMD)⁵, totalizando mais de 28880 hibridizações que abrangem 41 tipos de câncer. O dado de hibridização bruto de cada estudo é processado e normalizado independentemente por um método único que consiste na transformação logarítmica dos dados, extração da mediana, e normalização pelo desvio padrão para cada *array* hibridizado. Os estudos são nomeados seguindo a convenção: nome do primeiro autor_perfil do tipo de tecido (por exemplo, van't Veer_Breast). A descrição das amostras é padronizada de acordo com o NCI Thesaurus *terms*²³. Para cada estudo a descrição das amostras é revisada e as mesmas são agrupadas em duas ou mais classes comparáveis, com pelo menos 4 amostras cada uma, possibilitando análises de interesse que incluem tipos de câncer com seus respectivos tecidos normais, tão bem como uma variedade de subtipos de câncer e análises baseadas em dados clínicos e patológicos (Tabela 4). Após definição da análise de interesse, a expressão diferencial de cada gene é avaliada pelo teste *t* de *Student* e *False Discovery Rate* (FDR) como medida de correção da significância no caso de comparação entre duas classes de amostras ou por correlação de Pearson no caso de comparações entre mais de duas classes. Para cada uma das assinaturas de expressão diferencial do Oncomine são derivadas 3 listas dos genes melhor ranqueados pelo pvalor do teste *t* de *Student*, representando os genes 1, 5 e 10% que se apresentam mais ou menos expressos entre as análises de expressão diferencial.

Nas análises de associação entre as assinaturas gênicas diferenciais é utilizado o teste exato de Fisher.

Além dos dados de expressão gênica, o Oncomine disponibiliza ferramentas que permitem a integração desses dados com dados de localização cromossômica, domínios e famílias proteicas, anotações de ontologia gênica, vias de sinalização e metabolismo, redes de interação proteína-proteína e complexos proteicos provenientes de vários bancos de dados externos^{44,61,64,71,78,84,120} (www.ncbi.nlm.nih.gov/mapview) (www.biocarta.com). (Tabela 5).

Essas informações são armazenadas e podem ser manipuladas e visualizadas através de uma interface *web* para um gene específico ou para uma lista contendo vários genes.

Tabela 4 Resumo das análises de expressão diferencial.

Tipo	n	Significante [n (%)]
Câncer versus câncer	249	234 (93,9%)
Alteração molecular	113	96 (85%)
Normal versus normal	64	64 (100%)
Câncer versus normal	66	60 (94%)
Estádio tumoral	61	19 (31,1%)
Subtipo histológico	57	53 (93%)
Grau tumoral	56	30 (53,6%)
Prognóstico	39	18 (46,1%)
Miscelânea	415	273 (65,8%)

Fonte: Adaptado de RHODES et al.⁹¹. Dado disponibilizado para 264 estudos de *microarray* atualmente disponíveis.

Legenda: As análises de expressão diferencial foram realizadas para os dados de 264 estudos de *microarray* e são de 9 tipos diferentes. O número de análises realizadas é dado como o número e a porcentagem de análises que resultaram em assinaturas de expressão diferencial estatisticamente significativa. Análises “câncer versus câncer” compararam um tipo de câncer a um ou mais outros tipos. Análises de “alteração molecular” compararam alterações moleculares distintas em um dado tipo de câncer. Análise “normal versus normal” comparou um tipo de tecido normal a outros tipos de normais. Análises “câncer versus normal” compararam um tipo de câncer com seu respectivo tecido normal de origem. Análises de “estádio tumoral”, “subtipo histológico” e “grau tumoral” compararam tumores de um mesmo tipo com base nesses parâmetros patológicos e análises de “prognóstico” compararam tumores de pacientes de melhor prognóstico com tumores de pior prognóstico. Análises “miscelânea” compararam amostras baseadas em resposta a tratamento, status de infecção viral, status de biomarcadores entre outros.

Tabela 5 Fontes externas integradas ao banco de dados do Oncomine.

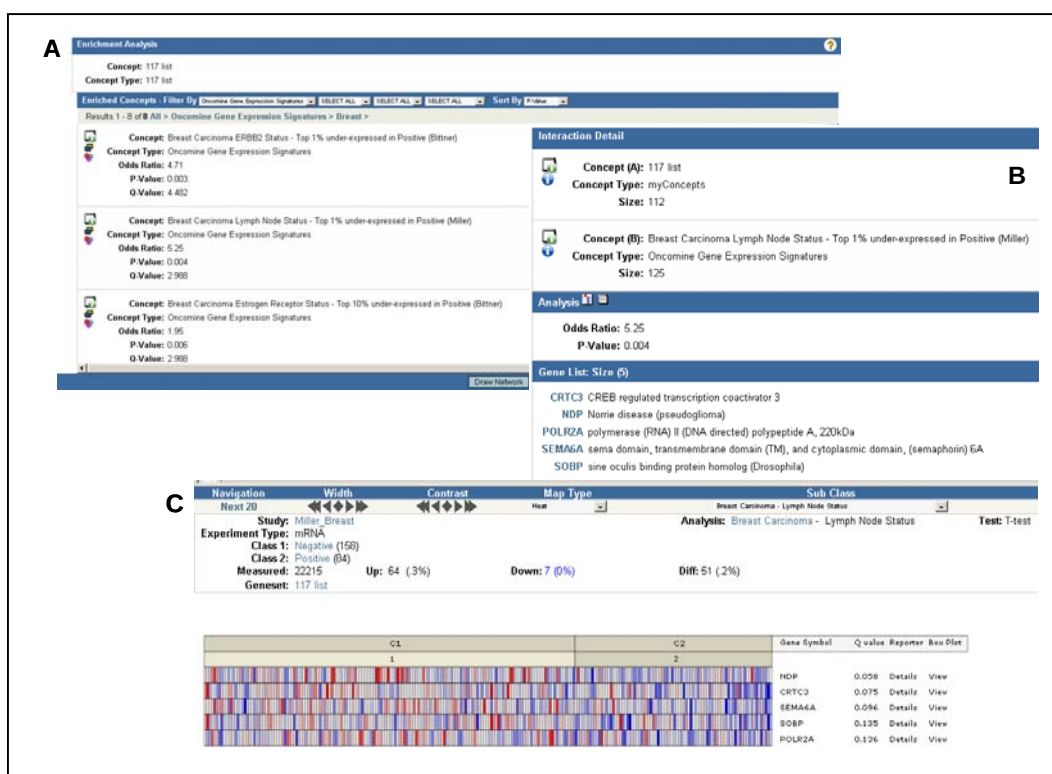
Fonte	Tipo
GenBank	Seqüência de nucleotídeos
Entrez Gene	Gene
Swisspro/Trembl	Proteína
Inparanoid	Ortólogos
Unigene	Cluster de EST
Gene Ontology	Processo biológico
	Componente celular
	Função molecular
InterPro	Famílias e domínios proteicas
Oncomine	Assinaturas de câncer
Connectivity Map	Assinaturas de drogas
TRANSFAC	Sítios de ligação ao promotor
picTAR	Alvos para miRNA
Broad	Motivo conservado no promotor
NCBI	Braço cromossômico
	Subregião cromossômica
Biocarta	Vias de sinalização
KEGG	Vias metabólicas
HPRD	Interação proteína-proteína
PINdb	Complexo protéico nuclear

Fonte: Adaptado de RHODES et al.⁹¹.

Em um primeiro momento, buscamos evidências de diminuição da expressão dos genes correspondentes aos nossos fragmentos diferencialmente metilados nas assinaturas de expressão gênica de câncer de mama disponibilizadas pelo Oncomine. Também utilizamos o Oncomine em busca de um enriquecimento de categorias de ontologia gênica (GO) e de regiões cromossômicas relevantes para o câncer de mama entre os genes diferencialmente metilados entre nossos dois grupos de pacientes.

Para a análise dos genes identificados como diferencialmente metilados no Oncomine, utilizamos o número de acesso no GenBank referente a cada um desses genes e o nome do gene e seus aliases. Após a submissão da lista contendo os números de acesso referente aos genes o programa realiza análises de associação pareadas entre a lista de genes de interesse e as assinaturas previamente definidas e armazenadas nos bancos de dados do Oncomine através do teste exato de Fisher. Essas análises levam em consideração o conjunto total de genes imobilizados no

microarray enriquecido para seqüências promotoras a fim de evitar um viés em relação aos genes representados na lâmina de *microarray* utilizada em nosso estudo. Para tanto, foi submetida ao programa a lista contendo os números de acesso referente aos genes da lâmina. As associações que representam um subgrupo de associações estatisticamente significativas são então disponibilizadas ($p < 1e-6$).



Legenda: A – Assinaturas de expressão de mama identificadas no Oncomine. B – Dados do estudo onde a assinatura foi identificada e lista dos genes em comum à minha assinatura. C – perfil de expressão dos genes comuns na assinatura relacionada (em azul = genes com expressão diminuída entre dois grupos de amostras comparados; em vermelho = genes com expressão aumentada entre dois grupos de amostras comparados).

Figura 10 Interface gráfica disponibilizada pelo programa Oncomine.

3.8 VALIDAÇÃO DA ASSINATURA DE METILAÇÃO UTILIZANDO O DADO DE EXPRESSÃO DO ESTUDO DESENVOLVIDO POR VAN'T VEER ET AL.

Em uma segunda etapa de validação dos nossos dados de metilação diferencial, utilizamos somente os dados de expressão gênica do estudo desenvolvido por VAN'T VEER et al.¹⁰⁸ disponibilizado pelo Oncomine. Nesse estudo, VAN'T VEER et al.¹⁰⁸ avaliou 78 amostras tumorais de pacientes com carcinoma ductal invasivo de mama, linfonodo negativos que não desenvolveram metástase (n=44) comparadas a pacientes que desenvolveram metástase em pelo menos 5 anos (n=34), exatamente como as pacientes avaliadas em nosso estudo. Nesse estudo a partir de 5000 genes diferencialmente expressos entre os dois grupos de pacientes ($p = 0.05$) foi estabelecida uma assinatura de expressão composta pelo perfil de expressão de 70 genes capaz de discriminar o grupo de pior prognóstico do grupo de melhor prognóstico.

Foram feitas duas análises distintas. Inicialmente, comparamos a nossa assinatura de metilação com a assinatura de expressão gênica de 70 genes de VAN'T VEER et al.¹⁰⁸ a fim de verificar quantos dos genes correspondentes aos nossos fragmentos diferencialmente metilados se encontravam diferencialmente expressos. Também verificamos se os genes que hipermetilados em cada um dos dois grupos de pacientes do nosso estudo apresentavam evidência de diminuição de expressão no grupo correspondente do estudo de VAN'T VEER et al.

Na segunda análise, dados de expressão para cada um dos genes correspondentes aos nossos fragmentos diferencialmente metilados foram obtidos a

partir do estudo de VAN'T VEER et al.¹⁰⁸. Essa informação de expressão foi então utilizada para clusterizar as pacientes do estudo desenvolvido por VAN'T VEER et al.¹⁰⁸. A clusterização hierárquica foi realizada utilizando um algoritmo de clusterização baseado em distância euclidiana (*Euclidian average with support*) e a significância da classificação foi testada através do método de reamostragem *Bootstrap* (n de reamostragem = 1000)¹⁰⁴.

RESULTADOS

4 RESULTADOS

4.1 CARACTERIZAÇÃO DO PERFIL DE METILAÇÃO DE AMOSTRAS TUMORAIS DE MAMA

4.1.1 Hibridização de fragmentos diferencialmente metilados

Para a caracterização do perfil de metilação das amostras tumorais de mama utilizamos a metodologia de hibridização de fragmentos diferencialmente metilados desenvolvida por WANG et al.¹¹⁵. Neste trabalho foram hibridizadas 21 amostras de carcinomas ductais invasivos de mama de pacientes sem comprometimento linfonodal das quais 13 não desenvolveram doença metastática e 8 desenvolveram metástase à distância em um período de pelo menos 5 anos. O tempo médio de seguimento foi de 10 anos para o grupo sem doença metastática e de 7 anos para o grupo que desenvolveu metástase à distância.

Resumidamente, o DNA genômico é digerido com a enzima *HpaII* que reconhece o sitio de restrição CGCG e possui atividade de clivagem sensível a metilação. Ilhas de CpG, em geral, possuem vários sítios para a enzima *HpaII* localizados uns próximos aos outros. Após a digestão do DNA genômico em regiões não metiladas, são geradas extremidades coesivas nas quais é possível ligar uma sequência adaptadora. Iniciadores complementares a essa sequência adaptadora são então utilizados para amplificar as regiões ricas em CpG não metiladas. Devido a proximidade entre os sítios da enzima *HpaII* os fragmentos gerados são pequenos (50-500pb) e podem ser eficientemente amplificados por PCR. Esses fragmentos

amplificados são então marcados com fluoróforos e utilizados como sonda na hibridização de *microarrays* contendo seqüências promotoras ricas em CpG.

A presença de metilação nos sítios da enzima *HpaII* impede a digestão do DNA e conseqüentemente a ligação dos adaptadores e a amplificação das regiões ricas em CpG. A digestão do DNA nesta situação ocorre fora das ilhas de CpG, gerando fragmentos de DNA mais longos que apresentam baixa eficiência de amplificação, estando portanto sub-representados na sonda utilizada para hibridização. Assim sendo, o sinal de hibridização é proporcional a representatividade daquela região na sonda, a qual por sua vez é inversamente proporcional ao nível de metilação.

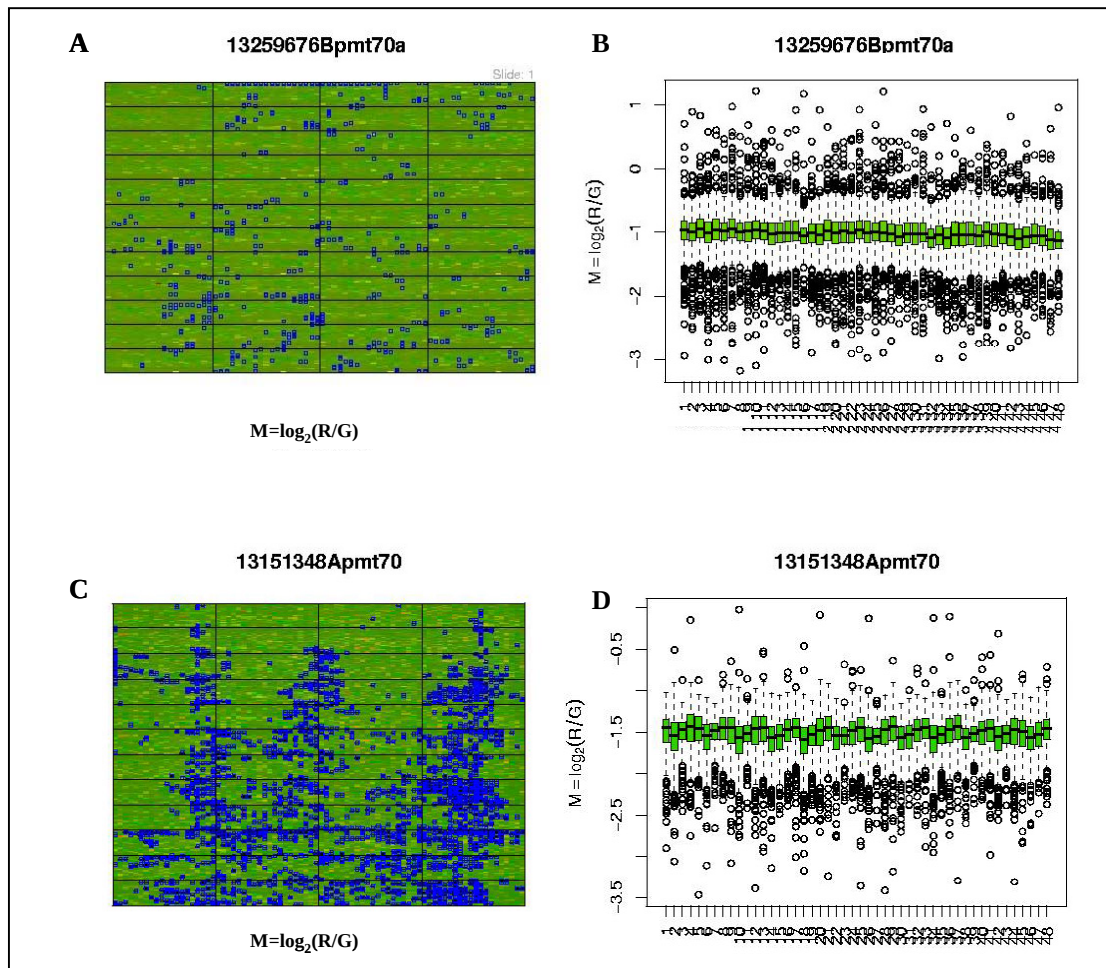
4.1.2 Análise de qualidade das hibridizações e normalização dos sinais de hibridização

O dado bruto da hibridização das 21 amostras tumorais de mama foi submetido à análise de qualidade e ao protocolo de normalização da intensidade do sinal de hibridização.

Na primeira etapa dessa análise, foram excluídos os pontos saturados e os pontos que apresentaram emissão de fluorescência igual ou menor ao sinal emitido pelo ruído local da lâmina (*background* local). O número de pontos excluídos foi adotado como um primeiro critério para avaliar a qualidade da hibridização das lâminas e foram consideradas de baixa qualidade lâminas com mais de 40% dos pontos excluídos. Na figura 11A, podemos observar que em uma hibridização de boa qualidade, a razão da intensidade do sinal de hibridização (amostra/referência) entre os pontos é constante e que menos de 20% dos pontos representados em azul foram

eliminados. No entanto, para uma hibridização de baixa qualidade, representada na 11C observamos que 1/3 do total de pontos da lâmina foram eliminados por estarem saturados ou com intensidade de sinal muito baixa. Um segundo critério adotado para avaliação da qualidade da hibridização foi a distribuição da razão de intensidade do sinal de hibridização entre os *sub-arrays* de uma mesma lâmina. Na figura 11B, notamos que a distribuição da razão de intensidade do sinal para um experimento de hibridização representado sob a forma de *Box plots* não apresentou grandes variações sistemáticas indicando boa qualidade da hibridização e também da imobilização das seqüências na lâmina (*print-tip grid*). Diferentemente do observado em 11B, a figura 11D representa uma lâmina com baixa qualidade de hibridização independente da boa imobilização das seqüências na lâmina.

Aplicados os critérios de qualidade, com exceção da hibridização de uma amostra do grupo de pacientes que desenvolveram metástase à distância, todas as demais hibridizações apresentaram boa qualidade e seguiram em nossas análises.



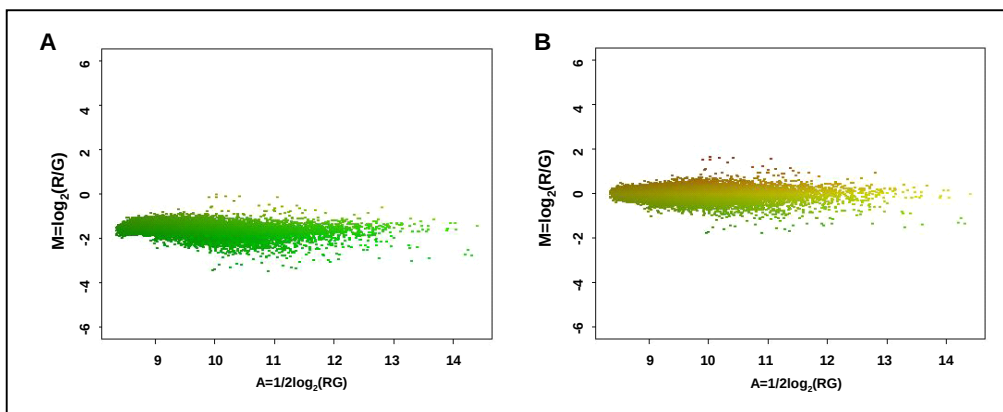
Legenda: **A e C:** Distribuição da intensidade de sinal dos pontos da lâmina (Cy5/Cy3). **B e D:** Box plots representando distribuição da razão de intensidade ($M = \log_2 R/G$) para cada um dos pentes de espotagem presentes nas lâminas (*print-tip grids*).

Figura 11 Análise da qualidade da hibridização de duas amostras tumorais independentes.

Uma vez realizada a análise de qualidade das hibridizações, foi feita a normalização da intensidade do sinal de hibridização de cada lâmina individualmente (Figura 12 gráfico *MA-plot*) para minimizar as variações experimentais. Para a normalização dos sinais de hibridização, utilizamos o método de *LOESS* desenhado para ajustar variações sistemáticas não-lineares, do tipo intensidade dependente, principalmente para pontos de baixa intensidade¹²².

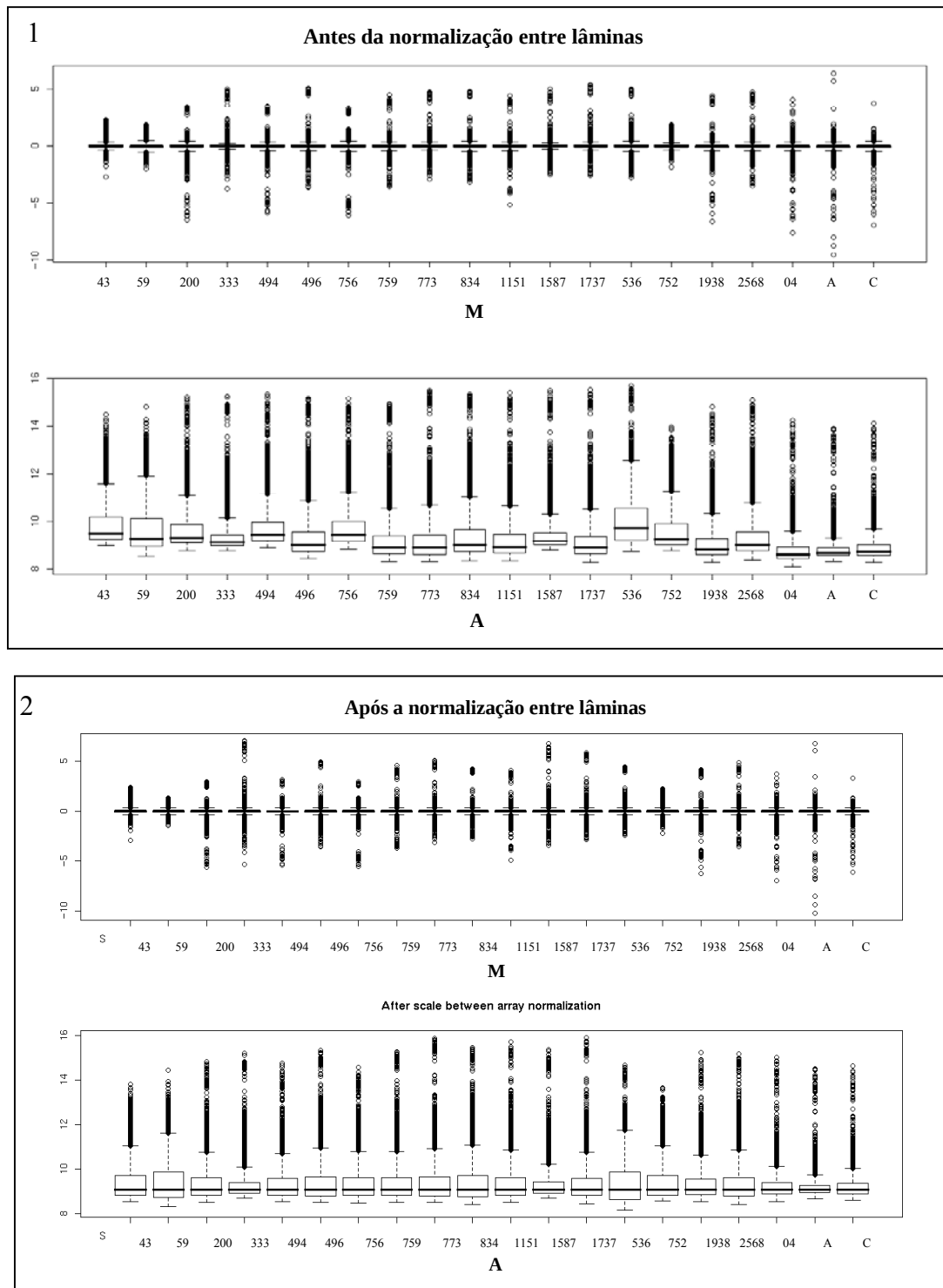
O painel A da figura 12 representa um *MA-plot* do dado de hibridização não normalizado de uma amostra representativa, ao passo que o painel B representa o mesmo dado normalizado. Este tipo de representação permite visualizar como variações experimentais sistemáticas, tais como a incorporação diferencial dos corantes Cy3 e Cy5 pode afetar de forma não linear sinais de hibridização de baixa e alta intensidade. Nos gráficos *MA-plot*, M representa a razão entre a intensidade R e G, em $\log_2$ ($M = \log_2 R/G$) e A (intensidade local) a média entre as intensidades R e G em $\log_2$ ($A = \frac{1}{2} \log_2(RG)$), onde R e G referem-se à incorporação dos fluoróforos Cy5 (*RED*) e Cy3 (*GREEN*). No painel 12B notamos que aplicada a normalização o desvio da nuvem de pontos em 12A pôde ser corrigido (trazida para o zero).

Após a normalização das lâminas individualmente foi feita a normalização entre todas as lâminas, permitindo que todas as hibridizações possam ser comparadas entre si. Na figura 13, podemos observar gráficos *boxplots* representativos de todas as hibridizações realizadas em nosso estudo. O gráfico *boxplot* dá uma idéia da posição, dispersão, assimetria, caudas e medidas discrepantes de um conjunto de dados. Na figura 13A, é possível observar que as hibridizações variam entre si e que uma vez feita a normalização, figura 13B, as variações dos *boxplots* são corrigidas.



Legenda: A e B: Gráfico tipo MA-plot que representa as razões e as médias da intensidade dos sinais do dado antes e depois da normalização, onde há uma correção da distorção dos pontos. Vale ressaltar que para esta representação, os pontos em vermelho são aqueles que apresentam razão de intensidade M , em escala logarítmica maior que zero (amplificação dos fragmentos); em amarelo estão representados os pontos com razão de intensidade $M = 0$ e sem hibridização diferencial; e por último, os pontos verdes aqueles em que a razão é menor que zero.

Figura 12 Normalização intra lâminas por *LOESS*.



Legenda: M representa a diferença entre a intensidade R e G, em $\log_2$ ($M = \log_2 R/G$) e A (intensidade local) a média entre as intensidades R e G em $\log_2$ ($A = 1/2 \log_2 RG$).

Figura 13 Representação gráfica da normalização entre lâminas por *LOESS*.

4.2 IDENTIFICAÇÃO DE FRAGMENTOS DE DNA DIFERENCIALMENTE METILADOS

Após a análise de qualidade das hibridizações e normalização das intensidades de sinal passamos para a identificação de fragmentos de DNA diferencialmente metilados entre o grupo que não apresentou doença metastática e o grupo que desenvolveu metástase à distância. Foram considerados apenas os sinais de hibridização dos fragmentos de DNA imobilizados na lâmina que correspondem a seqüências promotoras de genes humanos conhecidos. Para cada um desses fragmentos foi calculada, separadamente para os dois grupos de pacientes envolvidos em nosso estudo, o $\log_2$ R/G da média do sinal de hibridização das diferentes amostras. Foram identificados 234 fragmentos de DNA diferencialmente metilados cuja diferença na média das intensidades de hibridização entre os dois grupos de pacientes apresentou um $p < 0.01$ e FDR < 0.05 .

O padrão de metilação desses 234 fragmentos foi então utilizado para agrupar as 20 amostras incluídas em nosso estudo. O resultado da clusterização hierárquica das amostras está representado na figura 14. Os genes presentes a figura 14 estão descritos na mesma ordem no anexo 1. Podemos observar que o padrão de metilação dos 234 fragmentos de DNA foi capaz de discriminar com 100% de acerto as pacientes que desenvolveram metástase em relação as que não desenvolveram metástase. O valor de *bootstrap* observado para o ramo principal de separação dos dois grupos de amostras nessa clusterização hierárquica foi de 100% (BP).

Para aumentar a estringência da nossa análise, os resultados de hibridização dos 234 fragmentos de DNA diferencialmente metilados foram submetidos a uma

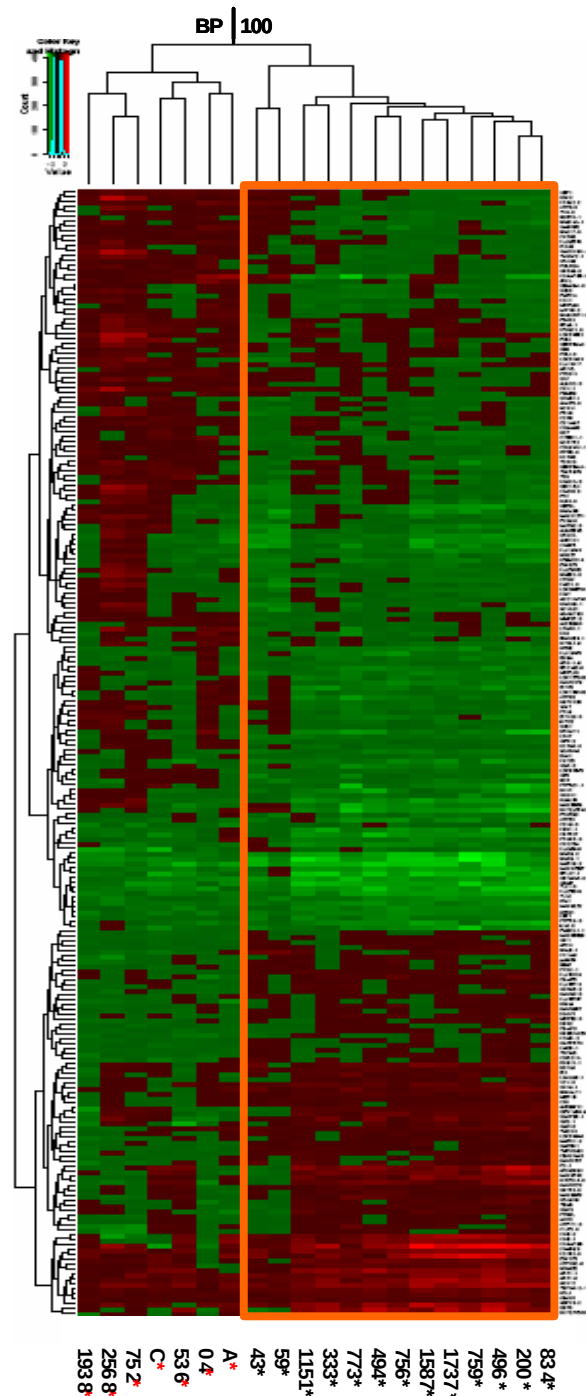
segunda análise estatística na qual foi calculado um *z score* para cada um dos 234 fragmentos. O *z score* reflete a divergência de um dado experimental em relação a média dos dados analisados, levando em conta o desvio padrão do conjunto de dados. Com esta abordagem identificamos 117 fragmentos de DNA diferencialmente metilados entre os dois grupos de pacientes com $z\ score = 3.0$, $p < 0.001$. Desses 117 fragmentos, 47 estão hipermetilados no grupo de amostras de pacientes que desenvolveram doença metastática e 70 estão hipermetilados no grupo de amostras de pacientes que não desenvolveram metástase.

Uma nova clusterização das amostras foi realizada utilizando os dados de hibridização dos 117 fragmentos de DNA diferencialmente metilados. Como podemos observar na figura 15, mesmo reduzindo o número de fragmentos de DNA analisados foi possível separar as amostras tumorais em dois grupos distintos que discriminaram com 100% de acerto as pacientes que desenvolveram metástase à distância das que não desenvolveram metástase à distância. Os genes presentes na figura 15 estão descritos na mesma ordem no anexo 2. O valor de *bootstrap* observado para a separação do ramo principal dos dois grupos de amostras nessa clusterização hierárquica foi de 92% (BP).

Para reduzir o número de fragmentos diferencialmente metilados entre os dois grupos de amostras, priorizando aqueles com alterações no padrão de metilação mais frequentes em seus respectivos grupos, selecionamos os fragmentos que estavam diferencialmente metilados em 75% das amostras de um dos grupos e em apenas 25% das amostras do outro grupo. Utilizando essa estratégia selecionamos 52 fragmentos de DNA diferencialmente metilados dos quais 18 estão hipermetilados no

grupo de pacientes com doença metastática e 34 hipermetilados no grupo de pacientes que não desenvolveram metástase.

Como podemos observar na figura 16, os 52 fragmentos de DNA diferencialmente metilados foram capazes de identificar dois grupos distintos de amostras tumorais que discriminaram com 100% de acerto as pacientes que desenvolveram metástase à distância em relação as que não desenvolveram metástase à distância. Os genes presentes a figura 16 estão descritos na mesma ordem no anexo 3. O valor de *bootstrap* observado para a separação dos dois grupos de amostras nessa clusterização hierárquica foi de 88% (BP) para o ramo principal.



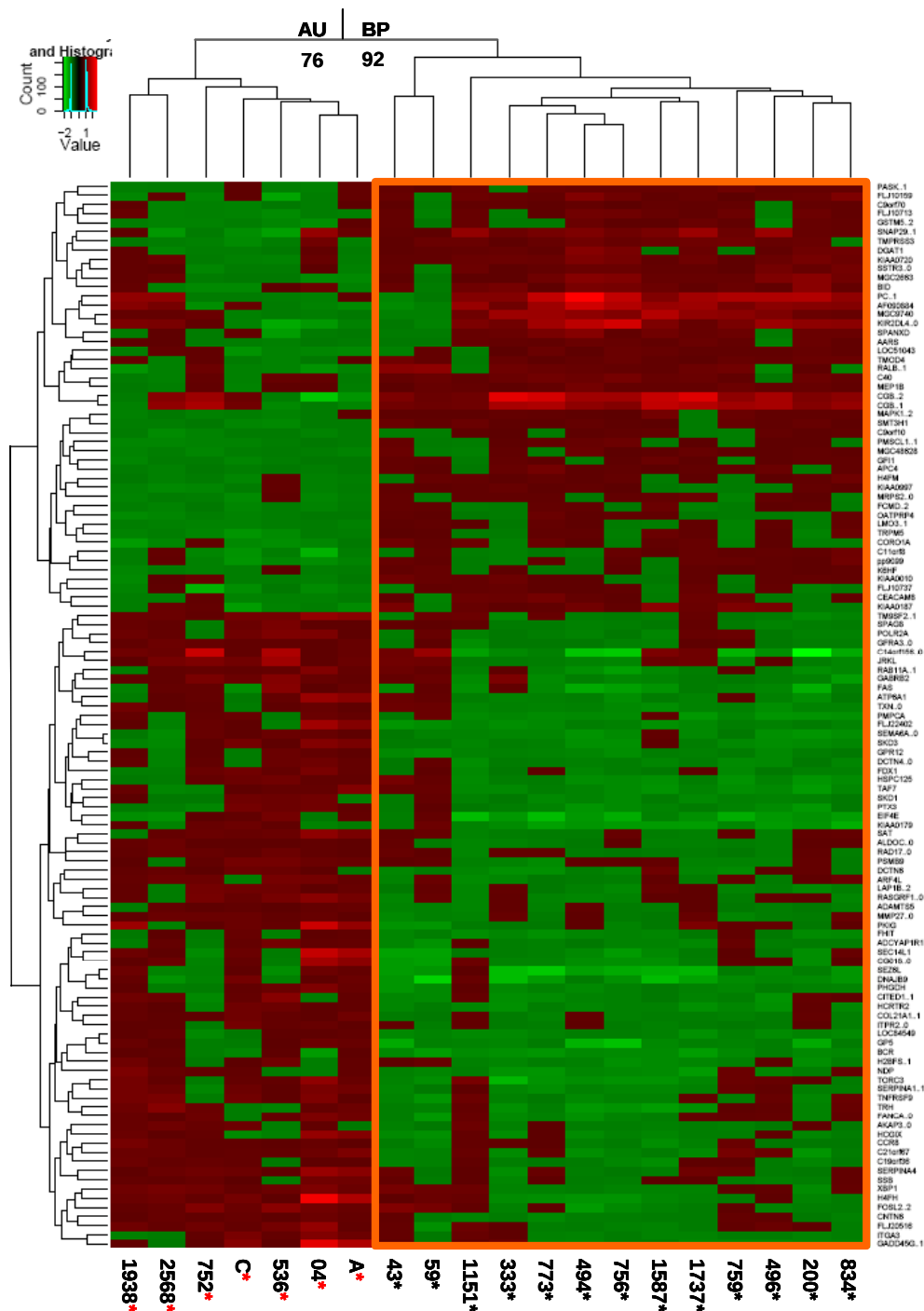
Legenda: Na escala de cores o vermelho representa valores positivos de intensidade de sinal de hibridização e o verde, valores negativos.

Os (*) representam amostras de pacientes linfonodos negativos com recidiva e (*) amostras de pacientes linfonodos negativos sem recidiva. O quadro laranja separa visualmente os dois grupos de amostra.

BP - valor de *bootstrap*.

A ordem dos genes apresentados nesta figura é a mesma apresentada na tabela do anexo 1.

Figura 14 Clusterização hierárquica das amostras com base no perfil de metilação dos 234 fragmentos de DNA diferencialmente metilados.



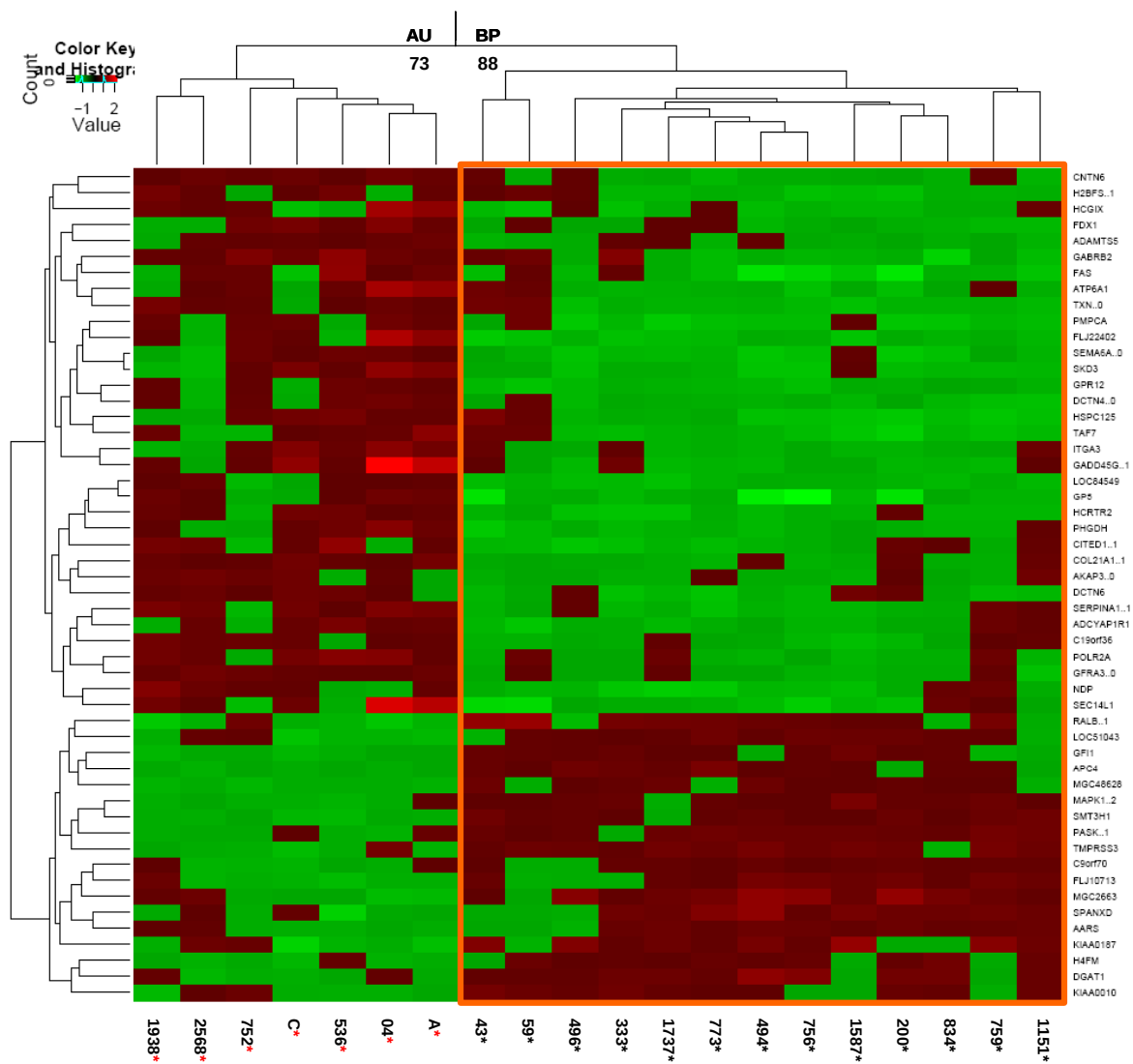
Legenda: Na escala de cores o vermelho representa valores positivos de intensidade de sinal e o verde, valores negativos.

Os (*) representam amostras de que desenvolveram metástase à distância e (*) amostras de pacientes que não desenvolveram metástase à distância. O quadro laranja separa visualmente os dois grupos de amostra.

BP - valor de *bootstrap*.

A ordem dos genes apresentados nesta figura é a mesma apresentada na tabela do anexo 2.

Figura 15 Clusterização hierárquica das amostras com base no perfil de metilação dos 117 fragmentos de DNA diferencionalmente metilados.



Legenda: Na escala de cores o vermelho representa valores positivos de intensidade de sinal e o verde, valores negativos.

Os (*) representam amostras de pacientes que desenvolveram metástase à distância (*) amostras de pacientes que não desenvolveram metástase à distância. O quadro laranja separa visualmente os dois grupos de amostra.

BP - valor de *bootstrap*.

A ordem dos genes apresentados nesta figura é a mesma apresentada na tabela do anexo 3.

Figura 16 Clusterização hierárquica das amostras com base no perfil de metilação dos 52 fragmentos de DNA diferencialmente metilados.

4.3 VALIDAÇÃO DOS FRAGMENTOS DE DNA DIFERENCIALMENTE METILADOS UTILIZANDO INFORMAÇÕES DISPONÍVEIS EM BANCO DE DADOS PÚBLICOS.

Para validar nossos dados de metilação diferencial utilizamos informações disponíveis nos bancos de dados PubMeth⁸⁰ e Oncomine⁹¹. Para essa análise optamos por trabalhar com a lista de genes correspondente aos 117 fragmentos de DNA diferencialmente metilados por se tratar de uma lista gerada a partir de uma análise estatística bastante estridente e ao mesmo tempo com um número suficiente de genes para a realização das análises.

4.3.1 Validação dos fragmentos de DNA diferencialmente metilados utilizando dados de metilação diferencial do PubMeth

O PubMeth é um banco de dados público que inclui informações sobre o status de metilação de aproximadamente 440 genes humanos em 43 tipos de tumores. Essas informações são extraídas dos resumos disponíveis no Medline/PubMed (www.ncbi.nlm.nih.gov.br/sites/entrez?db=PubMed) combinada à leitura e anotação manual dos resumos pré-selecionados. A pesquisa no PubMeth pode ser feita para um gene específico, buscando em quais tipos de câncer esse gene se apresenta metilado, ou para um tipo ou subtipo específico de tumor buscando quais genes estão metilados.

Utilizando o banco de dados PubMeth, verificamos, para cada um dos genes correspondentes aos 117 fragmentos diferencialmente metilados, se havia evidência de alteração no padrão de metilação em diferentes tipos de tumores. A busca por

informações de genes catalogados no PubMeth foi realizada utilizando tanto o número de acesso do GenBank quanto o nome oficial do gene e foi possível a identificação de 114 dos 117 fragmentos diferencialmente metilados pelo programa.

Dos 114 genes analisados apenas 1 (FHIT) havia sido descrito como freqüentemente metilado em tumores de mama, porém (RASGRF1, GADD45G, ADAMTS5, FAS, PLEKHG5) se mostraram metilados em outros tipos de tumores. Como podemos observar na tabela 6, o gene FHIT é o gene que está metilado em um maior número de tumores (15) seguido pelo gene GADD45G (4).

É importante ressaltar que, dados da literatura mostram que há uma correlação negativa entre o silenciamento do gene FHIT e a progressão tumoral de mama^{36,37,124}. Em um destes trabalhos, ZOCHBAUER-MULLER et al.¹²⁴ avaliaram o status de metilação por *Methylation Specific PCR* (MSP-PCR), a expressão da proteína FHIT por imunoistoquímica e perda de heterozigose (LOH) em 36 amostras de tumores de mama em diferentes estádios da doença (I, IIa, IIb, IIIa, IIIbe IV). Juntos os resultados destas análises demonstraram que a metilação do gene FHIT é um evento importante para o silenciamento deste gene em tumores de mama e está associada tanto a perda da expressão do RNA mensageiro quanto da proteína. No entanto, não houve correlação entre o status de metilação e o estágio tumoral.

Tabela 6 Frequência de metilação dos genes em diferentes tumores disponibilizada pelo PubMeth.

Tipo de câncer	<i>FHIT</i>	<i>RASGRF1</i>	<i>GADD45G</i>	<i>ADAMTS5</i>	<i>FAS</i>	<i>PLEKHG5</i>
Linfoma	20-40%		40-60%			
Gástrico	60-80%		0-20%			
Mama	40-60%					
Cervical	20-40%					
Esôfago	20-40%		0-20%			
Colorectal	0-20%				20-40%	
Fígado	60-80%					
Laringe						
Leucemia	20-40%			40-60%		
Nasofarínge			0-20%			
Vesícula biliar	20-40%					
Cérebro	20-40%					
Ovário	20-40%					
Próstata	0-20%				0-20%	
Pulmão	20-40%	20-40%				
Cabeça e pescoço						
Testículo	20-40%					
Tumor de Wilms						60-80%
Bexiga	20-40%				0-20%	80-100%

Legenda: 0-20%, 20-40%, 40-60%, 60-80% e 80-100% representam a porcentagem de amostras de tumores metilados dentre os analisados.

4.3.2 Validação dos fragmentos de DNA diferencialmente metilados utilizando dados de expressão gênica do Oncomine.

Em seguida, utilizando o banco de dados de expressão gênica em câncer Oncomine, buscamos evidências de diminuição de expressão em tumores de mama para os genes correspondentes aos 117 fragmentos de DNA diferencialmente metilados identificados no nosso estudo.

O Oncomine é um banco de dados de expressão gênica em câncer que contém informações referentes a 362 estudos de expressão gênica em câncer utilizando a técnica de *microarray* disponibilizados no *Gene Expression Omnibus* (GEO)⁷ e no *Stanford Microarray Database* (SMD)⁵, totalizando mais de 28880 hibridizações que abrangem 41 tipos de câncer.

Dos 117 genes 113 foram relacionados ao banco de dados do Oncomine. Como apresentado na tabela 7, 38 dos 113 genes analisados apresentaram evidências de diminuição de expressão em 5 assinaturas de mama de acordo com os dados armazenados no Oncomine. Desses 38 genes 6 estavam entre os 10% dos genes que se apresentam menos expressos em tumores ERBB2 positivos quando comparados a negativos, 5 estavam entre os 1% dos genes que se apresentam menos expressos em amostras de pacientes com comprometimento de linfonodos comparados a pacientes sem comprometimento de linfonodos, 23 e 21 estavam entre os 10% dos genes que se encontram menos expressos em amostras positivas para receptor de estrógeno em duas assinaturas independentes, e 23 estavam entre os 10% dos genes que se apresentam menos expressos em amostras positivas para receptor de progesterona. Dos 38 genes analisados, 27 apresentam evidência de diminuição de expressão em mais de um estudo e 5 genes (CORO1A, FANCA, HCRTR2, SEZ6L e SFXN3)

tiveram sua expressão induzida na linhagem tumoral de mama MCF7 tratada com ácido Valpróico, um inibidor de histona deacetilase, quando comparada à linhagem não tratada.

Nesta análise também observamos evidências de diminuição de expressão de aproximadamente 50 dos 113 genes diferencialmente metilados em pelo menos 3 outros tumores, tais como tumores de próstata, de cólon, gástrico, bexiga, endométrio, ovário, testículo, mesotelioma, além de leucemias e linfomas.

Vale ressaltar que 17 dos 113 genes analisados apresentaram evidências de diminuição de expressão em 2 assinaturas que estão envolvidas com recorrência local e metástase à distância. Desses 17 genes 4 estavam entre os 1% dos genes que se apresentam menos expressos em tumores de pacientes com câncer colorectal que desenvolveram recorrência local quando comparados a pacientes que não desenvolveram (EIF4E, FANCA, FAS, TIPIN) e 13 estavam entre os 5% dos genes que se apresentam menos expressos em amostras de pacientes com câncer de endométrio que desenvolveram metástase à distância comparados a pacientes que não desenvolveram doença metastática (ANAPC4, CCR8, COL21A, DNAJB9, GFRA3, NDP, NEIL, PKIG, SEC14L1, SERPINA 1, SERPINA 4, SSTR3, TXN).

Tabela 7 Lista de assinaturas de expressão gênica de mama do Oncomine com evidência de diminuição de expressão de genes da lista 117 genes.

Genes	ERBB2+	Linfonodo+	RE ⁺ ¹	RE ⁺ ²	RP+
AARS			X		X
ACY3				X	
ANAPC4	X				
BID			X	X	X
BMS1			X	X	X
C6orf66			X	X	X
CORO1A			X		X
CRTC3		X			
DNAJC11			X	X	X
FANCA			X	X	X
FAS			X	X	X
FDX1			X	X	X
GABRB2	X				
GFI1			X		X
GFRA3			X	X	
HIST1H4I					
JRKL			X	X	X
KIR2DL4				X	
MAPK1			X	X	X
MPPED2	X				
NDP	X	X			
PHGDH			X	X	X
POLR2A		X			
PSMB9			X		X
PTX3	X			X	X
RASGRIF1				X	X
RRP1B			X		X
SAT1			X	X	X
SEC14L1			X	X	X
SEMA6A	X	X			
SLCO5A1				X	
SOBP		X		X	
SUMO3			X		X
TIPIN			X	X	X
TNFRSF9			X	X	X
TXN			X		X
UBE3C			X		

Legenda: ERBB2+: 1% de genes menos expressos em amostras tumorais ERBB2+ quando comparadas a amostras ERBB2 ($p = 0.003$) (Bittner); Linfonodo+: 1% de genes menos expressos em amostras tumorais de pacientes linfonodo positivas quando comparadas a pacientes linfonodo negativas ($p = 0.004$) (Miller); RE⁺¹: 10% de genes menos expressos em amostras tumorais de mama positivas para RE quando comparadas a negativas ($p = 0.006$) (Bittner); RE⁺²: 10% genes menos expressos em amostras tumorais de mama positivas para RE quando comparadas a negativas ($p = 0.007$) (Ginestier); RP+: 10% de genes menos expressos em amostras tumorais de mama positivas para RP quando comparadas a negativas ($p = 0.01$) (Ginestier).

4.3.3 Validação da assinatura de metilação utilizando o dado de expressão do estudo desenvolvido por VAN'T VEER et al.¹⁰⁸

Em uma segunda etapa de validação dos nossos dados de metilação diferencial, utilizamos somente os dados de expressão gênica do estudo desenvolvido por VAN'T VEER et al. disponibilizado pelo Oncomine. Nesse estudo, VAN'T VEER et al. avaliaram 78 amostras tumorais de pacientes com carcinoma ductal invasivo de mama, linfonodo negativos que não desenvolveram metástase (n=44) comparadas a pacientes que desenvolveram metástase em pelo menos 5 anos (n=34), exatamente como as pacientes avaliadas em nosso estudo. Nesse estudo a partir dos 5000 genes diferencialmente expressos entre os dois grupos de pacientes ($p = 0.05$) foi estabelecida uma assinatura de expressão composta pelo perfil de expressão de 70 genes capaz de discriminar o grupo de pior prognóstico do grupo de melhor prognóstico.

Inicialmente, comparamos a nossa assinatura de metilação composta pelos 117 fragmentos de DNA diferencialmente metilados com a assinatura de expressão gênica de 70 genes do estudo de VAN'T VEER et al. Nessa comparação não identificamos genes comuns às duas assinaturas. Em seguida, verificamos se os genes correspondentes aos fragmentos hipermetilados em cada um dos dois grupos de pacientes do nosso estudo apresentavam evidência de diminuição de expressão no grupo correspondente do estudo de VAN'T VEER et al. Como descrito anteriormente, dos 117 genes diferencialmente metilados entre os dois grupos de amostras, 47 estavam hipermetilados nas amostras das pacientes que desenvolveram doença metastática e 70 hipermetilados nas amostras das pacientes que não desenvolveram metástase à distância. Dois genes (RALB e TMPRSS3) dos 47 genes

hipermetilados no grupo de pacientes com doença metastática, estavam entre os genes com expressão diminuída no grupo de pacientes com doença metastática do estudo de VAN'T VEER et al. Por outro lado, dentre os 70 genes hipermetilados nas amostras das pacientes que não desenvolveram metástase à distância, 10 (CLPB, FANCA, GFR3A, GP5, PHGDH, PTX3, RASGRF1, RRP1B, TIPIN e TXN) estavam entre os genes com expressão diminuída nos tumores de pacientes sem doença metastática do estudo da VAN'T VEER et al.¹⁰⁸

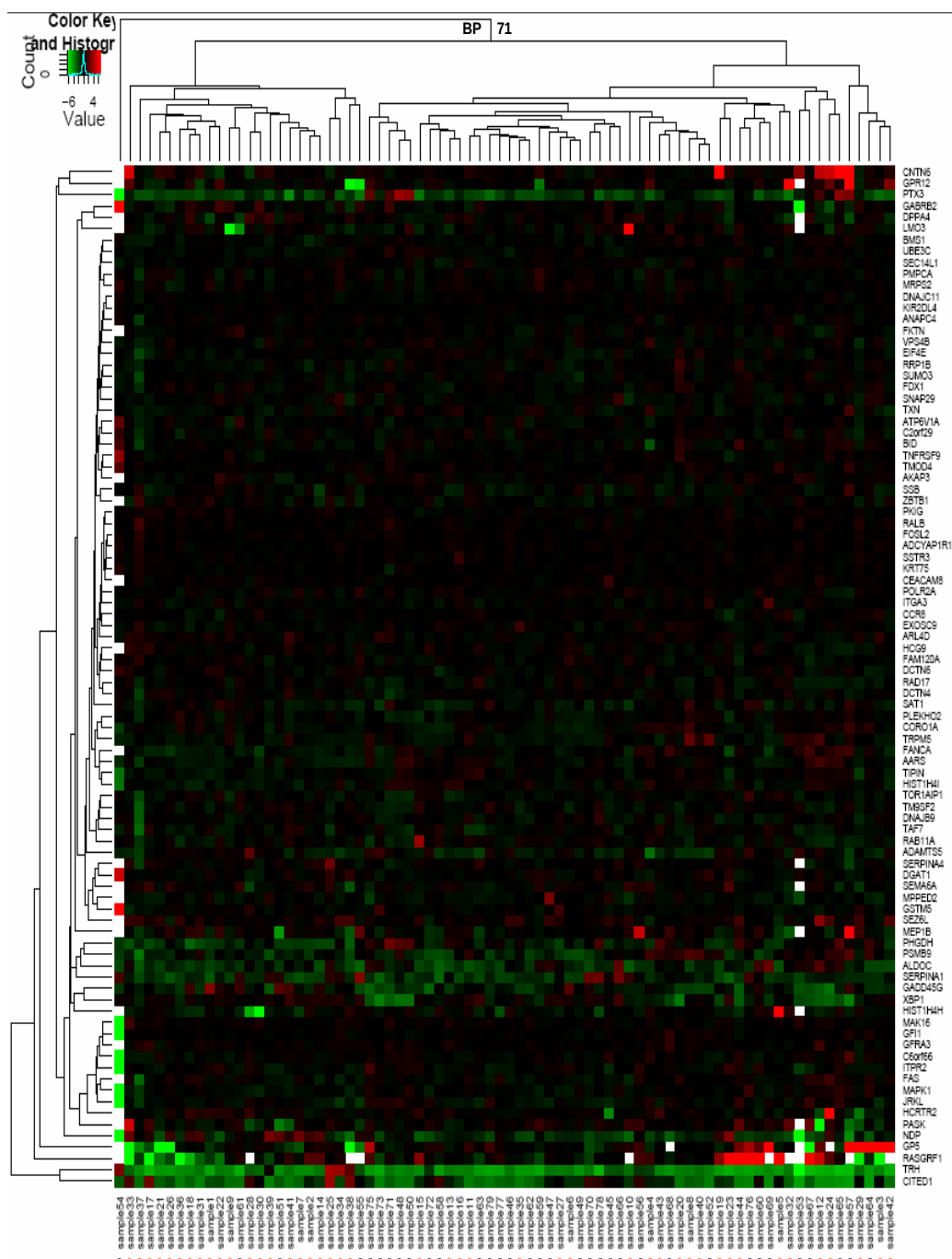
Esta análise nos mostra que apesar de estarmos comparando dados de metilação com dados de expressão gênica, fomos capazes de identificar genes que apresentam o mesmo comportamento entre os dois estudos.

Para verificar se a assinatura de metilação composta pelos 117 fragmentos de DNA diferencialmente metilados é capaz de identificar quais pacientes linfonodo negativos desenvolverão doença metastática, seria necessário validá-la experimentalmente em um grupo independente de amostras. Em nosso estudo, isso não foi possível por se tratar de um grupo de pacientes pouco freqüente na nossa população já que a maioria dos casos são diagnosticados em estadio avançado da doença quando as pacientes já apresentam comprometimento de linfonodos. Além disto, somente 30% das pacientes linfonodo negativos recaem.

Uma maneira indireta de verificar o valor prognóstico da nossa assinatura de metilação foi utilizá-la para classificar as pacientes do estudo desenvolvido por VAN'T VEER et al. Para tanto, o dado de expressão para cada um dos genes correspondentes aos nossos fragmentos diferencialmente metilados foram obtidos a partir do estudo da VAN'T VEER et al.¹⁰⁸ Esse dado de expressão foi então utilizado

para separar as pacientes. Nessa análise, verificamos que 27 dos 117 genes não estavam presentes no estudo de expressão de mama.

Como podemos observar na figura 17, o padrão de expressão dos genes correspondentes a 90 dos nossos fragmentos diferencialmente metilados foi capaz de separar corretamente 52 das 78 amostras utilizadas no estudo de expressão, o que representa 66,6% de acerto. O valor de *bootstrap* dessa clusterização foi de 71% no principal ramo de separação. Das amostras classificadas erroneamente, 3 amostras de pacientes que desenvolveram metástase e 23 amostras de pacientes que não desenvolveram metástase clusterizaram com o grupo oposto. No estudo original, a assinatura de 70 genes classificou corretamente 83% das amostras, com 5 pacientes de pior prognóstico e 8 de melhor prognóstico, atribuídas à categoria oposta.



Legenda: Na escala de cores o vermelho representa valores positivos de intensidade de sinal e o verde valor negativo.

Os (*) representam amostras de pacientes linfonodos negativos com recidiva e (*) amostras de pacientes linfonodos negativos sem recidiva.

Valor de *bootstrap* BP.

Figura 17 Clusterização hierárquica dos 90 fragmentos de DNA diferentemente metilados utilizando as amostras do estudo de expressão de mama¹⁰⁸.

DISCUSSÃO

5 DISCUSSÃO

A presença de células tumorais nos linfonodos axilares constitui o fator prognóstico mais importante para pacientes com câncer de mama, devendo virtualmente toda mulher com comprometimento linfonodal ser submetida a tratamento adjuvante. No entanto, pacientes portadoras de câncer de mama em estadios iniciais da doença e sem comprometimento de linfonodos podem apresentar diferenças marcantes na evolução da doença e cerca de 30% dessas pacientes morrem em decorrência do desenvolvimento de doença metastática. A necessidade de se determinar quais pacientes linfonodo negativas possuem um risco aumentado de desenvolver doença metastática tem conduzido estudos voltados para a identificação de marcadores moleculares que possam ser utilizados como ferramentas no prognóstico destas pacientes.

Portanto, a proposta deste trabalho foi analisar o perfil de metilação de amostras de pacientes linfonodo negativas que desenvolveram doença metastática ou não a fim de identificar marcadores moleculares de prognóstico que possam distinguir estes dois grupos de pacientes.

Alterações no padrão de metilação do DNA, destacando-se a hipermetilação da região promotora de genes supressores de tumor associada ao silenciamento gênico, têm sido amplamente utilizadas na caracterização de novos genes envolvidos no processo de formação e progressão tumoral e são encontradas em todos os tipos de neoplasias humanas^{9,58}.

A identificação de novas regiões promotoras diferencialmente metiladas presentes no genoma tumoral é um passo importante para se entender o papel da metilação do DNA na tumorigênese, no entanto a presença de metilação em tais regiões também pode ser utilizada como marcador molecular para o diagnóstico e prognóstico da doença.

A utilização da metilação diferencial como marcador tumoral é vantajosa, já que, ao contrário de moléculas de RNAs e de proteínas, o DNA é bastante estável e de fácil manipulação viabilizando a implantação de testes de detecção na rotina clínica. Além disto, devido à alta especificidade e sensibilidade das técnicas utilizadas para avaliar o padrão de metilação, geralmente baseadas em PCR, a metilação diferencial do DNA em células tumorais pode ser detectada mesmo em meio a um grande número de células normais. Ao contrário das alterações genéticas que ocorrem em diferentes regiões do gene e são de tipos diferentes, as alterações no padrão de metilação ocorrem sempre na mesma região, simplificando a metodologia de detecção^{30,83,113}.

Várias tecnologias que permitem avaliar o perfil de metilação global tem sido utilizadas para a análise de metilação diferencial entre tumores e tem possibilitado a correlação com os dados clínicos das pacientes^{1,86,110,116,121}. Entre estas, podemos mencionar as técnicas hibridização de metilação diferencial desenvolvidas por HUANG et al.⁴⁹ e por WANG et al.¹¹⁵.

Na metodologia desenvolvida por HUANG et al. uma população de fragmentos de DNA metilados é amplificada a partir de amostras tumorais a qual serão utilizadas como sonda contra os *microarrays* de seqüências enriquecidas para ilhas de CpG. Em um estudo piloto, desenvolvido em nosso laboratório, utilizamos

esta metodologia para caracterizar e comparar o perfil de metilação de ilhas de CpG do cromossomo 21 entre uma amostra de DNA de cérebro normal e duas linhagens de glioblastoma. Neste estudo, observamos que muito poucas ilhas de CpG apresentavam sinal de hibridização e que não era possível observar diferenças significativas no sinal de hibridização destas ilhas entre as linhagens analisadas mesmo para fragmentos sabidamente diferencialmente metilados entre as linhagens. Observamos que este fato não estava diretamente associado à ausência de seqüências diferencialmente metiladas e sim a baixa representatividade das ilhas de CpG na sonda obtida através do protocolo de amplificação devido a dificuldade de amplificação das ilhas de CpG por estas serem longas e apresentarem um alto conteúdo de CG. Além disto, verificamos que havia uma enorme competição com a população de fragmentos de DNA genômico irrelevantes para o estudo da metilação que também se ligam aos adaptadores, e são muitas vezes menores e apresentam baixo conteúdo de CG.

Ao contrário da técnica desenvolvida por HUANG et al.⁴⁹, a metodologia desenvolvida por WANG et al.¹¹⁵ permite a amplificação de fragmentos pequenos e não metilados no genoma o que aumenta substancialmente a chance de detecção de diferenças de metilação. Esta condição foi corroborada por um estudo promovido por SCHUMACHER et al.⁹⁵ que demonstraram as vantagens em se utilizar fragmentos de DNA não metilado aumentando assim substancialmente a chance de detecção de diferenças de metilação. A eficiência de amplificação de um determinado fragmento e sua representatividade na sonda utilizada na hibridização deverá permanecer inalterada entre diferentes amostras de DNA a não ser que haja diferença nos níveis

de metilação entre as mesmas. Essas diferenças podem então ser visualizadas através da intensidade do sinal de hibridização de cada uma das amostras.

Por este motivo, utilizamos essa metodologia de hibridização de metilação diferencial desenvolvida por WANG et al.¹¹⁵ para comparar o perfil de metilação global de amostras de carcinomas ductais invasivos de mama de pacientes linfonodo negativas que desenvolveram doença metastática, e pacientes que não desenvolveram metástase à distância com o objetivo de identificar uma assinatura de metilação que fosse capaz de discriminar as pacientes de diferentes prognósticos.

A etapa de identificação de fragmentos de DNA diferencialmente metilados entre o grupo que apresentou doença metastática e o grupo que não desenvolveu metástase a distância foi realizada em três etapas e foram considerados apenas os sinais de hibridização dos fragmentos de DNA imobilizados na lâmina que correspondem a seqüências promotoras de genes humanos conhecidos. Na primeira etapa desta análise, foram identificados 234 fragmentos de DNA diferencialmente metilados cuja diferença na média das intensidades de hibridização entre os dois grupos de pacientes apresentou um $p < 0.01$ e FDR < 0.05 . Esses 234 fragmentos representam 5% do total de fragmentos correspondentes a genes conhecidos. Essa porcentagem era esperada uma vez que se trata de um grupo bastante homogêneo de pacientes. No trabalho original desenvolvido por WANG et al. o percentual de genes diferencialmente metilados é maior (20%) uma vez que eles compararam linhagens celulares normais e tumorais de prótata. Entretanto, somente um gene foi identificado como diferencialmente metilado quando eles comparam duas linhagens tumorais, PC3M-LN4 metastática e PC3M-Pro4 menos metastática, derivadas da linhagem PC3.

Na segunda etapa, para aumentar a estringência da nossa análise, os resultados de hibridização dos 234 fragmentos de DNA diferencialmente metilados foram submetidos a uma segunda análise estatística na qual foi calculado um *z score* para cada um dos 234 fragmentos. Com esta abordagem identificamos 117 fragmentos de DNA diferencialmente metilados entre os dois grupos de pacientes com *z score* = 3.0, *p* <0.001, sendo que 47 estão hipermetilados no grupo de pacientes que desenvolveram doença metastática e 70 estão hipermetilados no grupo de pacientes que não desenvolveram metástase. Por fim, para reduzir o número de fragmentos diferencialmente metilados entre os dois grupos de amostras, priorizando aqueles com alterações no padrão de metilação mais freqüentes em seus respectivos grupos, selecionamos os fragmentos que estavam diferencialmente metilados em 75% das amostras de um dos grupos e em apenas 25% das amostras do outro grupo. Dessa forma, selecionamos 52 fragmentos de DNA diferencialmente metilados dos quais 18 estão hipermetilados no grupo de pacientes com doença metastática e 34 hipermetilados no grupo de pacientes que não desenvolveram metástase.

Todas as três assinaturas quando utilizadas para a clusterização hierárquica das amostras foram capazes de discriminar com 100% de acerto as pacientes que desenvolveram doença metastática das que não desenvolveram metástase à distância. Este resultado demonstra que a metodologia de escolha foi eficiente em identificar genes diferencialmente metilados entre os dois grupos de amostras e em estabelecer uma assinatura de metilação associada a uma pior evolução da doença.

Uma vez observado que estas assinaturas de metilação eram capazes de separar as amostras em dois grupos distintos era necessária a validação das mesmas em um grupo de amostras independentes. Entretanto, não foi possível conseguir

amostras independentes com as características clínicas e patológicas de interesse no nosso estudo por se tratar de pacientes pouco freqüentes na nossa população já que a maioria dos casos é diagnosticada em estágio avançado da doença quando as pacientes já apresentam comprometimento de linfonodos.

Para tanto, utilizamos dois tipos de informações disponíveis em bancos de dados públicos e comerciais para validar a assinatura de 117 genes diferencialmente metilados que envolvia dados de metilação e dados de expressão gênica.

Inicialmente, utilizamos o banco de dados PubMeth⁸⁰ para verificar para cada um dos 117 genes diferencialmente metilados em nossa análise, se havia evidência de alteração no padrão de metilação em diferentes tipos de tumores. Dentre os 114 genes identificados pelo programa apenas 1 (FHIT) foi descrito como freqüentemente metilado em tumores de mama e 4 (RASGRF1, GADD45G, ADAMTS5, FAS, PLEKHG5) se mostraram metilados em outros tipos de tumores.

Dentre os genes identificados, o gene FHIT foi o gene que se apresentou metilado em um maior número de tipos tumorais (15) incluindo tumor de mama. Este gene é um supressor de tumor, envolvido no controle da proliferação celular, apoptose e tem papel de modulador na resposta ao dano ao DNA⁵⁴.

O gene FHIT pode ser inativado tanto por alterações genéticas e cromossômicas (mutação, perda de heterozigosidade e deleção) quanto epigenéticas (hipermetilação do promotor) em vários tipos de câncer^{2,50,52,53,69,70}. Além disto, o nível da proteína FHIT pode ser regulado por degradação em proteassomo mediada por ativação, dependente de EGF, de membros da família FGFR, incluindo ERBB2,¹³. Embora, alterações no nível de expressão do gene FHIT têm sido observadas em

20 a 40% dos carcinomas primários de mama, a redução ou ausência da proteína pode ser vista em mais de 72% das amostras^{15,21,36,46}.

GATALICA et al.³⁶ avaliaram a expressão da proteína FHIT em carcinomas ductais de mama e lesões epiteliais e epitélio normal adjacentes ao tumor de 50 pacientes que sofreram mastectomia através da técnica de imunoistoquímica e observaram perda ou redução significativa de FHIT em 72% dos tumores. Eles também observaram que a expressão da proteína diminuía à medida que o tumor progredia para estágios mais avançados (epitélio normal, lesões hiperplásicas, hiperplasia atípica e carcinoma ductal *in situ* e carcinoma invasivo).

GINESTIER et al.³⁷, utilizando *tissue microarray*, observaram que em pacientes linfonodo negativas, com tumores menores que 2 cm, e que não receberam quimioterapia adjuvante a perda total da expressão da proteína FHIT estava significativamente associada a um pior prognóstico, com sobrevida livre de doença em 5 anos de 78,9% (60.6 – 97.3) quando comparadas a pacientes que apresentavam expressão intermediária ou total da proteína, 87,2% (80.4 – 94.0) ($p = 0.024$).

ZOCHBAUER-MULLER et al.¹²⁴ avaliaram o status de metilação (MSP) e expressão da proteína FHIT (imunoistoquímica) em 36 amostras de tumores de mama em diferentes estádios da doença (I, IIa, IIb, IIIa, IIIb e IV). Das 36 amostras, 20 estavam metiladas e destas 16 apresentaram perda de expressão da proteína. Dentre as 16 amostras não metiladas, 10 apresentaram perda de expressão da proteína. Neste trabalho, eles também analisaram a perda de heterozigosidade (LOH) de FHIT em 24 das 36 amostras e observaram LOH do locus de FHIT em 50% destas amostras. A correlação entre LOH e metilação foi de 48% nos tumores de mama. Juntos estes resultados demonstram que a metilação do gene FHIT é um evento

importante para o silenciamento deste gene em tumores de mama e está associada tanto a perda da expressão do RNA mensageiro quanto da proteína. No entanto, não houve correlação entre o status de metilação e o estágio tumoral.

Em uma segunda abordagem de validação, buscamos evidências de diminuição de expressão em tumores de mama para os genes correspondentes aos 117 fragmentos de DNA diferencialmente metilados, utilizando o banco de dados de expressão gênica em câncer do Oncomine⁹¹. Com esta abordagem, observamos evidência de diminuição de expressão para 38 dos 113 genes analisados em 5 assinaturas de tumores de mama de acordo com os dados armazenados no Oncomine. Dos 38 genes analisados, 27 apresentam evidência de diminuição de expressão em mais de um estudo e 5 genes (CORO1A, FANCA, HCRT2, SEZ6L e SFXN3) tiveram sua expressão induzida na linhagem tumoral de mama MCF7 tratada com ácido Valpróico, um inibidor de histona deacetilase, quando comparada à linhagem não tratada. Estes resultados demonstram que os genes diferencialmente metilados identificados em nosso estudo já foram descritos como diferencialmente expressos em câncer, incluindo câncer de mama e, portanto são bons candidatos a serem regulados por metilação.

Além disto, comparamos a nossa assinatura de 117 genes diferencialmente metilados com a assinatura de expressão gênica de 70 genes de VAN'T VEER et al.¹⁰⁸ Nessa comparação não identificamos genes comuns às duas assinaturas.

A tecnologia de *DNA microarrays* tem se mostrado promissora no que diz respeito à identificação de assinaturas de expressão gênica que possam prever a sobrevida de pacientes com câncer e assistir na seleção de condutas terapêuticas ótimas, permitindo que terapias sejam adaptadas de acordo com a severidade da

doença. Além da assinatura de expressão gênica identificada por VAN'T VEER et al. acima mencionada, outros estudos também identificaram assinaturas de expressão gênica capazes de prever o potencial metastático de tumores primários de mama^{87,99,108,114}. SORLIE et al.⁹⁹ e RAMASWAMY et al.⁸⁷, identificaram com sucesso assinaturas de expressão gênica capazes de separar 5 diferentes subtipos de carcinoma de mama e tumores primários de suas metástases, respectivamente. WANG et al. identificou uma assinatura de 76 genes com valor prognóstico, que pode ser aplicada para pacientes linfonodo negativos independentemente de idade, tamanho e grau do tumor e status de receptor hormonal¹¹⁴.

EIN-DOR et al.²⁸ realizaram um estudo no qual exploraram os dados do estudo de VAN'T VEER et al.¹⁰⁸ na tentativa de explicar a não concordância entre os grupos de genes derivados dos diferentes estudos. Eles observaram que muitos genes diferencialmente expressos no estudo desenvolvido por VANT VEER et al. tinham uma correlação apenas intermediária com o desenvolvimento da doença e que a diferença entre estas correlações era pequena. Além disto, o ranqueamento dos genes com base na correlação com a sobrevida das pacientes mudava drasticamente quando um grupo diferente de pacientes era utilizado para treinar o classificador. Desta forma, observaram que o classificador de 70 genes não era o único possível no estudo e que era influenciado pelo grupo de pacientes utilizados para a seleção dos genes.

Um estudo semelhante foi desenvolvido por MICHIELS et al.⁷⁵ que estenderam a análise para vários estudos de expressão gênica descritos na literatura entre janeiro de 1995 e abril de 2003. Neste estudo, eles também observaram que as assinaturas gênicas identificadas como preditores de prognóstico eram altamente

variáveis e que estas assinaturas moleculares dependiam da seleção dos pacientes do grupo de treinamento dos classificadores.

Apesar de não haver genes em comum entre nossa assinatura e a assinatura de VAN'T VEER et al., verificamos que dois genes (RALB e TMPRSS3) dentre os 47 genes hipermetilados no grupo de pacientes com doença metastática, e 10 genes (CLPB, FANCA, GFR3A, GP5, PHGDH, PTX3, RASGRF1, RRP1B, TIPIN e TXN) entre os 70 genes hipermetilados nas amostras das pacientes que não desenvolveram metástase à distância, apresentavam evidência de diminuição de expressão no grupo correspondente do estudo de VAN'T VEER et al.¹⁰⁸.

Dados da literatura confirmam a importância de alguns destes genes para o desenvolvimento tumoral do câncer de mama assim como para outros tipos de tumor. Como exemplo, podemos destacar o gene TXN que codifica uma proteína multifuncional que atua como um potente antioxidante e oxiredutor. A expressão deste gene está aumentada em diversos tumores em resposta direta ao stress e confere resistência a drogas quimioterápicas e a apoptose induzida por radiação⁵⁹. Tem sido demonstrado que a atividade redutora de TXN ativa vários fatores de transcrição envolvidos na regulação do crescimento e sobrevivência celular, incluindo NF- κ B, AP1 and SP1. Além disto, TXN aumenta a expressão de HIF1A possivelmente pela inibição da degradação de HIF1A^{62,117}. TXN se liga e inibe a atividade pró-apoptótica do regulador de apoptose ASK1 e do supressor de tumor PTEN^{74,93}. Quando PTEN é expresso nas células tumorais sua proteína atenua a atividade de PIP3/AKT na via de sobrevivência celular. TXN estimula a angiogênese induzida por VEGF em tumores através do aumento da atividade do fator de transcrição SP1 em condições de normoxia e aumenta os níveis de HIF1A em condições de

hipóxia¹¹⁷. Recentemente, observaram em câncer de pulmão que uma vez que TXN aumenta HIF1A este aumenta COX2 que faz de TXN um potente estimulador do crescimento tumoral aumentando proliferação e a angiogênese e inibindo a apoptose²². Uma construção dominante negativa que bloqueia a atividade de TXN inibe o crescimento de tumores xenográficos em camundongo³³.

Para tentar elucidar a relevância biológica dos genes diferencialmente metilados identificados em nosso estudo foi feita uma análise utilizando os dados de ontologia gênica (*Gene Ontology* - GO) disponibilizados pelo Oncomine²⁸. Nessa análise não foi possível identificar funções moleculares, processos biológicos ou localizações celulares enriquecidas em nossa assinatura. No entanto, identificamos genes cujas funções poderiam ter um papel relevante no processo tumoral, tais como genes envolvidos em processos como apoptose (FAS, BID, GADD45G), diferenciação celular (GADD45G, SEMA6), adesão celular (ITGA3) sobrevivência celular (RALB) e motilidade celular (CORO1A, TXN, DCTN4), proliferação celular (TXN), reparo de DNA (FANCA) e genes descritos como possíveis supressores de tumor em câncer (DNAJC11, FHIT).

Além disto, para verificar se a assinatura de metilação composta pelos 117 fragmentos de DNA diferencialmente metilados era capaz de discriminar as pacientes linfonodo negativas que desenvolveram doença metastática, a utilizamos para separar as pacientes do estudo desenvolvido por VAN'T VEER et al.¹⁰⁸. Para tanto, os dados de expressão para cada um dos genes correspondentes aos nossos fragmentos diferencialmente metilados presentes no estudo de VAN'T VEER et al. foram obtidos e utilizados para separar as amostras deste estudo de expressão. Os genes comuns aos dois estudos foram capazes de discriminar 52 das 78 amostras

utilizadas no estudo de expressão, o que representa 66,6% de acerto contra 83% de acerto da assinatura de expressão do estudo original. Das amostras classificadas erroneamente, 3 amostras de pacientes que desenvolveram metástase e 23 amostras de pacientes que não desenvolveram metástase clusterizaram com o grupo oposto.

Entretanto, não podemos deixar de mencionar que mesmo sendo possível validar a nossa assinatura de metilação com dados de expressão disponíveis na literatura e principalmente com os dados do estudo desenvolvido por VAN'T VEER et al.¹⁰⁸, a metodologia de hibridização de metilação diferencial utilizada em nosso estudo permite a detecção de alterações no número de cópias de um determinado segmento genômico o que provoca um aumento no sinal de hibridização sem necessariamente haver correlação com o grau de metilação do fragmento. No entanto, neste caso é possível verificar que o aumento de sinal abrange vários genes com localização próxima na região amplificada.

Desta forma, verificamos se as diferenças de hibridização identificadas em nossa análise poderiam estar relacionadas a alterações no número de cópias dos genes e não a diferenças de metilação. Para tanto, utilizamos o Oncomine para identificar regiões cromossômicas ou citobandas enriquecidas em nossa assinatura que apresentassem evidências de alterações no número de cópias já descritas na literatura. Nesta análise, observamos que 5 dos 117 genes presentes na assinatura (ADAMTS5, C21orf67, H2BFS, RRP1B, SUMO3 e TMPRSS3) encontram-se mapeados no braço longo do cromossomo 21 ($p = 0.002$) e com exceção do gene ADAMTS5, todos estão mapeados na citobanda 21q22.3 ($p = 0.004$).

Estudos recentes têm avaliado alterações no número de cópias em amostras tumorais de mama que possam ser utilizados como marcadores de prognóstico para

pacientes com câncer de mama através da técnica de *Comparative Genomic Hybridization* (CGH)^{42,51}. HWANG et al.⁵¹, avaliaram 62 amostras de pacientes com câncer de mama em diferentes estádios da doença, sendo que 31 desenvolveram doença metastática e 31 não desenvolveram metástase à distância. Neste estudo, eles observaram que a região 21q22.3 se encontrava dentre as 17 regiões que apresentaram um ganho freqüente no número de cópias entre as amostras de pacientes com câncer de mama linfonodo negativas que desenvolviam doença metastática em relação as pacientes que não desenvolveram metástase à distância e que tinha correlação estatisticamente significativa com a sobrevida livre de doença sistêmica nestas pacientes. Em um segundo estudo HAN et al.⁴² compararam ganhos e perdas de regiões cromossômicas entre 28 amostras de pacientes com câncer de mama triplo negativas (receptor de estrógeno, progesterona e ERBB2) e portanto de pior prognóstico, 62 amostras hormônio positivas e 24 amostras positivas somente para ERBB2. Neste estudo, eles observaram que ganhos em 8 regiões incluindo a região 21q22.3 eram exclusivos das amostras triplo negativas e que portanto estavam relacionados a um pior prognóstico.

Em nosso estudo observamos somente o enriquecimento da citobanda 21q22.3 em nossa assinatura. Isto demonstra que uma pequena porção de genes da nossa assinatura podem estar apresentando diferenças no número de cópias, mas que possivelmente, a maior parte dos genes diferencialmente metilados identificados em nossa análise estão alterados por metilação. No entanto, é necessário que estas diferenças de metilação sejam confirmadas por outras metodologias tais como, *Methylation Specific PCR* (MSP-PCR) e ou tratamento do DNA por bissulfito de sódio seguido de sequenciamento.

Como podemos observar, este trabalho gera perspectivas para muitos outros, nos quais estes genes possam ser validados quanto ao perfil de metilação, que esta assinatura possa ser validada em um grupo independente de amostras e que possa ser demonstrada sua aplicação na prática clínica.

CONCLUSÕES

6 CONCLUSÕES

- 1 Através da utilização da metodologia de hibridização de fragmentos diferencialmente metilados foi possível estabelecer o perfil de metilação global das amostras de pacientes com carcinoma ductal invasivo de mama linfonodo negativas que desenvolveram ou não metástase à distância;
- 2 Foi possível identificar, utilizando diferentes critérios, assinaturas de metilação compostas por 234, 117 e 52 fragmentos de DNA diferencialmente metilados capazes de discriminar os dois grupos de pacientes em estudo;
- 3 Evidências de alteração no perfil de metilação ou diminuição de expressão foram encontradas para 50 dos 117 genes correspondentes aos fragmentos de DNA diferencialmente metilados após consulta nos bancos de dados PubMeth e Oncomine.
- 4 O perfil de expressão dos 90 genes correspondentes a 90 dos nossos 117 fragmentos de DNA diferencialmente metilados foi capaz de prever com 66% de acerto o prognóstico de pacientes com câncer de mama linfonodo negativas do estudo desenvolvido por VAN'T VEER et al.¹⁰⁸.

***REFERÊNCIAS
BIBLIOGRÁFICAS***

7 REFERÊNCIA BIBLIOGRÁFICA

- 1 Ahluwalia A, Yan P, Hurteau JA, et al. DNA methylation and ovarian cancer. I. Analysis of CpG island hypermethylation in human ovarian cancer using differential methylation hybridization. **Gynecol Oncol** 2001; 82:261-8.
- 2 Ahmadian M, Wistuba II, Fong KM, et al. Analysis of the FHIT gene and FRA3B region in sporadic breast cancer, preneoplastic lesions, and familial breast cancer probands. **Cancer Res** 1997; 57:3664-8.
- 3 Antequera F, Bird A. Number of CpG islands and genes in human and mouse. **Proc Natl Acad Sci U S A** 1993; 90:11995-9.
- 4 Attwood JT, Yung RL, Richardson BC. DNA methylation and the regulation of gene transcription. **Cell Mol Life Sci** 2002; 241-57.
- 5 Ball CA, Awad IA, Demeter J, et al. The Stanford Microarray Database accommodates additional microarray platforms and data formats. **Nucleic Acids Res** 2005; 33:D580-D582.
- 6 Baracat, FF, Oliveira, AB. Câncer de mama: etiopatogenia, diagnóstico e estadiamento. In: Abrão, FS editor. **Tratado de oncologia genital e mamária**. São Paulo: Roca; 1995. p.549-579.
- 7 Barrett T, Suzek TO, Troup DB, et al. NCBI GEO: mining millions of expression profiles--database and tools. **Nucleic Acids Res** 2005; 33:D562-D566.
- 8 Baylin SB. Mechanisms underlying epigenetically mediated gene silencing in cancer. **Semin Cancer Biol** 2002; 12:331-7.
- 9 Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. **Trends Genet** 2000; 16:168-74.

- 10 Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP. Alterations in DNA methylation: a fundamental aspect of neoplasia. **Adv Cancer Res** 1998; 72:141-96.
- 11 Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. **Cell** 1999; 98:387-96.
- 12 Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. **Behav Brain Res** 2001; 125:279-84.
- 13 Bianchi F, Tagliabue E, Menard S, Campiglio M. Fhit expression protects against HER2-driven breast tumor development: unraveling the molecular interconnections. **Cell Cycle** 2007; 6:643-6.
- 14 Bird AP. Gene number, noise reduction and biological complexity. **Trends Genet** 1995; 11:94-100.
- 15 Campiglio M, Pekarsky Y, Menard S, Tagliabue E, Pilotti S, Croce CM. FHIT loss of function in human primary breast cancer correlates with advanced stage of the disease. **Cancer Res** 1999; 59:3866-9.
- 16 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.
- 17 Clark SJ, Melki J. DNA methylation and gene silencing in cancer: which is the guilty party? **Oncogene** 2002; 21:5380-7.
- 18 Coelho EC, Makdissi FBA. Detecção precoce do câncer de mama. In: Kowalski LP, Guimarães GC, Salvajoli JV, Feher O, Antonelli CBG editor. **Manual de condutas diagnósticas e terapêuticas em oncologia**. São Paulo: Âmbito Editores; 2006. p.755-758.
- 19 Costello JF, Fruhwald MC, Smiraglia DJ, et al. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns.

- 20 Craig JM, Kraus J, Cremer T. Removal of repetitive sequences from FISH probes using PCR-assisted affinity chromatography. **Hum Genet** 1997; 100:472-6.
- 21 Croce CM, Sozzi G, Huebner K. Role of FHIT in human cancer. **J Clin Oncol** 1999; 17:1618-24.
- 22 Csiki I, Yanagisawa K, Haruki N, et al. Thioredoxin-1 modulates transcription of cyclooxygenase-2 via hypoxia-inducible factor-1alpha in non-small cell lung cancer. **Cancer Res** 2006; 66:143-50.
- 23 de Coronado S, Haber MW, Sioutos N, Tuttle MS, Wright LW. NCI Thesaurus: using science-based terminology to integrate cancer research results. **Stud Health Technol Inform** 2004; 107:33-7.
- 24 De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. **Proc Natl Acad Sci U S A** 1996; 93:7149-53.
- 25 Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? **J Cell Mol Med** 2005; 9:208-21.
- 26 Ehrlich M. DNA methylation in cancer: too much, but also too little. **Oncogene** 2002; 21:5400-13.
- 27 Ehrlich M, Gama-Sosa MA, Huang LH, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. **Nucleic Acids Res** 1982; 10:2709-21.
- 28 Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? **Bioinformatics** 2005; 21:171-8.
- 29 Esteller M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. **Oncogene** 2002; 21:5427-40.

- 30 Esteller M. Relevance of DNA methylation in the management of cancer. **Lancet Oncol** 2003; 4:351-8.
- 31 Feinberg AP. Cancer epigenetics takes center stage. **Proc Natl Acad Sci U S A** 2001; 98:392-4.
- 32 Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. **Nature** 1983; 301:89-92.
- 33 Gallegos A, Gasdaska JR, Taylor CW, et al. Transfection with human thioredoxin increases cell proliferation and a dominant-negative mutant thioredoxin reverses the transformed phenotype of human breast cancer cells. **Cancer Res** 1996; 56:5765-70.
- 34 Gama-Sosa MA, Slagel VA, Trewyn RW, et al. The 5-methylcytosine content of DNA from human tumors. **Nucleic Acids Res** 1983; 11:6883-94.
- 35 Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. **J Mol Biol** 1987; 196:261-82.
- 36 Gatalica Z, Lele SM, Rampy BA, Norris BA. The expression of Fhit protein is related inversely to disease progression in patients with breast carcinoma. **Cancer** 2000; 88:1378-83.
- 37 Ginestier C, Bardou VJ, Popovici C, et al. Loss of FHIT protein expression is a marker of adverse evolution in good prognosis localized breast cancer. **Int J Cancer** 2003; 107:854-62.
- 38 Gonzalez-Zulueta M, Bender CM, Yang AS, et al. Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. **Cancer Res** 1995; 55:4531-5.
- 39 Graff JR, Herman JG, Lapidus RG, et al. E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. **Cancer Res** 1995; 55:5195-9.

- 40 Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. **Hum Genet** 1989; 83:155-8.
- 41 Hamada GS, Nishimoto IN, Torloni H. **Registro Hospitalar do Câncer, 1994.** Rio de Janeiro: Pro-Onco, 1997. (Monografias do Registro Hospitalar de Câncer-Hospital A.C. Camargo, vol1,1996.).
- 42 Han W, Jung EM, Cho J, et al. DNA copy number alterations and expression of relevant genes in triple-negative breast cancer. **Genes Chromosomes Cancer** 2008; 47:490-9.
- 43 Hanahan D, Weinberg RA. The hallmarks of cancer. **Cell** 2000; 100:57-70.
- 44 Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. **Nucleic Acids Res** 2004; 32:D258-D261.
- 45 Harris RA, Eichholtz TJ, Hiles ID, Page MJ, O'Hare MJ. New model of ErbB-2 over-expression in human mammary luminal epithelial cells. **Int J Cancer** 1999; 80:477-84.
- 46 Hayashi S, Tanimoto K, Hajiro-Nakanishi K, et al. Abnormal FHIT transcripts in human breast carcinomas: a clinicopathological and epidemiological analysis of 61 Japanese cases. **Cancer Res** 1997; 57:1981-5.
- 47 Hellman S. Karnofsky Memorial Lecture. Natural history of small breast cancers. **J Clin Oncol** 1994; 12:2229-34.
- 48 Herman JG, Latif F, Weng Y, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. **Proc Natl Acad Sci U S A** 1994; 91:9700-4.
- 49 Huang TH, Perry MR, Laux DE. Methylation profiling of CpG islands in human breast cancer cells. **Hum Mol Genet** 1999; 8:459-70.

- 50 Huebner K, Garrison PN, Barnes LD, Croce CM. The role of the FHIT/FRA3B locus in cancer. **Annu Rev Genet** 1998; 32:7-31.
- 51 Hwang KT, Han W, Cho J, et al. Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer. **Int J Cancer** 2008; 123:1807-15.
- 52 Iliopoulos D, Guler G, Han SY, et al. Fragile genes as biomarkers: epigenetic control of WWOX and FHIT in lung, breast and bladder cancer. **Oncogene** 2005; 24:1625-33.
- 53 Ingvarsson S, Agnarsson BA, Sigbjornsdottir BI, et al. Reduced Fhit expression in sporadic and BRCA2-linked breast carcinomas. **Cancer Res** 1999; 59:2682-9.
- 54 Ishii H, Mimori K, Inoue H, et al. Fhit modulates the DNA damage checkpoint response. **Cancer Res** 2006; 66:11287-92.
- 55 Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. **CA Cancer J Clin** 2008; 58:71-96.
- 56 Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. **CA Cancer J Clin** 2007; 57:43-66.
- 57 Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. **Nat Rev Genet** 2002; 3:415-28.
- 58 Jones PA, Laird PW. Cancer epigenetics comes of age. **Nat Genet** 1999; 21:163-7.
- 59 Kaimul AM, Nakamura H, Masutani H, Yodoi J. Thioredoxin and thioredoxin-binding protein-2 in cancer and metabolic syndrome. **Free Radic Biol Med** 2007; 43:861-8.
- 60 Kane MF, Loda M, Gaida GM, et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and

- mismatch repair-defective human tumor cell lines. **Cancer Res** 1997; 57:808-11.
- 61 Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. **Nucleic Acids Res** 2004; 32:D277-D280.
 - 62 Kim WJ, Cho H, Lee SW, Kim YJ, Kim KW. Antisense-thioredoxin inhibits angiogenesis via pVHL-mediated hypoxia-inducible factor-1 α degradation. **Int J Oncol** 2005; 26:1049-52.
 - 63 Klenova EM, Morse HC, III, Ohlsson R, Lobanenko VV. The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. **Semin Cancer Biol** 2002; 12:399-414.
 - 64 Krek A, Grun D, Poy MN, et al. Combinatorial microRNA target predictions. **Nat Genet** 2005; 37:495-500.
 - 65 Laird PW. The power and the promise of DNA methylation markers. **Nat Rev Cancer** 2003; 3:253-66.
 - 66 Lee YT. Breast carcinoma: pattern of metastasis at autopsy. **J Surg Oncol** 1983; 23:175-80.
 - 67 Lobanenko VV, Nicolas RH, Plumb MA, Wright CA, Goodwin GH. Sequence-specific DNA-binding proteins which interact with (G + C)-rich sequences flanking the chicken c-myc gene. **Eur J Biochem** 1986; 159:181-8.
 - 68 Loukinov DI, Pugacheva E, Vatolin S, et al. BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. **Proc Natl Acad Sci U S A** 2002; 99:6806-11.
 - 69 Man S, Ellis IO, Sibbering M, Blamey RW, Brook JD. High levels of allele loss at the FHIT and ATM genes in non-comedo ductal carcinoma in situ and grade I tubular invasive breast cancers. **Cancer Res** 1996; 56:5484-9.

- 70 Maruyama R, Toyooka S, Toyooka KO, et al. Aberrant promoter methylation profile of bladder cancer and its relationship to clinicopathological features. **Cancer Res** 2001; 61:8659-63.
- 71 Matys V, Fricke E, Geffers R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. **Nucleic Acids Res** 2003; 31:374-8.
- 72 McClelland M, Ivarie R. Asymmetrical distribution of CpG in an 'average' mammalian gene. **Nucleic Acids Res** 1982; 10:7865-77.
- 73 Merlo A, Herman JG, Mao L, et al. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. **Nat Med** 1995; 1:686-92.
- 74 Meuillet EJ, Mahadevan D, Berggren M, Coon A, Powis G. Thioredoxin-1 binds to the C2 domain of PTEN inhibiting PTEN's lipid phosphatase activity and membrane binding: a mechanism for the functional loss of PTEN's tumor suppressor activity. **Arch Biochem Biophys** 2004; 429:123-33.
- 75 Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. **Lancet** 2005; 365:488-92.
- 76 Ministério da Saúde. Secretaria de Assistência à Saúde. Instituto Nacional de Câncer. **Estimativa de incidência e mortalidade de câncer no Brasil**. Rio de Janeiro: INCA; 2005.
- 77 Ministério da Saúde. Secretaria de Assistência à Saúde. Instituto Nacional de Câncer. **Estimativa/2008: incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2009.
- 78 Mulder NJ, Apweiler R, Attwood TK, et al. InterPro, progress and status in 2005. **Nucleic Acids Res** 2005; 33:D201-D205.
- 79 Nan X, Ng HH, Johnson CA, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. **Nature** 1998; 393:386-9.

- 80 Ongenaert M, Van Neste L, De Meyer T, Menschaert G, Bekaert S, Van Criekinge W. PubMeth: a cancer methylation database combining text-mining and expert annotation. **Nucleic Acids Res** 2008; 36:D842-D846.
- 81 Page MJ, Amess B, Townsend RR, et al. Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties. **Proc Natl Acad Sci U S A** 1999; 96:12589-94.
- 82 Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. **N Engl J Med** 2004; 351:2817-26.
- 83 Patel A, Groopman JD, Umar A. DNA methylation as a cancer-specific biomarker: from molecules to populations. **Ann N Y Acad Sci** 2003; 983:286-97.
- 84 Peri S, Navarro JD, Amanchy R, et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. **Genome Res** 2003; 13:2363-71.
- 85 Radice D, Redaelli A. Breast cancer management: quality-of-life and cost considerations. **Pharmacoeconomics** 2003; 21:383-96.
- 86 Rahmatpanah FB, Carstens S, Guo J, et al. Differential DNA methylation patterns of small B-cell lymphoma subclasses with different clinical behavior. **Leukemia** 2006; 20:1855-62.
- 87 Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. **Nat Genet** 2003; 33:49-54.
- 88 Razin A, Cedar H. Distribution of 5-methylcytosine in chromatin. **Proc Natl Acad Sci U S A** 1977; 74:2725-8.
- 89 Razin A, Cedar H. DNA methylation and gene expression. **Microbiol Rev** 1991; 55:451-8.

- 90 Razin A, Riggs AD. DNA methylation and gene function. **Science** 1980; 210:604-10.
- 91 Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. **Neoplasia** 2007; 9:166-80.
- 92 Robertson KD. DNA methylation and human disease. **Nat Rev Genet** 2005; 6:597-610.
- 93 Saitoh M, Nishitoh H, Fujii M, et al. Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. **EMBO J** 1998; 17:2596-606.
- 94 Schmutte C, Jones PA. Involvement of DNA methylation in human carcinogenesis. **Biol Chem** 1998; 379:377-88.
- 95 Schumacher A, Kapranov P, Kaminsky Z, et al. Microarray-based DNA methylation profiling: technology and applications. **Nucleic Acids Res** 2006; 34:528-42.
- 96 Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. **Science** 1987; 235:177-82.
- 97 Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. **N Engl J Med** 2001; 344:783-92.
- 98 Sobin LH, Wittekind CH. Tumores de mama. In: INCA editor. **TNM classificação de tumores malignos**. 6. Rio de Janeiro: A L A Eisenberg; 2004. p.137-148.
- 99 Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. **Proc Natl Acad Sci U S A** 2001; 98:10869-74.

- 100 Srinivas PR, Kramer BS, Srivastava S. Trends in biomarker research for cancer detection. **Lancet Oncol** 2001; 2:698-704.
- 101 Steeg, PS. Control of invasion and metastasis. In: Harris JR, Lippman ME, Morrow M, Osborne OK editors. **Diseases of the breast**. 3rd. Philadelphia: Lippincott Williams e Wilkins; 2004. p.457-472.
- 102 Stirzaker C, Millar DS, Paul CL, et al. Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. **Cancer Res** 1997; 57:2229-37.
- 103 Styblo, MT, Wood, WC. Traditional prognostic factors for breast cancer. In: Bland KI, Copeland III EM editors. **The breast: comprehensive management of benign and malignant diseases**. 2nd. Philadelphia: W B Saunders Company; 1998. p.419-425.
- 104 Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. **Bioinformatics** 2006; 22:1540-2.
- 105 Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. **Proc Natl Acad Sci U S A** 2002; 99:3740-5.
- 106 Tavassoéli FA, Devilee P. **Tumours pathology and genetics of tumours of the breast and female genital organs**. Lyon: IARC Press; 2003.
- 107 Thomssen C, Janicke F, Harbeck N. Clinical relevance of prognostic factors in axillary node-negative breast cancer. **Onkologie** 2003; 26:438-45.
- 108 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.
- 109 van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. **N Engl J Med** 2002; 347:1999-2009.
- 110 van Doorn R, Zoutman WH, Dijkman R, et al. Epigenetic profiling of cutaneous T-cell lymphoma: promoter hypermethylation of multiple tumor

- suppressor genes including BCL7a, PTPRG, and p73. **J Clin Oncol** 2005; 23:3886-96.
- 111 Van Poznak C, Tan L, Panageas KS, et al. Assessment of molecular markers of clinical sensitivity to single-agent taxane therapy for metastatic breast cancer. **J Clin Oncol** 2002; 20:2319-26.
 - 112 Vanyushin BF, Tkacheva SG, Belozersky AN. Rare bases in animal DNA. **Nature** 1970; 225:948-9.
 - 113 Verma M, Srivastava S. Epigenetics in cancer: implications for early detection and prevention. **Lancet Oncol** 2002; 3:755-63.
 - 114 Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. **Lancet** 2005; 365:671-9.
 - 115 Wang Y, Yu Q, Cho AH, et al. Survey of differentially methylated promoters in prostate cancer cell lines. **Neoplasia** 2005; 7:748-60.
 - 116 Wei SH, Chen CM, Strathdee G, et al. Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. **Clin Cancer Res** 2002; 8:2246-52.
 - 117 Welsh SJ, Bellamy WT, Briehl MM, Powis G. The redox protein thioredoxin-1 (Trx-1) increases hypoxia-inducible factor 1alpha protein expression: Trx-1 overexpression results in increased vascular endothelial growth factor production and enhanced tumor angiogenesis. **Cancer Res** 2002; 62:5089-95.
 - 118 Widschwendter M, Jiang G, Woods C, et al. DNA hypomethylation and ovarian cancer biology. **Cancer Res** 2004; 64:4472-80.
 - 119 Widschwendter M, Jones PA. DNA methylation and breast carcinogenesis. **Oncogene** 2002; 21:5462-82.

- 120 Xie X, Lu J, Kulbokas EJ, et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. **Nature** 2005; 434:338-45.
- 121 Yan PS, Perry MR, Laux DE, Asare AL, Caldwell CW, Huang TH. CpG island arrays: an application toward deciphering epigenetic signatures of breast cancer. **Clin Cancer Res** 2000; 6:1432-8.
- 122 Yang YH, Buckley MJ, Speed TP. Analysis of cDNA microarray images. **Brief Bioinform** 2001; 2:341-9.
- 123 Yuille MR, Condie A, Stone EM, et al. TCL1 is activated by chromosomal rearrangement or by hypomethylation. **Genes Chromosomes Cancer** 2001; 30:336-41.
- 124 Zochbauer-Muller S, Fong KM, Maitra A, et al. 5' CpG island methylation of the FHIT gene is correlated with loss of gene expression in lung and breast cancer. **Cancer Res** 2001; 61:3581-5.

ANEXOS

ANEXO 1

Anexo 1. Assinatura de 234 genes diferencialmente metilados entre os dois grupos de amostras.

Nº Acesso	Símbolo	<i>p</i> valor
NM_004744	LRAT	0.000123
NM_006515	SETMAR	0.000173
NM_025010	KLHL18	0.000182
NM_016152	RARB	0.000199
NM_002891	RASGRF1	0.000232
NM_138467	LOC127253	0.000306
NM_006936	SMT3H1	0.000315
NM_004800	TM9SF2	0.000365
NM_001497	B4GALT1	0.000383
NM_004679	VCY	0.000437
NM_001526	HCRTR2	0.000450
NM_032509	LOC84549	0.000490
NM_002526	NT5E	0.000606
AF171235	PHGDH	0.000689
NM_019855	CABP5	0.000714
AK131307	PFKFB3	0.000740
X17122	SERPINA1	0.000827
NM_000965	RARB	0.000897
NM_001696	ATP6E	0.001028
NM_024608	FLJ22402	0.001035
AF217986	pp9099	0.001075
NM_003543	H4FH	0.001080
NM_007213	PRAF2	0.001122
NM_017840	MRPL16	0.001123
NM_203316	DPAGT1	0.001135
NM_015148	PASK	0.001168
NM_001418	EIF4G2	0.001170
NM_002841	PTPRG	0.001187
NM_016039	C14orf166	0.001195
NM_005288	GPR12	0.001203
NM_021115	SEZ6L	0.001224
NM_000937	POLR2A	0.001229
NM_000737	CGB	0.001230
NM_000851	GSTM5	0.001251
NM_021824	NIF3L1	0.001260
NM_005080	XBP1	0.001318
NM_015160	PMPCA	0.001448

Nº Acesso	Símbolo	<i>p</i> valor
NM_015056	KIAA0179	0.001505
NM_181874	GRM7	0.001771
NM_004693	K6HF	0.001783
NM_002113	HFL1	0.001783
NM_004126	GNG11	0.001816
NM_006705	GADD45G	0.001892
NM_005925	MEP1B	0.001935
NM_172171	CAMK2G	0.001958
NM_012141	DDX26	0.001982
NM_017858	FLJ20516	0.001996
NM_020796	SEMA6A	0.002037
NM_030958	OATPRP4	0.002072
NM_000138	FBN1	0.002175
NM_005725	TSPAN-2	0.002175
NM_021928	FLJ22649	0.002238
NM_018198	FLJ10737	0.002254
NM_005033	PMSCL1	0.002329
NM_015602	LAP1B	0.002386
NM_207491	MGC48628	0.002423
NM_012464	TLL1	0.002471
NM_030813	SKD3	0.002487
NM_001335	CTSW	0.002499
NM_152486	SAMD11	0.002571
NM_007074	CORO1A	0.002664
NM_002745	MAPK1	0.002676
NM_052878	C19orf36	0.002711
NM_002376	MARK3	0.002727
NM_032192	FLJ20940	0.002736
BC007366	C9orf70	0.002754
NM_002989	SCYA21	0.002768
NM_004109	FDX1	0.002806
NM_001496	GFRA3	0.002853
NM_024490	ATP10C	0.002853
NM_032404	TMPRSS3	0.002888
NM_016592	GNAS	0.002888
AL512687	NOMO2	0.002962
NM_002021	FMO1	0.002963
NM_004314	ART1	0.002995
NM_030820	COL21A1	0.003017
NM_014461	CNTN6	0.003088

Nº Acesso	Símbolo	<i>p</i> valor
NM_015872	LOC51043	0.003099
CDNA0174	cDNA	0.003099
NM_138387	G6PC3	0.003126
NM_014051	PTD011	0.003178
NM_052818	CG018	0.003214
NM_018189	FLJ10713	0.003223
AK001021	FLJ10159	0.003278
NM_002255	KIR2DL4	0.003357
NM_001005911	IHPK2	0.003366
NM_004488	GP5	0.003411
NM_007055	RPC155	0.003447
NM_004314	ART1	0.003456
NM_000183	HADHB	0.003521
NM_019851	FGF20	0.003532
NM_002012	FHIT	0.003538
NM_007117	TRH	0.003554
NM_001968	EIF4E	0.003568
NM_000266	NDP	0.003660
NM_144765	EVA1	0.003691
NM_014555	TRPM5	0.003735
NM_000895	LTA4H	0.003759
NM_024106	MGC2663	0.003803
NM_016269	LEF1	0.003819
NM_032747	MGC14697	0.003912
NM_012242	DKK1	0.003934
NM_003003	SEC14L1	0.003961
NM_197967	BID	0.003985
NM_198681	KIAA0720	0.004058
NM_002852	PTX3	0.004060
NM_006731	FCMD	0.004062
NM_014165	HSPC125	0.004078
NM_080658	MGC9740	0.004135
BC010444	MATN2	0.004237
NM_003329	TXN	0.004246
NM_000637	GSR	0.004273
NM_004869	SKD1	0.004273
CR457153	SPG3A	0.004322
NM_003495	H4FM	0.004373
NM_004505	USP6	0.004404
NM_005989	AKR1D1	0.004458

Nº Acesso	Símbolo	<i>p</i> valor
NM_001051	SSTR3	0.004538
NM_012328	DNAJB9	0.004566
NM_016221	DCTN4	0.004574
NM_032280	DKFZp761J139	0.004701
NM_004073	CNK	0.004716
NM_016013	LOC51103	0.004806
NM_000135	FANCA	0.004812
NM_001690	ATP6A1	0.004909
NM_002970	SAT	0.004911
CDNA0073	cDNA92	0.004959
NM_000970	RPL6	0.005018
NM_022777	FLJ14117	0.005069
NM_005165	ALDOC	0.005084
NM_000737	CGB	0.005171
NM_133332	WHSC1	0.005269
NM_197962	GLRX2	0.005272
NM_001010901	LOC388791	0.005360
NM_018640	LMO3	0.005378
NM_017445	H2BFS	0.005402
NM_004663	RAB11A	0.005463
NM_138796	LOC128153	0.005495
NM_002759	PRKR	0.005497
NM_015594	DKFZP434O047	0.005520
NM_003772	JRKL	0.005539
NM_003794	SNX4	0.005568
NM_148954	PSMB9	0.005724
NM_007038	ADAMTS5	0.005726
NM_002621	PFC	0.005741
AL353935	AL353935	0.005742
AF090884	AF090884	0.005777
NM_014950	KIAA0997	0.005907
NM_016034	MRPS2	0.005907
NM_024093	MGC5509	0.006015
NM_181804	PKIG	0.006054
NM_002851	PTPRZ1	0.006116
NM_016170	TLX2	0.006134
NM_001561	TNFRSF9	0.006158
NM_015235	CSTF2T	0.006225
NM_005642	TAF7	0.006234
NM_014753	KIAA0187	0.006269

Nº Acesso	Símbolo	<i>p</i> valor
NM_003834	RGS11	0.006270
NM_022078	FLJ12455	0.006324
NM_007284	PTK9L	0.006326
NM_002204	ITGA3	0.006349
NM_017546	C40	0.006366
AK026711	AK026711	0.006478
NM_032417	SPANXD	0.006480
NM_005844	HCGIX	0.006572
NM_012079	DGAT1	0.006594
NM_022095	ZNF335	0.006619
NM_201544	LGALS8	0.006692
NM_022557	GH2	0.006758
NM_024334	MGC3222	0.006799
NM_004143	CITED1	0.006972
NM_022748	TEM6	0.006987
NM_001584	C11orf8	0.006999
NM_006215	SERPINA4	0.007039
NM_003142	SSB	0.007081
NM_001661	ARF4L	0.007085
NM_007195	POLI	0.007120
NM_005201	CCR8	0.007241
CDNA0167	cDNA71	0.007258
NM_014278	APG-1	0.007262
CDNA0220	cDNA26	0.007355
NM_000959	PTGFR	0.007403
NM_000920	PC	0.007415
NM_000585	IL15	0.007416
NM_016503	MRPL30	0.007428
NM_005263	GFI1	0.007515
NM_002881	RALB	0.007531
NM_013367	APC4	0.007537
NM_005253	FOSL2	0.007540
NM_178191	ATPIF1	0.007601
NM_001605	AARS	0.007644
NM_014612	C9orf10	0.007669
NM_000513	OPN1MW	0.007677
NM_152877	FAS	0.007680
NM_004283	RAB3D	0.007862
NM_030971	BCR	0.007865
NM_152713	ITM1	0.007880

Nº Acesso	Símbolo	<i>p</i> valor
NM_006422	AKAP3	0.007902
NM_001730	KLF5	0.007919
NM_017514	HSSEXGENE	0.007953
NM_000813	GABRB2	0.007980
NM_000669	ADH1C	0.008070
NM_001452	FOXF2	0.008090
NM_003436	ZNF135	0.008179
NM_001687	ATP5D	0.008321
NM_024323	MGC11271	0.008405
NM_004782	SNAP29	0.008462
NM_016240	SCARA3	0.008593
AK027414	NOD27	0.008609
NM_001943	DSG2	0.008738
NM_058188	C21orf67	0.008792
NM_133343	RAD17	0.008850
NM_006571	DCTN6	0.008863
NM_021933	FLJ12438	0.008971
NM_207168	ENSA	0.009039
NM_004315	ASAH1	0.009045
NM_014671	KIAA0010	0.009081
NM_022122	MMP27	0.009115
NM_032954	WBSCR14	0.009115
NM_013353	TMOD4	0.009120
NM_000985	RPL17	0.009160
AJ310884	AJ310884	0.009179
NM_031210	C14orf156	0.009210
BC034926	CILP2	0.009295
NM_016058	LOC51002	0.009381
NM_002223	ITPR2	0.009491
NM_201567	CDC25A	0.009499
AK025521	TORC3	0.009540
NM_020685	HT021	0.009589
NM_001816	CEACAM8	0.009595
NM_002937	RNASE4	0.009629
NM_001118	ADCYAP1R1	0.009801
NM_172312	SPAG8	0.009852
NM_001260	CDK8	0.009894

Legenda: *p* valor teste *t* de *student*.

Anexo 2. Assinatura de 117 genes diferencialmente metilados entre os dois grupos de amostras.

Nº Acesso	Símbolo	<i>p</i> valor
NM_002891	RASGRF1	0.000232
NM_006936	SMT3H1	0.000315
NM_004800	TM9SF2	0.000365
NM_001526	HCRTR2	0.000450
NM_032509	LOC84549	0.000490
AF171235	PHGDH	0.000689
X17122	SERPINA1	0.000827
NM_024608	FLJ22402	0.001035
AF217986	pp9099	0.001075
NM_003543	H4FH	0.001080
NM_015148	PASK	0.001168
NM_005288	GPR12	0.001203
NM_021115	SEZ6L	0.001224
NM_000937	POLR2A	0.001229
NM_000737	CGB	0.001230
NM_000851	GSTM5	0.001251
NM_005080	XBP1	0.001318
NM_015160	PMPCA	0.001448
NM_015056	KIAA0179	0.001505
NM_004693	K6HF	0.001783
NM_006705	GADD45G	0.001892
NM_005925	MEP1B	0.001935
NM_017858	FLJ20516	0.001996
NM_020796	SEMA6A	0.002037
NM_030958	OATPRP4	0.002072
NM_018198	FLJ10737	0.002254
NM_005033	PMSCL1	0.002329
NM_015602	LAP1B	0.002386
NM_207491	MGC48628	0.002423
NM_030813	SKD3	0.002487
NM_007074	CORO1A	0.002664
NM_002745	MAPK1	0.002676
NM_052878	C19orf36	0.002711
BC007366	C9orf70	0.002754
NM_004109	FDX1	0.002806
NM_001496	GFRA3	0.002853
NM_032404	TMPRSS3	0.002888

Nº Acesso	Símbolo	<i>p</i> valor
NM_030820	COL21A1	0.003017
NM_014461	CNTN6	0.003088
NM_015872	LOC51043	0.003099
NM_052818	CG018	0.003214
NM_018189	FLJ10713	0.003223
AK001021	FLJ10159	0.003278
NM_002255	KIR2DL4	0.003357
NM_004488	GP5	0.003411
NM_002012	FHIT	0.003538
NM_007117	TRH	0.003554
NM_001968	EIF4E	0.003568
NM_000266	NDP	0.003660
NM_014555	TRPM5	0.003735
NM_024106	MGC2663	0.003803
NM_003003	SEC14L1	0.003961
NM_197967	BID	0.003985
NM_198681	KIAA0720	0.004058
NM_002852	PTX3	0.004060
NM_006731	FCMD	0.004062
NM_014165	HSPC125	0.004078
NM_080658	MGC9740	0.004135
NM_003329	TXN	0.004246
NM_004869	SKD1	0.004273
NM_003495	H4FM	0.004373
NM_001051	SSTR3	0.004538
NM_012328	DNAJB9	0.004566
NM_016221	DCTN4	0.004574
NM_000135	FANCA	0.004812
NM_001690	ATP6A1	0.004909
NM_002970	SAT	0.004911
NM_005165	ALDOC	0.005084
NM_000737	CGB	0.005171
NM_018640	LMO3	0.005378
NM_017445	H2BFS	0.005402
NM_004663	RAB11A	0.005463
NM_003772	JRKL	0.005539
NM_148954	PSMB9	0.005724
NM_007038	ADAMTS5	0.005726
AF090884	AF090884	0.005777
NM_014950	KIAA0997	0.005907

Nº Acesso	Símbolo	<i>p</i> valor
NM_016034	MRPS2	0.005907
NM_181804	PKIG	0.006054
NM_001561	TNFRSF9	0.006158
NM_005642	TAF7	0.006234
NM_014753	KIAA0187	0.006269
NM_002204	ITGA3	0.006349
NM_017546	C40	0.006366
NM_032417	SPANXD	0.006480
NM_005844	HCGIX	0.006572
NM_012079	DGAT1	0.006594
NM_004143	CITED1	0.006972
NM_001584	C11orf8	0.006999
NM_006215	SERPINA4	0.007039
NM_003142	SSB	0.007081
NM_001661	ARF4L	0.007085
NM_005201	CCR8	0.007241
NM_000920	PC	0.007415
NM_005263	GFI1	0.007515
NM_002881	RALB	0.007531
NM_013367	APC4	0.007537
NM_005253	FOSL2	0.007540
NM_001605	AARS	0.007644
NM_014612	C9orf10	0.007669
NM_152877	FAS	0.007680
NM_030971	BCR	0.007865
NM_006422	AKAP3	0.007902
NM_000813	GABRB2	0.007980
NM_004782	SNAP29	0.008462
NM_058188	C21orf67	0.008792
NM_133343	RAD17	0.008850
NM_006571	DCTN6	0.008863
NM_014671	KIAA0010	0.009081
NM_022122	MMP27	0.009115
NM_013353	TMOD4	0.009120
NM_031210	C14orf156	0.009210
NM_002223	ITPR2	0.009491
AK025521	TORC3	0.009540
NM_001816	CEACAM8	0.009595
NM_001118	ADCYAP1R1	0.009801

Nº Acesso	Símbolo	<i>p</i> valor
NM_172312	SPAG8	0.009852

Legenda: *p* valor teste *t* de *Student*.

Anexo 3. Assinatura de 52 genes diferencialmente metilados entre os dois grupos de amostras.

Nº Acesso	Símbolo	<i>p</i> valor
NM_006936	SMT3H1	0.000315
NM_001526	HCRTR2	0.000450
NM_032509	LOC84549	0.000490
AF171235	PHGDH	0.000689
X17122	SERPINA1	0.000827
NM_024608	FLJ22402	0.001035
NM_015148	PASK	0.001168
NM_005288	GPR12	0.001203
NM_000937	POLR2A	0.001229
NM_015160	PMPCA	0.001448
NM_006705	GADD45G	0.001892
NM_020796	SEMA6A	0.002037
NM_207491	MGC48628	0.002423
NM_030813	SKD3	0.002487
NM_002745	MAPK1	0.002676
NM_052878	C19orf36	0.002711
BC007366	C9orf70	0.002754
NM_004109	FDX1	0.002806
NM_001496	GFRA3	0.002853
NM_032404	TMPRSS3	0.002888
NM_030820	COL21A1	0.003017
NM_014461	CNTN6	0.003088
NM_015872	LOC51043	0.003099
NM_018189	FLJ10713	0.003223
NM_004488	GP5	0.003411
NM_000266	NDP	0.003660
NM_024106	MGC2663	0.003803
NM_003003	SEC14L1	0.003961
NM_014165	HSPC125	0.004078
NM_003329	TXN	0.004246
NM_003495	H4FM	0.004373
NM_016221	DCTN4	0.004574
NM_001690	ATP6A1	0.004909
NM_017445	H2BFS	0.005402
NM_007038	ADAMTS5	0.005726
NM_005642	TAF7	0.006234
NM_014753	KIAA0187	0.006269

Nº Acesso	Símbolo	<i>p</i> valor
NM_002204	ITGA3	0.006349
NM_032417	SPANXD	0.006480
NM_005844	HCGIX	0.006572
NM_012079	DGAT1	0.006594
NM_004143	CITED1	0.006972
NM_005263	GFI1	0.007515
NM_002881	RALB	0.007531
NM_013367	APC4	0.007537
NM_001605	AARS	0.007644
NM_152877	FAS	0.007680
NM_006422	AKAP3	0.007902
NM_000813	GABRB2	0.007980
NM_006571	DCTN6	0.008863
NM_014671	KIAA0010	0.009081
NM_001118	ADCYAP1R1	0.009801

Legenda: *p* valor teste *t* de *Student*.