

# **PERFIL GENÔMICO DE CARCINOMAS DE MAMA DO TIPO DUCTAL INVASIVO**

**FELIPE FIDALGO DE CARVALHO**

**Dissertação apresentada à Fundação Antônio  
Prudente para obtenção do título de Mestre em  
Ciências**

**Área de Concentração: Oncologia**

**Orientadora: Dra. Ana Cristina Victorino  
Krepischi**

**São Paulo**

**2012**

## FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Carvalho, Felipe Fidalgo de

**Perfil genômico de carcinomas de mama do tipo ductal invasivo /**  
Felipe Fidalgo de Carvalho – São Paulo, 2012.

96p.

Dissertação (Mestrado)-Fundação Antônio Prudente. Curso de  
Pós-Graduação em Ciências – Área de concentração: Oncologia.

Orientadora: Ana Cristina Victorino Krepschi

Descritores: 1. NEOPLASIAS DA MAMA. 2. ONCOGENES. 3. GENES  
SUPRESSORES. 4. GENES *DRIVERS*. 5. ALTERAÇÃO DO NÚMERO  
DE CÓPIAS DE DNA.

## DEDICATÓRIA

Dedico esse trabalho aos meus dois maiores orgulhos da vida, minhas maiores fontes de inspiração e aprendizado, esse trabalho é dedicado principalmente para minha mãe e minha avó, as duas jóias mais raras e preciosas desse universo. Amo vocês mais que a mim.

À minha família “Fidalgo” pelo incentivo e fé que sempre depositaram em mim.

Aos meus amigos/irmãos que estão sempre comigo nos momentos bons e ruins, mais bons do que ruins. E assim sempre será.

## **AGRADECIMENTOS**

Primeiramente agradeço à minha orientadora Ana por me acolher, pela imensa paciência comigo, pelos conselhos, ajuda, incentivo, crédito, pela liberdade e autonomia que me dá para fazer minhas coisas e pelo tempo (muito valioso por sinal) que gasta vendo milhões de coisas, me orientando. Você foi mais que uma chefe ou orientadora, foi praticamente uma mãe. MUITO OBRIGADO MESMO!!! DE CORAÇÃO!!! Ass.: Anjo! hahaha

A Amandinha que me puxou pra cá e vem desde minha iniciação me ajudando, ouvindo, comentando sobre os finais de semana e sendo mais que uma colega de trabalho e se tornando praticamente uma irmã mais velha que eu admiro e gosto demais!!!

Ao Departamento de Patologia do Hospital AC Camargo.

A Elô, Vera e Louise do Banco de Macromoléculas pela grande eficácia na extração e entrega (algumas delivery, outras drive thru) de material.

Dra. Carla Rosenberg e todo seu grupo de pesquisa também, pelas reuniões de sexta que por mais que sejam complicadas de entender pelo meu péssimo inglês sempre enriquece nosso conhecimento científico. E pelos lanchinhos também, claro!

Tatiane Rodrigues, parça de experimentos e congressos, quartos super-aquecidos pelos fornos da cozinha do hotel. É nós! Rs

Nescau, por ter me ajudado com o inglês no summary e por uma aula particular de como escrever direitinho!!! Vlw!!!!!!

Amigos do CIPE, que seu for citar todos vou desenvolver tendinite. Vocês foram MUITO importantes em toda essa caminhada. Pelos cafés na copa, almoços, baladinhas bizarras, digestão no sofá da Hebe, sinuca no Portuga, enfim, não imagino melhores pessoas para um ambiente de trabalho. Vocês são demais!!!

À minha mãe, sempre!!! Aquela que eu venero, olho ainda com olhos de criança que não sabe nada e busca em você o exemplo de tudo que há de vir por essa vida. Você e a vovó são o maior presente que alguém pode ganhar de Deus. Se não fosse você ia ser quase impossível véia.

Vovó, minha rainha maior!!! Você é sem dúvidas o maior exemplo de força, honestidade, humildade, garra, fé e de que tudo é possível mesmo quando não há luz visível. Você nos mostrou que a luz vem da gente, de dentro, e sem você nossa vida seria uma eterna escuridão. Obrigado por ser minha outra mãe durante toda minha vida. Minha véinha linda!!!

Minha princesa (namorada) Julia que me ajudou muuuito nos momentos de tensão pré-defesa e que me deu muita força quando eu precisei, além de também ser a pessoa maravilhosa por quem eu me apaixono todos os dias!!!

Família Fidalgo toda, principalmente minha “irmã” Tata, Inha, Lucca e Pietro, sem vocês certamente eu seria uma pessoa infeliz. Amo vocês!!!

Acho justo também agradecer ao meu papagaio e tartaruga, até pq são parte da família!!!

E claro, meus irmãos que escolhi, Texa, Le, Gordinha, Anã, Pipou, Musa, Peito, Pestana, Babs, Muricy, Vini e Renan. Sem demagogia, agradeço todos os dias a Deus por ter vocês ao meu lado. As melhores risadas, capotes, tédios, afogamentos, viagens, furadas, enfim, são com vocês!!!

## RESUMO

Fidalgo F. **Perfil genômico de carcinomas de mama do tipo ductal invasivo**. São Paulo; 2012. [Dissertação de Mestrado-Fundação Antônio Prudente].

O câncer é uma doença complexa, decorrente do acúmulo de múltiplas e sucessivas mutações somáticas, resultando em modificações no funcionamento de genes que atuam em vias moleculares específicas, tais como controle de ciclo celular, apoptose, diferenciação e angiogênese. No estudo do genoma de tumores, um dos objetivos é identificar novos genes relacionados à carcinogênese, muitos deles possivelmente contidos em alterações cromossômicas. O câncer de mama é um modelo importante para estudos genômicos por se tratar de uma doença complexa e heterogênea, ainda não totalmente compreendida. Neste estudo, propusemos a investigação de alterações genômicas como ganhos e perdas de segmentos de DNA em carcinomas de mama do tipo ductal invasivo, buscando a identificação de perfis genômicos característicos de alguns diferentes subtipos deste câncer. Adicionalmente, buscamos estabelecer correlação entre as alterações genômicas encontradas e a ocorrência ou não de comprometimento linfonodal axilar ao diagnóstico. Foram investigados 71 carcinomas de mama invasivos, sendo 57 tumores primários do tipo ductal. A análise inicial do grupo de tumores detectou a ocorrência frequente (>20%) de ganhos nos cromossomos 1q, 8q, 16p, e na região 17q12-q21, enquanto que perdas recorrentes foram detectadas principalmente em 8p, 11q, 16q, e nos cromossomos 13 e 18. Foram identificadas 3828 alterações somáticas no número de cópias de segmentos genômicos (SCNAs), com número médio de 53 SCNAs por genoma tumoral e tamanho médio de 6,9 Mb. O perfil genômico de SCNAs diferiu entre os carcinomas de mama de acordo com o *status* de expressão do receptor de estrógeno (ER): foi detectada diferença estatística no número de SCNA por

genoma tumoral e no tamanho dessas alterações, ambos maiores no grupo de tumores negativos quanto à expressão do receptor de estrógeno (ER(-)) quando comparado com o grupo expressando ER (ER(+)). Além disso, o perfil genômico de frequência de SCNA evidenciou que o grupo ER(-) apresenta uma variabilidade maior de regiões genômicas afetadas quando comparado ao grupo ER(+), com grandes regiões de perdas em 4q e 5q só detectadas nos tumores ER(-), além de ganhos focais em 10p e 17q. Por outro lado, o grupo de tumores ER(+) exibiu principalmente ganho em 16p. Na análise para a identificação de alterações genômicas relevantes para a presença de metástase em linfonodos axilares ao diagnóstico [Nx (linf(+))0], selecionamos apenas os 57 carcinomas de mama invasivos do tipo ductal. Verificamos ganhos mais frequentes nos tumores Nx (linf(+)) em 5p, 7p, 8q, 10p e no cromossomo 19; o grupo de tumores sem acometimento linfonodal exibe SCNAs mais concentradas em outras regiões do genoma, como perdas em 1p e 9p e ganho em 16p. A última etapa de nossa análise foi investigar as regiões genômicas afetadas por SCNAs de maior amplitude, como ganhos em alto número de cópias e deleções em homozigose, excluindo aquelas já associadas à tumorigênese em mama. Essa análise foi realizada apenas no grupo de carcinomas de mama com metástase linfonodal ao diagnóstico, uma vez que estávamos buscando identificar potenciais oncogenes e genes supressores tumorais associados à progressão tumoral. Sete genes mapeados nestas regiões afetadas por SCNAs de maior amplitude foram selecionados para validação técnica. Tais genes, candidatos a novos genes relevantes para a progressão do câncer de mama foram validados e são: *KISS1*, *ADAMTS3* e *HSD17B12* (detectados em *amplicons*) e *ZBTB20*, *MTAP*, *ELF5* e *RERGL* (afetados por deleção em homozigose).

## SUMMARY

Fidalgo F. [**Genomic profiling of ductal invasive breast carcinomas**]. São Paulo; 2012. [Dissertação de Mestrado-Fundação Antônio Prudente].

Cancer is a complex disease caused by the accumulation of multiple and successive somatic mutations, resulting in modifications in gene functions that act on specific molecular pathways, such as cycle cell control, apoptosis, differentiation and angiogenesis. On tumor genome studies one of the aims is to identify new genes related to carcinogenesis, a lot of them possibly restrained on the chromosomal alterations. The breast cancer is an important model to genomic studies because it is a complex and heterogeneous disease, not fully understood yet. In this study, we proposed to evaluate the genomic alterations as gains and losses of DNA segments in ductal invasive breast carcinomas, seeking to identify characteristic genomic profiles of some different subtypes of this cancer. Moreover, we seek to establish a correlation between the genomic alterations found and the occurrence or not of axillary lymph nodal involvement at diagnosis. It was investigated 72 breast carcinomas of ductal invasive type. The initial analysis of the tumor group detected the frequent occurrence (>20%) of gains on chromosomes 1q, 8q, 16p and 17q12-q21 region, while recurrent losses were detected mainly at 8p, 11q, 16q and on the chromosomes 13 and 18. It was identified 3828 copy number somatic alterations in genomic segments (SCNAs), with an average number of 53 per tumoral genome and an average size of 6,9 Mb. Loss events showed a higher significantly size ( $p < 0,0001$ ) than gain events, and high amplitude events (amplicons and homozygous losses) affected significantly smaller genomic segments. The genomic profile of SCNA differed also between the breast carcinomas according to the estrogen receptor (ER) expression status: it was detected significant difference on the SCNA number per tumoral genome and on the size of this alterations, that were both greater in the group of estrogen receptor negative tumors (ER(-)) when compared to the expressing group (ER(+)). Furthermore, the genomic profile of SCNAs showed that the ER(-) group has a greater variability of affected genomic regions when compared to ER(+) group, with large losses



mapped at 4q and 5q, besides focal gains that were identified at 10p and 17q; the ER(+) tumor group showed a high frequency of 16p gain. We also directed the analysis to identify relevant genomic alterations possibly related to the presence of metastasis at axillary lymph nodes at diagnosis [Nx (linf(+))]. Gains were frequent on Nx (linf(+)) tumors at 5p, 7p, 8q 10p and chromosome 19, while a frequent loss was detected at 8p; the carcinomas without lymph nodes metastasis exhibited a different profile, with SCNAs concentrated in other regions of the genome such as 1p and 9p losses and 16p gain. The last step of our analysis was to investigate genomic regions affected by high amplitude SCNAs (high copy number gains=amplicons, and homozygous deletions), excluding those already related to the breast tumorigenesis. This analysis was performed only on the breast carcinoma group with lymph node metastasis at diagnosis, since we were seeking to identify potential oncogenes and tumor suppressor genes associated with the cancer progression. Genes mapped on seven regions affected by high amplitude SCNAs were selected to technical validation. These new relevant candidate genes in breast cancer progression are: *KISS1*, *ADAMTS3* and *HSD17B12* (mapped at amplicons) and *ZBTB20*, *MTAP*, *ELF5* and *RERGL* (affected by homozygous deletion). The final purpose of a study like ours is to enlarge the knowledge to obtain defined gene signatures to each tumor, possibiliting more effective individualized treatments.

## LISTA DE FIGURAS

<b>Figura 1</b>	Alterações cromossômicas detectadas em câncer.....	11
<b>Figura 2</b>	Distribuição das mutações de ponto e SCNA detectadas em 40 genes de câncer no grupo de 100 tumores de mama do trabalho de STEPHENS et al. 2012.....	19
<b>Figura 3</b>	Técnica de array-CGH.....	26
<b>Figura 4</b>	Tamanho da alteração de acordo com o tipo de SCNA.....	45
<b>Figura 5</b>	Perfil global de ganhos e perdas de segmentos genômicos do grupo de carcinomas de mama invasivos.....	49
<b>Figura 6</b>	Análise GISTIC do grupo de carcinomas de mama invasivos.....	52
<b>Figura 7</b>	Padrão de SCNA dos carcinomas de mama de acordo com o status de expressão do receptor hormonal de estrógeno (ER).....	55
<b>Figura 8</b>	Padrão de SCNA dos carcinomas de mama de acordo com o status linfonodal.....	59
<b>Figura 9</b>	Análise GISTIC do grupo de carcinomas de mama linfonodo-positivo e do grupo linfonodo-negativo.....	61
<b>Figura 10</b>	Amplificação em 1q32.1.....	65
<b>Figura 11</b>	Amplificação em 4q13.3.....	66

<b>Figura 12</b>	Deleção em homozigose em 9p21.3.....	67
<b>Figura 13</b>	Deleção em homozigose em 12p12.3.....	68
<b>Figura 14</b>	Deleções em homozigose em 3q13.31 e 11p13.....	69
<b>Figura 15</b>	<i>Amplicons</i> em 11p11.2.....	70
<b>Figura 16</b>	Genes validados por <i>qPCR</i> .....	72

## LISTA DE QUADROS E TABELAS

<b>Quadro 1</b>	Parâmetros de qualidade do experimento de <i>microarray</i> .....	31
<b>Quadro 2</b>	Genes selecionados testados por <i>qPCR</i> : símbolo oficial, nome e mapeamento desses genes, e sequência dos pares de <i>primers</i> utilizados.....	36
<b>Tabela 1</b>	Caracterização clínica e anatomopatológica da amostra estudada de setenta e dois carcinomas de mama esporádicos do tipo ductal invasivo.....	39
<b>Tabela 2</b>	Frequência de variáveis clínicas e anatomopatológicas no grupo de 57 carcinomas de mama do tipo ductal invasivo ou <i>Non Special Type</i> (NST).....	43
<b>Tabela 3</b>	Tamanho da alteração de acordo com cada tipo de SCNA.	45
<b>Tabela 4</b>	Dados de <i>array-CGH</i> dos carcinomas de mama.....	47
<b>Tabela 5</b>	Segmentos genômicos mínimos de sobreposição de SCNAs significativamente associados ao grupo de carcinomas de mama em frequência igual ou superior 30% de acordo com análise STAC.....	50
<b>Tabela 6</b>	Regiões cromossômicas significativamente associadas ao grupo de carcinomas de mama do tipo ductal invasivo na análise GISTIC.....	53
<b>Tabela 7</b>	Descrição do padrão de SCNA dos carcinomas de mama em relação ao <i>status</i> do receptor hormonal de estrógeno (positivo ou negativo).....	54

<b>Tabela 8</b>	Regiões cromossômicas significativamente ( $p \leq 0,001$ ) associadas ao grupo de carcinomas de mama ER(-) de acordo com a comparação dos dados após análise STAC..	56
<b>Tabela 9</b>	Descrição do padrão de SCNA dos 57 carcinomas de mama do tipo ductal invasivo (número médio por tumor e distribuição do número médio de acordo com o tipo de alteração) em relação ao <i>status</i> linfonodal ao diagnóstico (positivo ou negativo).....	58
<b>Tabela 10</b>	Regiões cromossômicas significativamente ( $p < 0,05$ ) associadas aos grupos de carcinomas de mama linf(+) e linf(-) após análise STAC.....	60
<b>Tabela 11</b>	Regiões cromossômicas significativamente associadas ao grupo de carcinomas de mama linf(+) na análise GISTIC....	62
<b>Tabela 12</b>	Regiões cromossômicas afetadas e respectivos genes escolhidos para validação por <i>qPCR</i> .....	63

## LISTA DE SIGLAS E ABREVIATURAS

<b>array-CGH</b>	Hibridização genômica comparativa em microarranjos
<b>BRCA1</b>	<i>breast cancer 1, early onset</i>
<b>BRCA2</b>	<i>breast cancer 2, early onset</i>
<b>CGH</b>	Hibridização genômica comparativa
<b>CIPE</b>	Centro Internacional de Pesquisa e Ensino
<b>CIT</b>	Células isoladas de tumor
<b>CNA</b>	<i>Copy number alteration</i> (alteração de número de cópias de segmentos de DNA)
<b>CNV</b>	<i>Copy Number Variation</i> (variação de número de cópias de segmentos de DNA)
<b>C<sub>T</sub></b>	Limiar do ciclo
<b>Cy3-dCTP</b>	Dinucleotídeo C marcado com Cianina 3
<b>Cy5-dCTP</b>	Dinucleotídeo C marcado com Cianina 5
<b>DGV</b>	<i>Database of Genomic Variants</i>
<b>DNA</b>	Ácido desoxirribonucléico
<b>Eff</b>	Eficiência de amplificação
<b>EGFR</b>	<i>epidermal growth factor receptor</i>
<b>ER</b>	Receptor de estrógeno
<b>ERBB2</b>	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)
<b>FASST2</b>	<i>Fast Adaptative States Segmentation Technique</i>
<b>FISH</b>	hibridização <i>in situ</i> fluorescente
<b>GO</b>	<i>Gene Ontology</i>
<b>GSEA</b>	<i>Gene Set Enrichment Analysis</i>
<b>HCACC</b>	Hospital do Câncer A.C. Camargo
<b>HER2</b>	Receptor epitelial humano tipo 2
<b>HMM</b>	<i>Hidden Markov Model</i>
<b>INCA</b>	Instituto Nacional do Câncer
<b>Kb</b>	kilobase

<b>LOH</b>	Perda de heterozigose
<b>Mb</b>	Megabase
<b>mm</b>	Milímetros
<b>mRNA</b>	RNA mensageiro
<b>°C</b>	grau Celsius
<b>PCR</b>	Polimerase chain reaction (reação em cadeia da polimerase)
<b>PR</b>	Receptor de progesterona
<b>qRT-PCR</b>	<i>quantitative real-time PCR</i> (reação em cadeia da polimerase em tempo real com análise quantitativa)
<b>RNA</b>	Ácido ribonucleico
<b>SCNA</b>	<i>Somatic copy number alteration</i> (alteração somática de número de cópias de segmentos de DNA)
<b>SKY</b>	Cariotipagem espectral
<b>SNP</b>	Polimorfismo de nucleotídeo único
<b>UCSC</b>	Universidade da Califórnia, Santa Cruz
<b>YBX1</b>	Y box binding protein 1
<b>ΔDRn</b>	Magnitude do sinal de fluorescência gerado durante a reação de amplificação (PCR) a cada ponto de tempo
<b>ng</b>	Nanograma
<b>NST</b>	<i>Non Special Type</i>
<b>QT-PO</b>	Quimioterapia pré-operatória

# ÍNDICE

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
1.1	Carcinoma de mama .....	2
1.2	Genoma do câncer .....	7
1.3	Alterações cromossômicas em câncer de mama .....	12
<b>2</b>	<b>OBJETIVOS</b> .....	<b>22</b>
2.1	Objetivo Geral.....	22
2.2	Objetivos Específicos .....	22
<b>3</b>	<b>MATERIAL E MÉTODOS</b> .....	<b>24</b>
3.1	Material.....	24
3.2	Métodos.....	25
3.2.1	Hibridação genômica comparativa em microarranjos de oligonucleotídeos ( <i>array-CGH</i> ).....	25
3.2.2	Análise dos dados dos microarranjos.....	31
3.2.3	Validação de ganhos e perdas de genes por <i>qPCR</i> .....	33
<b>4</b>	<b>RESULTADOS</b> .....	<b>37</b>
4.1	Caracterização das amostras estudadas .....	37
4.2	Mapeamento das SCNAs recorrentes no grupo de carcinomas de mama esporádicos do tipo ductal invasivo .....	44
4.3	Delineamento do perfil genômico de acordo com o padrão de expressão do receptor de estrógeno .....	54
4.4	Correlação entre o perfil genômico de <i>scna</i> dos carcinomas com o <i>status</i> de metástase linfonodal ao diagnóstico .....	57
4.5	Mapeamento de amplicons e deleções em homozigose para identificação de potenciais oncogenes e supressores tumoriais relevantes para a progressão para metástase linfonodal .....	63



<b>5</b>	<b>DISCUSSÃO .....</b>	<b>73</b>
<b>6</b>	<b>CONCLUSÕES .....</b>	<b>86</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRAFICAS .....</b>	<b>88</b>

## **ANEXO**

**Anexo 1** Genes citados em trabalho de Stephens et al. 2012

# 1 INTRODUÇÃO

O câncer é uma doença complexa cuja etiologia reside no acúmulo de múltiplas e sucessivas alterações genéticas em células somáticas (HAHN e WEINBERG 2002). Este processo gera modificações no padrão de funcionamento de genes que atuam em vias moleculares específicas, tais como controle de ciclo celular, apoptose, diferenciação e angiogênese, por exemplo.

Alguns princípios gerais do processo oncogênico já foram descritos, principalmente as funções que a célula precisa adquirir, em qualquer tipo tumoral: independência de sinais de proliferação celular, escape da inibição de crescimento e da sinalização de apoptose, aquisição de potencial indefinido de proliferação, capacidade de invasão de tecidos adjacentes e, por fim, habilidade de colonizar sítios distantes (metástases) (HANAHAN e WEINBERG 2000). Adicionalmente, inúmeros trabalhos produziram *insights* a respeito das vias biológicas e redes gênicas funcionais afetadas durante a progressão do câncer (JONES 2008). Mais recentemente, evidencia-se a natureza do câncer como uma doença sistêmica, com envolvimento de um número crescente de tipos celulares no seu desenvolvimento e papel primordial do microambiente (WEIGELT e BISSELL 2008) e do sistema imune (DUNN et al. 2004).

## 1.1 CARCINOMA DE MAMA

O carcinoma de mama é o câncer mais frequente em mulheres, e o segundo tipo de câncer mais frequente no mundo. Entre as mulheres, é a primeira causa de morte por câncer. Aproximadamente 1,4 milhões de casos novos de câncer de mama foram registrados no ano de 2008 em todo o mundo, representando 23% de todos os tipos de câncer. De acordo com dados do Instituto Nacional de Câncer-INCA (Ministério da Saúde 2011), o número previsto de casos novos de câncer de mama para o Brasil em 2012 seria de 52.000.

Os tumores de mama são altamente heterogêneos, e essa diversidade está associada com aspectos específicos de morfologia e imunohistoquímica, entre outros. Como fator complicador, pacientes afetadas por câncer de mama do mesmo tipo e estadio podem apresentar progressão variável.

Os tipos mais comuns de tumores de mama são agrupados devido à sua localização nos lóbulos e ductos mamários. Tumores de mama confinados aos ductos ou lóbulos são chamados de carcinomas *in situ*, que podem ser divididos em:

(a) carcinoma lobular *in situ* – é restrito aos lóbulos, não sendo considerado um câncer invasivo, embora mulheres com este tipo de carcinoma possuam um maior risco de adquirir um tumor invasivo em ambas as mamas;

(b) carcinoma ductal *in situ* – é o tipo mais comum de carcinoma *in situ*, no qual as células do tumor estão restritas aos ductos. É tratado cirurgicamente, além de radioterapia quando necessário; pode evoluir para um câncer invasivo.

Os tumores que ultrapassam as paredes ductais ou lobulares e se espalham pelo tecido mamário adjacente são conhecidos como carcinomas invasivos. Os carcinomas ductais invasivos correspondem a aproximadamente 80% dos tumores invasivos da mama. Estes tumores podem também chegar a invadir linfonodos da mama ou próximos à mesma e, mais tardiamente, desenvolver metástase à distância. Existem muitos tipos dessa classe tumoral, como: carcinoma medular – este tipo tumoral possui uma borda bem definida separando-a do tecido mamário normal, com presença de células do sistema imune; carcinoma tubular – apresenta melhor prognóstico devido à característica de ser menos invasivo do que os outros tipos; tumores metaplásicos – é um tipo muito raro de carcinoma ductal invasivo, apresentando células que não são encontradas normalmente na mama, como células da pele e ósseas; carcinoma colóide ou mucinoso – é também um tipo raro de carcinoma, formado por células tumorais produtoras de muco; este último tipo possui melhor prognóstico e menor chance de metástase se comparado aos tumores invasivos mais comuns. Outro tipo relevante de tumores é o carcinoma lobular invasivo, que representa de 10 a 15% dos tumores invasivos da mama, e possui alta capacidade de invadir o tecido gorduroso mamário e causar metástase (National Comprehensive Cancer Network-NCCN 2012).

Dados clínicos e de análises anatomopatológicas podem determinar o grau de malignidade do tumor e fornecer indicações para tratamento e prognóstico por meio da chamada classificação TNM: Tamanho do tumor primário (T) (classifica em categorias e subcategorias de acordo com o tamanho do tumor); status linfonodal regional (N) (identifica se há metástase linfonodal); e, por último, presença de metástase à distância (M) (ARNONE et al. 2010). A partir da combinação destes parâmetros é possível definir o estadiamento da doença, que também é classificado em categorias: estadios 0, I, IIA e IIB, IIIA, IIIB e IIIC, e estadio IV (SOBIN et al. 2009).

Alguns fatores prognósticos tradicionais no carcinoma de mama são o *status* linfonodal axilar e o tamanho dessa metástase. O critério de acometimento linfonodal é derivado primeiramente da análise do linfonodo sentinela, e depende do tamanho do grupamento celular encontrado no mesmo. Se houve a detecção de um grupamento celular no linfonodo sentinela e este for <2 mm, considera-se como micro-metástase linfonodal; grupamentos >2 mm são considerados como linfonodo-sentinelas acometidos por metástase e, a partir disso, há indicação de retirada dos linfonodos axilares para investigação.

O status de acometimento linfonodal parece estar diretamente associado ao desenvolvimento de metástases à distância nas pacientes com câncer de mama. Aproximadamente 75% das pacientes que apresentam linfonodo-negativo nunca desenvolverão metástase à distância (KAHN et al. 2006). Em trabalho recente, GOBARDHAN et al. (2011) correlacionaram diversos parâmetros clínico-anatomopatológicos (dados de acometimento de

linfonodo, tamanho do tumor, grau de diferenciação, idade da paciente, índice mitótico e expressão de receptores hormonais) de pacientes submetidas a diferentes tipos de tratamentos. A análise global desses dados indicou que quanto maior o grau de acometimento linfonodal, menores os índices de sobrevida global e sobrevida livre de doença.

Mais recentemente, a biópsia de linfonodo-sentinela vem substituindo a dissecação dos linfonodo axilares. Em análises mais detalhadas dos linfonodos sentinela, as micro-metástases, se existentes, podem ser detectadas. Micro-metástases linfonodais ou células isoladas do tumor (CITs) são observadas em até 23% das pacientes com câncer de mama (GOBARDHAN et al. 2011). Micro-metástases e CITs diferem em relação a tamanho de grupamento celular (CITs  $\leq 0,2$  mm e micro-metástases variam entre  $>0,2$  mm a 2 mm) e indicam diferentes estadiamentos da doença. As CITs não são consideradas metástases, mas existe entre os especialistas a preocupação de que estejam associadas a metástases linfonodais (CSERNI et al. 2006; DE MASCAREL et al. 2008; GOBARDHAN et al. 2011). Outro fator prognóstico bem definido em tumores de mama é o grau histológico de Scarff-Bloom-Richardson modificado por Elston e Ellis (SBR modificado ou Nottingham) - a agressividade fenotípica é graduada pela soma de valores dados segundo critérios morfológicos de diferenciação glandular, pleomorfismo nuclear e contagem mitótica (Elston CW e Ellis IO. 1991). O índice SBR foi amplamente validado e é recomendado pela União Internacional Contra o Câncer (UICC 2009).

Carcinomas de mama exibem diferentes perfis de expressão gênica (ou assinatura gênica), que contribuem também para diferenciá-los quanto a prognóstico e resposta a quimioterapia (PUSZTAI et al. 2008). Com base na expressão diferencial de grupos de genes, os carcinomas mamários podem ser divididos em seis subtipos principais. Três desses subtipos (Luminal A, B e C) exibem expressão moderada ou alta de genes epiteliais luminais, como o gene do receptor de estrógeno (*ER*), por exemplo. Além desses, existem três subtipos principais que não são positivos para *ER*, descritos a seguir. O subtipo HER2 positivo apresenta níveis baixos de expressão de genes epiteliais luminais, alta expressão de laminina e das citoqueratinas 5 e 17, apresenta alta expressão de genes contidos no amplicon do gene *ERBB2*, e está associado a um prognóstico ruim (PEROU et al. 2000; SØRLIE et al. 2001). O subtipo *normal breast-like* corresponde a tumores que apresentam perfil de expressão gênica muito similar a células de mama normal e de fibroadenomas (REIS-FILHO e PUSZTAI 2011).

O receptor de estrógeno é um importante marcador prognóstico nos tumores de mama; tumores que não expressam esse receptor evoluem de maneira pior, sendo até considerados como uma doença distinta dentro dos carcinomas de mama em níveis de transcrição gênica (SØRLIE et al. 2001; WANG DY et al. 2011; REIS-FILHO e PUSZTAI 2011). Um grupo de grande importância prognóstica é o de tumores negativos quanto aos receptores de estrogênio e progesterona (*PR*), e também negativos quanto à expressão de *ERBB2*, conhecidos como triplo-negativos. Esse tipo de tumor está associado a um pior prognóstico, diferente dos tumores que são *ER* e *PR*-

positivos, que exibem boa resposta a terapias endócrinas, como tamoxifeno (WANG DY et al. 2011). Diversos estudos ressaltam a importância do status de expressão gênica do receptor de estrogênio, evidenciando que tumores de mama *ER*-positivo e *ER*-negativo são fundamentalmente doenças distintas em termos moleculares (REIS-FILHO e PUSZTAI 2011).

Entretanto, na prática clínica, o câncer de mama é geralmente classificado principalmente com base na expressão protéica de receptores de estrogênio e progesterona, e do gene *ERBB2* (PEROU et al. 2000; SØRLIE et al. 2001), além de marcadores de índice mitótico como KI67 e algumas citoqueratinas. Isto acontece em grande parte porque o perfil de expressão gênica de cada tumor ainda não pode ser obtido em tempo hábil para direcionar decisões clínicas, como o tipo de tratamento mais adequado. Mas outro aspecto importante é a falta do conhecimento necessário para usar essas informações com fins terapêuticos.

## **1.2 GENOMA DO CÂNCER**

As alterações genéticas exercem um papel importante não só na indução da tumorigênese em geral, como estão associadas à heterogeneidade inter e intra-tumoral (BAYANI et al. 2007). A falha na manutenção da integridade do genoma, com acúmulo de danos ao DNA, é altamente relevante na tumorigênese, sendo a instabilidade genômica uma característica de quase todos os cânceres humanos (BARTEK et al. 2007; NEGRINI et al. 2010). A grande instabilidade genética é evidenciada pela



aquisição de inúmeras alterações somáticas (STRATTON et al. 2009), como substituições de base, translocações, ampliações, duplicações e deleções, além de aneuploidias cromossômicas.

Uma consequência importante dos diversos tipos de mutação no câncer é a perda de função de genes supressores tumorais e a ativação de proto-oncogenes, eventos que disparam a proliferação celular descontrolada, a instabilidade genômica, e o concomitante aparecimento de tecidos neoplásicos. Em adição à iniciação da carcinogênese, a instabilidade genômica também direciona a progressão dos tumores de benignos para malignos, pela ocorrência de alterações genéticas adicionais que facilitam a evolução para um estágio mais agressivo. Alguns estudos, porém, mostram que a instabilidade genômica em muitos cânceres esporádicos humanos não ocorre devido à mutação na sequência de genes “protetores” do genoma, e que a frequência de mutações detectada nestes genes pode ser uma subestimativa da frequência real de inativação dos mesmos nos tumores, uma vez que a função gênica pode ser reprimida por outros mecanismos (NEGRINI et al. 2010).

Os resultados da análise genômica de tumores revelam, em conjunto, um grupo relativamente pequeno de genes que sofrem mutações somáticas em uma alta proporção dos cânceres, ocorrendo de maneira recorrente e característica para alguns tipos de tumor. No entanto, mutações não recorrentes foram detectadas em um número muito superior de genes, como variantes individualmente raras (STRAUSBERG e SIMPSON 2010). A recorrência de uma mutação pode, de fato, indicar sua relevância para a

oncogênese, mas a ausência de recorrência não deve ser tomada diretamente como uma evidência de mutação sem significado funcional. Tais variantes incomuns evidenciam a diversidade molecular do mesmo tipo tumoral durante o desenvolvimento de cada tumor individual. A análise desses padrões mutacionais únicos no contexto de vias biológicas e redes gênicas pode revelar, subjacentes à diversidade genômica tumoral, mecanismos comuns do processo carcinogênico (LEARY et al. 2008).

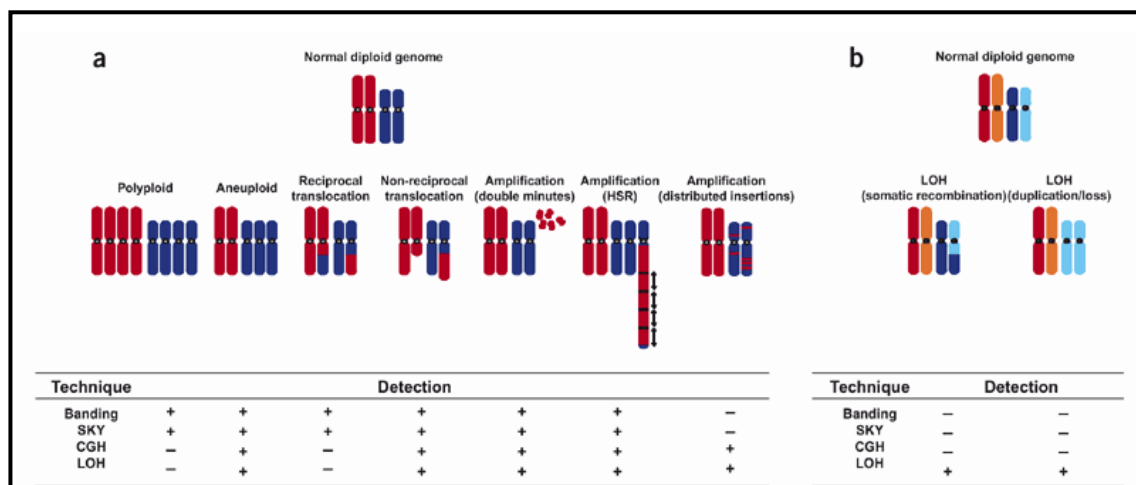
A maioria das classes de mutação afeta um único gene, ou no máximo dois genes (como os genes de fusão decorrentes de translocações cromossômicas), mas duplicações e deleções podem afetar muitos megabases e, portanto, inúmeros genes. Assim, a análise do genoma de um tumor pela investigação de ganhos e perdas de segmentos cromossômicos produz evidências genéticas da existência de genes provavelmente relevantes nas regiões afetadas. Em decorrência da instabilidade do genoma do tumor, o número de deleções e duplicações identificado em um mesmo tumor ou tipo tumoral pode ser grande o bastante para dificultar o reconhecimento de genes relevantes do ponto de vista funcional. Por isso, informações adicionais precisam ser obtidas posteriormente em outros níveis de análise, como estudo de expressão gênica, identificação de modificações epigenéticas, bem como ensaios funcionais, para determinar quais dos genes afetados são funcionalmente relevantes na carcinogênese. Diferentemente dos cânceres hereditários, para os quais o evento inicial da carcinogênese muitas vezes pode ser claramente identificado (mutação constitutiva em um gene supressor tumoral), os cânceres esporádicos

exibem um perfil heterogêneo e o processo de iniciação da instabilidade genômica ainda é pouco compreendido.

Os rearranjos cromossômicos são modificações genéticas freqüentemente encontradas no câncer. A citogenética clássica foi um instrumento muito importante na identificação de rearranjos cromossômicos associados a câncer, recorrentes ou não, em particular na detecção de translocações (FUTREAL et al. 2004; MITELMAN et al. 2007). Entretanto, o estudo citogenético não foi uma abordagem bem sucedida para alguns tipos tumorais, em especial tumores sólidos, principalmente devido à complexidade das alterações cariotípicas e também à dificuldade de cultivo das células tumorais para obtenção de metástases. Nestes casos, o advento da hibridação comparativa de genomas (*CGH*) (KALLIONIEMI et al. 1992) permitiu a utilização do DNA do tumor para investigar ganhos e perdas cromossômicos, produzindo um grande corpo de conhecimento acerca das alterações cromossômicas de tumores, em especial de tumores sólidos (GEBHART 2004). Como a técnica de *CGH* utiliza metáfases de células normais como alvo de hibridação para investigar alterações genômicas quantitativas no DNA do tumor, a resolução desta metodologia permaneceu limitada à estrutura cromossômica (5-10 Mb). O posterior desenvolvimento da hibridação comparativa de genomas em *microarrays* (*array-CGH*) permitiu investigar também as alterações cromossômicas submicroscópicas (deleções ou duplicações < 4 Mb), usando milhares de seqüências-alvo simultaneamente, o que aumentou extraordinariamente a resolução e a capacidade de investigação da citogenética (RODRIGUEZ-REVENGA et al.

2007). É importante ressaltar que as técnicas de *CGH* e *array-CGH* apresentam limitações, como a impossibilidade de detecção de rearranjos cromossômicos estruturais ou numéricos “equilibrados”, ou seja, não associados a perdas ou ganhos de material genético, como as alterações equilibradas, inversões e poliploidias.

A Figura 1 apresenta um resumo das aberrações cromossômicas em câncer e a capacidade de detecção de algumas técnicas, como bandamento cromossômico G (*G banding*), hibridação comparativa de genomas (*CGH*), cariotipagem molecular (*spectral karyotyping* - *SKY*), e métodos para investigação de perda de heterozigose (*loss of heterozygosity* – *LOH*).



**Figura 1** - Alterações cromossômicas detectadas em câncer. Ilustração esquemática exemplificando os vários tipos de aberrações cromossômicas que ocorrem no genoma de tumores, e resumo da capacidade de detecção de algumas técnicas comumente aplicadas: (a) aberrações cromossômicas estruturais que levam a desequilíbrios genômicos de ganhos ou perdas (como translocações não-recíprocas, deleções e ampliações, e aberrações de cromossomos inteiros, como aneuploidias e poliploidias); (b) alterações qualitativas que não geram ganho ou perda de material genômico (como perda de heterozigose). **Fonte:** LOBO (2008).

A análise global do genoma de diversos tipos tumorais utilizando *array-CGH* de alta resolução ou *SNP-arrays* revelou que variações no número de cópias de segmentos de DNA submicroscópicos (definidas como duplicações e deleções entre 1 Kb-4 Mb), além das alterações de segmentos citogeneticamente visíveis ao microscópio, também são eventos muito frequentes no câncer (WEIR et al. 2007; MULLIGHAN e DOWNING 2009). Em um extenso trabalho publicado por BEROUKHIM et al. (2010) analisaram mais de 3.000 tumores, pertencentes a 26 tipos histológicos. Grande parte das alterações somáticas de mudança no número de cópias de segmentos genômicos (chamadas pelos autores de *Somatic Copy Number Alteration*) era comum a diversos tipos tumorais. Ao longo do genoma, as alterações mais prevalentes eram de dois tipos: alterações de segmentos curtos de DNA (média 1,8 Mb, denominada de alteração focal), e alterações de quase exatamente o comprimento de um braço cromossômico/cromossomo inteiro. Em média, cada amostra de câncer apresentava 25% do genoma afetado por SCNA de braço cromossômico, ao passo que 10% exibiam rearranjos focais, com pouca sobreposição (2%) entre os grupos.

### **1.3. ALTERAÇÕES CROMOSSÔMICAS EM CÂNCER DE MAMA**

Alterações genômicas em número de cópias (ganhos e perdas de segmentos cromossômicos ou *somatic copy number alterations* - SCNA) são reconhecidamente importantes no câncer, principalmente porque sua

identificação pode ajudar no mapeamento de regiões contendo oncogenes e supressores tumorais. Tais alterações genômicas são mediadoras de alterações fenotípicas por meio do impacto na expressão gênica.

Tumores sólidos como os carcinomas de mama são sabidamente heterogêneos do ponto de vista da composição genômica de suas células, e mesmo alterações cromossômicas funcionalmente importantes para a carcinogênese podem estar presentes em apenas um subgrupo celular dentro do mesmo tumor. Assim, tumores sólidos frequentemente mostram multiclonalidade, sugerindo a existência de alto grau de heterogeneidade intratumoral, o que dificulta o estudo do genoma tumoral.

Em carcinomas de mama, estudos citogenéticos clássicos evidenciaram que alterações dos braços cromossômicos 1q, 3p, 6q e 8p estão frequentemente presentes; isocromossomos i(1q) e derivados de translocação t(1;16) são também comuns, além de ganhos dos cromossomos 7, 8 e 20 (BIRNBAUM et al. 1999; CARCANGIU et al. 2005). Sinais citogenéticos de amplificação de DNA, como as regiões homogeneamente coradas (HSR), são comumente detectados em carcinomas de mama e são preferencialmente associadas à amplificação de segmentos de DNA do cromossomo 8. No entanto, é importante salientar que muitas das alterações cromossômicas comumente observadas nos carcinomas de mama não são exclusivas deste tipo tumoral; como exemplo, rearranjos no cromossomo 1, principalmente ganhos de 1q, também são muito frequentes em tumores em geral, além de tumores de mama (COURJAL e THEILLET 1997; BIRNBAUM et al. 1999; CARCANGIU et al. 2005).

Estudos de perda de heterozigose (LOH) evidenciaram, entre outros mecanismos, a deleção de muitos segmentos genômicos contendo potenciais genes supressores tumorais. Muitas regiões afetadas por LOH foram mapeadas em câncer de mama, como 1p, 1q, 3p, 6q, 8p, 11q, 13q, 16q, 17p, 17q. Algumas dessas regiões exibiam perdas em mais de 50% dos tumores, como 8p, 16q e 17p (BIRNBAUM et al. 1999; CARCANGIU et al. 2005). Em geral, trabalhos mais recentes em tumores de mama mostram que as alterações frequentemente mapeadas afetam principalmente os cromossomos 8, 11q, 13q, 16p e 16q, 17p e 17q e 20q (DOWNING et al. 2010).

Mesmo considerando que múltiplos rearranjos somáticos são achados extremamente frequentes no genoma de tumores de mama, o padrão estrutural de tais rearranjos é muito complexo e sua contribuição para o desenvolvimento do câncer ainda pouco caracterizada. STEPHENS et al. (2009) sequenciaram 24 tumores de mama para identificar e caracterizar rearranjos somáticos; notavelmente, os rearranjos detectados, principalmente intra-cromossômicos, foram mais comumente duplicações *in tandem*, com grande parte dessas alterações sendo <2 Mb (i.e., indetectável pela citogenética tradicional), confirmando que o mecanismo de ganho e/ou amplificação de segmentos genômicos é um processo comum na etiologia do câncer de mama. Neste estudo de 2009, novos genes de fusão e genes internamente rearranjados também foram identificados, mas nenhum deles recorrente, achado interpretado pelos autores como indicativo de que eram eventos passageiros ou raros no desenvolvimento do câncer de mama. Como os tumores apresentaram diferenças marcantes dependendo do

número de duplicações *in tandem* que apresentavam, para tumores com alto número de duplicações os autores postulam a existência de um processo defeituoso de manutenção do DNA.

As proteínas HER2 e EGFR pertencem à família de receptores de fatores de crescimento epidérmicos, e seus genes codificadores são alvo-comum de amplificação no câncer de mama. O aumento no número de cópias destes genes pode ocorrer por amplificação gênica ou por polissomia (cópias cromossômicas adicionais). O subtipo de carcinoma mamário HER2 positivo (HER2+), que corresponde a cerca de 30% de todos os carcinomas de mama, resulta essencialmente de amplificação gênica dentro do *amplicon* contendo o gene *ERBB2* (SØRLIE et al. 2001). Por outro lado, a super-expressão de *EGFR* no carcinoma de mama basal raramente decorre de amplificação gênica, mas sim de polissomia do cromossomo 7 ou de indução transcricional pelo fator de transcrição *YBX1* (STRATFORD et al. 2007; GILBERT et al. 2008).

Em 2011, STAAF et al. utilizaram as técnicas de *array-CGH* e *SNP-arrays* para estudar 218 tumores HER2. Foram identificadas 34 regiões genômicas exibindo ganhos, sendo as de maior amplitude mapeadas em 1q, 8q e 17q, além de 46 regiões associadas a perdas, principalmente em 8p e 17p. Adicionalmente, este estudo realizou uma integração dos dados de *copy number alterations* dos segmentos amplificados em 17q com dados de expressão gênica destes tumores, revelando que alguns genes tinham um padrão concordante de amplificação (ganho de cópias da sequência) e aumento de expressão. Já os eventos de perdas de heterozigose (*loss of heterozigosity – LOH*) foram mais frequentes em 8p, 17p, 9p, 18q e 13q



neste trabalho. Outro achado interessante foi que a fração do genoma afetada por LOH foi menor nos tumores HER2 que eram positivos para *ER* quando comparados aos *ER*-negativos, mostrando que tumores *ER*-positivos apresentam menor instabilidade genômica.

Carcinomas de mama com poucos rearranjos ou rearranjos focais localizados em *amplicons* (outros além de *ERBB2*) são usualmente receptores de estrógeno/progesterona-positivos (BERGAMASCHI et al. 2006). Este grupo de tumores estrógeno/progesterona-positivos se caracteriza por ganhos frequentes em 16p (JÖNSSON et al. 2010; KWEI et al. 2010). No carcinoma de mama, alguns estudos mostram que a perda de 16q ocorre nestes tumores positivos quanto à expressão do receptor de estrógeno (ER+) e está associada a um melhor prognóstico, sendo detectável comumente no carcinoma ductal *in situ*, indicando que esse evento (perda de 16q) ocorreria no início da tumorigênese (LOO et al. 2004; DOWNING et al. 2010). Além disso, carcinomas ductais de baixo grau exibem ganho mais frequente de 16p e perda de 16q, enquanto que em tumores de alto grau o ganho de 16p e perda de 16q seriam eventos menos frequentes.

Tumores de mama com muitas duplicações genômicas *in tandem* de segmentos genômicos diferentes geralmente são receptores estrógeno/progesterona-negativos e classificados pelo perfil de expressão como *basal-like*. O subtipo Triplo-negativo (estrógeno/progesterona-negativos e HER2-negativo) demonstra grande instabilidade genômica, com inúmeras regiões diferentes do genoma afetadas por SCNAs, mas principalmente por um segmento de perda mais frequente em 4q (KWEI et

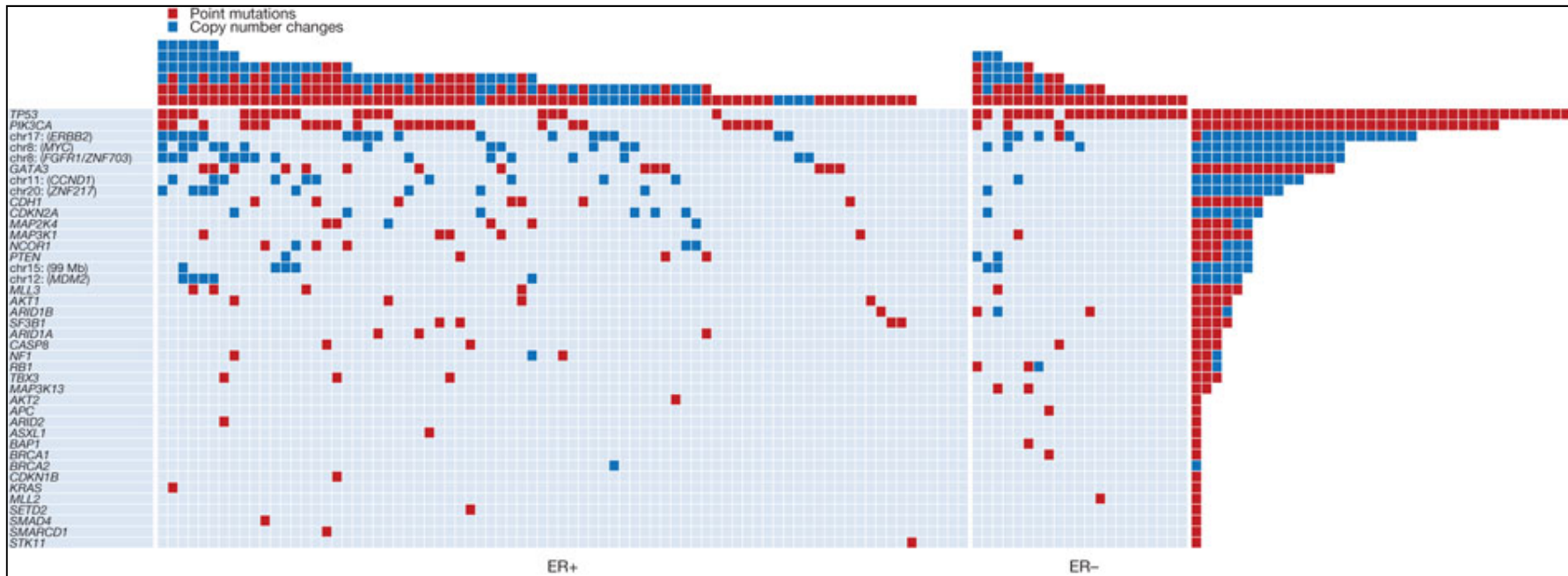
al. 2010). Eventos como perdas em 4q e 5q e ganhos em 10p e 17q12 também já foram identificados por outro trabalho utilizando a técnica de *array-CGH* em tumores ER(-) (MELCHOR et al. 2007). Ainda nesse grupo de tumores com receptores hormonais negativos ocorre um grande número de *SCNAs*, principalmente duplicações, apresentando grande quantidade de alterações cromossômicas em relação a outros carcinomas de mama (HAN et al. 2008; HU et al. 2009; STEPHENS et al. 2009).

Todos os tumores portam mutações somáticas em seus genomas, mas apenas um *subset* de tais mutações, conhecidas como *driver mutations*, conferem vantagem seletiva clonal para a célula tumoral e estão implicadas de maneira causal na oncogênese (STRATTON et al. 2009). Em teoria, as regiões com ganhos ou perdas de maior amplitude (portanto, presentes no genoma de um número maior de células daquele tumor e, algumas vezes, apresentando grande diferença no número de cópias em relação ao genoma) contêm os genes mais relevantes para a progressão tumoral, sendo mutações *drivers*. As principais *driver mutations* e processos mutacionais operando no câncer de mama ainda não são totalmente conhecidos.

Em 2012, um extenso trabalho de CURTIS et al. realizou uma análise integrada de alteração no número de cópias e expressão gênica de > 990 tumores de mama, mostrando que 40% das variantes genômicas estudadas (germinativas e aberrações somáticas) influenciam o complexo padrão de expressão gênica em tumores de mama. Neste trabalho os autores também identificaram novas *driver mutations* em genes como *MTAP*, *PPP2R2A* e *MAP2K4*.

Em outro trabalho recente, STEPHENS et al.(2012) examinaram o genoma de 100 tumores de mama investigando SCNAs e mutações em exons de genes codificadores de proteína. O número de mutações detectado foi extremamente variável entre tumores individuais, com forte correlação entre este número e a idade ao diagnóstico e grau histológico. Neste grupo de 100 tumores, substituições e pequenas inserções e deleções foram identificadas em genes previamente implicados no desenvolvimento do câncer de mama, incluindo *AKT1*, *BRCA1*, *CDH1*, *GATA3*, *PIK3CA*, *PTEN*, *RB1* e *TP53*. Prováveis *drivers mutations* também foram identificadas em genes já relacionados a outros tipos de câncer, incluindo *APC*, *ARID1A*, *ARID2*, *ASXL1*, *BAP1*, *KRAS*, *MAP2K4*, *MLL2*, *MLL3*, *NF1*, *SETD2*, *SF3B1*, *SMAD4* e *STK11*. Neste mesmo trabalho foram descritos novos potenciais genes recessivos de câncer de mama (*ARID1B*, *CASP8*, *MAP3K1*, *MAP3K13*, *NCOR1*, *SMARCD1* e *CDKN1B*), pelo padrão apresentado de mutações truncadas e frequentes inativações bialélicas.

A figura abaixo, extraída deste trabalho de STEPHENS et al. (2012) citado acima, mostra a distribuição das mutações de ponto e das SCNAs detectadas em 40 genes de câncer no grupo de 100 tumores de mama estudados; tais alterações foram consideradas pelos autores como *driver mutations*.



**Figura 2** - Distribuição das mutações de ponto e SCNA detectadas em 40 genes de câncer no grupo de 100 tumores de mama do trabalho de STEPHENS et al. 2012. Neste painel de mutações “drivers” do câncer de mama, cada um dos 40 genes nos quais uma mutação ou mudança no número de cópias foi identificada está listado no lado esquerdo da figura. O número de mutações em cada gene nos 100 tumores é mostrado (linhas), assim como o número de *driver mutations* em cada tumor de mama (colunas). Mutações pontuais e mudanças no número de cópias estão coloridas em vermelho e azul, respectivamente.

Dentre os ganhos rotineiramente detectados em tumores de mama estão as ampliações em 8q24 (contendo o gene *MYC*) e de 17q12 (contendo o gene *ERBB2*), esta última característica do subtipo de carcinoma de mama HER2. Ganhos em 8q são extremamente comuns em diversos tipos tumorais, em especial a amplificação em 8q24 abrangendo o oncogene *MYC* (YU et al. 2009).

Outro segmento de amplificação recorrente está mapeado em 8p12 (10-15% dos tumores de mama, contendo o gene *FGFR1*). A amplificação em 8p, principalmente na região 8p11-12, ocorre de maneira recorrente em 10 a 15% dos carcinomas de mama e parte destes casos são *amplicons* não abrangendo o gene *FGFR1*, mas contendo outros potenciais oncogenes como *PPAPDC1B* (BERNARD-PIERROT et al. 2008). O gene *PPAPDC1B* foi recentemente descrito como um novo alvo terapêutico; este gene regularia a sobrevivência e a transformação de linhagens celulares que portavam o *amplicon* 8p, podendo também potencializar a via do receptor de estrógeno regulando negativamente o gene *DUSP22*. Outros trabalhos descrevem diferentes potenciais oncogenes nessa mesma região, como *DDHD2*, *ERLIN2* e *WHSC1L1*, esse último gene mostrando-se importante para a sobrevivência e o crescimento celular (YANG et al. 2010).

A citobanda 11q13 também contém um segmento amplificado em 15-20% dos tumores de mama (contendo o gene *CCND1*), um evento não apenas já associado com expressão do receptor de estrógeno, mas também relacionado à resistência à terapia endócrina (KARLSSON et al. 2011). Co-amplificações em 8p12 e 11q13 são eventos comuns no genoma de tumores

de mama, e a associação com dados clínicos sugere que haja sinergia entre estes *amplicons*, colaborando para um pior resultado ao tratamento do que a ocorrência isolada de uma dessas amplificações. Tumores que possuem essa co-amplificação parecem ser raros entre tumores que expressam o gene *ERBB2* e teoricamente fazem parte de um subgrupo denominado “*Amplifier subtype*”, caracterizado por não expressar o gene *ERBB2* (JÖNSSON et al. 2010; KARLSSON et al. 2011).

Os carcinomas de mama do tipo ductal invasivo, atualmente chamados de *Non Special Type* (NST), correspondem a 85% dos tumores invasivos da mama, com altas taxas de incidência e mortalidade, e associação com um pior prognóstico. Portanto, a compreensão dos mecanismos moleculares subjacentes a esta tumorigênese é de fundamental importância. Considerando que um dos fatores prognósticos de maior impacto no câncer de mama é a presença ou não de metástase linfonodal axilar ao diagnóstico, direcionamos nosso estudo para a identificação de ganhos e perdas de segmentos genômicos no genoma de carcinomas primários que ocorram em associação à presença de metástase linfonodal. A investigação do genoma tumoral por *array-CGH* de alta resolução permite identificar SCNAs relevantes para a tumorigênese e também delimitar rearranjos focais afetando segmentos menores de DNA, contendo poucos genes candidatos a *drivers*. A identificação de *amplicons* e regiões afetadas por perdas em homozigose permite revelar potenciais oncogenes e genes supressores tumorais relevantes para a agressividade de carcinomas de mama.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

O principal objetivo deste projeto foi identificar genes e regiões cromossômicas relevantes para a tumorigênese de carcinomas de mama do tipo ductal invasivo, com ênfase na detecção de alterações genômicas específicas presentes no genoma dos tumores primários e associadas à ocorrência de metástase linfonodal.

### 2.2 OBJETIVOS ESPECÍFICOS

- a. Mapear regiões do genoma afetadas por alterações somáticas no número de cópias de segmentos de DNA (*somatic copy number alterations* – *SCNA*) em um grupo de carcinomas de mama do tipo ductal invasivo (*Non Special Type*), por meio da técnica de *array-CGH* em plataforma de *oligoarray* 60K;
- b. Delineamento do perfil genômico característico de carcinomas de mama de acordo com a expressão proteica do receptor de estrógeno;
- c. Estabelecer correlação entre o perfil genômico de *SCNA* dos carcinomas de mama com o *status* de linfonodos axilares ao

diagnóstico (presença ou ausência de metástase linfonodal axilar);

- d. Identificação de genes e/ou regiões afetados por alterações genômicas de grande amplitude (amplificação ou deleção em homozigose), para identificação de potenciais oncogenes e supressores tumorais.



### **3 MATERIAL E MÉTODOS**

#### **3.1 MATERIAL**

As amostras de carcinomas de mama estudadas neste projeto foram provenientes de pacientes atendidas no Hospital do Câncer A.C. Camargo (HCACC). Este foi um estudo retrospectivo, realizado com material extraído de tecido congelado a fresco, e as amostras encontravam-se armazenadas no Banco de Tumores do HCACC.

Os critérios de inclusão dos tumores estão listados a seguir:

- a. Carcinomas de mama do tipo ductal invasivo, tumor primário;
- b. Análise de imunistoquímica informativa quanto ao padrão de expressão proteica dos receptores hormonais de estrógeno e progesterona, e de HER2;
- c. Informação sobre acometimento linfonodal axilar ao diagnóstico.

Os critérios de exclusão utilizados encontram-se a seguir:

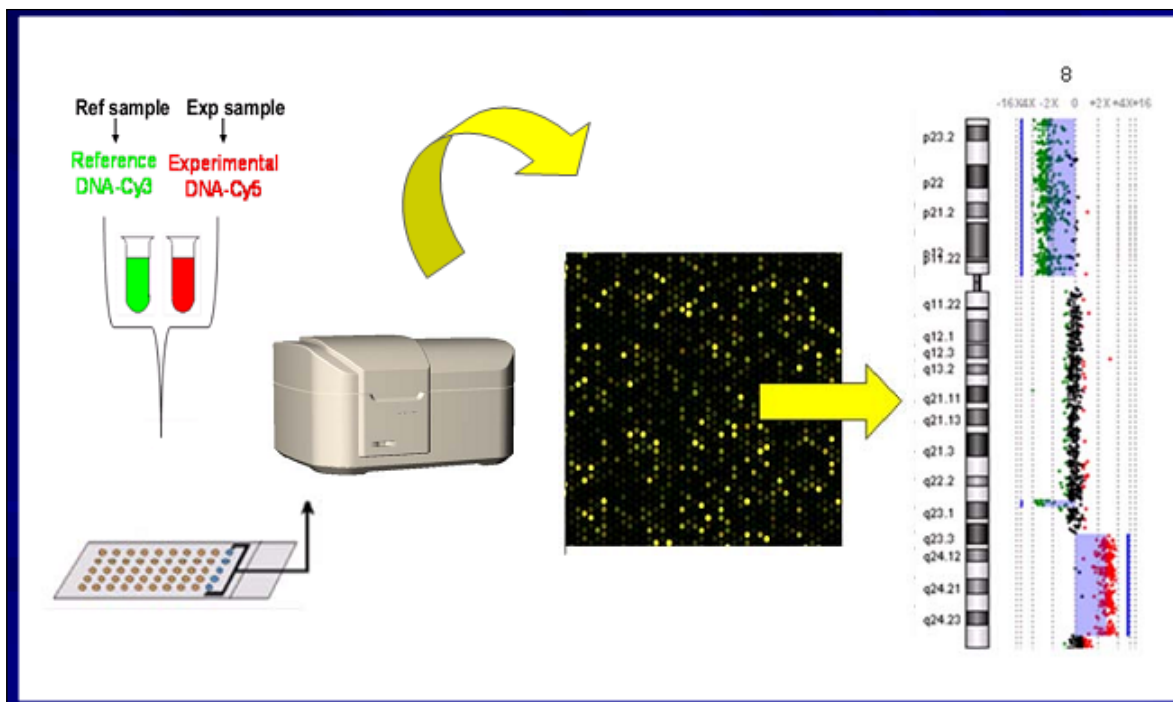
Utilizando o banco de dados do HCACC, foram selecionados 68 carcinomas de mama do tipo ductal invasivo. Também foram estudados 3 carcinomas classificados como subtipos neuroendócrino, medular e metaplásico. Alíquotas de DNA foram solicitadas ao Biobanco do HCACC; todo o procedimento de extração de biomoléculas, bem como controle de qualidade, foi realizado pelo laboratório do próprio Biobanco.

Os carcinomas de mama selecionados haviam sido previamente estudados por imunohistoquímica quanto ao padrão de expressão proteica de receptores hormonais (estrógeno e progesterona) e do gene *ERBB2* (HER2), dentro do procedimento de rotina do Departamento de Anatomia Patológica do HCACC. Parte desses carcinomas de mama também havia sido previamente avaliada por hibridação *in situ* fluorescente (FISH) quanto à amplificação do gene *ERBB2*, visando auxiliar a caracterização do subtipo tumoral. Os dados clínicos e de análise imunohistoquímica detalhados dessas amostras encontram-se descritos na seção Resultados.

## **3.2 MÉTODOS**

### **3.2.1 Hibridação genômica comparativa em microarranjos de oligonucleotídeos (*array-CGH*)**

A investigação dos tumores de mama para identificação de alterações somáticas de ganhos e perdas de segmentos de DNA foi realizada utilizando a técnica de hibridação genômica comparativa em microarranjos de oligonucleotídeos (*array-CGH*). Esta técnica permite a detecção de variações no número de cópias de segmentos de DNA da amostra a ser testada em relação a uma amostra referência (Figura 3).



Fonte: Adaptado Agilent Technologies (2012).

**Figura 3** - Técnica de array-CGH.

A plataforma utilizada neste estudo foi a *SurePrint G3 Human 8x60K* (Agilent Technologies), que contém cerca de 60.000 sequências de oligonucleotídeos (sondas), anotadas no banco de dados genômicos *UCSC Genome Browser* (<http://genome.ucsc.edu/>). A média de espaçamento entre estes oligonucleotídeos na sequência do genoma humano é de aproximadamente 50 Kb, mas a distribuição real dos mesmos é variável, com maior densidade em regiões codificadoras.

Utilizamos como referência um *pool* de DNA humano de alta qualidade disponível comercialmente (Promega), composto por amostras de 10 mulheres. Essa escolha foi feita porque as amostras de DNA dos tumores de mama obtidas no Biobanco do Hospital apresentaram excelente qualidade, com alto peso molecular e nenhuma degradação. A escolha de

DNAs qualitativamente compatíveis é fundamental na técnica de *array-CGH* para minimizar o ruído experimental.

- **Procedimento experimental**

Amostras de DNA teste (tumor) e DNA referência (Promega) são marcadas diferencialmente por *random priming* usando dinucleotídeos CTP com fluorocromos incorporados (Cy3- ou Cy5-dCTP), etapa realizada utilizando o *kit* de marcação fluorescente da empresa Oxford Gene Technologies (OGT).

A quantidade de DNA utilizada varia conforme o *design* do microarranjo em uso (número de sondas) e o procedimento de marcação (com ou sem amplificação do DNA). No caso da plataforma deste estudo (60K), foram utilizados 400 ng de DNA tanto para a amostra teste quanto para a amostra referência.

O procedimento de marcação consiste das seguintes etapas:

- as amostras de DNA foram fragmentadas a 95°C durante 15 minutos, para diminuir a complexidade das moléculas e aumentar a eficiência de marcação.

- após a fragmentação, foi adicionado às amostras o reagente contendo *random primers* (iniciadores de amplificação com sequência randômica), para subsequente desnaturação da dupla fita de DNA a 95°C por 3 minutos.

- as amostras foram mantidas no gelo por 5 minutos, com subsequente adição da enzima *Klenow* e de Cy3-dCTP em amostras teste e

Cy5-dCTP nas amostras referência.

- a marcação do DNA ocorreu durante 2 (duas) horas a 37°C, período no qual a polimerase *Klenow* inicia a extensão de novas fitas de DNA a partir dos *random primers* hibridados à fita de DNA, com incorporação à nova fita em síntese dos dinucleotídeos marcados com os fluorocromos Cy3 e Cy5.

- o processo foi interrompido pela inativação a 65°C por 10 minutos.

- purificação das amostras marcadas utilizando colunas *Illustra ProbeQuant G-50 Micro columns* (GE Healthcare), seguindo o protocolo do fabricante.

- avaliação das marcações no aparelho *ND-1000 Spectrophotometer* (NanoDrop) para verificar o nível de incorporação de fluorocromos e também a quantidade de DNA, utilizando o cálculo a seguir:

(pmol per  $\mu\text{L}$  dye)

---

( $\mu\text{g}$  per  $\mu\text{L}$  de DNA genômico)

Amostras que exibiram níveis de marcação e quantidade de DNA dentro dos níveis sugeridos pelo fabricante foram utilizadas para prosseguir o experimento (atividade específica para Cy3 entre 25 a 40; para Cy5, entre 20 a 35).

A hibridação das amostras marcadas aos microarranjos genômicos ocorreu seguindo as etapas a seguir:

- foi adicionado a cada amostra a ser testada (tumor) uma amostra

referência e 5 ug de *Human Cot-1 DNA* (Invitrogen; *pool* de DNA humano de placenta enriquecido em sequências repetitivas).

- precipitação das amostras a serem hibridadas, com posterior ressuspensão em 18 µl de TE.

- adição de 22,5 µl do tampão de hibridação *2x HI-RPM Hybridization Buffer* e 4,5 µl do *10x aCGH Blocking Agent* (ambos da Agilent Technologies).

- desnaturação das amostras a 95°C por 3 minutos e posterior incubação a 37°C por 30 minutos – período chamado de *pre-annealing*, no qual ocorre a supressão de sequências repetitivas do DNA marcado pela hibridação do DNA *Cot-1*.

- aplicação de cada amostra sobre um dos microarranjos da lâmina.

- hibridação das amostras às sondas dos microarranjos em forno de hibridização rotatório durante no mínimo 16 horas a 65°C, 20rpm.

- lavagem dos microarranjos para eliminação de moléculas de DNA não hibridizadas, utilizando os tampões *Agilent Oligo aCGH/Chip-on-Chip Wash Buffer* 1 (5 minutos a temperatura ambiente) e 2 (1 minuto a 37°C), acetonitrila por 1 minuto *CHROMASOLV* (Sigma) e *Stabilization and Drying Solution* por 30 segundos (Agilent Technologies), seguindo instruções do fabricante.

- **Extração de dados dos microarranjos**

As lâminas foram escaneadas usando um equipamento provido de 2 *lasers*, a *Agilent DNA microarray scanner* (com um sistema de 48 slides),

existente no Centro Internacional de Pesquisa e Ensino do HCACC. Este *scanner* tem capacidade de obter imagens de alta resolução (3 *microns*) dos *microarrays* de grande densidade, como necessário para a plataforma utilizada 8x60K.

Os dados de fluorescência das imagens dos microarranjos foram extraídos com o software *Feature Extraction 10.7.3.1* (Agilent), utilizando o arquivo *021924\_D\_F\_20100501* (identificação do *design* do microarranjo 60K utilizado) e protocolo de extração *CGH\_107\_Sep09*. Todos os experimentos foram avaliados com relação a parâmetros de qualidade como: intensidade de fluorescência total (Cy3 – canal verde; Cy5 – canal vermelho); intensidade de fluorescência do *background*; relação entre fluorescência das sondas e fluorescência do *background* (*signal to noise*); índice de avaliação da dispersão do valor obtido da razão  $\log_2 \text{Cy3/Cy5}$  para cada sonda em relação a todos os valores obtidos no experimento (*DLRS - derivative log ratio spread*). Todos os experimentos utilizados neste projeto atingiram os índices de qualidade mínimos indicados pelo fabricante. Em especial, utilizamos apenas experimentos que obtiveram  $DLRS < 0,3$ , que é o limiar sugerido para experimentos com qualidade boa o suficiente para gerar dados confiáveis. Os padrões de qualidade podem indicar possíveis erros de processamento da técnica ou mostrar que os dados daquele *microarray* em particular podem estar comprometidos.

Abaixo seguem os valores indicados pela empresa como ideais para alguns parâmetros de qualidade.

**Quadro 1** - Parâmetros de qualidade do experimento de *microarray*.

<b>Medida</b>	<b>Excelente</b>	<b>Bom</b>	<b>Ruim</b>
<b>Background Noise</b>	<5	5 a 10	>10
<b>Signal intensity</b>	>150	50 a 150	<50
<b>Signal to Noise</b>	>100	30 a 100	<30
<b>Reproducibility</b>	<0.05	0.05 a 0.2	>0.2
<b>DLRSpread</b>	<0.2	0.2 a 0.3	>0.3

### 3.3.2 Análise dos dados dos microarranjos

A análise dos dados derivados de *array-CGH* foi realizada utilizando o software *Nexus Copy Number 6.0* (BioDiscovery). Este *software* também avalia a qualidade experimental (*quality call report* – QC), considerando o desvio dos valores de  $\log_2$  de cada sonda em relação ao valor médio de  $\log_2$  de todas as sondas do experimento; assume-se que um desvio padrão pequeno indica menos artefatos nos sinais biológicos obtidos. Todos os experimentos previamente selecionados como de qualidade adequada também obtiveram  $QC < 0,2$  no software Nexus, que é o limiar sugerido.

As regiões genômicas de ganhos e perdas foram identificadas para cada amostra utilizando o algoritmo *Fast Adaptive States Segmentation Technique* (FASST2), com limiar de significância de  $1 \times 10^{-8}$ . O algoritmo FASST2 baseia-se na abordagem de *Hidden Markov Model* para estimar no genoma um grande número de níveis possíveis de segmentos situados em diferentes *status* de número de cópias. Um processamento subsequente é realizado para identificar dentre tais diferentes níveis de segmentos genômicos detectados aqueles que se situam acima de limiares previamente estabelecidos para alterações de ganhos e perdas, gerando assim os



eventos de alteração de número de cópias de segmentos genômicos (que podem ser denominados de *aberration calls* ou *SCNAs*).

Para minimizar a taxa de falso-positivos, foram consideradas como *SCNA* aqueles segmentos representados por no mínimo cinco sondas consecutivas com valor médio de  $\log_2$  Cy3/Cy5 (amostra tumoral/amostra teste) alterado acima dos limiares previamente estabelecidos: razão  $\log_2 >0,3$  e  $>1,4$  para ganhos genômicos e ganhos em alto número de cópias (aqui chamados de amplicons), respectivamente; razão  $\log_2 <-0,3$  e  $<-1,1$  para perdas genômicas e deleções em homozigose, respectivamente.

As alterações em número de cópias documentadas no banco de dados *Database of Genomic Variants* (DGV, <http://projects.tcag.ca/variation/>) como variantes constitucionais comuns na população em geral (*copy number variations – CNV*) foram excluídas da análise.

- **Análise estatística**

A distribuição do número total de *SCNA* e também de seus diferentes tipos (perdas, ganhos, deleções em homozigose e ganhos em alto número de cópias) foi avaliada no grupo de tumores e também comparada entre diferentes subgrupos (receptor de estrógeno positivo e receptor de estrógeno negativo; linfonodo-positivo e linfonodo-negativo). O teste utilizado nesta comparação foi o *test t* de Student, utilizando o software GraphPad PRISMA 5.

Foram também aplicadas diversas ferramentas do *software* Nexus para análises estatísticas das *SCNA* identificadas:

- (a) *Significance Testing for Aberrant Copy Number* (STAC) – método para testar a significância das alterações identificadas em um conjunto de múltiplas amostras do mesmo tipo biológico. Esse algoritmo busca identificar um grupo de SCNA que ocorre de maneira recorrente e não randomicamente.
- (b) *Genomic Identification of Significant Targets in Cancer* (GISTIC) – algoritmo usado para identificar regiões afetadas por SCNA em alta frequência no grupo testado em relação ao *background* geral de alterações. Após a identificação das alterações de ganhos ou perdas genômicos em cada amostra, um *G score* é calculado para cada SCNA; o *G score* é uma medida da frequência de ocorrência da SCNA e também da magnitude da mudança no número de cópias (*log ratio intensity*) no conjunto de dados. Cada região é marcada separadamente como ganho ou perda, e regiões alteradas em cada amostra são permutadas, simulando dados com alterações aleatórias. Essa distribuição aleatória (esperada) é comparada estatisticamente à distribuição observada para identificar SCNAs que provavelmente ocorrem não ao acaso (*score* ultrapassando o limiar de significância, aqui definido como *G-score*).

### **3.2.3 Validação de ganhos e perdas de genes por qPCR**

A técnica de *real-time PCR* quantitativo (qPCR) foi utilizada para validar a alteração no número de cópias de alguns genes selecionados a partir da análise dos dados obtidos no *array-CGH*. Foram testadas as

mesmas amostras nas quais a SCNA foi identificada para estimar o número de cópias de uma determinada sequência de DNA afetada pela alteração.

O sistema de PCR em tempo real é baseado na detecção e quantificação de uma molécula repórter fluorescente. O sinal fluorescente emitido aumenta proporcionalmente à quantidade de produto de PCR na reação. Quanto maior o número de cópias iniciais do DNA-alvo, mais cedo é observado um aumento significativo na fluorescência. Um limiar de fluorescência fixo é definido de maneira significativa acima do *baseline* que pode ser alterado. O parâmetro  $C_T$  (limiar do ciclo) é definido como o número do ciclo pelo qual a emissão de fluorescência excede o limiar fixado. Quando o valor obtido é correlacionado ao valor da fluorescência do gene controle, é possível avaliar se os genes testados exibem maior ou menor número de cópias. Os genes controles utilizados neste estudo foram *GAPDH* (12p13) e *P2RX7* (12q24). Estes genes foram escolhidos como controle por estarem mapeados em regiões do genoma para as quais não foram detectadas alterações no número de cópias nos tumores testados.

O sistema utilizado foi *SYBR Green* da Applied Biosystems. *SYBR Green* é um corante fluorogênico de ligação (*fluorogenic minor groove binding dye*) que emite fluorescência quando se liga ao DNA dupla fita. No procedimento, isso ocorre quando há extensão da fita complementar durante o ciclo do PCR. Utilizamos o método *SYBR® Green* baseado em  $2^{-\Delta\Delta C_t}$  (LIVAK e SCHMITTGEN 2001). No método  $2^{-\Delta\Delta C_t}$ , 20ng de cada amostra de DNA foi amplificada em 20µL da reação contendo 0,3µmol de cada primer e 1X *SYBR Green*. O equipamento utilizado para a reação foi *7500 Fast Real-*

*Time PCR System* (Applied Biosystems). É feita uma pré-incubação a 95°C por 10 minutos, seguido de 40 ciclos de desnaturação a 95°C por 15 segundos e o anelamento e extensão a 60°C por 1 minuto. As amostras foram feitas em duplicata para os genes testados e controle. A média de  $C_T$  da duplicada foi usada no cálculo do número da dosagem do gene. O número de dosagem dos genes foi calculado utilizando a seguinte equação:  $2^{-[\Delta Ct \text{ (target)} - \Delta Ct \text{ (norm)}]}$ , onde  $\Delta Ct$  do gene alvo é igual a diferença entre o valor de  $C_T$  por amostra, a média controle por gene testado e o  $\Delta Ct$  do controle é igual a diferença entre o valor de  $C_T$  por amostra testada e a média do controle pelo gene testado. Relações entre 0,82 e 1,22 foram consideradas normais (número de cópias igual do gene teste na amostra testada e na amostra controle); entre 0,41 e 0,61 há indicação de diminuição no número de cópias; acima de 1,4 indica aumento no número de cópias (com uma probabilidade de 10% de erro na concentração medida).

A seguir estão listados os genes que foram testados.

**Quadro 2** - Genes selecionados testados por *qPCR*: símbolo oficial, nome e mapeamento desses genes, e sequência dos pares de *primers* utilizados.

<b>Símbolo oficial</b>	<b>Nome do gene</b>	<b>Mapeamento</b>	<b>Sequência dos <i>primers</i></b>
<b><i>RERGL</i></b>	RERG/RAS-like	12p12.3	<i>Forward</i> CCCCACAAAGTTCCTTCA <i>Reverse</i> CTCATCTGCTCTGAAACT GG
<b><i>ADAMTS3</i></b>	ADAM metallopeptidase with thrombospondin type 1 motif, 3	4q21	<i>Forward</i> CGTAGAAAGCCTTTGGG <i>Reverse</i> GGTGCATGATGGAACG
<b><i>ZBTB20</i></b>	zinc finger and BTB domain containing 20	3q13.2	<i>Forward</i> CGCACTGCAAAGCAAC <i>Reverse</i> CTGAGCCTCTTCAAATTC G
<b><i>MTAP</i></b>	methylthioadenosine phosphorylase	9p21	<i>Forward</i> CACGATTCTGCATGTTG G <i>Reverse</i> CAAACAAGCTGAGGGCA
<b><i>ELF5</i></b>	E74-like factor 5 (ets domain transcription factor)	11p13-p12	<i>Forward</i> CAACCATTGCCAGGGA <i>Reverse</i> CAGCTTGGCCAAATGTC
<b><i>KISS1</i></b>	KiSS-1 metastasis-suppressor	1q32	<i>Forward</i> CCTCACCTAAAGCCTTTT GC <i>Reverse</i> CCAAGCGTGTCTTCTGT CC
<b><i>HSD17B12</i></b>	hydroxysteroid (17-beta) dehydrogenase 12	11q11	<i>Forward</i> CCCTTTAAGCCATTCCG <i>Reverse</i> GCCAATATTCAAACCGA GC

## 4 RESULTADOS

### 4.1 CARACTERIZAÇÃO DAS AMOSTRAS ESTUDADAS

Foi realizado um levantamento de dados clínicos das pacientes e também das características anatomopatológicas das amostras de carcinomas de mama estudadas. O detalhamento das informações obtidas encontra-se na Tabela 1.

Foram investigados 71 carcinomas invasivos, sendo 68 deles do tipo ductal invasivo (*Non Special Type* ou *NST*), um carcinoma medular, um carcinoma metaplásico e um carcinoma neuroendócrino. Dentre os 71 carcinomas de mama, 11 pacientes receberam quimioterapia pré-operatória (QT-PO).

Os tumores de mama estudados foram subdivididos em categorias de acordo com o padrão de expressão proteica do receptor de estrógeno ao diagnóstico:

#### 1. Padrão de expressão proteica do receptor de estrógeno:

- Expressão *negativa* de receptor de estrogênio [ER(-)]: 29 tumores

- Expressão *positiva* de receptor de estrogênio [ER(+)]: 42 tumores

Adicionalmente, após a exclusão dos 11 carcinomas de mama invasivos de pacientes submetidas a QT-PO e dos 3 carcinomas de subtipos especiais (medular, metaplásico e neuroendócrino), os 57 carcinomas de mama ductal invasivo (ou NST) foram classificados de acordo com as variáveis clínicas e anatomopatológicas (Tabela 2). No caso do *status* de linfonodo axilar ao diagnóstico, a classificação do grupo foi feita seguindo os critérios listados abaixo:

- N0 - nenhum linfonodo acometido
- N1 - 1 a 3 linfonodos acometidos
- N2 - 4 a 9 linfonodos acometidos
- N3 - 10 ou mais linfonodos acometidos

**2. Padrão de acometimento de linfonodo axilar ao diagnóstico:**

- Sem metástase linfonodal axilar – classe N0 [LINF(-)]: 20 tumores

- Metástase linfonodal axilar - classes N1/N2/N3 [LINF(+)]: 37 tumores

**Tabela 1** - Caracterização clínica e anatomopatológica da amostra estudada de 71 carcinomas de mama invasivos.

ID amostra	ER	PR	HER2 (IHQ)	ERBB2 (FISH)	TNM	Estadio clínico	Tamanho (cm)	Invasão Vascular	SBR	Observações	Linfonodo (n retirado/ n acometido)
<b>MIC56</b>	pos	pos	neg	não feito	T3N0M0	3b	2	neg	1	PÓS-QT, p53+	26/0
<b>MIC49</b>	pos	pos	3+	não feito	T2N2M0	3c	0,8 (5 pós-QT)	neg (pos pós-QT)	2	PÓS-QT, p53+	27/5
<b>MIC8</b>	pos	pos	neg	não feito	T4bN3M0	3c	8,5	pos	2	PÓS-QT	35/14
<b>MIC7 (2)</b>	pos	pos	neg	não feito	T4bN1M0	2a	3,5	pos	2	PÓS-QT	22/1
<b>MIC69 (2)</b>	pos	pos	neg	não feito	T4bN1M0	3b	2	neg	2	PÓS-QT	26/1
<b>MIC6</b>	pos	neg	neg	não feito	T4bN2M0	3c	nd	neg	1	PÓS-QT	15/4
<b>MIC35</b>	pos	neg	neg	não feito	T4bN2M0	3b	4,5	pos	3	PÓS-QT	11/6
<b>MIC25</b>	neg	neg	neg	não feito	T4bN2M0	2a	4,8	pos	3	PÓS-QT	19/8
<b>MIC18</b>	pos	pos	neg	não feito	cT4bN3M0	3a	6	pos	3	PÓS-QT	28/23
<b>MIC103</b>	neg	neg	3+	não feito	T3N3M0	3c	18	pos	3	PÓS-QT	30/17
<b>MIC 75</b>	neg	neg	2+	não feito	TxN1M0	2b	2	neg	3	PÓS-QT	6/2
<b>MIC 224</b>	neg	neg	neg	não feito	T2N0M0	2b	3	neg	3	p53+, ki-67(20%), EGFR+	8/0
<b>MIC68</b>	neg	neg	neg	não feito	pT3N1M0	3c	8	pos	3	p53+	18/2
<b>MIC43</b>	pos	pos	neg	não feito	T4bN3M0	2a	6	pos	3	p53+	22/17
<b>MIC41</b>	pos	neg	neg	não feito	T1cN1M0	2a	1,8	pos	3	p53+	24/2
<b>MIC37</b>	pos	pos	neg	não feito	T2N0M0	2b	3	neg	1	p53+	24/0
<b>MIC22</b>	pos	neg	2+	não feito	T1cN1M0	2a	1,8	neg	1	p53+	29/3



<b>MIC144</b>	neg	neg	neg	não feito	T2N3M0	2a	3,5	pos	3	p53+	46/34
<b>MIC141</b>	neg	neg	neg	não feito	pT2pN1	1	5	pos	3	p53+	26/3
<b>MIC 189</b>	neg	neg	neg	não feito	T2N2M0	2a	4,2	neg	3	p53+	43/4
<b>MIC73</b>	pos	pos	neg	não feito	pT2N2M0	2a	2,5	neg	2	Neuroendócrino	17/7
<b>MIC146</b>	neg	neg	neg	não feito	cT4bN3M0	2a	12	pos	3	Metaplásico/Coriocarcinoma , p53+	40/17
<b>MIC142</b>	neg	neg	neg	não feito	T2N0M0	2a	3,5	neg	3	Medular, p53+	22/0
<b>MIC 234</b>	neg	neg	neg	não feito	T2N0M0	3c	3,5	neg	3	ki-67(80%), p53+	36/0
<b>MIC 228</b>	neg	neg	neg	FISH Neg	T2N2M0	3b	3,5	neg	3	ki-67(60%), p53+, FISH com Polissomia do cr17	34/9
<b>MIC152</b>	neg	neg	3+	não feito	pT4bpN2M0	3c	2	pos	3	ki-67 (80%), EGFR+	24/6
<b>SM29</b>	neg	neg	neg	não feito	T2N0M0	3c	2	neg	3	ki-67 (70%), p53+	4/0 (sent)
<b>MIC 225</b>	pos	neg	3+	não feito	T1cN0M0	3a	2	neg	2	ki-67 (20%)	2/0
<b>MIC134</b>	neg	neg	neg	não feito	T2N1M0	4	3,8	neg	3		20/1
<b>MIC93</b>	pos	pos	2+	FISH Neg	T2N3M0	1	4	pos	2		31/23
<b>MIC31</b>	pos	neg	2+	FISH Neg	T2N0M0	3a	4,5	neg	2		21/0
<b>MIC96</b>	pos	pos	neg	não feito	pT1cpN1M0	3a	2	neg	2		28/1
<b>MIC88</b>	pos	pos	3+	não feito	T2N1M0	2a	1,5	neg	3		37/2
<b>MIC85</b>	pos	pos	neg	não feito	T1cN2M0	3c	2	pos	2		17/5
<b>MIC82</b>	neg	neg	2+	não feito	T4bN3M0	3b	7	pos	3		22/22
<b>MIC81</b>	pos	pos	3+	não feito	T1N2M0	2a	1,2	neg	1		25/8

<b>MIC66</b>	pos	neg	2+	não feito	T1CN0M0	1	2	neg	2	13/0
<b>MIC64</b>	pos	pos	3+	não feito	T2N0i+M0	2a	3	neg	2	10/0
<b>MIC63</b>	pos	pos	neg	não feito	T4bN1M0	3c	4	pos	2	15/3
<b>MIC60</b>	pos	pos	3+	não feito	T1CN0M0	3b	1,7	neg	2	5/0
<b>MIC59</b>	pos	pos	neg	não feito	T2N0M0	3b	3	neg	2	21/0
<b>MIC57</b>	pos	pos	neg	não feito	T1N1M0	3a	1,3	pos	2	14/1
<b>MIC55</b>	pos	neg	neg	não feito	pT1cpN0M0	2a	1,8	pos	2	17/0
<b>MIC48</b>	pos	pos	2+	não feito	T3N2M0	3a	6,5	pos	1	27/5
<b>MIC47</b>	neg	neg	3+	não feito	T4bN3M0	3b	16	pos	3	21/21
<b>MIC46</b>	neg	neg	3+	não feito	pT2pN3M0	2a	4,4	neg	3	31/17
<b>MIC44</b>	pos	pos	neg	não feito	T2N0M0	2a	6,5	neg	2	17/0
<b>MIC42</b>	pos	pos	neg	não feito	T1cN0M0	1	1,5	neg	2	23/0
<b>MIC40</b>	pos	pos	neg	não feito	T4bN1M0	2a	7	neg	3	16/2
<b>MIC38</b>	pos	pos	neg	não feito	T2N2M0	1	2,4	neg	1	19/5
<b>MIC34</b>	pos	pos	neg	não feito	T2N1M0	2a	3,2	neg	2	27/1
<b>MIC33</b>	pos	neg	neg	não feito	pT4cN1M0	3a	3,3	pos	1	34/1
<b>MIC21</b>	neg	neg	neg	não feito	pT1cN0M0	1	1,5	neg	3	20/0
<b>MIC20</b>	neg	neg	2+	não feito	T4bN2M0	2a	6	pos	3	38/8
<b>MIC149</b>	neg	neg	3+	não feito	T1cN2M0	1	1,2	pos	3	21/2
<b>MIC145</b>	neg	neg	2+	HER2/chr17=5	T2N0M0	3b	2,7	neg	3	32/0

<b>MIC143</b>	neg	neg	2+	HER2/chr17=3	pT1cN0M0	2b	1,6	neg	3	17/0
<b>MIC137</b>	neg	neg	neg	não feito	T1cN1M0	3c	1,3	neg	3	23/1
<b>MIC120</b>	pos	pos	neg	não feito	T2N3M0	2a	2,3	pos	3	19/17
<b>MIC10</b>	pos	neg	3+	não feito	T4bN1M0	1	11	pos	2	7/1
<b>MIC 99</b>	pos	neg	3+	não feito	T3N1M0	2b	6	pos	2	19/2
<b>MIC 9</b>	pos	neg	3+	não feito	T4N3M0	3b	2	pos	1	22/20
<b>MIC 61</b>	pos	neg	3+	não feito	T4bN1M0	2a	2,5	pos	1	4/1
<b>MIC 235</b>	pos	pos	2+	Her2/cr17 >5	pT2cpN2M0	3a	2	neg	3	20/6
<b>MIC 229</b>	neg	neg	2+	HER2/cr17=2,5	T1N1M0	3b	2	neg	3	15/1
<b>MIC 218</b>	neg	neg	neg	não feito	T2N0	3a	2,5	neg	3	28/0
<b>MIC 211</b>	pos	pos	3+	não feito	T2N0M0	3c	3	pos	3	2/0 (Sent)
<b>MIC 208</b>	neg	neg	3+	não feito	T4N1M1	3b	5,5	pos	3	16/3
<b>MIC 195 (2)</b>	pos	pos	3+	não feito	T1bN0M0	1	1	neg	1	8/0
<b>MIC 148</b>	neg	neg	3+	não feito	T1cN0M0	2a	1,5	neg	3	3/0 (sent.)
<b>MIC 147</b>	neg	neg	3+	não feito	T2N2M0	3a	2,3	pos	3	10/4

Abreviações: ER: receptor de estrógeno; PR: receptor de progesterona; HER2: expressão proteica do gene *ERBB2*; IHQ: dados de imunoistoquímica; Grau de SBR, sistema de graduação Scarff-Bloom Richardson modificado por Elston e Ellis; PÓS-QT: amostras obtidas pós-tratamento quimioterápico; Sent.: dados de linfonodo sentinela; cr17: cromossomo 17.

**Tabela 2** - Frequência de variáveis clínicas e anatomopatológicas no grupo de 57 carcinomas de mama do tipo ductal invasivo ou *Non Special Type* (NST).

<b>Variáveis</b>	<b>Categorias</b>	<b>N (Frequência)</b>
<b>ER</b>	Positivo	33 (57,9%)
	Negativo	24 (42,1%)
<b>Grau de SBR</b>	1	9 (15,8%)
	2	17 (29,8%)
	3	31 (54,4%)
<b>Estadiamento clínico</b>	I	9 (15,8%)
	II	21 (36,9%)
	III	26 (45,6%)
	IV	1 (1,7%)
<b>Invasão Vascular</b>	Positivo	25 (43,9%)
	Negativo	32 (56,1%)
<b>Linfonodo axilar</b>	N0	20 (35,1%)
	N1	19 (33,3%)
	N2	10 (17,5%)
	N3	8 (14%)

Abreviações: ER, receptor de estrógeno; Grau de SBR, sistema de graduação Scarff-Bloom Richardson.

## 4.2 MAPEAMENTO DE SCNAS RECORRENTES NO GRUPO DE CARCINOMAS DE MAMA ESPORÁDICOS DO TIPO DUCTAL INVASIVO

Utilizamos como controle nos experimentos de *array-CGH* um *pool* de DNA disponível comercialmente. Essa escolha foi feita porque as amostras de DNA dos tumores de mama obtidas no Biobanco do HCACC apresentaram excelente qualidade de integridade (alto peso molecular sem degradação) e pureza, conforme verificado respectivamente em eletroforese em gel de agarose e Nanodrop. A escolha de DNAs qualitativamente compatíveis é fundamental nos experimentos de hibridação comparativa para minimizar o ruído experimental.

De maneira geral, a análise de *array-CGH* evidenciou a complexidade e grande número dos eventos genômicos de ganhos e perdas associados ao carcinoma de mama. Duas categorias principais de alterações cromossômicas foram detectadas: (a) aneuploidia total parcial (ganhos e perdas) de cromossomos ou braços cromossômicos inteiros, e (b) rearranjos focais (segmentos afetados menores que um braço cromossômico, podendo ser submicroscópicos, ou seja, <4Mb); os dois tipos foram considerados como eventos de SCNAs neste estudo.

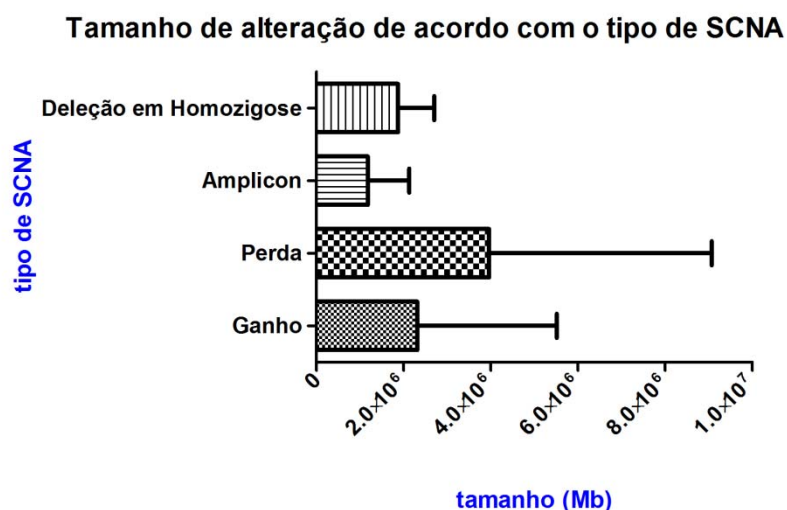
A análise de *array-CGH* do grupo de 71 tumores de mama identificou:

- número total de SCNAs: **3.828**
- média de SCNA por genoma tumoral: **53,2**
- tamanho médio de SCNA: **6,9 ±12 Mb** (mediana 2,8 Mb)

O tamanho da alteração genômica de acordo com os tipos de SCNA está descrito na Tabela 3 e pode ser visualizado no gráfico de barras abaixo (Figura 4).

**Tabela 3** - Tamanho da alteração de acordo com cada tipo de SCNA.

	<b>Tipo de SCNA</b>			
	<b>Ganho</b> (n=1980)	<b>Perda</b> (n=1664)	<b>Amplicon</b> (n=167)	<b>Deleção em Homozigose</b> (n=17)
<b>Tamanho de SCNA (Mb)</b>				
<b>Média (desvio-padrão)</b>	6,4 (12,7)	8,1 (11,8)	1,7 (1,7)	1,8 (1,3)
<b>Mediana</b>	2,3	3,9	1,2	1,9
<b>Percentil 25%</b>	1,1	1,9	0,6	0,4
<b>Percentil 75%</b>	5,5	9,1	2,1	2,7



**Figura 4** - Tamanho da alteração de acordo com o tipo de SCNA.

A distribuição do número total de eventos de ganhos foi comparada com a distribuição do número de perdas e não foi detectada diferença estatística ( $p=0,2189$ ; teste não-paramétrico Mann-Whitney). O tamanho das

SCNAs foi comparado utilizando o teste não-paramétrico Kruskal-Wallis (one-way ANOVA) e foi detectada diferença significativa entre os grupos ( $p < 0,0001$ ): os eventos de perda apresentaram tamanho significativamente maior ( $p < 0,0001$ ) do que os eventos de ganho (teste não-paramétrico Mann-Whitney). Além disso, os eventos de alta amplitude (amplicons e perdas em homozigose) afetaram segmentos genômicos significativamente menores ( $p < 0,0001$ ; teste não-paramétrico Mann-Whitney).

O software Nexus identifica os genes mapeados em cada SCNA identificada e também os relaciona com bancos de dados como *UCSC genes* e *Cancer Gene Census* do Sanger Center (este último contém genes com alterações cromossômicas e mutações já descritas em vários tipos de câncer). Nessa análise, é possível obter para cada SCNA quais são os genes e o número desses genes dentro daquela região. Considerando todas as SCNAs identificadas no grupo de tumores de mama estudado, foram detectados ~303.000 genes afetados - esses números indicam o número de vezes em que ocorreu a identificação em cada uma das SCNAs de um gene presente no banco de dados investigado (*UCSC genes*), e não 303.000 genes diferentes, havendo muita sobreposição pela ocorrência frequente das mesmas SCNAs em diferentes tumores do grupo. No entanto, cerca de 98.000 *hits* (aproximadamente 1/3 dos genes identificados) são genes descritos no banco *Cancer Gene Census*.

Os dados resumidos de *array-CGH* de cada um dos tumores estão sumarizados na Tabela 4.

**Tabela 4** - Dados de *array-CGH* dos carcinomas de mama: número total de *SCNA*/genoma tumoral e número de cada tipo de alteração: ganho, perda, ganho em alto número de cópias (*amplicon*), e perda em homozigose.

Amostra	SCNA#	Ganho	Perda	Amplicon&	Perda em homozigose
64T	18	9	9	0	0
66T	29	7	22	0	0
57T	3	0	3	0	0
229T	1	0	1	0	0
143T	69	33	34	2	0
44T	28	17	11	0	0
55T	25	10	15	0	0
211T	37	19	15	3	0
225T	59	33	22	4	0
21T	2	1	1	0	0
134T	30	23	6	1	0
31T	205	90	81	28	6
59T	13	12	1	0	0
142T	64	43	12	9	0
7T	64	41	23	0	0
42T	13	13	0	0	0
137T	48	30	18	0	0
37T	50	19	31	0	0
56T	35	15	15	5	0
69T	69	22	47	0	0
34T	47	33	14	0	0
218T	108	29	79	0	0
96T	42	9	33	0	0
145T	84	40	35	8	1
33T	19	7	7	5	0
234T	103	52	51	0	0
148T	62	29	30	3	0
SM29	95	46	44	5	0
61T	110	36	74	0	0
60T	20	9	7	4	0
10T	91	46	39	6	0
195T	21	19	0	2	0
224T	91	50	38	1	2
147T	22	12	8	2	0
35T	154	63	85	5	1
63T	42	29	12	1	0
6T	7	7	0	0	0
40T	22	12	10	0	0
208T	48	15	30	3	0
88T	0	0	0	0	0

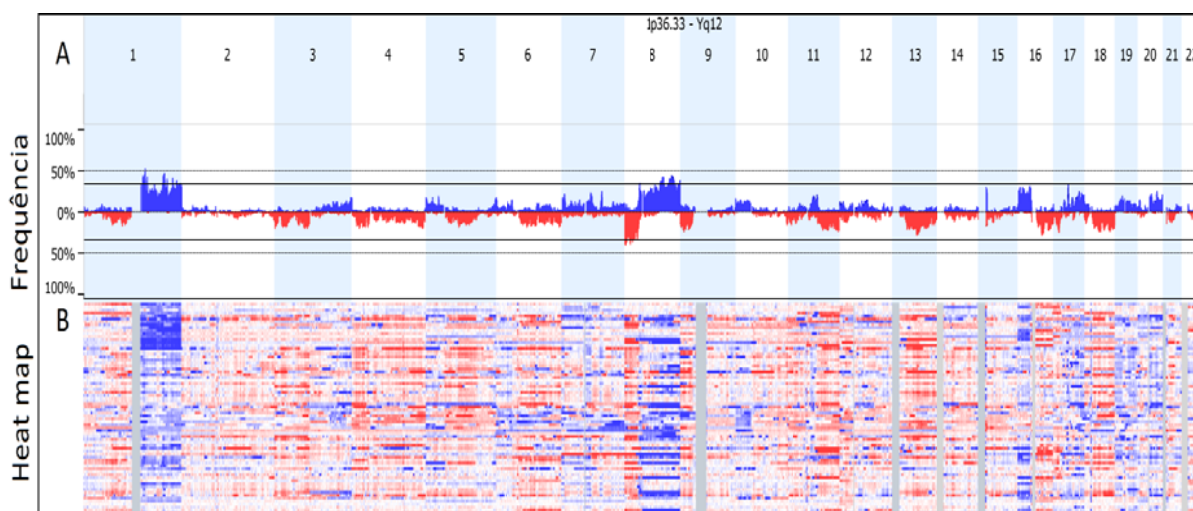


## Cont /Tabela 4

85T	1	1	0	0	0
73T	26	10	15	1	0
68T	94	43	51	0	0
120T	39	27	12	0	0
99T	5	4	0	1	0
38T	5	3	2	0	0
25T	11	9	2	0	0
235T	72	34	33	5	0
149T	70	39	22	9	0
47T	41	25	12	4	0
43T	128	68	57	3	0
9T	3	3	0	0	0
82T	216	102	103	8	3
41T	123	68	53	2	0
152T	8	7	0	1	0
81T	30	13	17	0	0
141T	59	36	22	0	1
48T	19	8	11	0	0
49T	2	0	2	0	0
18T	132	72	51	9	0
22T	45	24	21	0	0
103T	175	92	78	5	0
46T	82	50	28	4	0
93T	11	10	1	0	0
228T	81	40	37	4	0
8T	184	74	99	10	1
20T	119	58	58	3	0
146T	39	31	6	0	2
189T	92	50	42	0	0
144T	59	39	19	0	1
75T	33	25	8	0	0

#Número total de *somatic copy number alterations* (alterações somáticas no número de cópias de segmentos de DNA); *amplicon* – definido como alterações exibindo valor médio de  $\log_2 > 1,4$ .

A análise inicial do grupo de tumores detectou a ocorrência frequente (>20%) de ganhos nos cromossomos 1q, 8q, 16p, e na região 17q12-q21, enquanto que perdas recorrentes foram detectadas principalmente em 8p, 11q, 16q, e nos cromossomos 13 e 18; tais dados podem ser visualizados na figura abaixo.



**Figura 5** - Perfil global de ganhos e perdas de segmentos genômicos do grupo de carcinomas de mama invasivos. **(A)** No eixo X cada sonda está posicionada de acordo com seu mapeamento genômico nos cromossomos (de 1p a 22q); o eixo Y corresponde à frequência de detecção (%) de eventos de ganhos e perdas de segmentos cromossômicos específicos (SCNAs) em todos os tumores do grupo (acima da *baseline* 0%, ganhos em azul; abaixo, perdas em vermelho). **(B)** *Heat map* dos dados de *array-CGH* dos tumores estudados, que mostra o perfil genômico de SCNAs de cada tumor em uma imagem compacta.

Realizamos duas análises estatísticas para identificar as SCNAs significativamente recorrentes no conjunto de tumores.

O primeiro algoritmo utilizado para testar a significância das SCNAs mais frequentes neste grupo de carcinomas de mama foi o **STAC** (*Significance Testing for Aberrant Copy Number*). A lista das regiões do genoma identificadas por este algoritmo contendo aberrações cromossômicas significantes ( $p < 0.001$ ; teste *Fisher-exact two-tailed*) presentes em no mínimo 30% das amostras encontra-se na Tabela 5. É importante salientar que tais regiões identificadas neste teste de significância são segmentos mínimos de sobreposição e com maior amplitude (valores mais altos de ganho ou perda) das SCNAs detectadas no grupo de tumores, sendo chamados de picos. Nesta análise, as SCNAs mais

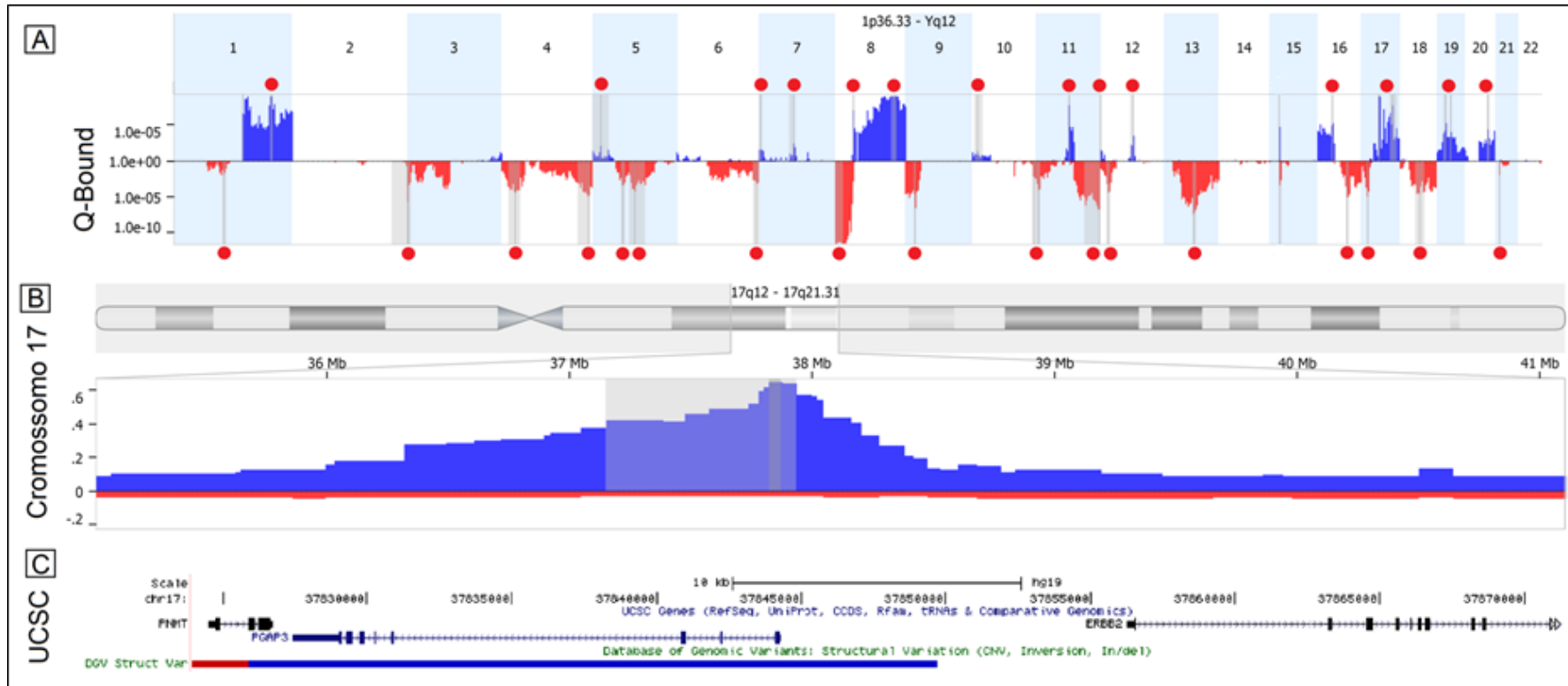
frequentes foram ganhos das regiões 1q21.2-q22, 1q32.1, 8p11.23-q11.22, e 17q12, sendo que variações populacionais comuns (CNVs) foram excluídas.

**Tabela 5** - Segmentos genômicos mínimos de sobreposição de SCNAs significativamente associados ao grupo de carcinomas de mama em frequência igual ou superior 30% de acordo com análise STAC.

Região cromossômica	Banda	Tamanho (pb)	Tipo de SCNA	Genes (n)	% de sobreposição com CNV (DGV)
chr1:149.839.441-150.538.257	q21.2 - q21.3	698816	Ganho	28	22
chr1:153.934.033-154.930.126	q21.3	996093	Ganho	31	0
chr1:154.977.967-156.197.373	q21.3 - q22	1219406	Ganho	79	6
chr1:201.421.914-202.127.128	q32.1	705214	Ganho	16	0
chr1:203.572.814-204.352.694	q32.1	779880	Ganho	20	1
chr1:204.886.021-205.578.790	q32.1	692769	Ganho	24	0
chr8:37.323.843-38.324.922	p11.23 - p11.22	1001079	Ganho	25	2
chr17:37.778.096-37.879.946	q12	101850	Ganho	9	4

O algoritmo **GISTIC** (*Genomic Identification of Significant Targets in Cancer*) foi outra abordagem utilizada para identificar, em relação ao *background* geral de SCNAs dos tumores de mama, quais regiões são afetadas em alta frequência no grupo testado e, portanto, seriam causalmente relacionadas à tumorigênese (*driver mutations*). Regiões cromossômicas com alterações no número de cópias são marcadas separadamente como ganhos ou perdas (SCNAs), e a análise estatística desse algoritmo identifica quais SCNAs provavelmente não ocorreram ao acaso (quando o valor ultrapassa o limiar de significância - *G-score*).

A figura a seguir mostra as SCNAs identificadas utilizando GISTIC, evidenciando as regiões de ganhos e perdas significativas no grupo de carcinomas de mama estudado (Figura 6A). A Figura 6B mostra em detalhes uma das regiões identificadas pela análise GISTIC como um evento frequente de SCNA neste grupo de tumores: o ganho em 17q12-q21. A associação de maior significância estatística (*q-bound* 9,52e-10; *G-score* 47) ocorre com um pequeno segmento de 47 Kb nesta região do cromossomo 17 (chr17:37823941-37871296), contendo o gene *ERBB2* (Figura 6C). De fato, analisando o perfil desta SCNA no grupo de tumores, 24/72 carcinomas de mama exibiram ganho desse segmento, sendo 18 deles classificados como sendo pertencentes ao subtipo HER2; adicionalmente, 17 desses eventos foram ganhos em alto número de cópias.



**Figura 6** - Análise GISTIC do grupo de carcinomas de mama invasivos. Em (A) o eixo Y corresponde ao valor de *q-bound* de eventos de ganhos e perdas de segmentos cromossômicos específicos (SCNAs); acima da linha de base estão representados ganhos genômicos (em azul) e abaixo, perdas genômicas (em vermelho). (A) Regiões identificadas pelo GISTIC como sendo estatisticamente relevantes no grupo de tumores estudados (barras verticais em cinza escuro, marcadas com pontos vermelhos). (B) Detalhamento de uma região frequente de SCNA de ganho no braço longo do cromossomo 17 (17q12-q21.31); o pico significativamente associado a este grupo de carcinomas de mama em 17q12 está marcado pela barra vertical em cinza escuro. (C) Segmento genômico em 17q12 correspondente ao pico identificado com maior significância estatística, com identificação dos genes nele mapeados; na imagem (extraída do Genome Browser UCSC: <http://genome.ucsc.edu/>) é possível verificar os genes mapeados na região, incluindo *ERBB2*.

A Tabela 6 contém a lista dessas regiões identificadas pela análise GISTIC como eventos de SCNA significativamente associados ao grupo de carcinomas de mama ( $q\text{-bound} < 0,05$  e  $G\text{-score} > 1$ ), em virtude da frequência de sua ocorrência e também da amplitude do ganho ou perda identificado. Regiões já descritas como CNVs populacionais foram excluídas.

**Tabela 7** - Regiões cromossômicas significativamente associadas ao grupo de carcinomas de mama invasivos na análise GISTIC.

Região cromossômica (Hg19)	Tipo de SCNA	$q\text{-bound}$	$G\text{-score}$
chr17:37823941-37871296	Ganho	9,52E-10	47
chr1:205334070-205511175	Ganho	9,52E-10	22
chr11:69606306-69761013	Ganho	1,95E-08	19
chr8:38007029-38113868	Ganho	8,87E-08	18
chr1:145566952-145706081	Ganho	2,18E-07	17
chr8:2760161-3154042	Perda	2,76E-12	16
chr19:18981993-19031955	Ganho	5,30E-06	15
chr16:29860542-29980045	Ganho	6,51E-06	15
chr20:48248845-48522017	Ganho	4,73E-05	14
chr19:29888739-30401470	Ganho	8,86E-05	13
chr17:65860878-65972485	Ganho	1,74E-04	13
chr11:132928485-133256550	Perda	1,51E-07	12
chr9:21057153-21209776	Perda	2,50E-07	12
chr2:242851179-243199373	Perda	1,58E-06	12
chr12:69166362-70025817	Ganho	3,05E-04	12
chr4:182227735-183588519	Perda	9,86E-06	11
chr17:13434338-13886159	Perda	1,03E-05	11
chr18:42180299-42365496	Perda	2,03E-05	11
chr7:72205407-72735841	Ganho	0,003605	11
chr4:27501585-31006398	Perda	2,85E-05	10
chr12:17950780-18879392	Perda	3,24E-05	10
chr5:88171471-90639368	Perda	4,49E-05	10
chr3:304303-426897	Perda	5,83E-05	10
chr11:4609990-5616285	Perda	7,27E-05	10
chr5:16605527-16975635	Ganho	0,007586	10
chr5:63280795-64290825	Perda	4,05E-04	9
chr6:163183503-165853076	Perda	6,25E-04	9
chr7:632740-1628634	Ganho	0,026084	9
chr12:481021-1017461	Ganho	0,034804	9
chr10:12056523-12219968	Ganho	0,049554	9
chr10:130920416-131468945	Perda	0,004208	8
chr1:104545267-107001026	Perda	0,010324	8

### 4.3 DELINEAMENTO DO PERFIL GENÔMICO DE ACORDO COM O PADRÃO DE EXPRESSÃO DO RECEPTOR DE ESTRÓGENO

O perfil genômico de SCNAs foi investigado para a identificação de possíveis diferenças entre os carcinomas de mama invasivos de acordo com o *status* de expressão do receptor de estrógeno (ER). A Tabela 7 apresenta a análise de distribuição das SCNAs detectadas no grupo de carcinomas de mama de acordo com o *status* de ER.

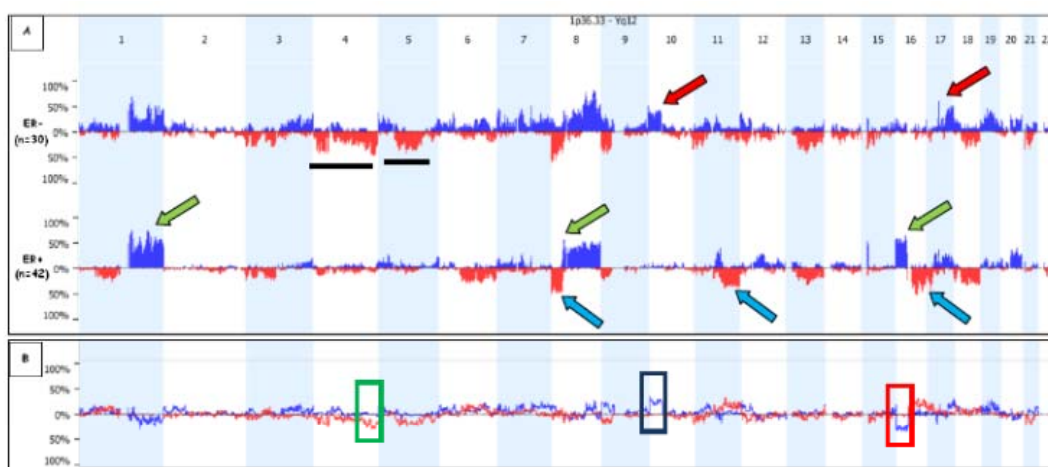
**Tabela 7** - Descrição do padrão de SCNA dos carcinomas de mama em relação ao *status* do receptor hormonal de estrógeno (positivo ou negativo): número médio de SCNA por genoma tumoral e distribuição do número médio de acordo com o tipo de alteração.

<b>SCNAs (média)</b>	<b>ER negativo</b> (n=29)	<b>ER positivo</b> (n=42)
Número SCNA/genoma tumoral (mediana)	67,1 (63)	48,8 (29,5)
Tamanho (Mb) (mediana)	6,9 (3,0)	6,8 (2,6)
Número de Ganhos	35,2	23,7
Número de Perdas	29,2	22,6
Número de <i>Amplicon</i>	2,4	2,2
Número de Perda em homozigose	0,3	0,2

A análise estatística, utilizando teste t para amostras não-paramétricas (Mann-Whitney) identificou diferença significativa apenas entre o número de SCNAs de ganhos por genoma tumoral do grupo ER(-) em relação ao grupo ER(+) ( $p=0,04$ ). Também identificou que as SCNAs dos tumores ER(-) são significativamente maiores do que ER(+).

Quando são consideradas quais regiões são afetadas por eventos de ganhos ou perdas no grupo ER(+) e ER(-), o padrão pode ser visualizado na

Figura 7A. As SCNAs mais frequentes nos grupos de carcinomas de mama ER(+) e ER(-) foram identificadas pela análise STAC e os conjuntos de SCNAs significativas de cada subgrupo foram comparados ( $p < 0.01$  e limiar diferencial de 25% de frequência no grupo). Essa análise pode ser visualizada na Figura 7B.



**Figura 7** - Padrão de SCNA dos carcinomas de mama de acordo com o status de expressão do receptor hormonal de estrógeno (ER).

(A) Acima, perfil genômico de frequência de SCNA do grupo de tumores ER(-); abaixo, perfil genômico do grupo ER(+), evidenciando que o grupo ER(-) apresenta uma variabilidade maior de regiões genômicas afetadas quando comparado ao grupo ER(+): na inspeção visual desses resultados evidenciam-se grandes regiões de perdas em 4q e 5q (barras pretas) só detectadas nos tumores ER(-), além de ganhos focais em 10p e 17q (setas vermelhas). O grupo de tumores ER(+) exibe SCNAs concentradas em regiões específicas do genoma, como ganhos em 1q, 8q e 16p (setas verdes), e perdas em 8p, 11q e 16q (setas azuis).

(B) Análise STAC identificando SCNAs mais significativas no grupo ER(-) *versus* o grupo ER(+): nessa comparação, evidenciam-se alterações focais de ganho em 10p (box preto) e perda em 4q (box verde) nos tumores ER(-), e ganho em 16p (box vermelho) nos tumores ER(+).



A Tabela 8 mostra as regiões genômicas contendo SCNA significativamente associadas ao subgrupo ER(-).

**Tabela 8** - Regiões cromossômicas significativamente ( $p \leq 0,001$ ) associadas ao grupo de carcinomas de mama ER(-) de acordo com a comparação dos dados após análise STAC.

Região	Banda	Tipo de SCNA	Diferença	Max p-valor	% de sobreposição com CNV (DGV)
chr10:34,964,766-35,567,952	p11.21	ganho	28	0,001	0
chr4:155,699,126-158,621,481	q32.1	perda	28	0,001	14
chr19:29,888,739-30,401,470	q12	ganho	28	0,001	2
chr10:31,696,240-33,441,454	p11.22	ganho	28	0,001	2
chr10:23,537,580-23,952,796	p12.2	ganho	28	0,001	2

#### 4.4 CORRELAÇÃO ENTRE O PERFIL GENÔMICO DE SCNA DOS CARCINOMAS COM O *STATUS* DE METÁSTASE LINFONODAL AO DIAGNÓSTICO

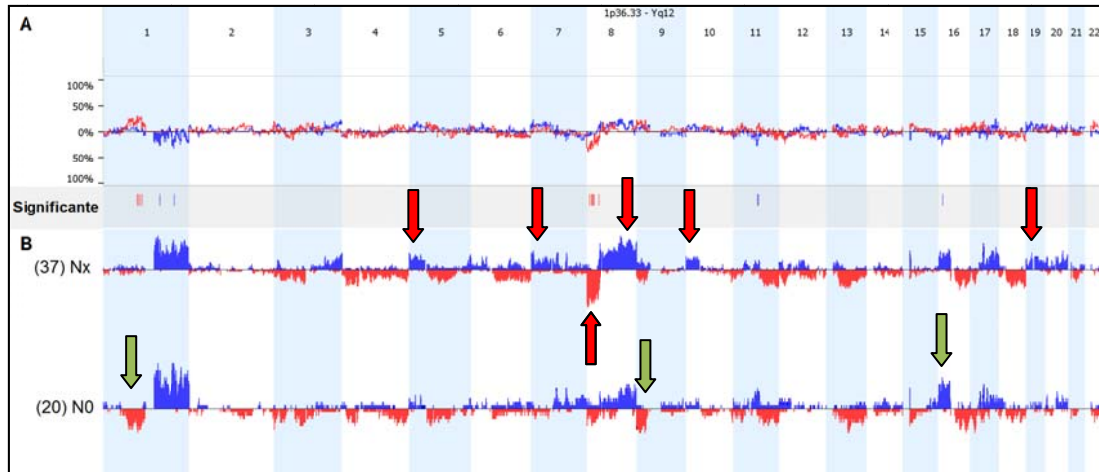
Foi compilada a informação clínica sobre o *status* linfonodal ao diagnóstico, que é um importante fator prognóstico. A partir desses dados, estabelecemos como critério para classificar os 57 carcinomas de mama ductal invasivos em *positivos para metástase linfonodal* os casos com  $\geq 1$  linfonodos acometidos (Nx = classes N1, N2 e N3), e como negativos aqueles com nenhum linfonodo acometido (N0). Separados os tumores em dois grupos - linfonodo-positivo (linf(+)) e linfonodo-negativo (linf(-)) -, realizamos uma comparação das SCNAs entre os mesmos para investigar diferenças nos padrões genômicos. A Tabela 9 apresenta os dados de SCNA dos tumores discriminados de acordo com o *status* linfonodal. Não foram detectadas diferenças significativas no número total de SCNAs e de seus diferentes subtipos (ganho, perda, amplicon e deleção em homozigose) entre os dois grupos de carcinomas de mama, com e sem metástase linfonodal ao diagnóstico.

**Tabela 9** - Descrição do padrão de SCNA dos 57 carcinomas de mama do tipo ductal invasivo (número médio por genoma tumoral e distribuição do número médio de acordo com o tipo de alteração) em relação ao *status* linfonodal ao diagnóstico (positivo ou negativo).

SCNA (número médio)	Linfonodo-Negativo	Linfonodo-Positivo
	N0 (n=20)	Nx (N1/N2/N3) (n=37)
Número de SCNA/genoma tumoral (mediana)	53,7 (42,5)	48,5 (38)
Número de Ganhos	26,3	25,8
Número de Perdas	24	20,9
Número de <i>Amplicons</i>	3	1,7
Número de Perdas em homozigose	0,4	0,1

Quando comparamos qualitativamente as alterações genômicas detectadas no grupo linf(+) com o padrão obtido no grupo linf(-), verificamos que o perfil genômico de SCNA é parcialmente distinto (Figura 8). O algoritmo STAC foi utilizado para identificar as SCNAs mais frequentes nesses dois grupos de carcinomas de mama: Nx (linf(+)) e N0 (linf(-)). Em seguida, o teste exato de Fischer foi utilizado para comparar o conjunto de SCNAs significativas de cada um dos grupos ( $p < 0.05$  e limiar diferencial de 25% de frequência no grupo).

Tais regiões significativamente associadas a Nx (linf(+)) e N0 (linf(-)) estão indicadas na Figura 8 e Tabela 10.



**Figura 8** - Padrão de SCNA dos carcinomas de mama de acordo com o status linfonodal. (A) Análise STAC identificando SCNAs mais significativas no grupo Nx (linf(+)) versus o grupo N0 (linf(-)); (B): No painel acima, o perfil genômico de frequência de SCNA do grupo de tumores Nx (linf(+)); abaixo, perfil genômico do grupo N0 (linf(-)); na inspeção visual desses resultados evidenciam-se regiões de ganhos mais frequentes nos tumores Nx (linf(+)) – 5p, 7p, 8q, 10p e 19 e perda frequente em 8p (setas vermelhas). O grupo de tumores N0 (linf(-)) exibe SCNAs em outras regiões do genoma, como perdas mais frequentes em 1p e 9p e ganho em 16p (setas verdes).

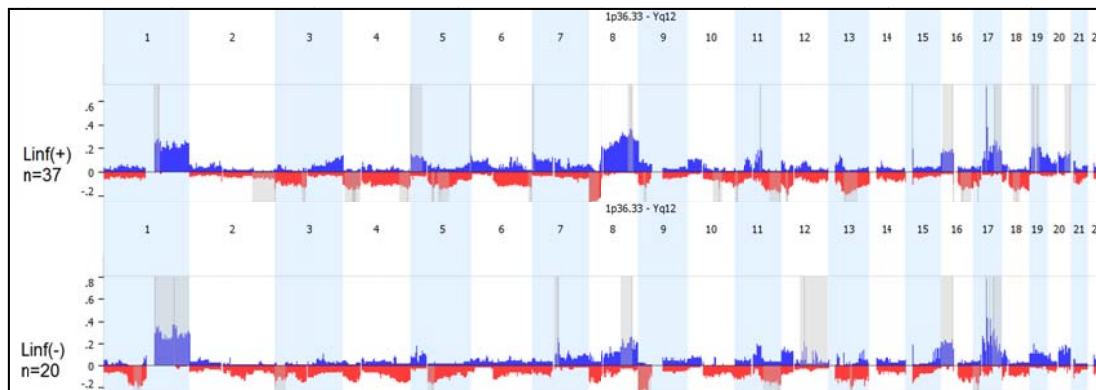
A principal característica na inspeção visual do perfil genômico dos 2 grupos (parte inferior da figura acima) é uma frequência significativamente aumentada no grupo Nx (linf(+)) de ganhos de baixa amplitude em 5p, 7p, 8q, 10p e no cromossomo 19 e perda frequente em 8p (setas vermelhas). Três regiões genômicas apresentaram frequência maior de SCNAs específicas no grupo de tumores N0 (linf(-)): perdas em 1p e 9p e ganho em 16p (setas verdes). A Tabela 10 lista regiões genômicas contendo SCNAs significativamente associadas aos grupos Nx(linf(+)) e N0 (linf(-)) segundo a análise STAC. É interessante notar os valores negativos de diferença de frequência entre os grupos correspondem a SCNAs mais frequentes no grupo linf(-).

**Tabela 10** - Regiões cromossômicas significativamente ( $p < 0,05$ ) associadas aos grupos de carcinomas de mama linf(+) e linf(-) após análise STAC.

Região	Banda	Tipo de SCNA	Tamanho (pb)	Diferença	p-value
<b>chr8:5,971,870-6,091,467</b>	p23.2	<b>Perda</b>	119597	40*	0,002
<b>chr8:11,823,901-12,212,580</b>	p23.1	<b>Perda</b>	388679	35*	0,007
<b>chr8:19,434,176-19,918,093</b>	p21.3	<b>Perda</b>	483917	33*	0,02
<b>chr8:11,003,802-11,101,380</b>	p23.1	<b>Perda</b>	97578	28*	0,04
<b>chr8:14,326,493-14,631,300</b>	p22	<b>Perda</b>	304807	28*	0,04
<b>chr8:17,168,289-17,407,064</b>	p22	<b>Perda</b>	238775	28*	0,04
<b>chr8:31,529,412-32,198,491</b>	p12	<b>Perda</b>	669079	28*	0,04
chr1:108,767,820-109,365,583	p13.3	<b>Perda</b>	597763	-25	0,003
chr1:96,762,645-97,412,333	p21.3	<b>Perda</b>	649688	-27	0,005
chr1:99,217,104-99,571,728	p21.3	<b>Perda</b>	354624	-27	0,005
chr1:103,312,707-104,081,862	p21.1	<b>Perda</b>	769155	-27	0,005
chr11:72,774,585-72,823,003	q13.4	<b>Ganho</b>	48418	-27	0,005
chr11:73,664,911-73,702,926	q13.4	<b>Ganho</b>	38015	-27	0,005
chr1:161,676,966-162,156,250	q23.3	<b>Ganho</b>	479284	-28	0,03
chr16:10,386,730-10,584,125	p13.2 - p13.13	<b>Ganho</b>	197395	-28	0,02
chr1:205,667,530-205,783,984	q32.1	<b>Ganho</b>	116454	-32	0,02

\* em negrito, SCNAs mais frequentes no grupo Nx (linf(+)).

O algoritmo **GISTIC** foi também utilizado para identificar, em relação ao *background* geral de SCNAs, quais regiões são afetadas em alta frequência no grupo de carcinomas derivados de pacientes com metástase linfonodal (Figura 9).



**Figura 9** - Análise GISTIC do grupo de carcinomas de mama linfonodo-positivo e do grupo linfonodo-negativo. O eixo X corresponde aos cromossomos 1 a 22; o eixo Y corresponde ao valor de  $q$ -bound de eventos de ganhos e perdas de segmentos cromossômicos específicos (SCNAs) nesses grupos de tumores; acima da linha de base estão representados ganhos genômicos (em azul) e abaixo, perdas genômicas (em vermelho). As barras verticais em cinza indicam as regiões com significância estatística.

A tabela a seguir contém as regiões identificadas como eventos de SCNA significativamente associados ao grupo de carcinomas de mama Nx (linf(+)) ( $q$ -bound  $< 0,01$  e  $G$ -score  $> 1$ ), em virtude da frequência de sua ocorrência e também da amplitude do ganho ou perda identificado.

**Tabela 11** - Regiões cromossômicas significativamente associadas ao grupo de carcinomas de mama linf(+) na análise GISTIC.

Região	Extended Region	Tipo de SCNA	G-Score
chr17:37,798,557-37,879,946	chr17:36,320,639-37,934,421	Ganho	26,6
chr8:126,511,533-127,318,289	chr8:118,127,984-133,866,081	Ganho	13,8
chr1:155,310,165-155,866,872	chr1:149,208,044-161,465,880	Ganho	10,6
chr8:2,170,104-3,673,702	chr8:0-29,220,115	Perda	9,7
chr1:145,448,491-145,728,126	chr1:144,927,778-147,786,706	Ganho	9,3
chr8:38,007,029-38,113,868	chr8:37,478,964-39,240,660	Ganho	8,8
chr19:29,888,739-30,401,470	chr19:28,272,497-36,780,962	Ganho	8,6
chr19:19,245,358-19,581,675	chr19:11,544,601-24,340,741	Ganho	8
chr11:76,479,899-76,787,185	chr11:74,690,255-78,916,502	Ganho	7,6
chr12:17,950,780-18,879,392	chr12:13,393,721-20,811,443	Perda	7,5
chr16:31,502,781-31,523,370	chr16:3,216,680-32,276,345	Ganho	7,3
chr13:46,949,291-47,327,645	chr13:45,838,581-83,296,785	Perda	7,2
chr15:20,102,541-22,129,154	chr15:20,102,541-22,459,813	Ganho	7,1
chr11:131,167,909-133,256,550	chr11:102,241,870-135,006,516	Perda	6,5
chr18:42,180,299-42,665,411	chr18:33,105,695-52,841,354	Perda	6,2
chr16:59,446,041-59,787,236	chr16:55,536,122-87,035,798	Perda	6
chr4:27,101,933-37,415,498	chr4:5,245,519-49,063,301	Perda	5,7
chrX:123,691,824-124,242,779	chrX:101,019,658-154,984,902	Perda	5,6
chr5:83,724,964-85,875,664	chr5:77,885,236-111,282,729	Perda	5,6
chr2:242,851,179-243,199,373	chr2:179,221,748-243,199,373	Perda	5,3
chr4:182,227,735-183,686,426	chr4:160,392,383-191,154,276	Perda	5,3
chr3:83,336,237-83,551,736	chr3:76,055,235-88,290,833	Perda	5,2
chr9:21,057,153-21,209,776	chr9:15,951,655-23,365,289	Perda	5,1
chr10:89,064,267-89,535,612	chr10:72,154,340-99,278,943	Perda	5,1

#### 4.5 MAPEAMENTO DE *AMPLICONS* E DELEÇÕES EM HOMOZIGOSE PARA IDENTIFICAÇÃO DE POTENCIAIS ONCOGENES E SUPRESSORES TUMORAIS RELEVANTES PARA A PROGRESSÃO PARA METÁSTASE LINFONODAL

A última etapa de nossa análise foi investigar as regiões genômicas afetadas por SCNAs de maior amplitude, como ganhos em alto número de cópias e deleções em homozigose. Essa análise foi realizada apenas no grupo de carcinomas de mama com metástase linfonodal ao diagnóstico, uma vez que estávamos buscando identificar potenciais oncogenes e genes supressores tumorais associados à progressão tumoral.

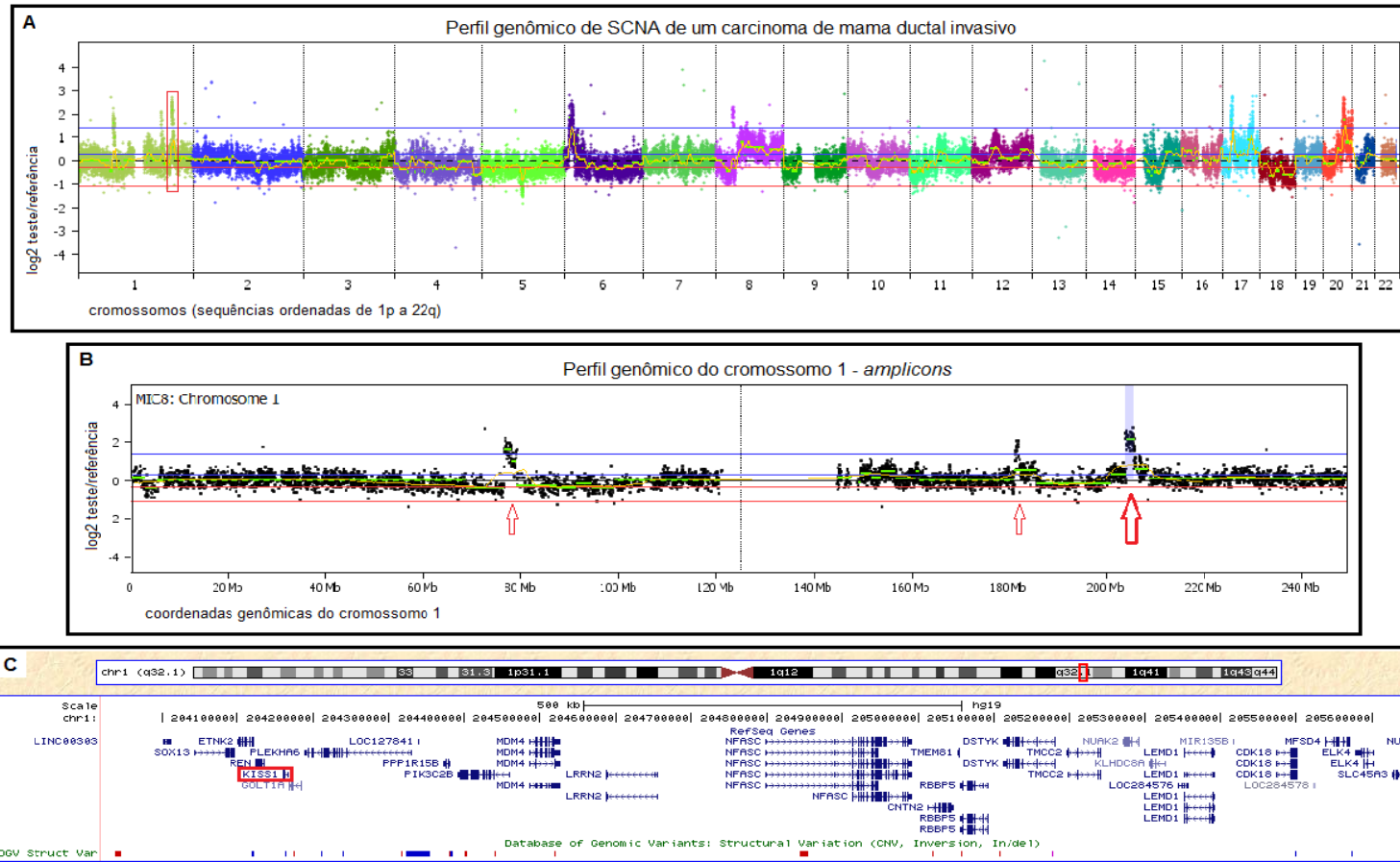
Sete regiões afetadas por SCNAs de maior amplitude foram selecionadas para validação técnica dos resultados. A Tabela 12 descreve essas regiões, seu mapeamento, tamanho, tipo de evento (*amplicon* ou perda em homozigose) e gene escolhido para validação.

**Tabela 12** - Regiões cromossômicas afetadas e respectivos genes escolhidos para validação por *qPCR*.

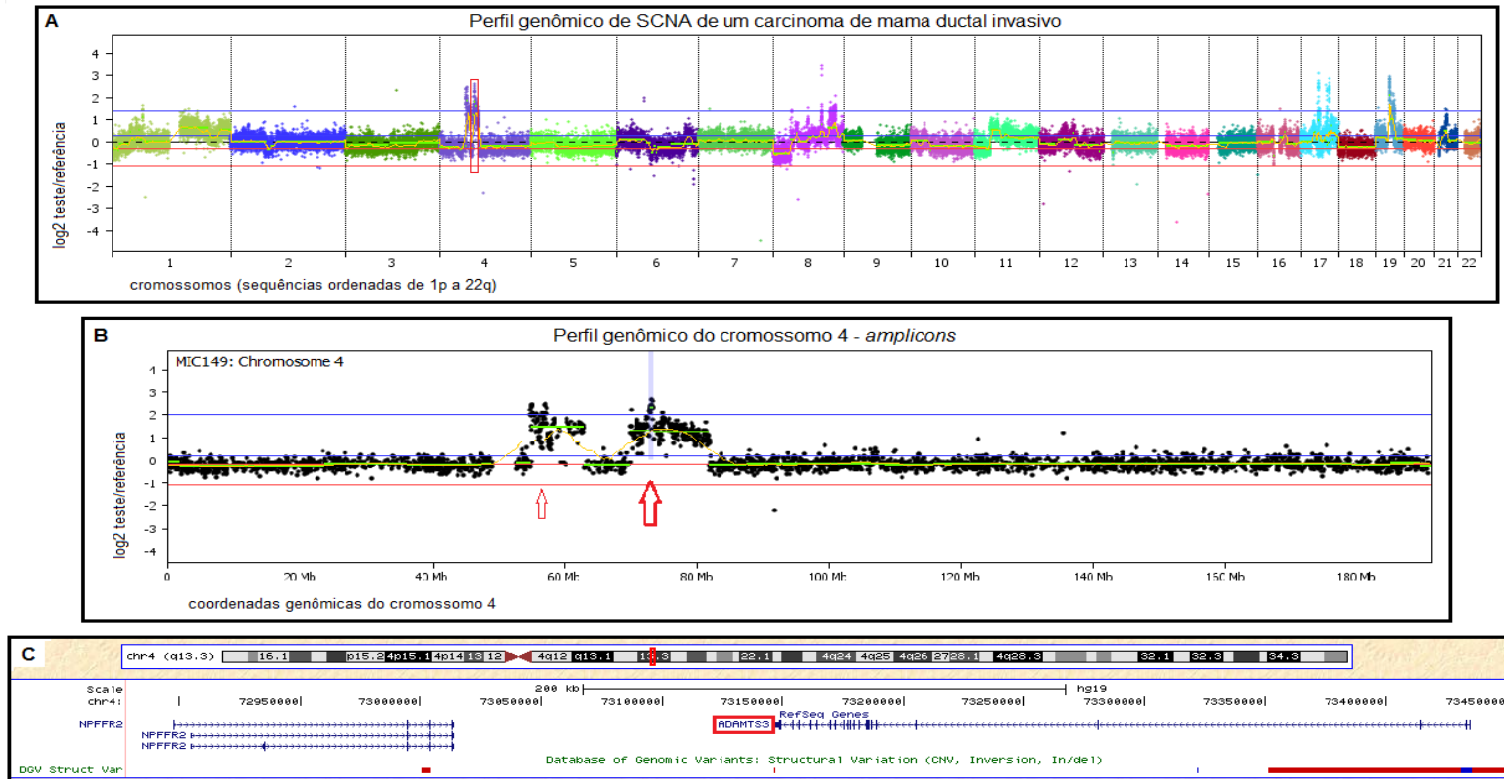
Região	Mapeamento	Tamanho (Kb)	Evento	Genes para validação
chr1q32.1	204159469 - 204165619	6,1	<i>amplicon</i>	<i>KISS1</i>
chr3q13.31	204165619 - 114866127	809,2	Perda em homozigose	<i>ZBTB20</i>
chr4q13.3	73146686 - 73434516	287,8	<i>amplicon</i>	<i>ADAMTS3</i>
chr9p21.3	21802635 - 22032985	230,3	Perda em homozigose	<i>MTAP</i>
chr11p13	34500340 - 34535352	35,0	Perda em homozigose	<i>ELF5</i>
chr11p11.2	43577986 - 43878169	300,2	<i>amplicon</i>	<i>HSD17B12</i>
chr12p12.3	18233803 - 18473041	239,2	Perda em homozigose	<i>RERGL</i>



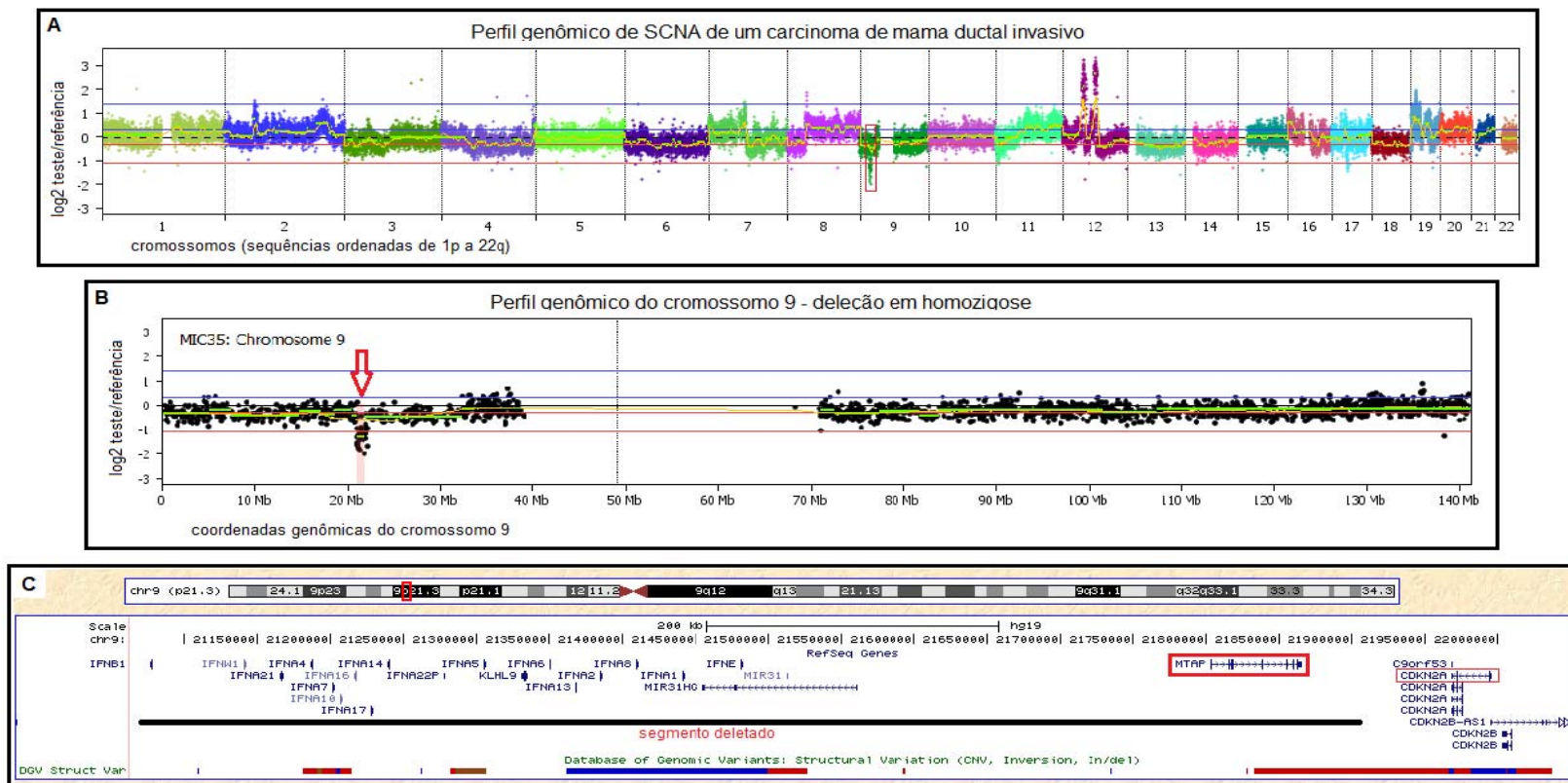
Abaixo, seguem figuras exemplificando como tais alterações foram visualizadas no perfil de *array-CGH*, a região afetada em detalhes no perfil genômico e no *Genome Browser*.



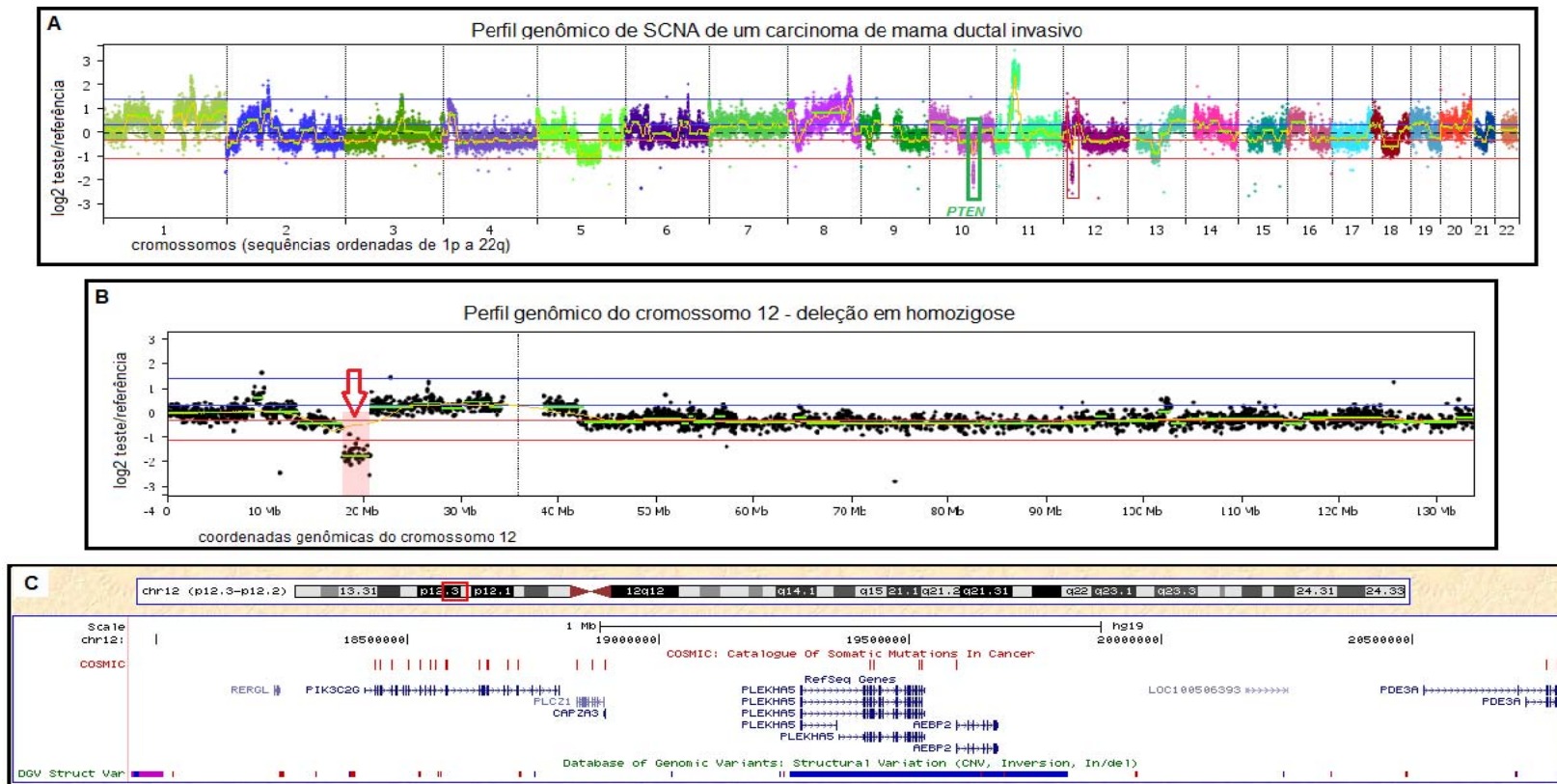
**Figura 10** - Amplificação em 1q32.1. (A) Perfil genômico de *array-CGH* do carcinoma de mama no qual a alteração foi identificada, com destaque (box vermelho) para o *amplicon* investigado; (B) Perfil do cromossomo 1, destacando três regiões de SCNAs de ganho (setas vermelhas, sendo a seta maior o *amplicon* em investigação); (C) Segmento genômico do *amplicon* detectado em 1q32.1, mostrando o mapeamento de diversos genes, dentre eles *KISS1* (box vermelho).



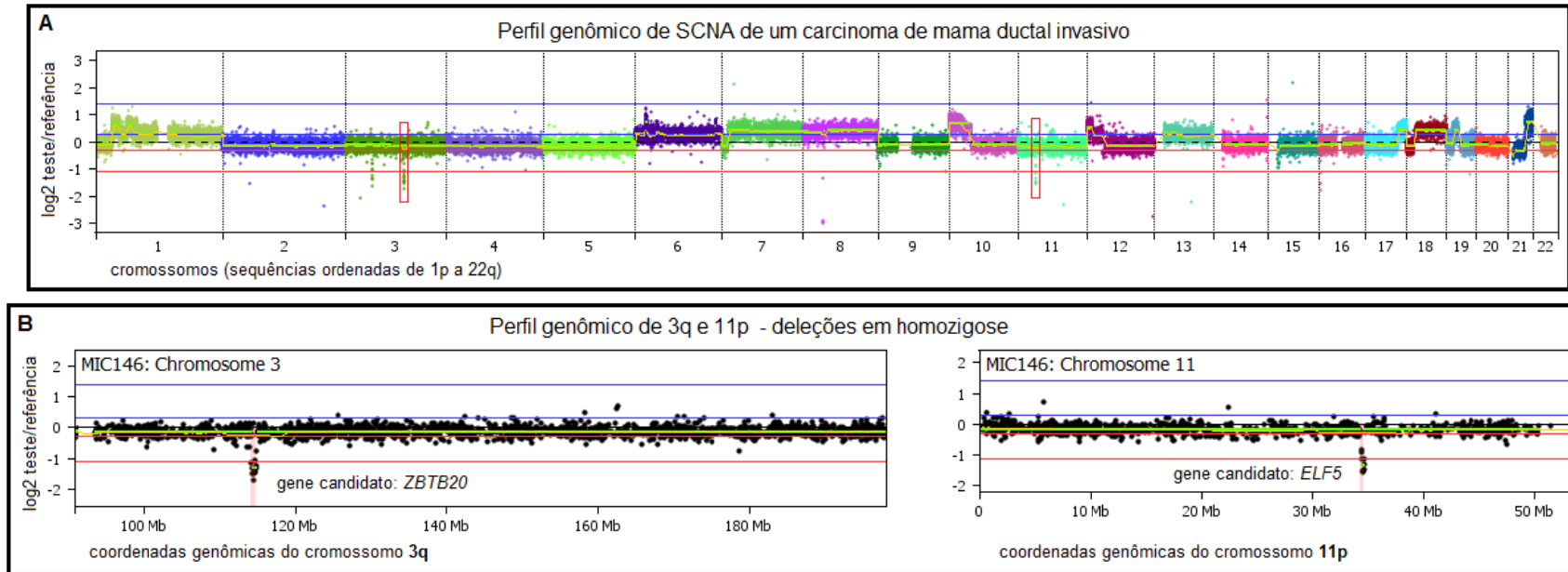
**Figura 11** - Amplificação em 4q13.3. (A) perfil genômico de *array-CGH* do carcinoma de mama no qual a alteração foi identificada, com destaque (box vermelho) para o *amplicon* investigado; (B) Perfil do cromossomo 4, destacando duas regiões de SCNAs de ganho (setas vermelhas, sendo a seta maior o *amplicon* em investigação); (C) Segmento genômico do *amplicon* detectado em 4q13.3, mostrando o mapeamento de diversos genes, dentre eles *ADAMTS3* (box vermelho).



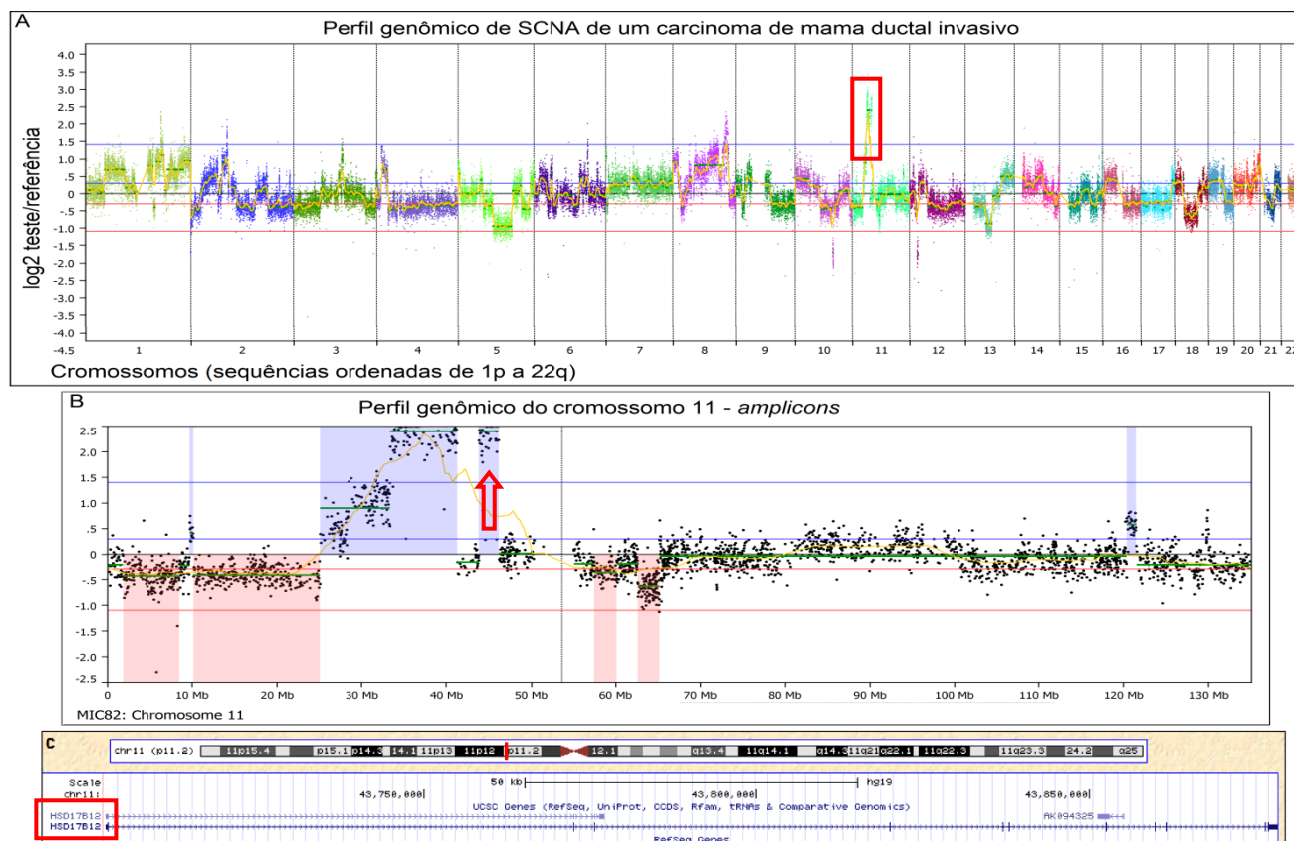
**Figura 12** - Deleção em homozigose em 9p21.3. (A) perfil genômico de *array-CGH* do carcinoma de mama no qual a alteração foi identificada, com destaque (box vermelho) para a perda em homozigose investigada; (B) Perfil do cromossomo 9, destacando a região de SCNAs de deleção em investigação (seta vermelha); (C) Segmento genômico da deleção em homozigose detectada em 9p21.3, mostrando o mapeamento de diversos genes, dentre eles *CDKN2A* (box vermelho).



**Figura 13** - Deleção em homozigose em 12p12.3. (A) Perfil genômico de *array-CGH* do carcinoma de mama no qual a alteração foi identificada, com destaque (box vermelho) para a deleção investigada, e também para uma deleção em homozigose no cromossomo 10 afetando o conhecido supressor tumoral *PTEN*; (B) Perfil do cromossomo 12, destacando a região de SCNA de perda (seta vermelha); (C) Segmento genômico da microdeleção detectada em 12p12.3, mostrando o mapeamento de diversos genes, dentre eles *RERGL* (box vermelho).



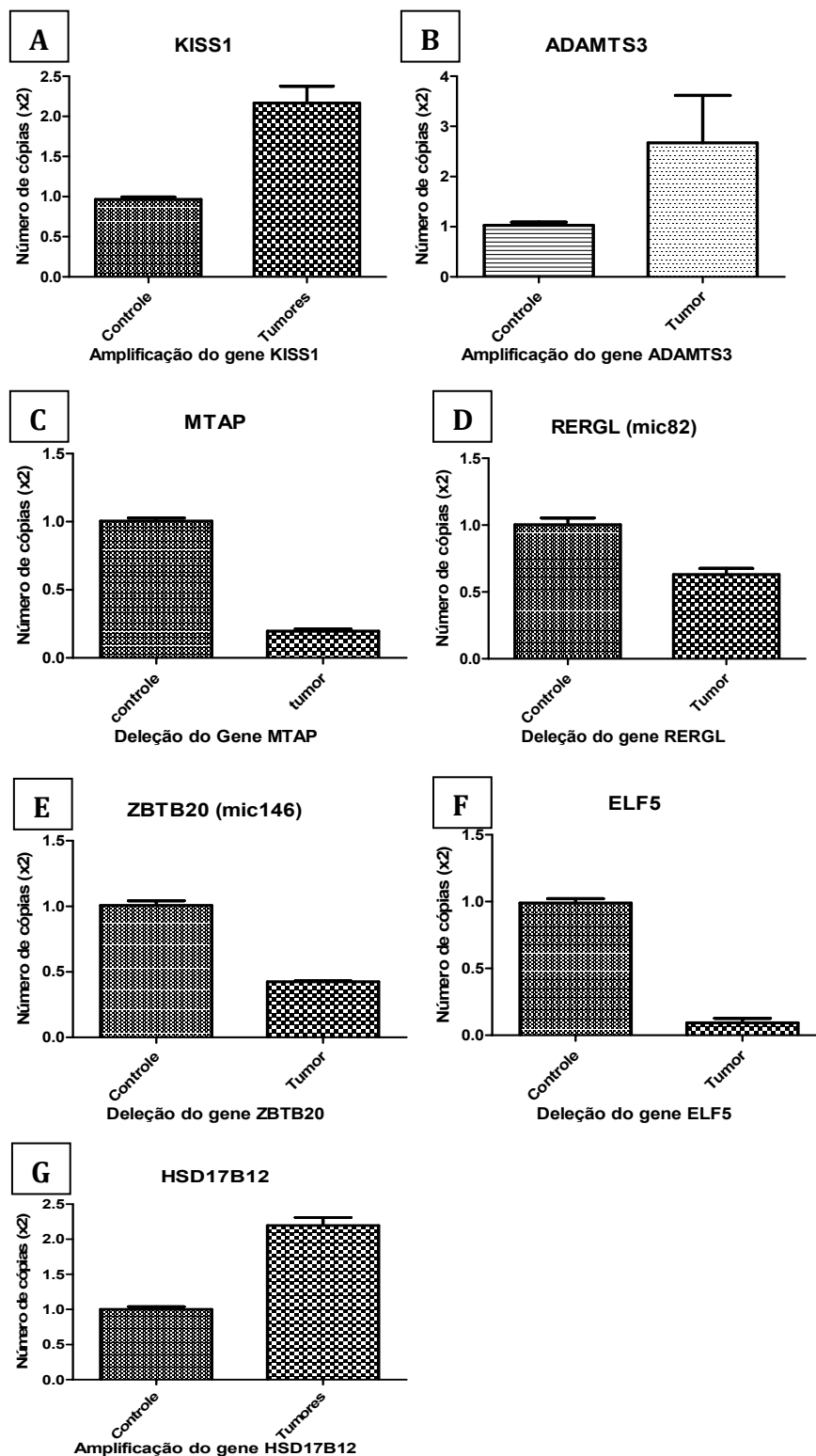
**Figura 14** - Deleções em homozigose em 3q13.31 e 11p13. (A) perfil genômico de *array-CGH* do carcinoma de mama no qual as alterações foram identificadas, com destaque (box vermelho) para as deleções investigadas; (B) Perfil das regiões cromossômicas 3q e 11p, mostrando as duas regiões de SCNAs de deleção abrangendo os genes para validação *ZBTB20* e *ELF5*.



**Figura 15 - Amplicons em 11p11.2.** (A) perfil genômico de *array-CGH* do carcinoma de mama no qual as alterações foram identificadas, com destaque (box vermelho) para o *amplicon* investigado; (B) Perfil das regiões cromossômicas de 11p, mostrando as regiões de SCNAs de *amplicons* abrangendo o gene para validação *HSD17B12*. (C) Segmento genômico do amplicon detectado em 11p11.2 e o gene em investigação *HSD17B12* (box vermelho).

Os sete genes seleccionados em amplicons ou deleções em homozigose detectados na análise de array-CGH foram todos validados no *qPCR* quantitativo quanto à alteração detectada no número de cópias, como pode ser visualizado na Figura 16.





**Figura 16** - Genes validados por *qPCR*. Relação tumores testados/validados: (A) Amplificação do gene *KISS1* (32/27 tumores); (B) Amplificação do gene *ADAMTS3* (4/4 tumores); (C) Deleção do gene *MTAP* (12/11 tumores); (D) Deleção do gene *RERGL* no tumor MIC82; (E) Deleção do gene *ZBTB20* no tumor MIC146; (F) Deleção do gene *ELF5* nos tumores (2/2 tumores); (G) Amplificação do gene *HSD17B12* nos tumores (3/3 tumores).

## 5 DISCUSSÃO

Alterações genômicas em número de cópias são reconhecidamente importantes no câncer, principalmente porque sua identificação pode ajudar no mapeamento de regiões contendo oncogenes e supressores tumorais. Tais alterações genômicas são mediadoras de alterações fenotípicas por meio do impacto na expressão gênica.

Neste trabalho, foram considerados como alterações somáticas no número de cópias de sequências de DNA (*SCNA – somatic copy number alteration*) aqueles segmentos representados por no mínimo cinco sondas consecutivas com o valor alterado de  $\log_2$  da razão Cy3/Cy5 (*amostra tumoral/DNA teste*), ou seja, acima do limiar que estabelecemos como *background* experimental. Considerando a plataforma 60K aqui utilizada, são 60.000 sondas distribuídas de maneira aproximadamente regular ao longo do genoma humano, com espaçamento médio entre cada sonda de 50 Kb. Portanto, o tamanho mínimo médio de perdas e ganhos genômicos a serem detectados pela análise que fizemos fica em torno de 200 Kb. No entanto, esse limite de 200 Kb é um valor médio da resolução para identificação de alterações genômicas individuais nos tumores. Na plataforma 60K usada, parte das sondas está preferencialmente alocada em sequências codificadoras do genoma, o que torna possível a detecção de rearranjos <200 Kb em segmentos genômicos com maior densidade gênica e, portanto, maior cobertura.

Outra consideração importante a ser feita é que os valores de limiar para detecção de ganhos e perdas genômicos são sempre determinados de maneira arbitrária em todos os *softwares* de análise de dados de *array-CGH*; isso significa que tais valores são definidos pelo usuário. Tal característica, inerente à análise de alterações genômicas a partir de dados de *array-CGH*, torna fundamental a avaliação inicial da qualidade experimental dos ensaios para obtenção de dados confiáveis. Além disso, a possibilidade de escolha e modificação dos critérios faz com que a análise seja fortemente dependente de experiência prévia em citogenômica, e também de conhecimento teórico sólido para avaliação da adequação dos dados obtidos ao contexto biológico.

Assim, inicialmente realizamos diversas simulações de análises no *software* Nexus, variando os parâmetros de detecção de alterações, utilizando experimentos de *array-CGH* de amostras com aberrações cromossômicas já conhecidas e validadas independentemente. A partir disso, os limiares para detecção de alterações genômicas nos tumores foram definidos como valores médios de razão  $\log_2$  do segmento alterado  $>0,3$  para ganhos genômicos e  $>1,4$  para ganhos em alto número de cópias (aqui denominados de *amplicons*), e razão  $\log_2 <-0,3$  e  $<-1,1$  para perdas genômicas e deleção em homozigose, respectivamente. Estes limiares são comumente adotados na análise de dados de *array-CGH* de tumores; qualquer viés eventualmente introduzido no resultado da análise pela adoção destes critérios será comum a todas as amostras do grupo, já que foram analisadas da mesma maneira.

Tumores sólidos como os carcinomas de mama são sabidamente heterogêneos do ponto de vista da composição genômica de suas células, e mesmo alterações cromossômicas funcionalmente importantes para a carcinogênese podem estar presentes em apenas um subgrupo celular dentro do mesmo tumor. Assim, optamos pela adoção de valores de detecção de ganhos e perdas que simultaneamente permitissem a exclusão do ruído experimental, que introduziria artefatos na análise, e garantissem a detecção de alterações cromossômicas presentes em mosaico na amostra tumoral. Por outro lado, definimos um limiar de significância bastante alto ( $10^{-8}$ ), o que garante que serão detectadas apenas aquelas alterações com valores significativamente alterados em relação ao *background* genômico daquela amostra tumoral, mesmo que em amplitude reduzida de ganho ou perda (ou seja, em mosaico na amostra).

O programa *Nexus Copy Number* permite a análise dos dados de *array-CGH* em dois diferentes níveis. No primeiro deles, é possível uma análise por amostra para identificar SCNAs de acordo com critérios de análise definidos pelo usuário. O nível seguinte permite a combinação dos resultados de todas as amostras para responder questões mais elaboradas, dentro de um contexto de população ou grupo de amostras e suas características clínicas ou anatomopatológicas.

Mesmo considerando que múltiplos rearranjos somáticos são achados extremamente frequentes no genoma de tumores de mama, o padrão estrutural de tais rearranjos é muito complexo e sua contribuição para o desenvolvimento do câncer ainda pouco caracterizada. Na análise do nosso

grupo de tumores identificamos um grande desvio-padrão no tamanho das SCNAs, com uma mediana muito menor do que a média das alterações; ou seja, identificamos que 50% dos rearranjos detectados são  $< \sim 3$  Mb (ou seja, submicroscópicos), sendo o restante de tamanho muito variável, incluindo neste espectro a ocorrência das aneuploidias de braços cromossômicos inteiros. Também verificamos que ocorre uma diferença significativa entre o tamanho dos eventos de baixa amplitude em relação ao tamanho de amplicons e perdas em homozigose, estes últimos sendo principalmente rearranjos focais ( $< 1,8$  Mb). Outro dado importante é a grande diferença verificada entre as categorias desses rearranjos focais de grande amplitude: o número de *amplicons* foi 10 vezes maior do que o número de perdas em homozigose no grupo dos tumores de mama. Em STEPHENS et al. (2009) sequenciaram 24 tumores de mama para identificar e caracterizar rearranjos somáticos; notavelmente, similar a nossos dados, os rearranjos detectados, principalmente intra-cromossômicos, foram mais comumente duplicações *in tandem*, com grande parte dessas alterações sendo  $< 2$  Mb (i.e., indetectável pela citogenética tradicional), confirmando que o mecanismo de amplificação de segmentos genômicos é um processo comum na etiologia do câncer de mama. Neste estudo de 2009, novos genes de fusão e genes internamente rearranjados também foram identificados, mas nenhum deles recorrente, achado interpretado pelos autores como indicativo de que são eventos passageiros ou raros no desenvolvimento do câncer de mama. Como os tumores apresentaram diferenças marcantes dependendo do número de duplicações *in tandem* que apresentavam, para tumores com alto número de

duplicações os autores postulam a existência de um processo defeituoso de manutenção do DNA.

Cariótipo de tumores sólidos frequentemente mostram multiclonalidade, sugerindo a existência de alto grau de heterogenetidade intratumoral. Grande parte das SCNAs detectadas em nosso trabalho em frequência alta no grupo estudado de tumores exhibe em geral valores de  $\log_2$  baixos (baixa amplitude) nos tumores individualmente, indicando que ocorrem em mosaico. Qualitativamente, as regiões que apareceram em alta frequência no grupo de carcinomas de mama foram ganhos nos cromossomos 1q, 8q, 16p, e na região 17q12-q21, e perdas recorrentes foram detectadas principalmente em 8p, 11q, 16q, e afetando os cromossomos 13 e 18. Isso condiz com os dados da literatura citados acima e outros, que mostram que as alterações frequentemente mapeadas ocorrem principalmente nos cromossomos 8, 11q, 13q, 16p e 16q, 17p e 17q e 20q (DOWNING et al. 2010).

O receptor de estrógeno é um importante marcador prognóstico nos tumores de mama; tumores que não expressam esse receptor evoluem de maneira pior, sendo até considerados como uma doença distinta dentro dos carcinomas de mama em níveis de transcrição gênica (SØRLIE et al. 2001; WANG DY et al. 2011; REIS-FILHO e PUSZTAI 2011). Considerando a relevância deste marcador, analisamos também o grupo de tumores de acordo com o padrão de expressão do receptor de estrógeno, em busca de regiões genômicas diferentemente afetadas por rearranjos. No carcinoma de mama, alguns estudos mostram que a perda de 16q ocorre em tumores

positivos quanto à expressão do receptor de estrógeno (ER+) e está associada a um melhor prognóstico, sendo detectável comumente no carcinoma ductal *in situ*, indicando que esse evento (perda de 16q) ocorreria no início da tumorigênese (LOO et al. 2004; DOWNING et al. 2010). Além disso, carcinomas ductais de baixo grau exibem ganho mais frequente de 16p e perda de 16q, enquanto que em tumores de alto grau o ganho de 16p e perda de 16q seriam eventos menos frequentes. Nas nossas amostras, o ganho de 16p e a perda de 16q ocorreram em conformidade com os dados da literatura científica, associando-se principalmente com a positividade para o ER. Eventos como perdas em 4q e 5q e ganhos em 10p e 17q12 foram significativos em tumores ER(-), resultados também condizentes com estudos anteriores de tumores de mama usando *array-CGH* como técnica de investigação (MELCHOR et al. 2007). É importante ressaltar que no grupo de tumores negativos para receptores de estrógeno e progesterona observou-se um grande número de SCNAs, principalmente duplicações, confirmando trabalhos anteriores que mostraram uma grande quantidade de alterações cromossômicas em alguns carcinomas de mama com receptores hormonais negativos (HAN et al. 2008; HU et al. 2009; STEPHENS et al. 2009).

O *status* linfonodal axilar é outro importante fator prognóstico em carcinomas de mama e parece estar diretamente associado ao desenvolvimento de metástases à distância nas pacientes com câncer de mama. Aproximadamente 75% das pacientes que apresentam linfonodo-negativo nunca desenvolverão metástase à distância (KAHN et al. 2006).

Em trabalho recente, GOBARDHAN et al. (2011) correlacionaram diversos parâmetros clínico-anatomopatológicos (dados de acometimento de linfonodo, tamanho do tumor, grau de diferenciação, idade da paciente, índice mitótico e expressão de receptores hormonais) de pacientes submetidas a diferentes tipos de tratamentos. A análise global desses dados indicou que quanto maior o grau de acometimento linfonodal, menores os índices de sobrevida global e sobrevida livre de doença. Em nossas análises não foram detectadas diferenças significativas comparando o número total de SCNAs e seus subtipos entre os grupos Nx (linf+) e N0 (linf-). Porém a análise visual do perfil genômico dos grupos identificou alterações frequentes em cada um. O grupo Nx apresentou regiões de ganhos mais frequentes em 5p, 7p, 8q, 10p e no cromossomo 19 e perda frequente em 8p. Já o grupo de tumores N0 (linf(-)) exibiu SCNAs concentradas em outras regiões do genoma, perdas mais frequentes em 1p e 9p e ganho em 16p.

Todos os tumores portam mutações somáticas em seus genomas, mas apenas um *subset* de tais mutações, conhecidas como *driver mutations*, conferem vantagem seletiva clonal para a célula tumoral e estão implicadas de maneira causal na oncogênese (STRATTON et al. 2009). EM um trabalho de 2012, CURTIS et al identificaram novas *driver mutations* em genes como *MTAP*, *PPP2R2A* e *MAP2K4*, um deles também detectado como alterado em nossas amostras: o gene *MTAP* foi selecionado em nosso estudo para validação por *real time PCR* quantitativo, e sua deleção foi confirmada em 11 dos tumores de mama investigados por *array-CGH*.



Analisamos em nossas amostras os genes apontados nos trabalhos de CURTIS et al (2012) e STEPHENS et al.(2012) como *drivers* na carcinogênese de mama (ver Anexo 1); verificamos que todos os genes citados como afetados por *copy number alterations* também apresentaram alteração (perdas e ganhos discriminados na Tabela do Anexo 1) na nossa casuística. Os genes menos frequentemente alterados no nosso grupo de tumores sofreram alterações em apenas duas das amostras (*CASP8* e *SF3B1* - 2,8%). Já os genes que apresentaram SCNAs listados na Figura 2 foram também genes com maior frequência de alteração na nossa amostra: *MYC* (43%), *ERBB2* e *FGFR1/ZNF703* (37,5%), *CDKN2A* (26%), *MAP2K4* (21%), *CCND1* e *MDM2* (19%), *ZNF217* (18%), *RB1* e *BRCA2* (17%), *ARID1B* e *NCOR1* (14%), *NF1* (12,5%) e *PTEN* (6%). Essa alta frequência nos mesmos genes com alteração em número de cópias confirma que eles são fortes candidatos a genes *drivers* em tumores de mama.

Em teoria, as regiões com ganhos ou perdas de maior amplitude (portanto, presentes no genoma de um número maior de células daquele tumor e, algumas vezes, apresentando grande diferença no número de cópias em relação ao genoma) contêm os genes mais relevantes para a progressão tumoral, sendo mutações *drivers*. Dentre os ganhos rotineiramente detectados em tumores de mama, as amplificações em 8q24 (gene *MYC*), 17q12 (gene *ERBB2*), 11q13 (gene *CCND1*) e 8p12 (gene *FGFR1*) são prevalentes. Em nossa casuística, 39% dos tumores de mama apresentaram ganho em número de cópias do gene *MYC*, sendo 4% dos casos eventos do tipo *amplicons*. A amplificação em 8p, principalmente na

região 8p11-12, ocorre de maneira recorrente em 10 a 15% nos carcinomas de mama e parte desses *amplicons* não abrange o gene *FGFR1*, contendo potenciais oncogenes como *PPAPDC1B* (BERNARD-PIERROT et al. 2008). No nosso grupo de 71 tumores, 25 amostras (35%) apresentaram ganho da região que contém o gene *PPAPDC1B*, das quais 2 exibiram ganho considerado de alta amplitude (*amplicon*).

A região 11q13 é um segmento amplificado em 15-20% dos tumores de mama, um evento já relacionado à resistência à terapia endócrina (KARLSSON et al. 2011). Co-amplificações das regiões 11q13 e 8p12 são comuns, sugerindo que haja um evento sinérgico entre estes *amplicons*, e que esse evento colabore para um pior resultado ao tratamento do que a ocorrência isolada de uma dessas amplificações. Com base nessas informações, inspecionamos nossos dados quanto à ocorrência dessas amplificações, detectando que 19% dos tumores de mama estudados neste projeto apresentam ganho da região 11q13. Ganho de 8p12 ocorreu, como já mencionado, em 35% dos tumores, ao passo que a co-amplificação de 11q13 e 8p12 se mostrou presente em sete tumores (~10%). A maioria dos tumores exibindo co-amplificação em 8p12 e 11q13 (5/7) eram linfonodo-positivos ao diagnóstico, mas sem associação com o status de expressão de ER (4 eram ER(+) e 3 ER(-)). Tumores que possuem essa co-amplificação parecem ser raros entre tumores que expressam o gene *ERBB2* e teroricamente fazem parte de um subgrupo denominado “*Amplifier subtype*”, caracterizado por não expressar o gene *ERBB2* (JÖNSSON et al. 2010; KARLSSON et al. 2011); de fato, das 7 amostras que estudamos que

exibem essa co-amplificação, apenas uma delas expressa esse gene, mas todas exibem um perfil genômico marcado por ampliações diversas. Essas observações de co-amplificações mostram a complexidade e as consequências funcionais dos rearranjos genômicos que ocorrem nos amplicons do câncer de mama, incluindo “*transcriptional cross-talk*” entre diferentes genes, bem como sua cooperação com vias maiores da tumorigênese.

Na última etapa de nossa análise, selecionamos deleções em homozigose e ampliações detectadas em tumores linfonodo-positivos; tais rearranjos eram mapeados em regiões ainda não associadas fortemente na literatura ao câncer de mama, com o intuito de investigar nestes segmentos novos genes supressores tumorais ou oncogenes, potencialmente importantes para a progressão da doença.

Um dos genes escolhidos foi o gene *MTAP* (*methylthioadenosine phosphorylase*), que identificamos como deletado em algumas amostras. Perdas deste gene foram previamente detectadas em muitos tipos de tumorais como de endométrio, pâncreas e ductos biliares; em carcinomas hepatocelulares, a baixa expressão de *MTAP* está relacionada a um estágio tumoral avançado e maior tumorigenicidade (WONG et al. 1998; HUSTINX et al. 2005; KIROVSKI et al. 2011). Na finalização de nosso estudo, CURTIS et al. (2012) sugeriram este gene como afetado por *driver mutations* em tumores de mama. O gene *MTAP* codifica uma enzima que desempenha função importante no metabolismo de poliaminas, sendo relevante para a via de síntese de adenosina (HUSTINX et al. 2005; Gene Cards 2012). Estudos

já sugerem este gene como um potencial alvo terapêutico em tumores nos quais o mesmo está deletado, parcialmente ou em homozigose.

Outro gene que exibiu deleção em nosso estudo e selecionado para validação foi *REGL/RAS-like (Ras-related and estrogen-regulated growth inhibitor-like protein)*. Este gene possui atividade de GTPase e foi visto em associação a tumores luminais e com marcadores de bom prognóstico em carcinoma de mama como ER(+), citoqueratinas 19, 18 e 7/8e, e inibidores do ciclo celular p27 e p21. A alta expressão de *REG* também mostrou associação com maior sobrevida específica e intervalo livre de metástases a distância em câncer de mama (HABASHY et al. 2011).

O gene *HSD17B12 (Hydroxysteroid (17 $\beta$ ) dehydrogenase type 12)* codifica uma enzima multifuncional que atua na conversão de estrona para estradiol (E2) e no alongamento da cadeia longa de ácidos graxos, em particular na conversão do ácido palmítico para ácido araquidônico, o precursor de esteróides e o mediador inflamatório prostaglandina E (VISUS et al. 2011). Este mesmo trabalho de 2011 relata que a superexpressão de *HSD17B12* ocorre em carcinomas de mama associada à superexpressão do gene *COX-2*. Isso condiz com nossos resultados, pois nas 4 amostras em que detectamos ganho em número de cópia do gene *HSD17B12* (11p11.2), também há ganho de do gene *PTGS2 (COX-2)* (1q31.1). Nossos dados validaram o ganho deste gene, sugerindo que este ganho poderia estar associado a consequente superexpressão e importância na progressão dos carcinomas de mama.

O gene *ADAMTS3 (ADAM metalloproteinase with thrombospondin*

*type 1 motif*, 3) foi outro gene selecionado para validação e pertence a uma família de metaloproteinases (*ADAMTS*) que estão relacionadas a eventos celulares como: clivagem de proteoglicanos, inibição da angiogênese, desenvolvimento gonadal, organogênese e também no processo de síntese de pró-colágeno II. Alguns genes desta família estão implicados diretamente na progressão tumoral por causar a degradação da matrix extracelular e permitir a invasão tumoral em tecidos adjacentes (FERNANDES et al. 2001; PORTER et al. 2004; MOLOKWU et al. 2010). Além disso, alguns membros dessa família já foram descritos com um padrão de expressão alterado em tumores de mama. No caso de *ADAMTS3*, o trabalho de PORTER et al. (2004) mostrou uma redução da expressão desse gene em carcinomas de mama. Em nossas amostras, identificamos ganho em número de cópias deste gene, evento que foi validado por *qPCR*; o significado deste ganho é incerto.

Outro gene importante selecionado foi *KISS1*, devido à função de supressor metastático em melanoma e tumores de mama. A perda ou redução da expressão de *KISS1* foi associada com maior progressão tumoral e metástase (BECK e WELCH 2010). Curiosamente, em nossos dados detectamos ganhos deste gene em alguns tumores linfonodopositivos, sugerindo que este gene pode exibir funções adicionais ou mesmo não ser sensível à mudança de dose gênica. O gene *ZBTB20* foi um gene selecionado já descrito como um regulador negativo de  $\alpha$  feto-proteína e de outros genes específicos do fígado, relacionado ao metabolismo de glicose e à regulação do eixo hormonal somatotrópico. Também é descrito, no

entanto, como um fator de transcrição que pode estar envolvido na hematopoiese, oncogênese e respostas imunes (WANG Q et al. 2011; Gene Cards 2012). Identificamos perda em homozigose deste gene em no mínimo um dos tumores de mama, dados confirmados pela validação por *qPCR*.

Por último, testamos a deleção detectada do gene *ELF5*, já associado a perdas genômicas em tumores. Este gene pertence a uma família de fatores de transcrição conhecida como ETS, que estão associados à proliferação celular e tumorigênese. *ELF5* está localizado no cromossomo 11p terminal, região que frequentemente sofre perdas em carcinomas de mama, rins e próstata, além de ter sido detectada ausência de expressão do mesmo em linhagens celulares de carcinomas (ZHOU et al. 1998). A deleção deste gene foi confirmada nos tumores de mama nos quais este evento havia sido detectado na análise de arra-CGH, corroborando dados da literatura que mostram deleções de 11p em tumores de mama. Portanto, o gene *ELF5* pode ser um candidato a gene supressor tumoral para um *subset* de carcinomas de mama.

## 6 CONCLUSÕES

Do presente estudo de investigação de alterações no número de cópias de grandes segmentos de DNA no genoma tumoral de um grupo de carcinomas de mama invasivos do tipo ductal pode-se concluir:

- a. 50% das alterações genômicas identificadas foram rearranjos focais submicroscópicos (< 3 Mb). Os eventos de perda afetaram regiões de tamanho significativamente maior do que os eventos de ganho, enquanto alterações de alta amplitude (deleção em homozigose e *amplicons*) foram principalmente rearranjos < 1,8 Mb. Os *amplicons* foram 10 vezes mais frequentes do que as perdas em homozigose. Detectou um padrão de ganho frequente em 1q, 8q, 16p, e 17q12-q21, enquanto que perdas recorrentes foram mapeadas principalmente em 8p, 11q, 16q, e nos cromossomos 13 e 18.
  
- b. O número médio de SCNAs por genoma tumoral considerando os eventos de ganhos é maior no grupo de carcinomas estrógeno-negativos [ER(-)] em relação ao grupo estrógeno-positivo [ER(+)]. As SCNAs dos tumores ER(-) são significativamente maiores do que no grupo ER(+). A análise visual do perfil genômico evidenciou regiões de perdas em 4q e 5q e ganhos focais em 10p e 17q nos carcinomas

de mama ER(-), ao passo que o grupo de tumores ER(+) exibiu ganhos em 1q, 8q e 16p e perdas em 8p, 11q e 16q.

- c. Os tumores do grupo de 57 carcinomas de mama invasivos do tipo ductal foram classificados quanto à presença (linf(+)) ou não (linf(-)) de metástase linfonodal axilar ao diagnóstico. Investigamos o perfil genômico de *SCNAs* para a identificação de regiões associadas à presença de metástase linfonodal. Não foram detectadas diferenças significativas no número total de *SCNAs* e de seus diferentes subtipos entre os dois grupos. A análise visual identificou ganhos frequentes nos tumores (linf(+)) em 5p, 7p, 8q, 10p e no cromossomo 19 e perda frequente em 8p; o grupo de tumores (linf(-)) exibiu *SCNAs* de perdas mais frequentes em 1p e 9p e ganho recorrente em 16p.
- d. No grupo (linf(+)), alterações genômicas de grande amplitude (*amplicons* e deleções em homozigose) ainda não descritas na literatura foram validadas por técnica independente. Os genes contidos nestes segmentos, tais como *KISS1*, *ADAMTS3* e *HSD17B12* (detectados em *amplicons*) e *ZBTB20*, *MTAP*, *ELF5* e *RERGL* (afetados por deleção em homozigose), são candidatos a oncogenes e genes supressores tumorais relevantes para a progressão dos tumores de mama e/ou para o próprio processo de carcinogênese.



## 7 REFERÊNCIAS BIBLIOGRÁFICAS

Agilent Technologies. **Técnica de array-CGH.** Available from:<URL:http://www.home.agilent.com> [2012 jun 05].

Arnone P, Zurrída S, Viale G, et al. The TNM classification of breast cancer: need for change. **Updates Surg** 2010; 62:75-81.

Bartek J, Bartkova J, Lukas J. DNA damage signalling guards against activated oncogenes and tumour progression. **Oncogene** 2007; 26:7773-9.

Bayani J, Selvarajah S, Maire G, et al. Genomic mechanisms and measurement of structural and numerical instability in cancer cells. **Semin Cancer Biol** 2007; 17:5-18.

Beck BH, Welch DR. The KISS1 metastasis suppressor: a good night kiss for disseminated cancer cells. **Eur J Cancer** 2010; 46:1283-9.

Bergamaschi A, Kim YH, Wang P, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. **Genes Chromosomes Cancer** 2006; 45:1033-40.

Bernard-Pierrot I, Gruel N, Stransky N, et al. Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. **Cancer Res** 2008; 68:7165-75.

Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. **Nature** 2010; 463:899-905.

Birnbaum D, Eisinger F, Jacquemier J, Longy M, Sobol H, Theillet C. **Breast tumors: an overview: atlas genetic cytogenetic oncology haematology**. 1999. Available from:<URL:<http://atlasgeneticsoncology.org/Tumors/breastID5018.html>> [2012 maio 20]

Carcangiu ML, Casalini P, Ménard S. **Breast tumors: an overview: atlas genetic cytogenetic oncology haematology**. 2005. Available from:<URL:<http://atlasgeneticsoncology.org/Tumors/breastID5018.html>> [2012 maio 20].

Courjal F, Theillet C. Comparative genomic hybridization analysis of breast tumors with predetermined profiles of DNA amplification. **Cancer Res** 1997; 57:4368-77.

Cserni G, Sapino A, Decker T. Discriminating between micrometastases and isolated tumor cells in a regional and institutional setting. **Breast** 2006; 15:347-54.

Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. **Nature** 2012; 486:346-52.

de Mascarel I, MacGrogan G, Debled M, Brouste V, Mauriac L. Distinction between isolated tumor cells and micrometastases in breast cancer: is it reliable and useful? **Cancer** 2008; 112:1672-8.

Downing TE, Oktay MH, Fazzari MJ, Montagna C. Prognostic and predictive value of 16p12.1 and 16q22.1 copy number changes in human breast cancer. **Cancer Genet Cytogenet** 2010; 198:52-61.

Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. **Immunity** 2004; 21:137-48.

Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. **Histopathology**. 1991 Nov;19(5):403-10.

Fernandes RJ, Hirohata S, Engle JM, et al. Procollagen II amino propeptide processing by ADAMTS-3. Insights on dermatosparaxis. **J Biol Chem** 2001; 276:31502-9.

Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. **Nat Rev Cancer** 2004; 4:177-83.

Gebhart E. Comparative genomic hybridization (CGH): ten years of substantial progress in human solid tumor molecular cytogenetics. **Cytogenet Genome Res** 2004; 104:352-8.

Gene Cards. **GeneCards** **statistic** **page**. Available from:<URL:[http://www.genecards.org/stat\\_graphs.shtml#RNA](http://www.genecards.org/stat_graphs.shtml#RNA)>. [2012 maio 20]

Gilbert JA, Goetz MP, Reynolds CA, et al. Molecular analysis of metaplastic breast carcinoma: high EGFR copy number via aneusomy. **Mol Cancer Ther** 2008; 7:944-51.

Gobardhan PD, Elias SG, Madsen EV, et al. Prognostic value of lymph node micrometastases in breast cancer: a multicenter cohort study. **Ann Surg Oncol** 2011; 18:1657-64.

Habashy HO, Powe DG, Glaab E, et al. RERG (Ras-like, oestrogen-regulated, growth-inhibitor) expression in breast cancer: a marker of ER-positive luminal-like subtype. **Breast Cancer Res Treat** 2011; 128:315-26.

Hahn WC, Weinberg RA. Rules for making human tumor cells. **N Engl J Med** 2002; 347:1593-603.

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.

Hanahan D, Weinberg RA. The hallmarks of cancer. **Cell** 2000; 100:57-70.

Hu X, Stern HM, Ge L, et al. Genetic alterations and oncogenic pathways associated with breast cancer subtypes. **Mol Cancer Res** 2009; 7:511-22.

Hustinx SR, Hruban RH, Leoni LM, et al. Homozygous deletion of the MTAP gene in invasive adenocarcinoma of the pancreas and in periampullary cancer: a potential new target for therapy. **Cancer Biol Ther** 2005; 4:83-6.

Jones D. Pathways to cancer therapy. **Nat Rev Drug Discov** 2008; 7:875-6.

Jönsson G, Staaf J, Vallon-Christersson J, et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. **Breast Cancer Res** 2010; 12:R42.

Kahn HJ, Hanna WM, Chapman JA, et al. Biological significance of occult micrometastases in histologically negative axillary lymph nodes in breast cancer patients using the recent American Joint Committee on Cancer breast cancer staging system. **Breast J** 2006; 12:294-301.

Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. **Science** 1992; 258:818-21.

Karlsson E, Waltersson MA, Bostner J, et al. High-resolution genomic analysis of the 11q13 amplicon in breast cancers identifies synergy with 8p12 amplification, involving the mTOR targets S6K2 and 4EBP1. **Genes Chromosomes Cancer** 2011; 50:775-87.

Kirovski G, Stevens AP, Czech B, et al. Down-regulation of methylthioadenosine phosphorylase (MTAP) induces progression of hepatocellular carcinoma via accumulation of 5'-deoxy-5'-methylthioadenosine (MTA). **Am J Pathol** 2011; 178:1145-52.

Kwei KA, Kung Y, Salari K, Holcomb IN, Pollack JR. Genomic instability in breast cancer: pathogenesis and clinical implications. **Mol Oncol** 2010; 4:255-66.

Leary RJ, Lin JC, Cummins J, et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. **Proc Natl Acad Sci U S A** 2008; 105:16224-9.

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

Lobo I. Chromosome abnormalities and cancer cytogenetics. **NatureEduc** 2008;1(1). [periodico on-line]. Available from: <URL:<http://www.nature.com/scitable/topicpage/chromosome-abnormalities-and-cancer-cytogenetics-879>>. [2012 maio 20]

Loo LW, Grove DI, Williams EM, et al. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. **Cancer Res** 2004; 64:8541-9.

Melchor L, Honrado E, Huang J, et al. Estrogen receptor status could modulate the genomic pattern in familial and sporadic breast cancer. **Clin Cancer Res** 2007; 13:7305-13.

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

Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. **Nat Rev Cancer** 2007; 7:233-45.

Molokwu CN, Adeniji OO, Chandrasekharan S, Hamdy FC, Buttle DJ. Androgen regulates ADAMTS15 gene expression in prostate cancer cells. **Cancer Invest** 2010; 28:698-710.

Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. **Nat Genet** 2003; 34:267-73.

Mullighan CG, Downing JR. Global genomic characterization of acute lymphoblastic leukemia. **Semin Hematol** 2009; 46:3-15.

[NCCN] National Comprehensive Cancer Network. **NCCN Guidelines for patients: breast cancer**. Available from: <URL:<http://www.nccn.com/files/cancer-guidelines/breast/index.html#/2/>> [2012 maio 05].

Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability--an evolving hallmark of cancer. **Nat Rev Mol Cell Biol** 2010; 11:220-8.

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

Porter S, Scott SD, Sassoan EM, et al. Dysregulated expression of adamalysin-thrombospondin genes in human breast carcinoma. **Clin Cancer Res** 2004; 10:2429-40.

Pusztai L, Broglio K, Andre F, Symmans WF, Hess KR, Hortobagyi GN. Effect of molecular disease subsets on disease-free survival in randomized adjuvant chemotherapy trials for estrogen receptor-positive breast cancer. **J Clin Oncol** 2008; 26:4679-83.

Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. **Lancet** 2011; 378:1812-23.

Rodriguez-Revena L, Mila M, Rosenberg C, Lamb A, Lee C. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. **Genet Med** 2007; 9:600-6.

Sobin LH, Gospodarowicz MK, Wittekind C. **TNM classification of malignant tumor**. 7<sup>th</sup> ed. Oxford: Wiley-Blackwell; 2009. Breast tumors; p.181-93.

Staaf J, Jönsson G, Ringnér M, Baldetorp B, Borg A. Landscape of somatic allelic imbalances and copy number alterations in HER2-amplified breast cancer. **Breast Cancer Res** 2011; 13:R129.

Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. **Nature** 2009; 462:1005-10.

Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. **Nature** 2012; 486:400-4.

Stratford AL, Habibi G, Astanehe A, et al. Epidermal growth factor receptor (EGFR) is transcriptionally induced by the Y-box binding protein-1 (YB-1) and can be inhibited with Iressa in basal-like breast cancer, providing a potential target for therapy. **Breast Cancer Res** 2007; 9:R61.

Stratton MR, Campbell PJ, Futreal PA. The cancer genome. **Nature** 2009; 458:719-24.

Strausberg RL, Simpson AJ. Whole-genome cancer analysis as an approach to deeper understanding of tumour biology. **Br J Cancer** 2010; 102:243-8.

Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. **Proc Natl Acad Sci U S A** 2005; 102:15545-50.

Sørliie 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.

Visus C, Ito D, Dhir R, et al. Identification of Hydroxysteroid (17 $\beta$ ) dehydrogenase type 12 (HSD17B12) as a CD8<sup>+</sup> T-cell-defined human tumor antigen of human carcinomas. **Cancer Immunol Immunother** 2011; 60:919-29.

Wang DY, Done SJ, McCready DR, Boerner S, Kulkarni S, Leong WL. A new gene expression signature, the ClinicoMolecular Triad Classification, may improve prediction and prognostication of breast cancer at the time of diagnosis. **Breast Cancer Res** 2011; 13:R92.

Wang Q, Tan YX, Ren YB, et al. Zinc finger protein ZBTB20 expression is increased in hepatocellular carcinoma and associated with poor prognosis. **BMC Cancer** 2011; 11:271.



Weigelt B, Bissell MJ. Unraveling the microenvironmental influences on the normal mammary gland and breast cancer. **Semin Cancer Biol** 2008; 18:311-21.

Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. **Nature** 2007; 450:893-8.

Wong YF, Chung TK, Cheung TH, Nobori T, Chang AM. MTAP gene deletion in endometrial cancer. **Gynecol Obstet Invest** 1998; 45:272-6.

Yang ZQ, Liu G, Bollig-Fischer A, Giroux CN, Ethier SP. Transforming properties of 8p11-12 amplified genes in human breast cancer. **Cancer Res** 2010; 70:8487-97.

Yu W, Kanaan Y, Bae YK, Baed YK, Gabrielson E. Chromosomal changes in aggressive breast cancers with basal-like features. **Cancer Genet Cytogenet** 2009; 193:29-37.

Zhou J, Ng AY, Tymms MJ, et al. A novel transcription factor, ELF5, belongs to the ELF subfamily of ETS genes and maps to human chromosome 11p13-15, a region subject to LOH and rearrangement in human carcinoma cell lines. **Oncogene** 1998; 17:2719-32.

**Anexo 1** - Frequência no grupo de tumores estudados de SCNAs contendo os genes citados no trabalho de Stephens et al. (2012) afetados por alterações de número de cópias.

<b>Gene</b>	<b>Ganho</b>	<b>Amplicons</b>	<b>Perda</b>	<b>Perda em Homozigose</b>	<b>Nº de Tumores acometidos (%)</b>
<i>MYC</i>	27 tumores	3 tumores	1 tumor	0	31 (43%)
<i>ERBB2</i>	7 tumores	17 tumores	3 tumores	0	27 (37,5%)
<i>FGFR1/ZNF703</i>	22 tumores	1 tumor	4 tumores	0	27 (37,5%)
<i>PPP2R2A*</i>	2 tumores	0	24 tumores	0	26 (36,1%)
<i>MTAP*</i>	5 tumores	0	14 tumores	1 tumor	20 (27,8%)
<i>CDKN2A</i>	4 tumores	0	15 tumores	0	19 (26,4%)
<i>MAP2K4**</i>	1 tumor	0	14 tumores	0	15 (20,8%)
<i>CCND1</i>	8 tumores	5 tumores	1 tumor	0	14 (19,4%)
<i>MDM2</i>	8 tumores	3 tumores	3 tumores	0	14 (19,4%)
<i>ZNF217</i>	13 tumores	0	0	0	13 (18%)
<i>RB1</i>	1 tumor	0	11 tumores	0	12 (16,7%)
<i>ASXL1</i>	12 tumores	0	0	0	12 (16,7%)
<i>BRCA2</i>	4 tumores	1 tumor	7 tumores	0	12 (16,7%)
<i>SMAD4</i>	1 tumor	0	11 tumores	0	12 (16,7%)
<i>PIK3CA</i>	7 tumores	1 tumor	3 tumores	0	11 (15,3%)
<i>MAP3K1</i>	4 tumores	0	7 tumores	0	11 (15,3%)
<i>ARID1B</i>	0	0	10 tumores	0	10 (13,9%)
<i>CDH1</i>	4 tumores	0	6 tumores	0	10 (13,9%)
<i>NCOR1</i>	0	0	10 tumores	0	10 (13,9%)
<i>TP53</i>	1 tumor	0	8 tumores	0	9 (12,5%)
<i>GATA3</i>	9 tumores	0	0	0	9 (12,5%)
<i>MLL3</i>	7 tumores	0	2 tumores	0	9 (12,5%)
<i>NF1</i>	8 tumores	1 tumor	0	0	9 (12,5%)
<i>MAP3K13</i>	8 tumores	0	1 tumor	0	9 (12,5%)
<i>APC</i>	1 tumor	0	8 tumores	0	9 (12,5%)
<i>ARID2</i>	4 tumores	0	4 tumores	0	8 (11,1%)
<i>KRAS</i>	4 tumores	1 tumor	3 tumores	0	8 (11,1%)

**Cont./Anexo 1**

---

<i>SMARCD1</i>	8 tumores	0	0	0	8 (11,1%)
<i>BRCA1</i>	5 tumores	0	2 tumores	0	7 (9,7%)
<i>MLL2</i>	7 tumores	0	0	0	7 (9,7%)
<i>AKT2</i>	6 tumores	0	0	0	6 (6,9%)
<i>CDKN1B</i>	2 tumores	1 tumor	3 tumores	0	6 (6,9%)
<i>STK11</i>	6 tumores	0	0	0	6 (6,9%)
<i>AKT1</i>	4 tumores	0	1 tumor	0	5 (7%)
<i>TBX3</i>	2 tumores	0	3 tumores	0	5 (7%)
<i>SETD2</i>	1 tumor	0	4 tumores	0	5 (7%)
<i>PTEN</i>	0	0	3 tumores	1 tumor	4 (5,6%)
<i>ARID1A</i>	1 tumor	0	2 tumores	0	3 (4,2%)
<i>BAP1</i>	0	0	3 tumores	0	3 (4,2%)
<i>SF3B1</i>	1 tumor	0	1 tumor	0	2 (2,8%)
<i>CASP8</i>	2 tumores	0	0	0	2 (2,8%)

---

\* *PPP2R2A* e *MTAP* são genes citados como *drivers* no estudo de Curtis et al. 2012.

\*\* *MAP2K4* é um gene citado em ambos (Stephens et al. 2012; Curtis et al. 2012).