**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO**

Marcelo Guimarães Tiezzi

**Análise da expressão de RNAs longos não codificantes no
câncer de mama localmente avançado**

**Ribeirão Preto
2019**

Marcelo Guimarães Tiezzi

**Análise da expressão de RNAs longos não codificantes no
câncer de mama localmente avançado**

Tese apresentada à Faculdade de Medicina de
Ribeirão Preto da Universidade de São Paulo
para obtenção do Título de Doutor em Ciências
Médicas.

Área de concentração: Biologia da reprodução.

Orientador: Prof. Dr. Hélio Humberto Angotti Carrara

Ribeirão Preto

2019

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Catálogo na Publicação

Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

Tiezzi, Marcelo Guimarães

Análise da expressão de RNAs longos não codificantes no câncer de mama localmente avançado / Marcelo Guimarães Tiezzi; orientador Helio Humberto Angotti Carrara – Ribeirão Preto, 2019.

51p.; il.; 30 cm.

Tese (Doutorado – Programa de Pós Graduação em Ciências Médicas. Área de Concentração: Biologia da Reprodução) Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

1. Câncer de mama, 2. *lncRNAs*, 3. Quimioterapia neoadjuvante, 4. Fatores prognósticos.

FOLHA DE APROVAÇÃO

Marcelo Guimarães Tiezzi

Análise da expressão de RNAs longos não codificantes no câncer de mama localmente avançado.

Tese apresentada à Faculdade de Medicina de
Ribeirão Preto da Universidade de São Paulo
para obtenção do Título de Doutor em Ciências
Médicas.
Área de concentração: Biologia da reprodução.

Aprovado em:

Banca Examinadora

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

DEDICATÓRIA

Dedico este trabalho aos meus pais e minha esposa e meus filhos, com amor, admiração e gratidão por seu apoio, carinho e presença ao longo do período de elaboração deste trabalho.

E em especial ao meu Tio, Professor e amigo, Prof. Dr. Edgar Guariento Guimarães (in memoriam).

AGRADECIMENTOS

Ao Prof. Dr. Hélio Humberto Angotti Carrara, meu orientador, pela ajuda e a compreensão.

Ao Prof. Dr. Daniel Guimarães Tiezzi, meu irmão, amigo e maior incentivador, pela paciência e ajuda na análise genômica.

Aos Drs. Dauto de Almeida Campos e Antonio Plácido Pereira pelo constante apoio e estímulo.

Ao Dr. Anderson da Costa Lino Costa, meu mestre e amigo, pelo ensino da Patologia.

Aos meus Pais, Roberto e Lígia pelo apoio incondicional.

À minha esposa e meus filhos, Maria Paola, Pietra e Lorenzo, pelo carinho e apoio.

À Suelen Soares Bezerra, Secretária do PPG em Ginecologia e Obstetrícia – FMRP/USP pela ajuda e orientação durante todo este trabalho.

À Faculdade de Medicina de Ribeirão Preto/USP e ao Departamento de Ginecologia e Obstetrícia pela oportunidade.

“Quem conhece a sua ignorância revela a mais profunda sapiência. Quem ignora a sua ignorância vive na mais profunda ilusão”

LaoTsé.

RESUMO

TIEZZI, M.G. **Análise da expressão de RNAs longos não codificantes no câncer de mama localmente avançado.** 2019. Tese (Doutorado) – Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto, 2019.

O carcinoma de mama é uma doença altamente prevalente e incidente. Em nosso meio cerca de metade dos casos são diagnosticados em estádios localmente avançados e/ou disseminados. Nesta situação o índice de sucessos terapêuticos é pequeno. A quimioterapia tem sido empregada como tratamento inicial para estas pacientes. Sabemos que o maior benefício do tratamento citotóxico é atingido em pacientes com resposta patológica completa ao tratamento sistêmico neoadjuvante em pacientes sem doença sistêmica. Este grupo de pacientes representa cerca de 20% dos casos. Identificar previamente um grupo de pacientes que mais se beneficiariam da terapia neoadjuvante pode reduzir custos de tratamento bem como evitar efeitos adversos no grupo de pacientes com baixa probabilidade de resposta. No entanto, ainda não é possível fazer tal predição, mesmo com a utilização de técnicas avançadas de biologia molecular. O uso da biologia molecular tem avançado rapidamente e recentemente vem sendo citado na literatura as ações dos RNAs não codificantes nas células neoplásicas e na progressão tumoral. Essas moléculas atuam em vários mecanismos celulares e estão relacionados com a ocorrência de doença metastática no câncer de mama. Identificar e caracterizar diferentes RNAs longos não codificantes (lncRNAs) que atuem na progressão tumoral do câncer de mama em pacientes com respostas distintas à terapia neoadjuvante, permitiria criar um meio de selecionar melhor as pacientes submetidas a este tipo de tratamento. Este estudo teve por objetivo avaliar a expressão de lncRNAs por hibridização genômica em microarranjos de cDNA, utilizando o chip Human Gene 2.0 ST (Affymetrix®) em pacientes com carcinoma mamário localmente avançado e metastático submetidas à quimioterapia primária e procurar identificar uma “assinatura genética” de tumores por meio da análise de expressão de lncRNAs capaz de identificar previamente grupos de pacientes que se beneficiariam com esta terapia. Encontramos 1003 transcritos diferencialmente expressos, entre os quais 366 ligados a um gene conhecido e três lncRNAs (*PVT1*, *GAS5* e *Linc00657*). Identificamos esses lncRNAs no banco de dados do TCGA e os mesmos se mostraram diferentemente expressos nas pacientes com resposta completa ao tratamento quimioterápico em relação àquelas sem resposta completa.

Palavras chave: Câncer de mama, *lncRNAs*, Quimioterapia neoadjuvante, Fatores prognósticos.

ABSTRACT

TIEZZI, M.G. - **Analysis of the expression of long non-coding RNAs in locally advanced breast cancer**. 2019. Tese (Doutorado) – Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto, 2019.

Breast carcinoma is a highly prevalent and incident disease. In our country about half of the cases are diagnosed in locally advanced and disseminated stages. In this situation the rate of therapeutic success is small. Chemotherapy has been used as the initial treatment for these patients. We know that the greatest benefit of cytotoxic treatment is achieved in patients with complete pathological response to neoadjuvant systemic treatment in patients without systemic disease. This group of patients represents about 20% of the cases. Previously identifying a group of patients who would most benefit from neoadjuvant therapy may reduce treatment costs as well as avoid adverse effects in the group of patients with a low probability of response. However, it is not yet possible to make such a prediction, even with the use of advanced molecular biology techniques. The use of molecular biology has advanced rapidly and recently the literature has been cited the actions of non-coding RNAs in neoplastic cells and tumor progression. These molecules act on several cellular mechanisms and are related to the occurrence of metastatic disease in breast cancer. Identifying and characterizing different non-coding long RNAs (lncRNAs) that act on the tumor progression of breast cancer in patients with distinct responses to neoadjuvant therapy, would create a way to better select the patients submitted to this type of treatment. This study aimed to evaluate the expression of lncRNAs by genomic hybridization in cDNA microarrays using the Human Gene 2.0 ST (Affymetrix®) chip in patients with locally advanced and metastatic mammary carcinoma submitted to primary chemotherapy and try to identify a "genetic signature" of tumors by the analysis of lncRNA expression capable of previously identifying groups of patients who would benefit from this therapy. We found 1003 differentially expressed transcripts, among them 366 linked to a known gene and three lncRNAs (*PVT1*, *GAS5* and *Linc00657*). We identified these lncRNAs in the database of the TCGA and they were shown differently expressed in the patients with complete response to the chemotherapy treatment in relation to those without a complete response.

Key words: Breast cancer, lncRNAs, neoadjuvant chemotherapy, prognostic factors.

ÍNDICE DE FIGURAS E TABELAS

Figura 1. Colorações imunoistoquímicas. A) HER2 positivo(3+); B) Ki67 positivo em cerca de 25% das células; C) RE positivo forte em 95% das células; D) RP positivo moderado em 90% das células.	23
Figura 2. CISH para pesquisa de amplificação do <i>ERBB2</i> , marcado em preto e em vermelho o cromossomo 17.	23
Figura 3. O material obtido das pacientes foi submetido a exame imunoistoquímico (IMH) e classificação molecular pelo PAM50. Este heatmap demonstra a clusterização não supervisionada baseado na expressão dos 50 genes utilizados pelo modelo para predizer a classificação molecular.	26
Figura 4. Os genes à esquerda estão mais expressos nas pacientes que não responderam ao tratamento neoadjuvante e os em verde são aqueles onde o valor do $p < 0.05$ e o log da variação da expressão é $> 0,7$. Nota-se que centenas de genes estão diferentemente expressos. Entre eles estão três LncRNAs.	27
Figura 5. A expressão de <i>PVT1</i> está aumentada nos carcinomas em relação ao parênquima mamário não tumoral.	28
Figura 6. O <i>PVT1</i> está superexpresso nos carcinomas RE positivos.	29
Figura 7. O <i>PVT1</i> também está superexpresso nos carcinomas RP positivos.	30
Figura 8. O <i>PVT-1</i> também está mais expresso nos carcinomas subtipo IC9.	30
Figura 9. A expressão do <i>PVT1</i> tem correlação significativa como número de cópias gênicas ($p < 0,0001$).	31
Tabela 1. Alguns exemplos de lncRNAs associados ao desenvolvimento de tumores.	14
Tabela 2. As pacientes incluídas tinham média de idade de 50 anos e tumores de grau e estadios avançados.	25

SUMÁRIO

1 INTRODUÇÃO	11
2 OBJETIVOS	17
2.1 Objetivo geral	18
2.1 Objetivos específicos	18
3 MATERIAIS E MÉTODOS	19
3.1. Seleção de pacientes	20
3.2. Técnica para obtenção das amostras, extração de RNA e qualidade da amostra	20
3.3. Tratamento neoadjuvante.....	20
3.4. Resposta ao tratamento	21
3.5. Expressão de receptores hormonais e HER2	21
3.6. Análise da expressão gênica global	21
3.7. Classificação molecular de acordo com o PAM50 e integrative cluster (IC).....	22
3.8. Análise de bioinformática nas amostras de câncer de mama do TCGA	22
4 RESULTADOS.....	24
5 DISCUSSÃO	32
6 CONCLUSÃO	37
7 REFERÊNCIAS BIBLIOGRÁFICAS	39
8 ANEXOS	44
Anexo 1. Graduação histológica dos carcinomas mamários:	45
Anexo 2. Sistema TNM	46
Anexo 3. Comparação do número de cópias dos genes MYC e PVT1 no TCGA.	49
Anexo 4. Sobrevida global conforme subtipos segundo a classificação pelo PAM50.	50
Anexo 5. Sobrevida global conforme subtipos segundo a classificação pelo Integrative clustering.	51

1 INTRODUÇÃO

O câncer de mama é a doença maligna mais frequente na mulher em todo o mundo. Estima-se que em 2019 ocorrerão 268.600 novos casos nos EUA e 41760 óbitos decorrentes da doença (“U.S. Breast Cancer Statistics”,). A disseminação metastática é o principal fator relacionado com a morte pela doença (JEMAL et al., 2009). Características histológicas, expressão de proteínas e perfil de expressão gênica intrínsecos às células neoplásicas permitem a caracterização de diferentes subtipos de carcinomas de mama (CALZA et al., 2006) (SØRLIE et al., 2001) (SØRLIE et al., 2003) (SØRLIE et al., 2006). Embora estes marcadores biológicos tenham sido descritos como fatores de prognóstico, os mecanismos pelos quais tumores semelhantes apresentam comportamento biológico distinto ainda não foram elucidados.

Novos avanços tecnológicos em biologia molecular e celular têm evidenciado mecanismos regulatórios da expressão gênica que eram desconhecidos até alguns anos atrás e ainda foram pouco explorados. Com o desenvolvimento e redução do custo dos sequenciadores de nova geração e as novas tecnologias de microarranjos de cDNA, evidências recentes de que rearranjos gênicos pós-transcricionais denominados splicing alternativo, e a transcrição diferencial de fragmentos intragênicos não codificantes (lincRNAs, do inglês longintragenic non-codingRNAs) estejam intrinsecamente relacionados com a progressão de doenças malignas vêm vindo à tona (KUNG; COLOGNORI; LEE, 2013). Algumas evidências recentes sugerem que estes mecanismos possam ser importantes para a manutenção do fenótipo maligno e resistência às drogas citotóxicas (CHEETHAM et al., 2013).

A quimioterapia neoadjuvante tem sido empregada como tratamento padrão para o câncer de mama localmente avançado (ANDRE et al., 2008) (OSAKO et al., 2010). Foi demonstrado que pacientes com resposta objetiva ao tratamento sistêmico, em especial com resposta patológica completa, apresentam taxas de sobrevida livre de doença e sobrevida global maiores do que o grupo de pacientes sem resposta ao tratamento neoadjuvante (LI et al., 2008) (MITRA; MITRA; TRICHE, 2012). Estudos recentes demonstraram que pacientes com tumores indiferenciados ou com expressão negativa de receptores hormonais (tumores triplo negativos – TNBC) e de proteína Her2/neu ou aqueles com amplificação do gene *ERBB2* (HER2 positivos) apresentam maiores taxas de resposta patológica completa ao tratamento neoadjuvante (GUPTA et al., 2010) (TAMINAU et al., 2012). Esta observação é paradoxal, visto que, pacientes portadoras de neoplasias dos subtipos triplo

negativos (TNBC) e HER2 positivas apresentam pior prognóstico quando comparadas com pacientes com tumores com expressão positiva de receptores hormonais. Assim, o fator determinante da sensibilidade à quimioterapia ainda é pouco compreendido.

Investigações utilizando ferramentas de análise genética em larga escala, em especial os microarranjos de cDNA, levaram a identificação de “assinaturas genéticas” capazes de melhor prever a resposta ao tratamento neoadjuvante (SØRLIE et al., 2003) (PHIPSON et al., 2016) (GENDOO et al., 2016) (LIU et al., 2016) (TAMINAU et al., 2012). Esta metodologia é, de certa forma, eficaz em identificar grupos de pacientes de risco, bem como, pode ser utilizada para selecionar pacientes de bom prognóstico que não necessitariam de tratamento sistêmico adjuvante. Atualmente, existem testes baseados em plataformas de microarranjos de cDNA comercialmente disponíveis que auxiliam o clínico e a paciente na decisão da indicação de quimioterapia adjuvante (VAN DE VIJVER et al., 2002) (ROSS et al., 2008). A utilização desta tecnologia é uma opção para o estudo em larga escala da expressão de lncRNAs.

De acordo com o tamanho, os RNA não codificantes são divididos em dois grupos, RNAs não codificantes pequenos com menos de 200 nucleotídeos e os RNAs não codificantes longos, que são o objeto de estudo deste trabalho. Nos últimos anos os RNAs pequenos foram muito estudados e mostraram algumas ações, principalmente os microRNAs na patogênese do câncer (LU et al., 2005) (GARTEL; KANDEL, 2008). Entretanto, está cada vez mais claro que o genoma dos mamíferos codifica numerosos lncRNAs longos, definidos como RNAs celulares endógenos de mais de 200 nucleotídeos de comprimento, que não possuem uma estrutura de leitura aberta de comprimento significativo (menos que 100 aminoácidos), que são um grupo muito heterogêneo de moléculas de RNA, com um largo espectro de funções moleculares e celulares através de diferentes ações (GUTSCHNER; DIEDERICHS, 2012). Muitos lncRNAs ainda têm sua função desconhecida, entretanto já existem evidências de lncRNAs associados com a patogênese do câncer e outras doenças (Tabela 1) (ZHANG et al., 2013).

Tabela 1. Alguns exemplos de lncRNAs associados ao desenvolvimento de tumores.

LncRNA	Origem da neoplasia	Expressão	Função
<i>MVIH</i>	Fígado	Aumentada	Biomarcador
<i>DD3 (PCA3)</i>	Próstata	Aumentada	Biomarcador
<i>HOTAIR</i>	Mama, fígado, colorretal, estroma GI	Aumentada	Biomarcador e oncogênico
<i>MALAT1</i>	Pulmão, fígado, rim. Mama, estroma endometrial, cervical, colorretal, ves. biliar e osteosarcoma	Aumentada	Biomarcador e oncogênico
<i>PCAT-1</i>	Próstata	Aumentada	Oncogênico
<i>H19</i>	Próstata, fígado e ves. biliar	Diminuída	Biomarcador
<i>PCGEM1</i>	Próstata	Aumentada	Oncogênico
<i>TUC338</i>	Fígado	Aumentada	Oncogênico
<i>BANCR</i>	Melanoma	Aumentada	Oncogênico
<i>YIYA</i>	Fígado, ovário, mama e esôfago	Aumentada	Oncogênico
<i>CRNDE</i>	Colorretal, fígado, pâncreas, próstata, ovário, gliomas, leucemias	Aumentada	Oncogênico
<i>CCAT1</i>	Estômago	Aumentada	Oncogênico
<i>HULC</i>	Fígado	Aumentada	Oncogênico
<i>CUDR</i>	Ves. Biliar	Aumentada	Oncogênico
<i>GAS5</i>	Mama	Diminuída	Supressor de tumor
<i>Linc-p21</i>	Pulmão, sarcoma e linfoma	Diminuída	Supressor de tumor
<i>MEG3</i>	Meningeoma, glioma, fígado, leukemia	Diminuída	Supressor de tumor
<i>PTENP1</i>	Próstata e colon	Diminuída	Supressor de tumor
<i>PTCSC3</i>	Tireóide	Diminuída	Supressor de tumor

Os lncRNAs podem atuar em vários mecanismos biológicos e sua importância em muitos aspectos da diferenciação celular e homeostase está bem esclarecida. Muitos destes mecanismos estão envolvidos na progressão do câncer e envolvem processos moleculares responsáveis pela remodelação da cromatina, repressão ou coativação transcricional, inibição proteica, alterações pós transcricionais, e regulação transcricional por competição com miRNAs (um mecanismo denominado de efeito chamariz ou decoy, do inglês). Evidências recentes demonstram que um dos papéis dos lncRNAs é direcionar o local específico para que os complexos

modificadores da cromatina possam causar efeitos epigenéticos. Pelo menos 38% dos lncRNAs presentes em tecidos humanos estão vinculados à um complexo denominado PRC2 (do inglês polycomb repressive complex 2) ou à proteínas modificadoras da cromatina como a CoREST e SMCX (KUNG; COLOGNORI; LEE, 2013) (LI et al., 2008). Desta forma, os lncRNAs desempenham um papel biológico importante na regulação da expressão gênica e podem estar envolvidos no processo de progressão da doença maligna (CHEETHAM et al., 2013). Na verdade, uma série de lncRNAs já foram identificados como moléculas relacionados à progressão de neoplasias malignas da próstata, mama, ovários, cérvix e intestino grosso e já são vistos como potenciais biomarcadores para diagnóstico precoce e prognóstico da doença (CHEETHAM et al., 2013).

O lncRNA nomeado *HOTAIR* é um gene localizado no cromossomo 12 e é o primeiro exemplo de que um RNA expresso em um cromossomo possa influenciar a transcrição de um outro cromossomo. O *HOTAIR* está altamente expresso no câncer de mama, estômago, intestino grosso, nasofaringe e colo uterino, sendo que a sua superexpressão está altamente relacionada com um maior risco de metástase e morte pela doença. O *HOTAIR* está envolvido na carcinogênese e afeta a mobilidade, proliferação, apoptose, invasão, agressão e metástases celulares. Há ainda indícios de que este lncRNA induza a resistência das células cancerosas à cisplatina (GUPTA et al., 2010) (HAJJARI; SALAVATY, 2015).

Prensner et al, em recente estudo, identificou 121 lncRNAs diferentemente expressos no câncer prostático e os diferentes padrões de expressão distinguem entre doença benigna, câncer localizado e metastático. Além disso, eles caracterizaram um lncRNA, o *PCAT-1* (prostate cancer associated transcript 1), que está altamente expresso nos cânceres prostáticos localizados de alto grau e nos metastáticos. Estudos de recentes indicam que o *PCAT-1* está envolvido no processo de reparo DNA e no desenvolvimento do carcinoma prostático inibindo o mecanismo do *BRCA2* (PRENSNER et al., 2014) (GUTSCHNER; DIEDERICH, 2012).

Ji e Diederichs et al, identificaram o *MALAT1* (metastasis-associated lung adenocarcinoma transcript-1) posteriormente chamado de *NEAT2* (nuclear- enriched abundant transcript 2) como um marcador prognóstico de metástases e sobrevida em pacientes com câncer de pulmão tipo não pequenas células. O *MALAT1* é um RNA não codificante extremamente abundante em muitas células humanas e

altamente conservado entre várias espécies, ressaltando a sua importância. Ele está relacionado a fatores fisiológicos, como o splicing alternativo, organização nuclear e modulação epigenética da expressão gênica, além de processos patológicos, que envolvem a diabetes e alguns tipos de cânceres. Ele regula o splicing alternativo dos pré mRNAs controlando os níveis de fatores ativos de splicing serina/arginina. A depleção do *MALAT1* altera o processo de um subproduto de um pré mRNA que é importante na biologia do câncer e regula a expressão de genes associados a metástases. Altos níveis de *MALAT1* são encontrados em pulmões e pâncreas saudáveis, e em muitos cânceres humanos, incluindo, o pulmonar, o sarcoma do estroma endometrial, o cervical e o hepatocelular (GUTSCHNER; DIEDERICHS, 2012).

Os lncRNAs recém descobertos são cada vez mais reconhecidos como moléculas ativas, em vez de “ruído transcricional” e acumulam-se evidências de que eles têm papéis críticos na carcinogênese por influenciar a proliferação de células tumorais (GUTSCHNER; DIEDERICHS, 2012). Futuros estudos para elucidar os mecanismos genéticos envolvidos na carcinogênese terão de incluir lncRNAs nas análises para conseguir uma visão completa sobre ativadores e inibidores destes mecanismos.

2 OBJETIVOS

2.1 Objetivo geral

Este estudo tem por objetivo avaliar a expressão diferencial de lncRNAs por hibridização genômica em microarranjos de cDNA em pacientes com carcinoma mamário localmente avançado e metastático submetidos a quimioterapia primária.

2.1 Objetivos específicos

- Identificar lncRNAs diferentemente expressos entre carcinomas de mama com resposta patológica completa à quimioterapia primária quando comparados com tumores sem resposta ao tratamento;
- Analisar a influência da expressão de lncRNAs na expressão de outros genes sabidamente envolvidos na resistência à droga e progressão do câncer de mama.
- Caracterizar a expressão dos lncRNAs relacionados com a resistência ao tratamento quimioterápico em amostras de tumores disponíveis em bancos de dados públicos

3 MATERIAIS E MÉTODOS

3.1. Seleção de pacientes

Foram incluídas prospectivamente todas as pacientes com diagnóstico histológico de carcinoma invasivo da mama que foram submetidas a tratamento primário com quimioterapia no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto – USP no período de janeiro de 2010 a dezembro de 2012. Todas as pacientes foram informadas sobre os objetivos da pesquisa antes da inclusão. Um termo de consentimento livre e esclarecido foi aplicado no momento da visita da paciente ao ambulatório de Mastologia. O estudo foi aprovado pelo Comitê de Ética em Pesquisa do HCFMRP-USP – protocolo 2467/09.

3.2. Técnica para obtenção das amostras, extração de RNA e qualidade da amostra

Um total de 38 amostras de tumores foram obtidas por biópsia percutânea com agulhas de 12G guiada por ultrassonografia. Foram colhidas 2 amostras: uma colocada em formol a 10% e posteriormente processada em parafina para confirmação do diagnóstico histológico, gradação (arquitetural e nuclear) e imunoistoquímica e a outra amostra foi congelada em nitrogênio (*snapfreezing*) e foi utilizada para a análise da expressão gênica.

O RNA foi isolado e purificado de amostras de tumores à fresco, congeladas em nitrogênio líquido e preservadas em freezer -80°C. Utilizamos o kit comercial Mirvana Kit (Ambion, USA) seguindo as orientações do fabricante. A quantificação e a integridade do RNA total foram analisadas por eletroforese capilar com o equipamento 2100 Bioanalyzer da Agilent e utilizamos as amostras com RIN ≥ 6 , e a pureza foi analisada por espectrofotometria com o equipamento Thermo Scientific Nono Drop 1000 e somente amostras com relação 280/260 $>1,80$ e relação 260/230 $\geq 2,0$ e $\leq 2,2$ foram utilizadas. Das 38 amostras obtidas, somente 17 foram consideradas adequadas para serem utilizadas na análise de expressão.

3.3. Tratamento neoadjuvante

Todas as pacientes foram submetidas ao esquema de tratamento proposto pelo protocolo do Setor de Mastologia e Oncologia Ginecológica do HCFMRP-USP. Os esquemas inicialmente propostos foram ECT com 4 ciclos de epirrubicina

75mg/m² e ciclofosfamida 600mg/m² em infusão endovenosa no D1 a cada 21 dias seguidos de 4 ciclos de docetaxel 100mg/m² também no D1 a cada 21/21 dias. Em pacientes com tumores HER2 positivos, o trastuzumabe foi associado ao docetaxel. A dose de ataque foi de 6mg/kg de peso e mais três doses de manutenção de 4mg/kg a cada 21/21 dias. Em três casos o esquema de tratamento foi modificado devido à progressão da doença com o esquema inicialmente proposto.

3.4. Resposta ao tratamento

Consideramos a resposta patológica completa como parâmetro de resposta, quando não encontramos neoplasia invasiva residual no sítio primário incluindo os linfonodos axilares após o tratamento primário com quimioterapia. As pacientes foram separadas em dois grupos: pacientes com resposta patológica completa (grupo pRC) e sem resposta patológica completa (SR).

3.5. Expressão de receptores hormonais e HER2

Os receptores hormonais foram analisados por imunoistoquímica (IHQ) segundo a escala de Allred e considerados positivos os casos com escore maior ou igual a 3 (HARVEY et al., 1999). O HER2 foi analisado pela IHQ conforme recomendação da American Society of Clinical Oncology / College of American Pathologists (os casos classificados como HER2 2+ foram analisados através do CISH para estudo da amplificação do gene (WOLFF et al., 2013).

3.6. Análise da expressão gênica global

O RNA das 17 amostras foi hibridizado com o chip Human Gene 2.0 ST (Affymetrix) seguindo as recomendações do fabricante. A hibridização e leitura dos chips foram realizadas em dois momentos diferentes (dois lotes, ou do inglês “batches”). Toda as análises foram realizadas utilizando o software Rstudio para Mac utilizando o R v3.3.2. Os pacotes “oligo” e “affy” foram utilizados para a leitura dos dados brutos (arquivos CEL) e a normalização dos dados utilizando o protocolo RMA. Os dois lotes foram normalizados separadamente e utilizamos o pacote “in

Silico Merging” com o método COMBAT da função *merge* para criar o objeto *Expression Set* final e reduzir o “efeito de lote” (TAMINAU et al., 2012).

Utilizamos a análise de componentes principais para verificar a redução do “efeito de lote” após a combinação dos ensaios. A análise de expressão diferencial dos lncRNAs foi realizada utilizando um modelo linear do pacote “limma” e o método empírico de Bayes como moderador dos erros padrão para estimativa da variação do log da expressão (“log-foldchange”) de cada probeset (PHIPSON et al., 2016). Utilizamos a expressão do RE, HER2 e o lote no modelo linear para identificação dos genes diferencialmente expressos (DE) comparando os tumores com resposta patológica completa e os sem resposta completa. Para a anotação dos probesets, utilizamos os pacotes “Annotate” e “hugene20sttranscriptcluster.db” e os lncRNAs identificados foram selecionados para análise de bioinformática nas amostras publicamente disponíveis de carcinomas de mama do The Cancer Genome Atlas (TCGA).

3.7. Classificação molecular de acordo com o PAM50 e integrative cluster (IC).

Utilizamos o pacote “genefu” com a função molecular subtyping para estimar os subtipos moleculares do carcinoma de mama de acordo com o PAM50 tanto nas amostras de microarray do nosso estudo como nas amostras do TCGA (RNAseq) (GENDOO et al., 2016) (LIU et al., 2016). Inferimos também a classificação de acordo com o Integrative Clustering (IC) em 10 subtipos (METABRIC GROUP et al., 2012) em todas as amostras do TCGA utilizando como input os dados de expressão gênica (RNAseq) e variação o número de cópias.

3.8. Análise de bioinformática nas amostras de câncer de mama do TCGA

Os dados clínicos, expressão gênica (RNAseq) e variação do número de cópias foram obtidos do repositório de dados públicos no web site <https://portal.gdc.cancer.gov>. A análise de expressão diferencial entre tumor primário e a glândula mamária normal foi realizada seguindo o protocolo do pacote *DESeq2* (https://bioconductor.org/help/course-materials/2017/CSAMA/labs/2-tuesday/lab-03_maseq/maseqGene_CSAMA2017.html). Para análise estatística, a expressão do *PVT1* foi normalizada e transformada em log2 e o teste *t* de Student foi utilizado para análise comparativa da expressão gênica entre duas variáveis independentes. A razão do log do número de cópias do gene *PVT1* e *MYC* foi

utilizada para o teste de correlação com a expressão gênica. Para a criação do gráfico, a razão do log do número de cópias foi categorizada em valores -2 (deleção profunda), -1 (deleção), 0 (diploide), 1 (ganho) e 2 (amplificação).

Figura 1. Colorações imunoistoquímicas. A) HER2 positivo(3+); B) Ki67 positivo em cerca de 25% das células; C) RE positivo forte em 95% das células; D) RP positivo moderado em 90% das células.

Figura 2. CISH para pesquisa de amplificação do *ERBB2*, marcado em preto e em vermelho o cromossomo 17.

4 RESULTADOS

Os tumores foram graduados histologicamente pelo sistema de Nottingham que leva em conta a formação tubular, pleomorfismo nuclear e índice mitótico, em graus de 1 a 3, sendo o grau I, tumores bem diferenciados, o II moderadamente, e o III pouco diferenciados (Anexo I). As pacientes tinham idade entre 32 e 63 anos (média de 50 anos), dezesseis delas com diagnóstico de carcinoma ductal invasor, e uma de carcinoma metaplásico da mama. Oito desses tumores eram de grau 2 histológico e nove, de grau 3. Segundo o estágio da doença, sete eram IIB, três IIIA, quatro IIIB e três IV (TNM - Anexo II). A tabela 2 resume os dados das pacientes.

Tabela 2. As pacientes incluídas tinham média de idade de 50 anos e tumores de grau e estágio avançados.

Caso	Idade (anos)	Tipo histológico	Grau histológico	Estadio
1	50	CDI	3	IV
2	61	CDI	3	IV
3	46	CDI	2	IIIA
4	58	CDI	3	IIB
5	49	CDI	2	IIB
6	40	CDI	3	IIB
7	52	CDI	2	IIIB
8	58	CDI	2	IIB
9	32	CDI	3	IIB
10	35	CDI	3	IIIA
11	41	CDI	2	IIB
12	63	CDI	3	IIIA
13	62	Metaplásico	3	IIIB
14	58	CDI	2	IIB
15	46	CDI	2	IIIB
16	59	CDI	3	IIIB
17	34	CDI	2	IV

Classificação molecular nos tumores localmente avançados. Segundo a classificação molecular baseada no PAM50, o subtipo molecular mais prevalente foi o Luminal B (7 casos), seguido do subtipo Basal (5 casos) e Her2 (3 casos). Apenas duas amostras foram classificadas como Luminal A, sendo que uma delas a IHQ foi compatível com tumor triplo negativo. Os tumores Luminais B são, na maioria, HER2 positivos na IHQ e os tumores triplo negativos formam os mais prevalentes entre os Basais. A Figura 3 mostra o heatmap da clusterização não supervisionada das amostras analisadas pelo microarray.

Figura 3. O material obtido das pacientes foi submetido a exame imunoistoquímico (IMH) e classificação molecular pelo PAM50. Este heatmap demonstra a clusterização não supervisionada baseado na expressão dos 50 genes utilizados pelo modelo para prever a classificação molecular.

Análise de expressão diferencial de lncRNAs. A análise de expressão diferencial de todo o transcriptoma identificou 1003 transcritos diferencialmente expressos comparando os grupos pRC vs SR. Dos 1003 transcritos, 366 estavam associados a um gene conhecido. Dentre eles, 40 eram não codificantes, sendo treze microRNAs, um pseudogene, quatro RNAs *antisense*, dezenove pequenos RNAs nucleolares e três LncRNAs (*PVT1*, *GAS5* e *Linc00657*). Os três LncRNAs estão superexpressos no grupo de tumores sem resposta patológica completa. O gráfico abaixo resume a diferença da expressão gênica global entre as pacientes do grupo pRC e SR (Figura 4).

Figura 4. Os genes à esquerda estão mais expressos nas pacientes que não responderam ao tratamento neoadjuvante e os em verde são aqueles onde o valor do $p < 0.05$ e o log da variação da expressão é $> 0,7$. Nota-se que centenas de genes estão diferencialmente expressos. Entre eles estão três LncRNAs.

Caracterização da expressão dos lncRNAs *PVT1*, *GAS5* e *LINC00657* em amostras de tumores e glândula mamária do TCGA. Nós analisamos a expressão comparativa global de todos os transcritos em 112 amostras pareadas de carcinoma de mama e tecido mamário do TCGA. Na análise pareada entre tumor primário e o parênquima mamário adjacente observamos que o *PVT1* está superexpresso no tecido neoplásico ($p < 0,0001$). Não existe diferença de expressão entre o tecido normal e a neoplasia maligna para os lncRNAs *GAS5* ($p = 0.4$) e *LINC00657* ($p = 0.8$). A Figura 5 resume a expressão diferencial dos lncRNAs.

Figura 5. A expressão de *PVT1* está aumentada nos carcinomas em relação ao parênquima mamário não tumoral.

Expressão do *PVT1*, de acordo com os subtipos moleculares nas amostras do TCGA. Analisamos a expressão do *PVT1* de acordo com a expressão das proteínas RE, RP e HER2. A expressão do *PVT1* está levemente aumentada nos tumores positivos para receptores hormonais (RE e RP). Com relação aos subtipos moleculares do câncer de mama, observamos um aumento da expressão do *PVT1* no subtipo IC9 (Figuras 6, 7 e 8). O subtipo IC9 é descrito como um grupo de tumores RE positivos com amplificação no braço longo do cromossomo 8, na mesma localização genômica do *PVT1*.

Figura 6. O *PVT1* está superexpresso nos carcinomas RE positivos.

Figura 7. O *PVT1* também está superexpresso nos carcinomas RP positivos.

Figura 8. O *PVT-1* também está mais expresso nos carcinomas subtipo IC9.

Para analisarmos a correlação da expressão do *PVT1* com o número de cópias do gene, utilizamos os dados de variação de número de cópias do TCGA. Observamos que existe uma correlação significativa entre a expressão do *PVT1* e a variação do número de cópias ($r = 0,55$ e $p < 0,0001$). A figura 9 resume a expressão do gene em relação ao número de cópias estimado.

Figura 9. A expressão do *PVT1* tem correlação significativa como número de cópias gênicas ($p < 0,0001$).

5 DISCUSSÃO

O advento de novas técnicas de biologia molecular como os microarrays em larga escala e o sequenciamento de DNA/RNA proporcionou um grande avanço no entendimento da patogênese de muitas doenças, em especial na oncogênese. Acreditava-se que o material genético não codificante não exercia funções celulares já que esses genes não eram produtores de uma proteína. Cerca de 98% do DNA humano não codifica proteínas e até o aparecimento das novas técnicas moleculares pouco se sabia a respeito da função desse enorme material genético. Com o mapeamento do genoma humano e o conhecimento dos miRNAs, criamos uma nova perspectiva a respeito da função dos genes não codificantes de proteínas. Atualmente já existem muitos miRNAs conhecidos que atuam em múltiplas funções celulares e estão presentes no mecanismo de patogênese de vários tumores. Mais recente ainda são os estudos dos lncRNAs, que eram ignorados até poucos anos atrás. Como já citado anteriormente agora conhecemos também uma quantidade significativa dessas moléculas atuando direta ou indiretamente na carcinogênese e esse número cresce a cada dia.

Em nosso estudo avaliamos a expressão de lncRNAs em carcinomas mamários localmente avançados e identificamos três genes não codificantes de proteínas que podem estar relacionados à uma pior resposta ao tratamento quimioterápico, o *PVT1*, o *GAS5* e o *LINC00657*.

O *PVT-1* está localizado na mesma região cromossômica, 8q24, que o oncogene *C-MYC*, com uma distância de apenas 55 Kb entre eles, e esta região cromossômica é uma das mais frequentemente amplificadas no câncer de mama (CHO et al., 2018). O *C-MYC* apresenta funções já conhecidas na patogênese tumoral. A expressão do *C-MYC* é indispensável durante o desenvolvimento embrionário normal. A atividade do *C-MYC* em células normais é regulada por sinais externos, como fatores de crescimento e componentes da matriz extracelular, bem como por sinais internos, como a maquinaria de ativação do ciclo celular. Células diferenciadas normalmente expressam baixos níveis de *C-MYC*, embora possam potencialmente transcrever a proteína *C-MYC* quando estimuladas (FARIA, 2006).

A ativação inadequada do gene *C-MYC*, que contribui para o desenvolvimento das neoplasias humanas, pode ocorrer por diferentes mecanismos: translocações cromossômicas, como verificado no linfoma de Burkitt, através da justaposição da região promotora do gene de cadeia pesada de imunoglobulina, altamente expressa em células B, ao lado do sítio gênico do *C-MYC*; amplificação gênica, pelo aumento

do número de cópias do *C-MYC* e, conseqüentemente, de sua expressão; estímulo da transcrição gênica, como observado nos carcinomas do cólon; inserção de retrovírus adjacente ao gene *C-MYC*, ativando sua expressão via sequências regulatórias virais, entre outros (FARIA, 2006).

A família do oncogene *MYC* está anormal em mais da metade de todos os cânceres e está frequentemente associada a pior prognóstico e sobrevida. O *MYC* está envolvido em diversos processos oncogênicos (CHEN; LIU; QING, 2018).

Há sinais de associação entre o *c-MYC* e o *PVT-1* em diversas neoplasias. A expressão e o número de cópias dos genes *MYC* e os *PVT-1* estão aumentados em cânceres humanos e a ablação do *PVT-1* em tumores colônicos diminuiu o potencial tumorigênico da neoplasia (TSENG et al., 2014).

No banco de dados do TCGA a análise do número de cópias do gene *PVT-1* e *c-MYC* nos carcinomas invasivos de mama mostra um aumento concomitante desses genes (Anexo 3), ou seja, a expressão do *PVT-1* parece estar estritamente relacionada a expressão do *c-Myc*, conforme mostrou Tseng et al (TSENG et al., 2014).

O *PVT-1* está desregulado em várias neoplasias e a sua expressão está aumentada em muitas doenças e cânceres, como o hepatocarcinoma, carcinoma de colo uterino, carcinoma de células renais do tipo células claras e no carcinoma tireoidiano. Entretanto, o mecanismo de atuação do *PVT-1* ainda é pouco conhecido (LU et al., 2017).

Em nosso estudo o *PVT-1* mostrou maior expressão nos tumores que não apresentaram boa resposta ao tratamento neoadjuvante e no tecido tumoral em comparação ao parênquima mamário não neoplásico. Ainda notamos uma maior expressão de *PVT-1* nos tumores que expressam receptores hormonais e nos carcinomas subtipo IC9, que são predominantemente carcinomas luminais B e têm prognóstico desfavorável (Anexos 4 e 5). Os carcinomas subtipo IC9 fazem parte de uma nova classificação realizada com base em uma análise da expressão gênica e número de cópias gênicas no banco de dados do METABRIC que associada ao comportamento biológico de 997 carcinomas mamários permitiu identificar 10 subgrupos diferentes de carcinomas mamários, nomeados Integrative Clusters 1-10.

Os carcinomas IC9 são predominantemente compostos por tumores RE positivos do subtipo luminais B. A análise de sobrevida demonstra que essas neoplasias são de mal prognóstico (METABRIC GROUP et al., 2012). Esta

observação sugere que a amplificação concomitante do *PVT1* e do *MYC* possa ser importante para a progressão do carcinoma da mama. Cho et al, recentemente mostraram que o promotor do *PVT-1* tem função supressora de tumor e inibe a expressão do *C-MYC* no mesmo cromossomo, e mutações do lncRNA *PVT-1* promovem o crescimento celular neoplásico (CHO et al., 2018).

O *GAS-5* (*growth arrest specific 5*) é um inibidor da proliferação celular e promotor da apoptose, a sua expressão está diminuída em várias neoplasias, como nos tumores da mama, próstata, pulmão e intestino grosso e a sua baixa expressão confere uma elevada capacidade de proliferação às células tumorais (KOLDEMIR; ÖZGÜR; GEZER, 2017).

Asghar A. et al identificaram níveis menores de expressão do *GAS5* em pacientes jovens (menores de 45 anos de idade) com câncer de mama quando comparadas com pacientes de mais idade, o que indicaria um pior prognóstico quando há redução da expressão desse lncRNA (ARSHI et al., 2018), já que pacientes jovens geralmente apresentam um pior prognóstico (AZIM et al., 2015).

Nós identificamos uma maior expressão do *GAS5* nas pacientes que não apresentaram boa resposta ao tratamento quimioterápico e em tumores HER2 negativos, além de uma menor expressão desse gene no carcinoma mamário quando comparado ao parênquima não neoplásico, o que indica que a perda da expressão do *GAS5* favorece a carcinogênese mamária e piora o prognóstico da doença.

Em pacientes portadoras de câncer de mama o *GAS 5* tem uma maior expressão em pacientes com menarca tardia (depois dos 13 anos) e é menos expresso em pacientes com sobrepeso (MANSOORI et al., 2018).

Há ainda evidências que o *GAS 5* contribui para a resistência ao trastuzumabe nos tumores HER2 positivos (LI et al., 2016).

Por fim identificamos o lncRNA *linc00657*, um gene recentemente descrito, que é induzido após dano ao DNA, chamado de RNA não codificante ativado por *DNA Damage* ou *NORAD* (LEE et al., 2016).

O *linc00657* se mostrou mais expressos nos tumores não responsivos à quimioterapia, nos carcinomas luminais, especialmente nos luminais B e nos IC6, que expressam receptores de estrogênio e são predominantemente luminais B, que são tumores também de prognóstico desfavorável e sobrevida semelhante aos IC9, que apresentaram alteração do *PVT-1* (RUSSNES et al., 2017).

O *linc00657* se mostrou aumentado em células tratadas com irradiação e em câncer de pâncreas sob condições de hipóxia, o que indicam um aumento da expressão em situações de stress (SUN et al., 2018) (MICHALIK et al., 2014).

Está claro que existem diversos lncRNAs envolvidos na patogênese do câncer e ainda conhecemos muito pouco a respeito dos mecanismos celulares nos quais essas moléculas estão envolvidas. Com o uso dos novos métodos de investigação molecular poderemos entender melhor a função desses RNAs, até pouco tempo desconhecidos. Já sabemos que os lncRNAs são os principais elementos do transcriptoma e controlam diversos mecanismos celulares. A maioria dos lncRNAs bem estudados até o momento são considerados cruciais na regulação de processos celulares, como ciclo celular, crescimento e apoptose, que asseguram a homeostase. Devemos aprofundar os estudos desses genes e sua possível aplicação como potenciais alvos para a terapêutica, que é uma grande promessa para a futura terapia do câncer (RENGANATHAN; FELLE-Y-BOSCO, 2017).

6 CONCLUSÃO

Os carcinomas mamários que respondem ao tratamento quimioterápico têm expressão gênica diferente daqueles que não respondem ao tratamento.

Os tumores que respondem ao tratamento superexpressam os RNAs longos não codificantes *PVT-1*, o *GAS 5* e o *Inc00657*.

O *PVT-1* está localizado na mesma região cromossômica que o gene *c-MYC* e parece atuar conjuntamente com ele na patogênese do carcinoma mamário. Esse lncRNA está superexpresso nos tumores mamários em relação ao parênquima mamário não neoplásico e nos carcinomas mamários com expressão de receptores hormonais e nos subtipos IC9.

Já o *GAS 5* está *downregulado* nos carcinomas subtipo HER-2.

O *Inc00657* não apresenta diferença de expressão nos tumores mamários quando comparados com o parênquima não tumoral, entretanto está superexpresso nos carcinomas luminais, em especial, os luminais B e nos subtipos IC6.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ANDRE, F. et al. HER2 Expression and Efficacy of Preoperative Paclitaxel/FAC Chemotherapy in Breast Cancer. **Breast Cancer Research and Treatment**, v. 108, n. 2, p. 183–190, 1 mar. 2008.
- ARSHI, A. et al. Expression Analysis of MALAT1, GAS5, SRA, and NEAT1 LncRNAs in Breast Cancer Tissues from Young Women and Women over 45 Years of Age. **Molecular Therapy - Nucleic Acids**, v. 12, p. 751–757, set. 2018.
- AZIM, H. A. et al. Genomic Aberrations in Young and Elderly Breast Cancer Patients. **BMC Medicine**, v. 13, n. 1, dez. 2015. Disponível em: <<http://www.biomedcentral.com/1741-7015/13/266>>. Acesso em: 7 maio. 2019.
- CALZA, S. et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. **Breast Cancer Research**, v. 8, n. 4, p. R34, 2006.
- CHEETHAM, S. W. et al. Long noncoding RNAs and the genetics of cancer. **British Journal of Cancer**, v. 108, n. 12, p. 2419–2425, 25 jun. 2013.
- CHEN, H.; LIU, H.; QING, G. Targeting Oncogenic Myc as a Strategy for Cancer Treatment. **Signal Transduction and Targeted Therapy**, v. 3, n. 1, dez. 2018. Disponível em: <<http://www.nature.com/articles/s41392-018-0008-7>>. Acesso em: 7 maio. 2019.
- CHO, S. W. et al. Promoter of LncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. **Cell**, v. 173, n. 6, p. 1398–1412.e22, 31 2018.
- FARIA, M. H. G. Impacto do oncogene C-MYC no câncer. p. 165–171, 2006.
- GARTEL, A. L.; KANDEL, E. S. miRNAs: Little known mediators of oncogenesis. **Seminars in Cancer Biology**, Postgenetic Oncology - MicroRNA and Cell Proliferation. v. 18, n. 2, p. 103–110, 1 abr. 2008.
- GENDOO, D. M. A. et al. Genefu: An R/Bioconductor Package for Computation of Gene Expression-Based Signatures in Breast Cancer. **Bioinformatics**, v. 32, n. 7, p. 1097–1099, 1 abr. 2016.
- GUPTA, R. A. et al. Long noncoding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. **Nature**, v. 464, n. 7291, p. 1071–1076, 15 abr. 2010.
- GUTSCHNER, T.; DIEDERICH, S. The hallmarks of cancer. **RNA Biology**, v. 9, n. 6, p. 703–719, 1 jun. 2012.

HAJJARI, M.; SALAVATY, A. HOTAIR: an oncogenic long non-coding RNA in different cancers. **Cancer Biology & Medicine**, v. 12, n. 1, p. 1–9, mar. 2015.

HARVEY, J. M. et al. Estrogen Receptor Status by Immunohistochemistry Is Superior to the Ligand-Binding Assay for Predicting Response to Adjuvant Endocrine Therapy in Breast Cancer. **Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology**, v. 17, n. 5, p. 1474–1481, maio 1999.

JEMAL, A. et al. Cancer Statistics, 2009. **CA: a cancer journal for clinicians**, v. 59, n. 4, p. 225–249, ago. 2009.

KOLDEMIR, O.; ÖZGÜR, E.; GEZER, U. Accumulation of GAS5 in Exosomes Is a Marker of Apoptosis Induction. **Biomedical Reports**, v. 6, n. 3, p. 358–362, mar. 2017.

KUNG, J. T. Y.; COLOGNORI, D.; LEE, J. T. Long Noncoding RNAs: Past, Present, and Future. **Genetics**, v. 193, n. 3, p. 651–669, mar. 2013.

LEE, S. et al. Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. **Cell**, v. 164, n. 1–2, p. 69–80, jan. 2016.

LI, W. et al. Downregulation of LncRNA GAS5 Causes Trastuzumab Resistance in Breast Cancer. **Oncotarget**, v. 7, n. 19, 10 maio 2016. Disponível em: <<http://www.oncotarget.com/fulltext/8413>>. Acesso em: 7 maio. 2019.

LI, X. et al. Intrinsic Resistance of Tumorigenic Breast Cancer Cells to Chemotherapy. **JNCI: Journal of the National Cancer Institute**, v. 100, n. 9, p. 672–679, 7 maio 2008.

LIU, M. C. et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). **NPJ Breast Cancer**, v. 2, p. 15023, 6 jan. 2016.

LU, D. et al. LncRNA PVT1 in Cancer: A Review and Meta-Analysis. **Clinica Chimica Acta**, v. 474, p. 1–7, nov. 2017.

LU, J. et al. MicroRNA Expression Profiles Classify Human Cancers. **Nature**, v. 435, n. 7043, p. 834–838, jun. 2005.

MANSOORI, Y. et al. Expression Levels of Breast Cancer-Related GAS5 and *LSINCT5* LncRNAs in Cancer-Free Breast Tissue: Molecular Associations with Age at Menarche and Obesity. **The Breast Journal**, v. 24, n. 6, p. 876–882, nov. 2018.

METABRIC GROUP et al. The Genomic and Transcriptomic Architecture of 2,000 Breast Tumours Reveals Novel Subgroups. **Nature**, v. 486, n. 7403, p. 346–352, jun. 2012.

MICHALIK, K. M. et al. Long Noncoding RNA MALAT1 Regulates Endothelial Cell Function and Vessel Growth. **Circulation Research**, v. 114, n. 9, p. 1389–1397, 25 abr. 2014.

MITRA, S. A. M.; MITRA, A. P. M.; TRICHE, T. J. M. A Central Role for Long Non-Coding RNA in Cancer. **Frontiers in Genetics**, v. 3, 2012. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fgene.2012.00017/full>>. Acesso em: 8 maio. 2019.

OSAKO, T. et al. High-Grade Breast Cancers Include Both Highly Sensitive and Highly Resistant Subsets to Cytotoxic Chemotherapy. **Journal of Cancer Research and Clinical Oncology**, v. 136, n. 9, p. 1431–1438, 1 set. 2010.

PHIPSON, B. et al. ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION. **The annals of applied statistics**, v. 10, n. 2, p. 946–963, jun. 2016.

PRENSNER, J. R. et al. PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. **Cancer research**, v. 74, n. 6, p. 1651–1660, 15 mar. 2014.

RENGANATHAN, A.; FELLE-Y-BOSCO, E. Long Noncoding RNAs in Cancer and Therapeutic Potential. **Advances in Experimental Medicine and Biology**, v. 1008, p. 199–222, 2017.

ROSS, J. S. et al. Commercialized Multigene Predictors of Clinical Outcome for Breast Cancer. **The Oncologist**, v. 13, n. 5, p. 477–493, 5 jan. 2008.

RUSSNES, H. G. et al. Breast Cancer Molecular Stratification. **The American Journal of Pathology**, v. 187, n. 10, p. 2152–2162, out. 2017.

SØRLIE, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 19, p. 10869–10874, 11 set. 2001.

SØRLIE, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 14, p. 8418–8423, 8 jul. 2003.

SØRLIE, T. et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. **BMC Genomics**, v. 7, p. 127, 26 maio 2006.

SUN, Y. et al. LINC00657 Played Oncogenic Roles in Esophageal Squamous Cell Carcinoma by Targeting MiR-615-3p and JunB. **Biomedicine & Pharmacotherapy**, v. 108, p. 316–324, dez. 2018.

TAMINAU, J. et al. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. **BMC Bioinformatics**, v. 13, p. 335, 24 dez. 2012.

TSENG, Y.-Y. et al. PVT1 Dependence in Cancer with MYC Copy-Number Increase. **Nature**, v. 512, n. 7512, p. 82–86, ago. 2014.

U.S. Breast Cancer Statistics. Disponível em:

<https://www.breastcancer.org/symptoms/understand_bc/statistics>. Acesso em: 6 maio. 2019.

VAN DE VIJVER, M. J. et al. A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. **The New England Journal of Medicine**, v. 347, n. 25, p. 1999–2009, 19 dez. 2002.

WOLFF, A. C. et al. Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. **Journal of Clinical Oncology**, v. 31, n. 31, p. 3997–4013, 7 out. 2013.

ZHANG, H. et al. Long non-coding RNA: a new player in cancer. **Journal of Hematology & Oncology**, v. 6, p. 37, 31 maio 2013.

Anexo 1. Graduação histológica dos carcinomas mamários:

Todos os carcinomas invasivos da mama devem ser graduados. Recomenda-se o grau histológico combinado ou final de Nottingham (modificação de Elston-Ellis do sistema de graduação de Scarff-Bloom-Richardson), que leva em conta a formação de túbulos, o pleomorfismo nuclear e o índice mitótico. Um sistema de escore de 1 a 3 é dado para cada um desses parâmetros avaliados separadamente, da seguinte maneira:

- Formação tubular: escore 1 (na maior parte do tumor, > 75%), escore 2 (entre 10 e 75%) e escore 3 (em menos de 10%).
- Grau nuclear (pleomorfismo): escore 1 – núcleos pequenos, com pouco aumento ou variação no tamanho em comparação com os núcleos das células epiteliais ductais normais, com contornos regulares e uniformidade da cromatina nuclear; escore 2 – os núcleos são maiores do que os núcleos das células ductais normais, aspecto vesiculoso, nucléolos mais visíveis, geralmente únicos e com variação moderada no tamanho e na forma; e escore 3 – variação acentuada no tamanho e na forma nuclear, nos núcleos volumosos e bizarros, nos vesiculosos, com nucléolos grandes proeminentes, geralmente múltiplos.
- Índice mitótico: contar apenas as figuras de mitose bem definidas, ignorando-se núcleos hipercromáticos e picnóticos. As figuras de mitose são contadas em 10 campos de maior aumento na periferia do tumor, onde a atividade proliferativa for intensa. Se houver variações do número de mitoses observadas em áreas diferentes do tumor, devem ser contados outros grupos de 10 campos para se obter a contagem correta (mais alta). É necessário calibrar o microscópio para produzir dados comparáveis, pois o tamanho do campo de maior aumento varia até 6 vezes de acordo com cada fabricante de microscópio. Portanto, a avaliação dos pontos para contagens mitóticas é feita de acordo com a área do campo de aumento maior (400 x) do microscópio. A última edição da OMS (2012) forneceu uma tabela para contagem mitótica de acordo com os vários campos do microscópio. Por exemplo: campo de 0,40 mm: contagem mitótica escore 1: ≤ 4 , escore 2: 5-9, escore 3: ≥ 10 ; campo de 0,60 mm: escore 1: ≤ 10 , escore 2: 11-20, escore 3: ≥ 21 .
- Importante: o retardo na fixação pode reduzir a contagem mitótica.
- Grau histológico combinado ou final: os escores de cada elemento são somados, resultando em escore de 3 a 9 pontos. A graduação final é expressa da seguinte forma:
 - 3 a 5 pontos: baixo grau, grau 1 (bem diferenciado);
 - 6 a 7 pontos: grau intermediário, grau 2 (moderadamente diferenciado);
 - 8 a 9 pontos: alto grau, grau 3 (pouco diferenciado).

Anexo 2. Sistema TNM

Tumor primário:

TX. O tumor primário não pode ser avaliado.

T0. Sem evidências de tumor primário.

Tis. Carcinoma in situ.

pT1. O tumor $\leq 2,0$ cm.

pT1mi – microinvasão (foco ou focos $\leq 0,1$ cm).

pT1a - $> 0,1$ cm e $\leq 0,5$ cm.

pT1b - $> 0,5$ cm e $\leq 1,0$ cm.

pT1c - $> 1,0$ cm e $\leq 2,0$ cm.

pT2. O tumor > 2 cm e ≤ 5 cm.

pT3. O tumor > 5 cm.

pT4. O tumor tem qualquer tamanho com invasão da parede torácica e/ou a pele.

pT4a – Extensão da parede torácica.

pT4b – Edema, ulceração ou nódulos satélites ipsilaterais na pele da mama.

pT4c – Presença de ambos, pT4a e pT4b.

pT4d – Carcinoma inflamatório.

Linfonodos regionais:

pNX. Os linfonodos regionais não podem ser avaliados.

pN0. Ausência de metástases ao exame histológico.

N0(i+). A área de disseminação da doença contém menos de 200 células e é menor do que 0,2 mm. A abreviatura "i +" significa que uma pequena porcentagem de células cancerosas, denominadas células tumorais isoladas, são observadas no exame imunoistoquímico.

pN0(mol+). Não são observadas células cancerosas nos linfonodos axilares, mas foram diagnosticados vestígios de células cancerosas com o teste molecular RT-PCR, que permite detectar um número muito pequeno de células.

pN1. O tumor se disseminou para 1 ou 3 linfonodos axilares e/ou linfonodos mamários internos.

pN1mi. Micrometástases em 1 a 3 linfonodos sob o braço. As áreas de disseminação do tumor nos linfonodos são de 2 mm ou menos.

pN1a. O tumor se disseminou para 1 a 3 linfonodos no braço com pelo menos uma área de doença disseminada com mais de 2 mm de diâmetro.

pN1b. O tumor se disseminou para os linfonodos mamários internos, mas essa disseminação só pode ser diagnosticada na biópsia do linfonodo sentinela.

pN1c. Se aplicam N1a e N1b.

pN2. O tumor se disseminou para 4 ou 9 linfonodos axilares ou para os linfonodos mamários internos.

N2a. O tumor se disseminou para 4 a 9 linfonodos axilares, com pelo menos uma área maior que 2 mm.

pN2b. O tumor se disseminou para um ou mais linfonodos mamários internos, detectados clinicamente, na ausência de comprometimento axilar.

pN3. Qualquer um dos seguintes:

pN3a. O tumor se disseminou para 10 ou mais linfonodos axilares, com pelo menos uma área de câncer disseminada maior que 2mm; ou, o tumor se disseminou para os linfonodos infraclaviculares com pelo menos uma área de câncer maior que 2mm.

pN3b. O tumor é encontrado em pelo menos um linfonodo axilar com pelo menos uma área de disseminação de câncer maior que 2 mm e aumentou os linfonodos mamários internos; ou, o tumor se disseminou para 4 ou mais linfonodos axilares com pelo menos uma área de disseminação maior que 2 mm e pequenas quantidades de são encontradas nos linfonodos mamários internos na biópsia do linfonodo sentinela.

pN3c. O tumor se disseminou para os linfonodos supraclaviculares, com pelo menos uma área maior do que 2 mm.

Metástases a distância:

pMX. A disseminação não pode ser avaliada

pM0. Ausência de metástases à distância.

pM1. Metástases à distância.

Estadiamento Patológico:

Estágio 0 - Tis, N0, M0.

Estágio IA - T1, N0, M0..

Estágio IB - T0 ou T1, N1mi, M0.

Estágio IIA - T0 ou T1, N1 (mas, não N1mi), M0; T2, N0, M0.

Estágio IIB - T2, N1, M0; T3, N0, M0.

Estágio IIIA - T0 a T2, N2, M0; T3, N1 ou N2, M0.

Estágio IIIB - T4, N0 a N2, M0.

Estágio IIIC - Qualquer T, N3, M0.

Estágio IV - Qualquer T, qualquer N, M1.

Anexo 3. Comparação do número de cópias dos genes MYC e PVT1 no TCGA.

Anexo 4. Sobrevida global conforme subtipos segundo a classificação pelo *PAM50*.

Anexo 5. Sobrevida global conforme subtipos segundo a classificação pelo Integrative clustering.