UNIVERSIDADE DE SÃO PAULO FACULDADE DE MEDICINA DE RIBEIRÃO PRETO PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

ISABELA ICHIHARA DE BARROS

Análise da expressão de RNAs longos não codificantes em tumores metastáticos de câncer de mama

RIBEIRÃO PRETO

2016

ISABELA ICHIHARA DE BARROS

Análise da expressão de RNAs longos não codificantes em tumores metastáticos de câncer de mama

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências

Área de concentração: Genética

Orientador: Prof. Dr. Wilson Araújo da Silva Junior

Ribeirão Preto

2016

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.

FICHA CATALOGRÁFICA

Ichihara de Barros, Isabela

Análise da expressão de RNAs longos não codificantes em tumores metastáticos de câncer de mama. / Isabela Ichihara de Barros; Orientador. Wilson Araujo da Silva Junior. – Ribeirão Preto, 2016.

92f.:il; 30 cm

Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Genética; Orientador: Araujo da Silva Junior, Wilson

1. IncRNAs. 2. Câncer de mama. 3. Metástase

NOME: BARROS, Isabela Ichihara de

TÍTULO: Análise da expressão de RNAs longos não codificantes em tumores metastáticos de câncer de mama

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências

Aprovada em:____/___/____

Banca examinadora

Dr(a):	_Instituição
Julgamento:	_Assinatura:
	Instituição
Julgamento:	_Assinatura:
- / .	
Dr(a):	_Instituição
Julgamento:	_Assinatura:

APOIO E SUPORTE FINANCIERO

Este trabalho bem como a sua apresentação em diversos congressos foi realizado com o apoio financeiro das seguintes instituições e entidades:

- > Fundação de Amparo à Pesquisa do Estado de São Paulo FAPESP.
- Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq.
- > Coordenação de Aperfeiçoamento de Pessoal de Nível Superior CAPES.
- > Fundação de Apoio ao Ensino, Pesquisa e Assistência FAEPA.
- > Faculdade de Medicina de Ribeirão Preto FMRP/USP.
- > Fundação Hemocentro de Ribeirão Preto.

AGRADECIMENTOS

Agradeço primeiramente à minha mãe, por ser meu porto seguro e minha guia. Ao meu pai, por todo o apoio e por todo o carinho de sempre. Ao meu irmão, pelo companheirismo e por me tornar sempre mais leve.

À minha família, que sempre torceu por mim. Especialmente à minha avó, Maria Eleonora, por toda sua fé e amor, e aos meus tios Ana e Gilson e primos Cauê e Luisa, por todo suporte e ajuda e por serem meus segundos pais e irmãos.

A Dilo, por estar sempre ao meu lado, me apoiando em todas as minhas decisões, e por toda a compreensão, companheirismo e carinho que tem me dedicado.

Ao professor Wilson Araújo da Silva Junior, por ter me recebido de portas abertas e pela orientação durante a realização deste trabalho.

Aos meus companheiros de laboratório, sem os quais o dia-a-dia não teria sido tão agradável e divertido. É muito bom fazer parte da família LGMB.

Aos professores Daniel Tiezzi, Daniela Tirapelli e Gilberto Carlotti pela colaboração.

À professora Vanessa Silveira, ao professor Tiago Pereira e ao professor Rodrigo Panepucci pelas contribuições no exame de qualificação.

Ao Programa de Pós-Graduação em Genética e ao Departamento de Genética da Faculdade de Medicina de Ribeirão Preto. Um agradecimento especial à Susie, por toda a ajuda.

À Fundação Hemocentro da Faculdade de Medicina de Ribeirão Preto, onde este trabalho foi realizado.

À CAPES pela concessão da bolsa de mestrado e apoio financeiro para realização deste trabalho.

À FAPESP, ao CNPq e à FAEPA, pelo apoio financeiro.

Às meninas da Casina, Ayling e Melina, por todo o suporte, torcida e pelos ótimos momentos que passamos juntas.

Às meninas da Rep Hour, Estela, Mariana e Carla, que me acolheram tão bem em Ribeirão.

Aos amigos que fiz em Ribeirão e àqueles que estavam distantes, por tornarem tudo mais feliz. Agradeço especialmente a Luiza, Thyago, Leandro, Rafael e Laís, amigos de toda uma vida, que, independente da distância, estão comigo aonde eu for; a Maria e a Thiago, queridos conterrâneos que sempre me acompanham; a Pedro, Vívian, Patrícia e Samir, por toda a torcida e a certeza de que sempre teremos uns aos outros; a Eudes, Isabela, Jessica, Ivan, Ayling, Rafaela e José, por todo o carinho e amizade.

A todos que contribuíram de alguma maneira para a realização deste trabalho.

RESUMO

BARROS, I. I. Análise da expressão de RNAs longos não codificantes em tumores metastáticos de câncer de mama. 2016. 92 f. Dissertação (Mestrado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2016.

O câncer de mama é um dos que mais afeta mulheres em todo o mundo e em 2016 estima-se que haverá aproximadamente 58 mil novos casos no Brasil. É uma doença caracteristicamente heterogênea, o que dificulta o diagnóstico, prognóstico e a abordagem terapêutica utilizada. Os biomarcadores já existentes para câncer de mama não são suficientes para explicar a ocorrência de metástase, que é a principal causa de morte entre os pacientes. Neste sentido, os RNAs longos não codificadores (IncRNAs) vêm se estabelecendo como importantes moléculas regulatórias em diversos processos biológicos e alguns têm sido associados com o desenvolvimento e a progressão do câncer de mama. Apesar disso, muito ainda precisa ser elucidado a respeito dessa nova classe de ncRNAs e seu envolvimento na metástase. Desta forma, determinar uma assinatura de IncRNAs em tumores metastáticos de mama pareados com seus respectivos tumores primários pode sugerir novas moléculas envolvidas no mecanismo de metástase. Neste estudo, foram analisadas amostras de cinco tumores primários de mama e seus correspondentes metastáticos em Sistema Nervoso Central (n=2) e em linfonodos (n=3). O RNA total de cada amostra foi extraído e avaliado quanto à sua qualidade para obtenção do transcriptoma pelo método de RNA-Seq. Para o tratamento dos dados foram usados os aplicativos: Tophat para o alinhamento das reads, o Cufflinks para a montagem e quantificação dos transcritos e o pacote DESeq2 para análise de expressão diferencial. O resultado da análise demonstrou que os IncRNAs com íntrons retidos são os mais abundantes e que as amostras de tumores primários e metastáticos compartilham a maioria dos IncRNAs expressos, tornando-os qualitativamente bastante semelhantes entre si. Dentre os IncRNAs diferencialmente expressos, doze são comuns entre os tumores primários, independente dos subtipos moleculares das amostras, e não há transcritos comuns às amostras metastáticas. A análise de agrupamento hierárquico definiu três assinaturas de expressão de IncRNAs para as amostras de tumores primários, metástase em cérebro e em linfonodos, podendo indicar IncRNAs potencialmente envolvidos no mecanismo de metástase.

Palavras-chave: IncRNA, metástase, câncer de mama.

ABSTRACT

BARROS, I. I. Analysis of the expression of long noncoding RNAs in breast cancer metastatic tumors. 2016. 92 f. Dissertação (Mestrado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2016.

Breast cancer is the principall cancer that affects women in the world and it is expect to Brazil, 58.000 new cases in 2016. It is a heterogeneous disease, what makes diagnosis, prognosis and therapeutical approach more difficult. Existing biomarkers for breast cancer are not sufficient to explain metastasis occurrence, which is the main cause of mortality. In this way, long non-coding RNAs (IncRNAs) have been establishing as important regulatory molecules in several biological processes and some of them have been associated to breast cancer development and progression. Nevertheless, many aspects about this new ncRNA class and its involvement in metastasis need to be elucidated. Thus, determining a IncRNA signature of metastatic tumors paired with their related primary tumors may indicate new molecules involved in the mecanism of metastasis. In the present study, Five samples of primary breast tumors and their related metastasis in central nervous system (n=2) and in lymph nodes (n=3) were analyzed. Total RNA of each sample was extracted and had its quality evaluated for transcriptome obtainment by RNA-Seq method. Data processing was performed utilizing the applications: Tophat, for read alignment; Cufflinks, for transcripts assembly and quantification; and DESeq2 package, for differential expression analysis. The results revealed that IncRNAs containing retained introns are the most abundant transcripts and that primary and metastatic samples share most of the expressed IncRNAs, what makes them qualitatively similar to each other. Differential expression analysis showed that twelve IncRNAs are common among primary tumors, apart from samples mollecular subtypes, and metastatic tumors do not have any transcript in common. Hierarchical clustering analysis defined three IncRNA expression signatures, for primary tumors, brain and lymph node metastatis, what may indicate lncRNAs potentially involved in metastasis mechanism.

Key words: IncRNAs, metastasis, breast cancer.

LISTA DE FIGURAS

Figura 3 - Amostras tumorais primárias e metastáticas obtidas para o presente estudo......36

LISTA DE TABELAS

Tabela 1 - Características histológicas e clínico-patológicas dos tumores de mama42
Tabela 2 - Avaliação da integridade dos RNAs extraídos das amostras tumorais43
Tabela 3 - Total de reads geradas pelo RNA-seq das amostras tumorais primárias e metastáticas e quantidade de reads mapeadas contra o genoma de referência GRCh37.p13

LISTA DE ABREVIATURAS E SIGLAS

- CDIS Carcinoma ductal *in situ*
- cDNA DNA complementar
- CLIS Carcinoma lobular in situ
- CTC Célula tumoral circulante
- DNA Ácido desoxirribonucléico
- EMT Transição epitélio-mesenquimal
- HCFMRP Hospital das Clínicas de Ribeirão Preto
- HER2 Receptor 2 do fator de crescimento epidermal humano
- lincRNA RNA não codificador longo intergênico
- IncRNA RNA não codificador longo
- MEC Matriz extracelular
- MET Transição mesenquimal-epitelial
- miRNA micro RNA
- MMP Metaloproteinase de matriz
- mRNA RNA mensageiro
- ncRNA RNA não codificador
- PCR Reação em cadeia da polimerase
- RE Receptor de estrógeno
- RIN Número de integridade do RNA
- RNA Ácido ribonucléico

- RP Receptor de progesterona
- UTR Região não traduzida

LISTA DE SÍMBOLOS

- °C grau celsius
- mg miligrama
- ml mililitro
- μl microlitro
- ng nanograma
- nm nanômetro
- pb pares de base
- RPM rotações por minuto
- % porcentagem

SUMÁRIO

1 INTRODUÇÃO	
1.1 Câncer de mama	18
1.1.1 Epidemiologia e aspectos clínicos do câncer de mama	
1.1.2 Heterogeneidade tumoral no câncer de mama	20
1.1.3 Genética e biologia molecular do câncer de mama	21
1.1.4 Mecanismos moleculares da metástase	24
1.1.5 Transição epitélio-mesenquimal	27
1.1.6 Disseminação linfática e hematógena	28
1.2 RNAs longos não codificadores (IncRNAs)	29
1.2.1 Características e funções dos IncRNAs	29
1.2.2 LncRNAs e câncer de mama	32
2 OBJETIVOS	35
3 MATERIAL E MÉTODOS	
3.1 Amostras tumorais	36
3.2 Isolamento de RNA total	36
3.3 Sequenciamento de RNA	
3.3.1 Preparo das bibliotecas	38
3.3.2 Clusterização	38
3.3.3 Sequenciamento	
3.4 Análise dos dados	40
3.4.1 Chamada de bases	40

3.4.2 Alinhamento das sequências40
3.4.3 Análise da expressão gênica40
3.4.4 Caracterização funcional dos IncRNAs41
3.5 Validação de expressão de IncRNA por PCR semi-quantitativa41
4 RESULTADOS
4.1 Dados clínicos das amostras tumorais42
4.2 Qualidade dos RNAs para sequenciamento42
4.3 Análise do RNA-seq43
4.3.1 Qualidade do sequenciamento43
4.3.2 Mapeamento
4.3.3 Análise dos IncRNAs expressos nas amostras tumorais45
4.3.4 Análise da expressão diferencial dos IncRNAs51
4.3.5 Validação de IncRNA54
5 DISCUSSÃO
6 CONCLUSÕES
REFERÊNCIAS64
ANEXOS

1 INTRODUÇÃO

1.1 Câncer de mama

1.1.1 Epidemiologia e aspectos clínicos do câncer de mama

O câncer de mama é o tipo de câncer mais comum entre as mulheres de todo o mundo, depois do câncer de pele não melanoma, sendo responsável pela morte de mais de 14.200 mulheres em 2013, em contraste com o número de mortes em homens (181) no mesmo ano (INCA, 2016). Mais alarmantes ainda são as previsões do Instituto Nacional do Câncer, que estima o surgimento de 57.960 novos casos no Brasil, e da *International Agency for Research on Cancer*, que prevê o surgimento de mais de 307.600 novos casos nos Estados Unidos em 2016. Dentre os casos de câncer de mama diagnosticados nos últimos anos, 20% a 30% deles podem estar associados à ocorrência de múltiplos fatores de risco, sendo os principais a idade acima de 40 anos, o histórico de doenças da glândula mamária, o histórico de câncer em parentes de primeiro grau, menarca precoce e gravidez tardia (Kamińska et al., 2015). Além dos fatores de risco, a raça/etnia também tem influência importante na incidência desse tipo de câncer, tendo sido observado um maior número de casos em mulheres brancas (Smigal et al., 2006).

O câncer de mama se constitui em um grupo de neoplasias originadas das células epiteliais que revestem os dutos de leite (Polyak, 2011). Se as células cancerosas não passaram pela membrana basal, o câncer é denominado *in situ* ou não invasivo e pode ser classificado em carcinoma lobular *in situ* (CLIS) e carcinoma ductal *in situ* (CDIS) (Sibbering e Courtney, 2016). A principal diferença entre os dois tipos é o padrão de expressão de E-caderina, que é praticamente ausente no CLIS (Buerguer et al., 2000). Acredita-se que o CDIS seja o precursor do carcinoma ductal invasivo - aquele onde há invasão do tecido adjacente - sendo a perda da organização da camada celular mioepitelial e da membrana basal os critérios que definem a transição de um para o outro (Hu, et al., 2008). De uma forma geral, o prognóstico do carcinoma de mama é influenciado pelo estágio do tumor, grau, superexpressão do homólogo 2 do oncogene viral de leucemia eritroblástica aviária v-erb-b2 (ERBB2), e receptores hormonais (Vymetalkova et al., 2015).

De acordo com o perfil de expressão gênica, o câncer de mama pode ser dividido em cinco subtipos biologicamente distintos: luminal A, luminal B, HER2-enriquecido (HER2⁺), basal e mama-normal símile (Liu et al., 2014). Cada um desses grupos apresenta diferentes fatores de risco para a sua incidência, riscos de progressão e sítios preferenciais de metástase (Polyak, 2011). Tanto o HER2⁺ quanto o câncer de mama basal possuem desenvolvimento ruim, sendo que este último é bastante frequente e tende a acometer mulheres mais jovens (Liu et al., 2014). Ambos os subtipos luminais A e B expressam o receptor de estrogênio (RE⁺), porém o tipo B tem uma evolução relativamente pior, e as características clínicas do câncer de mama luminal A parecem ser mais relevantes para o tempo de sobrevida do paciente e terem o maior poder prognóstico (Liu et al., 2014). Além dos cinco subtipos supracitados, outros já foram descritos, como alguns subgrupos de câncer de mama triplo-negativos (Lehmann et al., 2011). Em 2007, Herschkowitz e colaboradores descreveram um subtipo adicional, denominado *claudin-low*, caracterizado pela baixa expressão de genes envolvidos em tight junctions e adesão celular. Suas diferenças em relação aos demais subtipos e suas particularidades, porém, só foram melhor descritas recentemente (Sabatier et al., 2014).

Atualmente, o método padrão para triagem e diagnóstico do câncer de mama é a mamografia, onde os principais sinais indicadores da doença são a presença de nódulo espiculado, calcificações pleomórficas agrupadas e área de assimetria focal com distorção do parênquima mamário (Nazário et al., 2015). A escolha da terapia adequada para cada paciente é feita com base nos fatores prognósticos histopatológicos e imunohistoquímicos, sendo primeiramente realizado o tratamento cirúrgico para retirada do tumor, seguindo-se de radioterapia (para tumores maiores de 5 cm) (Nazário et al., 2015). Todos os pacientes submetidos a cirurgia para remoção de câncer de mama maligno devem passar por procedimento em nódulo axilar, tanto para fins de estadiamento quanto para tratamento de axila nodo positiva (Sibbering e Courtney, 2016). A quimioterapia é indicada apenas em casos de doença localmente avançada ou metastática, ou quando nos estágios iniciais o tumor mede mais de 2 cm ou ainda se há comprometimento da axila (Nazário et al., 2015). O objetivo da quimioterapia é justamente diminuir o estágio do tumor, tornando-o operável (Sibbering e Courtney, 2016). O tratamento endócrino, por sua vez, é indicado para

pacientes RE⁺ que não podem ser operados (devido a co-morbidades ou recusa à operação) (Sibbering e Courtney, 2016).

1.1.2 Heterogeneidade tumoral no câncer de mama

A mama é um dos poucos órgãos que passa por ciclos de crescimento, diferenciação, regressão e remodelamento durante a vida do organismo (Bissell et al., 2003). A glândula mamária é derivada de células ectodérmicas, formando uma árvore de ductos rudimentares durante a vida intra-uterina, que é induzida a crescer e formar a massa de gordura mamária durante a puberdade (Koren e Betires-Alj, 2015). Já na fase adulta, a mama apresenta hierarquia epitelial e população heterogênea de células-tronco (Visvader e Stingl, 2014), porém no tumor não há necessariamente uma correlação direta entre o fenótipo e a célula de origem (Melchor et al., 2014).

O câncer de mama é caracteristicamente uma doença bastante heterogênea, tanto do ponto de vista morfológico, quanto em relação ao microambiente tumoral, a expressão protéica e ao genoma (Badve et al., 2015), o que dificulta o diagnóstico, o prognóstico e a abordagem terapêutica utilizada (Koren e Betires-Alj, 2015). Existem duas hipóteses para explicar a heterogeneidade intertumoral, que dá origem aos diferentes subtipos de câncer de mama: a primeira é baseada na célula de origem, onde cada tipo de tumor se originaria de um tipo celular; a segunda considera que a célula de origem pode ser a mesma para todos os subtipos, porém o fenótipo de cada um é determinado por eventos genéticos e epigenéticos adquiridos (Polyak, 2007). Esta última hipótese é corroborada por Melchor e cols. (2014), que, em trabalho utilizando modelos murinos geneticamente modificados, demonstraram que a lesão genética inicial é o determinante primordial do perfil molecular do tumor subsequente, sugerindo que a formação tumoral depende dos eventos iniciais da carcinogênese.

A heterogeneidade intratumoral pode ser parcialmente explicada pela teoria da célulatronco cancerosa (CTC), que sugere que uma subpopulação de CTCs pode se auto-renovar e se diferenciar em uma determinada linhagem para originar a massa de células do tumor (Zhang M. et al., 2015). Ela ainda pode ser resultado do acúmulo de mudanças genéticas cumulativas nas células-tronco tumorais ou então da combinação do tipo celular que originou o tumor, alterações genéticas e epigenéticas adquiridas, e sinais parácrinos das células circundantes (Polyak, 2007). Em estudo de Zhang e cols. (2015), foi sugerido que a heterogeneidade intratumoral não deve ser exclusiva de diferenças genéticas entre várias subpopulações, mas que modificações epigenéticas devem permitir que essas subpopulações se desenvolvam em células com diferentes potenciais tumorigênicos, e que elas se tornem diferentes tipos celulares, que darão origem a tumores fenotipicamente diferentes.

1.1.3 Genética e biologia molecular do câncer de mama

A maioria dos casos de câncer de mama é esporádica, porém 5 a 10% de todos os casos está relacionada à herança de mutações genéticas e tem como característica o acometimento de mulheres jovens (Amendola e Vieira, 2005). Atualmente, o câncer de mama é considerado hereditário se há pelo menos três casos de câncer de mama/ovário na família, dois casos de carcinoma mamário em parentes próximos antes dos 50 anos e ancestralidade judia Ashkenazi com câncer de mama triplo negativo antes dos 60 anos (Branham et al., 2016). Variantes de suscetibilidade incluem mutações raras de alto risco, variantes mais moderadas, e variantes de suscetibilidade comuns que possuem pouca penetrância (Nickels et al., 2013).

Os genes BRCA1 e BRCA2 são supressores tumorais e cerca de 30% dos cânceres de mama são atribuídos a mutações nesses genes (Sana e Malik, 2015). A expressão do gene *BRCA1* é importante para o processo de reparo do DNA, ativação de pontos de checagem do ciclo celular e manutenção da estabilidade cromossômica (Sotiriou e Pusztai, 2009). De forma similar, a presença na célula do gene *BRCA2* não funcional faz com que o reparo de quebras na dupla fita de DNA seja propenso ao erro, levando à instabilidade cromossômica (Sana e Malik, 2015). Além de BRCA1/2, outros genes importantes estão associados com alta penetrância familial do câncer de mama, como o TP53, o gene mais frequentemente alterado nos cânceres humanos (Yadav et al., 2015). Ele é herdado de forma autossômica dominante e mutações nesse gene levam a um risco de aproximadamente 49% de desenvolvimento de câncer de mama em mulheres com 60 anos (Scalia-Wilbur et al., 2016).

A proteína p53, codificada pelo gene TP53, é crucial na preservação da integridade genômica, agindo em processos de hipóxia e ativação de oncogenes, além de interagir com proteínas que controlam a morte celular programada (Vymetalkova et al., 2015). Outro gene relevante é o PTEN, um supressor tumoral que, quando perdido em tumores mamários, aumenta a atividade da via PI3K (Liu et al., 2016), que por sua vez está envolvida na regulação do crescimento e da diferenciação celular e do desenvolvimento (Vara et al., 2004). Em trabalho publicado por Bubien et al. (2013), foi encontrado um risco cumulativo de câncer de mama de 77% aos 70 anos para mulheres com mutações no PTEN.

Entre os genes relacionados com moderada penetrância familial, tem sido observadas mutações de perda de função no PALB2 em 0,6% a 3,9% das famílias com histórico de câncer de mama (Antoniou et al., 2014). A proteína de PALB2 interage com as proteínas dos genes BRCA1/2, ativando a resposta de reparo do DNA e garantindo que não haja erros durante a recombinação homóloga – a interrupção dessa interação, geralmente devido a mutações, aumenta significativamente o risco de câncer (Poumpouridou et al., 2015). Aos 70 anos, o risco de desenvolvimento do carcinoma de mama em mulheres portadoras de mutação em PALB2 vai de 33%, para mulheres que não possuem parentes afetados, até 58% para mulheres com dois parentes de primeiro grau diagnosticados com câncer de mama aos 50 anos (Antoniou et al., 2014). De forma similar, o produto do gene CHEK2 é ativado quando há dano no DNA e interage com outras proteínas, como a dos genes BRCA1 e TP53, interrompendo o ciclo celular (Sheikh et al., 2015; Massink et al., 2015). A anormalidade de CHEK2 deve-se à mutação 1100*delC, que gera um códon de parada prematura (Massink et al., 2015) e aumenta em duas vezes o risco de câncer de mama nos seus portadores (CHEK2 Breast Cancer Case-Control Consortium, 2004). O gene ATM também é atuante no reparo de quebras na dupla fita de DNA e um dos seus papéis é a fosforilação do BRCA1 (Scalia-Wilbur et al., 2016). Mutações em heterozigose em ATM levam a um risco duas vezes maior de desenvolvimento de carcinoma mamário na população em geral, o que chega a ser cinco vezes maior em mulheres com menos de 50 anos de idade (Economopoulou et al., 2015). Juntamente com CHEK2, PALB2 e BRCA1/2, ATM é indicado como um dos melhores candidatos para inclusão em um painel multigênico para câncer de mama, uma vez que esses genes têm alta prevalência e penetrância de mutações patogênicas (Lerner-Ellis et al., 2015).

Além dos genes de alta e moderada penetrância familial, mutações em outros genes têm sido identificadas, o que pode ajudar na compreensão da herdabilidade do câncer de mama. Em 2015, Michailidou e colaboradores identificaram 15 novas variantes, ajudando a explicar, juntamente com os demais *loci*, aproximadamente 16% do risco familial de câncer de mama. No mesmo ano, Aloraifi e cols., em estudo de uma população irlandesa, identificaram novas variantes potencialmente patogênicas em 30 outros genes, o que poderia explicar a etiologia de 35% do mecanismo de herança dessa doença.

O câncer de mama esporádico, por sua vez, é assim considerado quando as mudanças genéticas acontecem nas células somáticas com o passar do tempo, tendo sido proposto que em um terço desses tumores há redução da expressão do BRCA1, sem que este apresente qualquer mutação (Branham et al., 2016). Apesar disso, já há descrição na literatura de carcinoma de mama esporádico em uma mulher com mutação na linhagem germinativa de BRCA1 (Curtit et al., 2015), o que torna a biologia desse tipo de câncer ainda mais complexa. O gene ATM também possui variantes identificadas em pacientes sem história familiar de câncer de mama, de acordo com um estudo realizado em uma amostra da população brasileira (Mangone et al., 2015). Muitos estudos vêm sendo realizados a fim de melhor compreender as causas do início e da progressão do câncer de mama, incluindo aqueles com RNAs não codificantes, como os micro RNAs (miRNAs) (Danza et al., 2014; Bastos et al., 2014), e aqueles que levam em conta a metilação dos genes (Zhang e Long, 2015; Zhu et al., 2015), demonstrando que as mudanças epigenéticas estão ganhando cada vez mais importância nesse contexto.

O gene HER2, receptor 2 do fator de crescimento epidermal humano, está envolvido com a regulação do crescimento e desenvolvimento da mama normal (Yarden, 2001). Ele está envolvido com crescimento e progressão tumoral e encontra-se amplificado em cerca de 30% dos cânceres de mama, além de estar positivamente correlacionado com metástase (Li et al., 2004). O produto de HER2 é uma potente oncoproteína que, quando ativada, resulta em transdução de sinal para o núcleo, levando a ativação de genes e à divisão celular. Devido ao seu elevado nível nas células tumorais, baixos níveis nos tecidos normais, localização acessível na superfície celular e seu papel na carcinogênese, HER2 é utilizado com alvo terapêutico, como por exemplo, a utilização de anticorpos monoclonais contra o seu domínio extracelular (Yarden, 2001). Quimiocinas também possuem papel importante no câncer de mama, como o fator 1 α derivado de célula estromal (SDF-1 α , também conhecida como CXCL12), cuja interação com seu receptor (CXCR4) leva à polimerização de actina e formação de pseudopodia em câncer de mama, induzindo respostas quimiotáticas e invasivas (Müller et al., 2001). Já foi demonstrado que HER2 aumenta a expressão do receptor de quimiocina CXCR4, por meio do aumento da taxa traducional, o que é necessário para invasão in vitro induzida por SDF-1 α (Li et al., 2004).

1.1.4 Mecanismos moleculares da metástase

A metástase é a diferença fundamental entre o crescimento celular benigno e o maligno, representando o principal problema clínico do câncer (Pecorino, 2012). Tal mecanismo consiste no desprendimento de células do tumor primário que podem invadir os tecidos adjacentes e colonizar sítios distantes, e é considerado uma das marcas do câncer, por ser característico de todos os tipos de tumores humanos (Hanahan e Weinberg, 2000). A metástase é responsável pela morte de aproximadamente 90% dos pacientes com câncer (Reymond et al., 2013), e esse alto índice é reflexo da obstrução física, competição com células normais por nutrientes e oxigênio, além de interferência da função normal do órgão devido ao espalhamento das células cancerosas (Pecorino, 2012). Apesar do alto índice de mortalidade associado à metástase, esse processo é, de uma forma geral, ineficiente, uma vez que as células tumorais falham em executar um ou mais passos necessários para a cascata metastática (Labelle e Hynes, 2012).

Classicamente, a metástase pode ser dividida em alguns passos básicos sequenciais: invasão local, intravasamento, circulação, extravasamento e colonização (Nguyen et al., 2009). As células podem migrar em conjunto ou individualmente, sendo as estratégias de migração celular individual classificadas como amebóides ou mesenquimais (Brábek et al., 2010). Durante a invasão e a migração, células cancerosas se destacam do tumor primário e invadem o tecido adjacente saudável, através da secreção de enzimas líticas que degradam a matriz extracelular (MEC) (Leber e Efferth, 2009). Dentre elas, destacam-se as metaloproteinases de matriz (MMPs), que exercem atividade proteolítica, facilitando a metástase, além de tornarem fatores de crescimento mais acessíveis às células cancerosas e de modularem interações célula-célula e célula-MEC (Gialeli et al., 2011). No contexto do câncer de mama, a transcrição de MMP-1 e MMP-13, por exemplo, é aumentada como consequência de mutação no gene supressor tumoral p53, além de muitas outras MMPs estarem associadas com as demais fases de progressão tumoral (Koujan et al., 2015). Ainda nessa etapa de invasão, acredita-se que o mecanismo de transição epitélio-mesenquimal (*EMT*) tem um importante papel na conversão das células tumorais em uma população migratória (Bravo-Cordero et al., 2012).

O intravasamento consiste na entrada da célula tumoral em um vaso sanguíneo ou linfático, o que requer a aderência da célula cancerosa à face estromal do vaso, o rompimento das junções endoteliais e a migração até o lúmen do vaso (Pecorino, 2012). Alguns fatores, como o fator de crescimento transformante β (TGF- β) e o fator de crescimento endotelial vascular (VEGF), diminuem a função da barreira endotelial, o que aumenta o número de células que entram no vaso (Reymond et al., 2013). Em trabalho publicado em 2014, Pignatelli e colaboradores demonstraram que células de câncer de mama primário que expressam uma isoforma associada a invasão de Mena (Mena^{INV}), proteína regulatória do citoesqueleto, são mais propensas à migração transendotelial, promovida pela presença de macrófagos. Porém, ainda há discussão sobre o intravasamento ser um processo passivo ou ativo e se as células tumorais atravessam a barreira endotelial pelas junções célula-célula (paracelular) ou pelo corpo celular (transcelular) (Zervantonakis et al., 2012). A maioria das células cancerosas parece adotar a rota paracelular in vitro, porém ainda não é claro qual rota seria adotada *in vivo* e se a escolha de um tipo de migração ou de outro depende do leito vascular ou do tipo de câncer (Reymond et al., 2013). Alternativamente, em uma linhagem celular de câncer de mama, Khuon e cols. (2010) demonstraram que as células tumorais invasivas medeiam o intravasamento transcelular ativo, pela ativação local de miosina endotelial.

Após o intravasamento, as células do tumor alcançam a circulação, sendo então denominadas células tumorais circulantes (CTCs), onde ficam expostas ao fluxo sanguíneo e vulneráveis à morte celular induzida por estresse ou por células imunes (Labelle e Hynes, 2012). Em parte, o sucesso da sobrevivência das CTCs na circulação se deve à ação de plaquetas e fibrinogênio, que impedem a remoção intravascular das células tumorais pelas células *natural killers* (Palumbo et al., 2005), uma vez que as plaquetas se agregam em torno das células cancerosas - estas, por sua vez, aumentam a coagulação das plaquetas, o que

aumenta a metástase (Reymond et al., 2013). As células tumorais podem ativar as plaquetas pela produção de trombina, induzida pelo fator tecidual associado ao tumor, que está altamente expresso em sítios de hipóxia, como os encontrados no tumor (Konstantopoulos e Thomas, 2009). Tipicamente, as CTCs não permanecem muito tempo na corrente sanguínea, inclusive a maioria das células de carcinoma tem diâmetro muito grande para conseguir passar por pequenos capilares e muitas delas ficam presas no primeiro capilar que encontram, o que ocorre minutos após entrarem na circulação (Labelle e Hynes, 2012).

O extravasamento é o passo menos conhecido da cascata metastática, porém já foi demonstrado que é um processo altamente dinâmico e coordenado e que requer a projeção de protrusões pela camada endotelial (Leong et al., 2014). Além disso, estudo em linhagens celulares mostrou que a superexpressão dos genes Twist e VEGFA aumentou significativamente o número de células tumorais extravasantes, e que estas induzem o agrupamento das células endoteliais e o remodelamento das suas junções celulares no sítio de extravasamento (Stoletov et al., 2010). Kanada e cols. (2014), em estudo com peixezebra, sugeriram que a passagem das células tumorais da circulação para o tecido adjacente pode ocorrer de dois modos: ativamente, pela invasão da parede do vaso; e passivamente, onde massas de células quiescentes seriam eventualmente recobertas por uma nova lâmina de células endoteliais na parede do vaso e haveria, simultaneamente, o gradual desaparecimento da camada endotelial original, permitindo que o grupo de células cancerosas se espalhe pelo tecido fora do vaso.

O próximo e último passo da cascata metastática é a colonização, que se refere à formação de um novo tumor. As células que extravasaram tanto podem morrer, quanto entrar em dormência ou formar colônias pela proliferação contínua – o que vai determinar o destino celular são as interações das células tumorais com os elementos do parênquima do órgão alvo (como células estromais), a vascularização e a vigilância imune (Shibue e Weinberg, 2011). O acúmulo de neutrófilos, por exemplo, age como um nicho prémetastático no tecido alvo para a disseminação metastática (Wculek e Malanchi, 2015). A habilidade das células tumorais de formarem colônias em órgãos distantes também depende em grande parte da sua capacidade de proliferação, o que requer interações de adesão de integrinas com a matriz extracelular, já tendo sido demonstrado que a fibrina é eficiente no suporte à formação de colônias (Knowles et al., 2015).

1.1.5 Transição epitélio-mesenquimal

A transição epitélio-mesenquimal (*EMT*) é o processo pelo qual as células epiteliais perdem as suas junções e suas polaridades apico-basais, reorganizam seus citoesqueletos, sofrem mudanças nos seus programas de sinalização que definem seus formatos e reprogramam sua expressão gênica, aumentando a sua motilidade e desenvolvendo um fenótipo invasivo (Lamouille et al., 2014). Esse processo simula o evento de gastrulação normal do desenvolvimento e pode ser classificado em três subtipos: o primeiro está associado com o desenvolvimento embrionário original e também ocorre durante o crescimento pós-natal; o segundo subtipo é iniciado em resposta a ferimentos e gera fibroblastos para reconstruir tecidos feridos; o terceiro é justamente o subtipo oncogênico, que recapitula características típicas da *EMT* do desenvolvimento, porém de forma menos ordenada e coordenada (Samatov et al., 2013).

A perda da expressão de E-caderina é considerado um evento fundamental da *EMT*, assim como a perda da polaridade celular (Thiery et al., 2009). A mudança na expressão de genes relacionados às junções aderentes previne que haja formação *de novo* de junções célula-célula e resulta na perda da função da barreira epitelial (Lamouille et al., 2014). A reorganização do citoesqueleto, por sua vez, permite que a célula adquira uma motilidade direcional, gerando projeções de membrana que facilitam o movimento celular e que possuem função proteolítica, o que leva à degradação da matriz extracelular e facilita a invasão (Lamouille et al., 2014).

A ativação do programa *EMT* nas células cancerosas facilita a progressão do câncer por diversos mecanismos: iniciando processos invasivos e antiapoptóticos que levam à metástase; gerando células estromais ativadas que levam à progressão do câncer por alterações bioquímicas e estruturais no microambiente tumoral; e estimulando uma maior malignidade associada ao fenótipo de célula-tronco cancerosa (Nisticò et al., 2012). A regulação da *EMT* conta ainda com o envolvimento de diversos fatores de transcrição, dentre os quais o ZEB1, ZEB2, Snail, Slug e Twist são os mais bem caracterizados (Samatov et al., 2013).

Um fato interessante, é que os tumores metastáticos apresentam um aumento na expressão de E-caderina, quando comparada com a expressão aberrante ou perda da sua expressão nos tumores primários, levando-os a se assemelharem ao fenótipo epitelial do tumor de origem (Chao et al., 2010). Tal característica deve-se em grande parte ao processo inverso da *EMT*, a transição mesenquimal-epitelial (*MET*), que é crítica para os últimos estágios da cascata metastática (Yao et al., 2011).

1.1.6 Disseminação linfática e hematógena

A disseminação metastática pode ocorrer pelo intravasamento das células tumorais em vasos linfáticos ou sanguíneos, sendo a decisão por um ou outro caminho influenciada por restrições físicas impostas aos tumores invasivos, assim como pela acessibilidade das vasculaturas sanguínea e linfática (Wong e Hynes, 2006). A maioria dos cânceres epiteliais primeiramente desenvolve metástase pela dispersão via vasos linfáticos até seus linfonodos de drenagem – inclusive, a detecção de metástases no linfonodo sentinela (primeiro linfonodo que o tumor atinge) tem grandes implicações prognósticas (Karaman e Detmar, 2014).

Os vasos linfáticos são estruturas de fundo cego perfeitamente adaptadas para captação de fluido, partículas e células, e são compostas de uma única camada endotelial e uma membrana basal descontínua (Nathanson, 2003; Rahman e Mohammed, 2015). Eles transportam o excesso de fluido extravascular, que passa pelos linfonodos e retorna para o sistema venoso, o que torna os linfonodos que estão drenando o tumor primário bastante propensos a desenvolver um tumor secundário (Rizwan et al., 2015). Além disso, os linfonodos possuem sua própria vascularização, o que possibilita a sua invasão por células tumorais que escapam dos vasos sanguíneos (Rizwan et al., 2015). Diferentemente destes, os vasos linfáticos são altamente permeáveis, o que os torna superiores em permitir a disseminação das células tumorais, sendo, inclusive, a via preferencial de metástase de tumores de mama (Rahman e Mohammed, 2015).

Os vasos sanguíneos, por sua vez, possuem junções interendoteliais típicas, assim como camadas circundantes de pericitos e músculo liso, além de membranas basais, o que

os tornam menos permeáveis às células tumorais (Wong e Hynes, 2006). Apesar disso, os novos vasos gerados por estímulo do tumor são mais propensos à invasão pelas células tumorais, além de serem mais tortuosos, o que permite um acesso mais fácil se comparado aos vasos de tecidos saudáveis (Pecorino, 2012). Os pericitos são importantes para os vasos sanguíneos, pois fornecem a eles suporte estrutural, já tendo sido demonstrado que uma baixa cobertura por essas células está significativamente associada com a presença de metástases distantes (Cooke et al., 2012).

1.2 RNAs longos não codificadores (IncRNAs)

1.2.1 Características e funções dos IncRNAs

Diversos genes codificadores de proteína já são utilizados como biomarcadores em diversos tipos de câncer (Henry e Haynes, 2012; Sana e Malik, 2015), porém eles compõem menos de 3% de todo o genoma, enquanto mais de 85% dele é transcrito (Hangauer et al., 2013). Estudos já indicaram a existência de milhares de transcritos não codificadores expressos no genoma de mamíferos e que esses transcritos compreendem uma rede complexa a partir de ambas as fitas do DNA (Mattick, 2005). Além disso, um número cada vez maior de estudos vem mostrando que processos epigenéticos cooperam com os processos genéticos em todos os estágios do desenvolvimento desse grupo de doenças (Jones e Baylin, 2007). Nesse contexto, os lncRNAs oferecem vantagens sobre as proteínas em algumas formas de regulação epigenética, como a capacidade de se prenderem ao sítio de transcrição, permitindo uma especificidade regulatória (Lee, 2012).

Os IncRNAs são definidos como moléculas de RNA não codificadoras de proteínas maiores que 200 nt, que podem ou não ser poliadeniladas e que podem ser encontradas tanto no núcleo quanto no citosol (Shi et al., 2013). Eles constituem um grupo bastante heterogêneo de RNAs, o que os permite cobrir um amplo espectro de funções celulares e moleculares pelos seus diferentes modos de ação (Gutschner e Diederichs, 2012). O modo como essas moléculas são transcritas e processadas é similar ao dos RNAs codificadores de proteína: a maioria delas é transcrita pela RNA polimerase II (algumas são transcritas pela RNA polimerase III); a maioria é processada, poliadenilada e tem a adição do *cap*-5'; e

grande parte delas tem uma sequência promotora proximal altamente conservada, éxons, íntrons e estruturas secundárias (Nie et al., 2012). Uma característica importante dos IncRNAs é que eles geralmente são expressos de maneira específica, quer seja em relação a uma doença, um tecido ou um estágio do desenvolvimento, o que os torna moléculas atrativas como alvos terapêuticos (Gutschner e Diederichs, 2012).

Os IncRNAs podem ser classificados com base em diferentes critérios, como por exemplo: no tamanho do transcrito; na associação ou na semelhança com genes codificadores de proteínas; e na expressão em diferentes estados biológicos (St. Laurent et al., 2015). Genericamente, os IncRNAs são classificados com base nas suas origens, uma vez que suas funções são difíceis de serem estudadas (Baker, 2011). O consórcio GENCODE, englobado pelo projeto ENCODE, é o maior catálogo de anotação de IncRNAs humanos e classifica esses transcritos com base na sua localização em relação a genes codificadores de proteína (Figura 1): RNAs antisenso são aqueles que interceptam qualquer éxon de um lócus codificador de proteína na fita oposta, ou que possuem evidência de regulação antisenso de um gene codificador de proteína; RNAs longos intergênicos (lincRNAs), como o próprio nome diz, são aqueles provenientes de loci de RNAs não codificadores intergênicos de comprimento maior que 200 pb; os transcritos senso sobrepostos contêm um gene codificador dentro de um íntron na mesma fita; IncRNAs intrônicos, por sua vez, estão contidos dentro de um íntron de um gene codificador de proteína, mas não interceptam nenhum éxon; e os transcritos processados são aqueles que não contêm uma janela aberta de leitura (do inglês open reading frame – ORF) e que não podem ser alocados em nenhuma das outras categorias (Derrien et al., 2012).

Em 2010, Khachane e Harrison propuseram que transcritos que surgem de partes de genes codificadores de proteína, como por exemplo, aqueles compostos somente de regiões não traduzidas (*UTR*) ou que incluem íntrons retidos, fossem também incluídos na categoria de lncRNAs. Recentemente, Braunschweig e cols. (2014) demonstraram que grande parcela dos genes multiexônicos contém um ou mais íntrons retidos, os quais encontram-se enriquecidos em regiões não traduzidas e RNAs não codificantes. Eles são significativamente associados com alto conteúdo CG, comprimento reduzido e sítios de *splicing* 5' e 3' relativamente fracos. O quanto as sequências dos sítios de splicing divergem da sequência consenso determina se os mesmos são fortes ou fracos, e suas posições relativas geram



diferentes modos de *splicing* alternativo, como a retenção de íntrons alternativos (Kornblihtt et al., 2013).



Figura 1 - LncRNAs em relação a genes codificadores de proteínas e suas denominações. E, éxon; UTR, região não traduzida.

Os IncRNAs podem atuar em trans, ou seja, regulando a expressão de genes localizados em domínios cromossômicos distantes ou mesmo em outros cromossomos, ou podem atuar em cis, quando são transcritos do mesmo lócus que seus genes alvo (Guil e Esteller, 2012). Um dos seus possíveis modos de ação é por interferência da transcrição, que pode ser positiva ou negativa (Ard et al., 2013). Eles também podem atuar como guias para o ancoramento de modificadores de cromatina, para alvejar regiões genômicas, ou como chamarizes para sequestrar modificadores de cromatina de sítios específicos do genoma (Han e Chang 2015). Há evidências de que os IncRNAs possuem a capacidade de atuar como esponjas de micro RNAs (miRNAs), afetando a expressão dos alvos desses miRNAs (Kallen et al., 2013), como reguladores pós-transcricionais, influenciando o *splicing* alternativo de mRNAs precursores, como precursores de pequenos RNAs, como estabilizadores de mRNA (Quan et al., 2015) e ainda podem agir como moléculas sinal (quando as células sofrem um determinado estímulo, os IncRNAs vão representar uma especificidade tecidual correspondente) (Malih et al., 2015).

1.2.2 LncRNAs e câncer de mama

Assim como os genes codificadores de proteínas, os IncRNAs podem participar na progressão do câncer como oncogenes ou supressores tumorais (Malih et al., 2015). Alguns deles já foram diretamente associados à iniciação e ao desenvolvimento do câncer de mama, como H19, HOTAIR, MALAT1, LSINCT5, ZFAS1 e UCA1 (Vikram et al., 2014), como pode ser exemplificado na figura 2.



Figura 2 - Visão geral da progressão do câncer de mama. O esquema retrata uma sessão transversal de um ducto mamário dividido em quadrantes que representam cada estágio da progressão do ducto normal até o *IDC*. Características epiteliais e potenciais IncRNAs associados com cada fase também são mostrados (modificada de Shore e Rosen, 2014).

O IncRNA H19 foi primeiramente descrito em 1990, em trabalho realizado por Brannan e colaboradores, no qual foi identificado que a função do gene H19 não era a de síntese protéica e que seu produto deveria atuar como um RNA. Posteriormente foi demonstrado que ele poderia servir como um transcrito primário de miRNA (Cai e Cullen, 2007), além de estar envolvido na promoção da progressão do ciclo celular em células de câncer de mama (Berteaux et al., 2005). Ele encontra-se superexpresso em mais de 70% de adenocarcinomas mamários e sua expressão em células de câncer de mama aumenta o crescimento independente de ancoragem, além de diminuir o tempo de aparecimento de tumor *in vivo* (Lottin et al., 2002).

O HOTAIR é transcrito a partir da fita antisenso do gene *HoxC* e é capaz de interagir com o complexo repressivo policombe 2 (*PRC2*) – uma histona metiltransferase que gera silenciamento epigenético – guiando-o para diferentes sítios no genoma (Hajjari e Salavaty, 2015), sendo o primeiro lncRNA implicado na modulação do estado da cromatina em *trans* (Bergmann e Spector, 2014). Esse lncRNA está altamente expresso em tumores de mama primários e metástase, o que o torna um potencial preditor de eventuais metástases e morte (Malih et al., 2015) e é um dos componentes essenciais da transição epitéliomesenquimal em linhagem celular de câncer de mama (Pádua Alves et al., 2013). O promotor de HOTAIR possui múltiplos elementos funcionais de resposta a estrógeno, o qual induz a sua transcrição, e uma de suas formas de atuação no carcinoma mamário é pela competição por ligação com genes supressores tumorais, como o BRCA1 (Hansji et al., 2014).

Envolvido na regulação do *splicing* alternativo de mRNAs precursores, o MALAT1 é um dos lncRNAs mais abundantes (Arun et al., 2016). Recentemente, o aumento da sua expressão foi sugerido como vital para a metástase em câncer de mama, uma vez que promove migração e invasão celular *in vitro*, além de funcionar como um RNA endógeno competitivo (*ceRNA*), induzindo *EMT* (Chou et al., 2016). Também já foi demonstrado que MALAT1 possui alta expressão em tecidos de câncer de mama triplo negativo, quando comparado com tecidos mamários adjacentes, assim como em linhagens celulares triplo negativas quando comparadas com linhagem de mama normal (Li, L. et al., 2016).

O LSINCT5, descrito em 2010 por Silva e cols., pertence a um grupo de IncRNAs induzidos por estresse e é um dos que se encontra superexpresso em linhagens celulares de câncer de mama. Esse transcrito está envolvido na proliferação celular e a diminuição da sua expressão está associada com a alteração da expressão de genes de quinases, transporte de membrana, associados à carcinogênese, motilidade e estresse (Silva et al., 2011).

O ZFAS1 é um IncRNA que encontra-se expresso em diversos tecidos, porém seus maiores níveis de expressão são encontrados em mama, onde contribui para o desenvolvimento da glândula mamária (Hansji et al., 2014). Ele atua na regulação da via

intracelular da proliferação e produção de leite e também na diminuição da proliferação e da diferenciação das células do epitélio mamário, o que sugere a sua função como supressor tumoral (Askarian-Amiri et al., 2011). No sentido contrário, o lncRNA UCA1 é necessário para manter a atividade tumorigênica em células de câncer de mama metastático (Hiemer et al., 2014). Isso pode ser decorrente da expressão inversa em relação ao micro RNA 143 (miR-143), já que a superexpressão de UCA1 diminui a expressão do miR-143, que por sua vez está relacionado com supressão de proliferação e invasão de células do câncer de mama (Tuo et al., 2015).

Em trabalho recente, Su e cols. (2014) extraíram dados de sequenciamento de RNA (RNAseq) de amostras de câncer de mama do TCGA e traçaram um perfil de IncRNAs expressos potencialmente relevantes para essa doença. A partir dos níveis de expressão dos transcritos, o grupo de pesquisadores identificou a ocorrência de quatro subgrupos: grupo I, relacionado ao subtipo basal-símile de câncer de mama; grupo II, correspondente ao subtipo HER2; e os grupos III e IV, relacionados aos subtipos luminal A e B. Esses resultados revelam que o câncer de mama também pode ser caracterizado molecularmente a partir da avaliação da expressão dos lncRNAs. Com base nessas informações e no recente trabalho publicado pelo nosso grupo, onde foi demonstrado que o IncRNA HOTAIR é requerido para a transição epitélio-mesenquimal em linhagem celular de câncer de mama (Pádua Alves et al., 2013), hipotetizamos que tumores primários de mama e suas respectivas metástases possuem perfis de expressão de IncRNAs que podem tanto caracterizar ambos os tipos tumorais quanto sugerir possíveis moléculas envolvidas com o mecanismo de metástase do câncer.

2 OBJETIVOS

2.1 Objetivo geral

Identificar novos IncRNAs associados com tumores metastáticos.

2.2 Objetivos específicos

2.2.1 Determinar o perfil de expressão de lncRNAs de tumor de mama metastático em cérebro e em linfonodo;

2.1.2 Classificar funcionalmente os IncRNAs preferencialmente expressos nos tumores metastáticos;

2.1.3 Selecionar IncRNAs potencialmente envolvidos com a metástase.

3 MATERIAL E MÉTODOS

3.1 Amostras tumorais

Amostras de tumores primários de mama de 20 pacientes, juntamente com seus correspondentes metastáticos em linfonodo e cérebro (Figura 3), foram obtidas junto ao Banco de Tumores e ao Banco de Tumores de Mama do Hospital das Clínicas de Ribeirão Preto da Universidade de São Paulo. Os tumores, coletados entre os anos de 2008 e 2014, foram imediatamente congelados em nitrogênio líquido e armazenados a -80 °C. Todas as amostras possuem o Termo de Consentimento Livre e Esclarecido (TCLE) devidamente assinado pelos pacientes e foram submetidas a análise histopatológica . O projeto foi aprovado pelo Comitê de Ética em Pesquisa do HCFMRP/USP (nº 829151/2014).



Figura 3 - Amostras tumorais primárias e metastáticas obtidas para o presente estudo.

3.2 Isolamento de RNA total

Para extração do RNA, foram retirados fragmentos de aproximadamente 30 mg das amostras tumorais, os quais foram manualmente macerados utilizando pistilos esterilizados e descontaminados. Foi adicionado nitrogênio líquido aos tecidos para impedir o descongelamento durante a maceração e todo o processo foi realizado em gelo seco. Após a maceração, foi adicionado 1 ml de Trizol (Thermo Fischer Sientific, Waltham, MA, USA) em
cada amostra, a fim de romper as membranas celulares, sendo cada uma então ressuspendida vigorosamente utilizando micropipeta.

Os tecidos em Trizol foram transferidos para tubos de 1,5 ml livres de RNase e adicionou-se 200 μ l de clorofórmio para separação da fase orgânica da aquosa, seguindo-se de homogeneização por vortex e centrifugação a 13.200 RPM por 15 minutos a 4 °C. A fase aquosa foi coletada e transferida para um novo tubo de 1,5 ml livre de RNase (a fase orgânica foi estocada a -80 °C) e a ela adicionou-se 500 μ l de isopropanol, para precipitação do RNA, sendo os tubos homogeneizados por inversão incubados em temperatura ambiente por 15 minutos, ao que se seguiu centrifugação a 13.200 RPM por 10 minutos a 4 °C. A fase aquosa foi descartada, 1 ml de etanol 75% gelado foi adicionado e o tubo foi agitado por 10 segundos em vortex para lavagem do botão de RNA. Após centrifugação a 10.000 RPM por 5 minutos a 4 °C, o sobrenadante foi descartado e o RNA foi lavado com etanol 75% novamente e centrifugado. A fase aquosa foi descartada e o excesso de etanol foi cuidadosamente retirado com o auxílio de micropipeta. Por fim, adicionou-se água ultrapura para ressuspender o botão de RNA (mínimo de 13 μ l) e os tubos foram estocados a -80 °C.

A quantificação do RNA foi realizada em espectrofotômetro NanoDrop (Thermo Fischer Sientific, Waltham, MA, USA). Taxas de absorbância a 260 nm e 280 nm (utilizada para medir a pureza do RNA) e a 260 nm e 230 nm (utilizada para uma medida secundária da pureza) entre 1.7 e 2.0 e entre 1.8 e 2.2, respectivamente, foram utilizadas para avaliação da qualidade da extração. A qualidade do RNA foi acessada com o Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), que realiza a eletroforese em gel em um *chip* e que utiliza o Número de Integridade do RNA (do inglês *RNA Integity Number – RIN*) para descrever a integridade da amostra. Essa ferramenta leva em conta todo o perfil eletroforético e classifica o RNA de acordo com um sistema numérico de 1 a 10, sendo 1 o perfil mais degradado e 10 o mais intacto. Amostras pareadas, com RIN igual ou maior que 6, foram selecionadas para sequenciamento.

3.3 Sequenciamento de RNA

3.3.1 Preparo das bibliotecas

A fim de obter o perfil dos transcritos expressos e de identificar novos transcritos nas amostras tumorais, foi realizado o sequenciamento do RNA total das mesmas pela técnica de RNA-Seq. Foram utilizados 300 ng de RNA para o preparo das amostras, que foi realizado com o TruSeq Stranded Total RNA LT Sample Prep Kit (Illumina Inc.,San Diego, CA, USA), de acordo com o protocolo do fabricante. Primeiramente, o RNA ribossômico (rRNA) foi removido do RNA total por oligos alvo-específicos biotinilados combinados com esferas de remoção de rRNA citoplasmático Ribo-Zero Gold. Após a purificação, o RNA foi fragmentado em segmentos de 200 a 500 pb, utilizando cátions divalentes sob alta temperatura. Os fragmentos foram então utilizados para síntese da primeira fita de DNA complementar (cDNA), usando a enzima transcriptase reversa e primers randômicos, seguindo-se a síntese da segunda fita, onde foram utilizadas as enzimas DNA Polimerase I e RNase H (ribonuclease que degrada a fita de RNA em híbridos DNA-RNA). Após a síntese do cDNA, uma única base adenina (A) foi adicionada a cada fragmento e em sequência foi feita a ligação dos adaptadores. Os produtos foram então purificados e enriquecidos por Reação em Cadeia da Polimerase (PCR) para geração das bibliotecas finais de cDNA. As mesmas foram quantificadas por PCR quantitativa (qPCR) e a qualidade foi avaliada com o Bioanalyzer 2100 (Agilent Technologies, EUA).

3.3.2 Clusterização

A geração de *clusters* foi realizada utilizado o *TruSeq Cluster Generation Kit v5* (Illumina Inc.), de acordo com instruções do fabricante, e consistiu na amplificação das bibliotecas de cDNA, produzindo agrupamentos clonais para subsequente sequenciamento. A clusterização ocorreu no cBot (Illumina, Inc.) em célula de fluxo contendo 8 faixas, que permitem a análise simultânea das amostras. As bibliotecas foram colocadas na célula de fluxo, onde se hibridizaram à camada de oligonucleotídeos imobilizados da superfície da célula, sendo em seguida copiados por uma DNA polimerase de alta fidelidade. Os moldes iniciais foram então desnaturados, deixando apenas as cópias imobilizadas na superfície da célula de fluxo. Em seguida, essas cópias foram amplificadas por amplificação em ponte, onde os moldes se curvam para se hibridizar com oligonucleotídeos adjacentes presentes na camada superficial (Figura 4). Pontes de DNA dupla fita (dsDNA) foram formadas a partir da cópia desses moldes pela DNA polimerase, sendo em seguida desnaturadas para formar duas fitas de DNA fita simples (ssDNA). Tal processo foi repetido em cada molde por ciclos de desnaturação e amplificação isotérmica para criar milhões de agrupamentos clonais. Cada grupo de pontes de dsDNA foi desnaturado e a fita reversa foi removida e, em seguida, os *primers* de sequenciamento foram hibridizados às suas sequências complementares nos adapatadores.



Figura 4 - Representação da amplificação em ponte. As fitas de cDNA se hibridizam à camada de oligonucleotídeos presente na superfície da célula de fluxo, curvando-se em formato de ponte. As fitas são copiadas por DNA polimerase, formando pontes de dupla fita que são subsequentemente desnaturadas. A repetição dos ciclos de amplificação e desnaturação geram milhares de agrupamentos clonais (figura modificada de Shendure e Ji, 2008).

3.3.3 Sequenciamento

O sequenciamento por síntese dos grupos clonais foi realizado em aparelho Genome Analyzer IIx (Illumina, Inc.), utilizando o *TruSeq SBS kit v5* (Illumina Inc.). Os moldes de DNA foram copiados base a base utilizando nucleotídeos associados a um terminador reversível marcado fluorescentemente. Após cada passo de síntese, os agrupamentos foram excitados por um *laser*, levando à emissão de fluorescência da última base adicionada, e então o terminador foi removido, permitindo a incorporação do próximo nucleotídeo. Os sinais fluorescentes de cada ciclo de síntese foram, então, capturados por uma câmera do próprio aparelho.

3.4 Análise dos dados

3.4.1 Chamada de bases

Após o sequenciamento, as imagens obtidas a partir da captura da florescência dos agrupamentos clonais foram analisadas, gerando a intensidade dos grupos e uma estimativa do ruído em cada leitura. Essa análise gerou o arquivo de entrada para a chamada de bases, que utiliza justamente as intensidades e a estimativa do ruído dos agrupamentos para gerar uma sequência de bases lida de cada grupo clonal, um nível de confiança para cada base e se as *reads* passaram pela filtragem. A chamada de bases gerou arquivos ".bcl" (chamada de base binária – do inglês *binary base call*), que foram convertidos em arquivos FASTQ, compatíveis com os algoritmos de alinhamento, utilizando o programa CASAVA 1.8 (Illumina, Inc.). Para análise da qualidade dos dados do sequenciamento, foi utilizado o programa FastQC (Babraham Institute, 2014).

3.4.2 Alinhamento das sequências

As sequências foram alinhadas contra o genoma de referência GRCh37.p13 (GENCODE) com o algoritmo Tophat v. 2.0.3 (Trapnell et al. 2009). Essa é uma boa ferramenta para alinhamento de *reads* para identificação de sítios de *splicing* entre éxons. O arquivo de saída desta análise é em formato BAM (forma binária, indexada e comprimida do SAM)

3.4.3 Análise da expressão gênica

Em seguida, utilizou-se o HTSeq (Anders; Pyl; Huber, 2014) da linguagem de programação Python, o qual permite o desenvolvimento rápido de roteiros para o processamento e análise de dados de sequenciamento de larga escala. Desse pacote, utilizou-se a ferramenta HTSeq-count para contar, para cada gene, quantas *reads* alinhadas se sobrepuseram a éxons (aquelas que se alinharam em múltiplas posições ou que se sobrepuseram a mais de um gene foram desconsideradas).

Essas contagens foram, em seguida, utilizadas para análise de expressão diferencial com o pacote DESeq2 (Love; Huber; Anders, 2014) do Bioconductor, que permite uma

análise mais quantitativa de dados de RNA-seq pelo melhoramento das estimativas de dispersão e *fold change*.

3.4.4 Caracterização funcional dos IncRNAs

Para a caracterização funcional dos IncRNAs diferencialmente expressos nas amostras tumorais, foi realizada a busca da função dessas moléculas na literatura e, adicionalmente, foram utilizadas as ferramentas IncRNA2function (http://mlg.hit.edu.cn/Incrna2function), IncRNAtor (http://lncrnator.ewha.ac.kr/) e LNCipedia (http://www.Incipedia.org/).

3.5 Validação de expressão de IncRNA por PCR semi-quantitativa

Foi selecionado o IncRNA LINCO0472 para validação por PCR semi-quantitativa. Esta técnica é uma técnica simples e rápida, e que permite a avaliação qualitativa da expressão gênica pela comparação entre as amostras.

Primeiramente, foi sintetizado cDNA de seis pares tumorais (três mama-linfonodo e três mama-cérebro), utilizando *primers* randômicos e a enzima *Multiscribe* (Thermo Fischer Sientific, Waltham, MA, USA), a partir de 500 ng de RNA de cada amostra. Em seguida, 20 µl dos cDNAs foram serialmente diluídos em água água estéril, gerando sete diluições (1:4, 1:8, 1:16, 1:32, 1:64 e 1:128). De cada diluição foi utilizado 2 µl para avaliação da expressão do LINC00472, utilizando os *primers:* F- 5'- GGGCTGCGGGAACATTAAAA-3' e R- 5'- GTCCACATCCTTCTCCTATC-3'. Como endógeno, foi utilizado o gene da β-tubulina (TUBB) (*primers:* F- 5'-TCAACACTTTCTTCAGTGAAACG-3' e R- 5'- AGTGCCAGTGCGAACTTCATC -3').

Para verificação da amplificação da sequência do IncRNA e do endógeno, foi utilizado 5 μl do produto de PCR para eletroforese em gel de agarose 1% e a visualização foi feita em transluminador UV (UVP, USA).

4 RESULTADOS

4.1 Dados clínicos das amostras tumorais

Todas as amostras tumorais obtidas foram avaliadas pelo serviço de patologia do HCFMRP e os dados clínicos e histopatológicos dos tumores de mama estão contidos na tabela 1.

Tumor de mama	n (20)	%		
Idade da paciente				
<50 anos	8	40,00		
>50 anos	12	60,00		
Tipo de Câncer				
CDI	16	80,00		
Outros	4	20,00		
Receptor de Estrogênio				
Positivo	13	65,00		
Negativo	7	35,00		
Receptor de Progesterona				
Positivo	9	45,00		
Negativo	11	55,00		
Receptor do Fator 2 Epidermal Humano (<i>Her2/neu</i>)				
Positivo	13	65,00		
Negativo	7	35,00		
NI	1	5,00		
Classificação Molecular				
Luminal A e B	5	25,00		
HER2	13	65,00		
Basal – TN	1	5,00		
NI	1	5,00		

Tabela 1 - Características histológicas e clínico-patológicas dos tumores de mama.

CDI, Carcinoma Ductal Invasivo; TN, Triplo Negativo, Ausência do receptor de estrogênio, progesterona e Fator 2 Epidermal Humano; NI, Não Informado.

4.2 Qualidade dos RNAs para sequenciamento

Todas as amostras tiveram seu RNA extraído e avaliado quanto à sua qualidade (Tabela 2). A partir desse resultado, foram selecionados os pares com RNAs mais íntegros possíveis para posterior sequenciamento.

Amostra tumor	DIN	Amostra metastática	DIN	Órgão da
primário	RIN	pareada	KIN	metástase
PM1	-	MC1	6.60	
PM2	7.70	MC2	9.00	- / .
PM3	8.90	MC3	5.50	Cérebro
PM4	-	MC4	7.80	
PM5	6.00	MC5	8.30	
PM6	7.70	ML1	4.20	
PM7	-	ML2	3.00	
PM8	7.50	ML3	3.70	
PM9	5.90	ML4	2.70	
PM10	5.40	ML5	6.70	
PM11	2.60	ML6	5.90	
PM12	2.60	ML7	3.100	
PM13	3.00	ML8	7.20	Linfonodo
PM14	6.00	ML9	7.50	
PM15	3.100	ML10	3.90	
PM16	4.50	ML11	7.00	
PM17	7.80	ML12	6.30	
PM18	7.30	ML13	7.100	
PM19	5.30	ML14	7.40	
PM20	6.20	ML15	6.90	

Tabela 2 - Avaliação da integridade dos RNAs extraídos das amostras tumorais.

RIN, Número de Integridade do RNA; PM, tumor primário de mama; MC, tumor metastático em cérebro; ML, tumor metastático em linfonodo.

A partir da análise dos RINs, foram selecionados os pares PM2 – MC2 e PM5 – MC5, correspondentes aos tumores de mama que metastizaram para o cérebro, sendo eles dos subtipos luminal e Her2, respectivamente. Os pares PM17 – ML12, PM18 – ML13 e PM20 – ML15 correspondem aos tumores primários de mama que metastizaram para linfonodo e os subtipos moleculares dos carcinomas primários são, respectivamente: triplo negativo, luminal e Her2.

4.3 Análise do RNA-seq

4.3.1 Qualidade do sequenciamento

A qualidade do sequenciamento foi acessada a partir do escore de qualidade Phred (escore Q), que reflete a precisão da chamada de bases, ou seja, indica a probabilidade de determinada base ter sido incorretamente chamada pelo sequenciador. Em todas as amostras sequenciadas, o escore Q estava acima de 30, o que indica que a probabilidade da chamada de base estar correta é de 99,9%.

4.3.2 Mapeamento

O sequenciamento das amostras tumorais gerou mais de 31.000 *reads* por amostra, sendo mais de 84% delas mapeadas contra o genoma de referência (Tabela 3).

Tabela 3 - Total de *reads* geradas pelo RNA-seq das amostras tumorais primárias e metastáticas e quantidade de *reads* mapeadas contra o genoma de referência GRCh37.p13.

Amostra	Total de <i>reads</i>	Reads mapeadas	<i>Reads</i> mapeadas (%)
MC2	61.160.613	59.377.307	97,1
MC5	36.335.915	34.966.789	96,2
ML12	33.653.331	31.332.027	93,1
ML13	43.908.604	41.676.353	94,9
ML15	39.621.444	37.667.088	95,1
PM2	66.716.734	64.333.825	96,4
PM5	40.016.389	33.794.185	84,5
PM17	32.902.987	31.020.038	94,3
PM18	41.227.383	39.488.143	95,8
PM20	38.769.275	35.125.650	90,6

Nos tumores primários de mama pareados com suas respectivas metástases em linfonodo, aproximadamente 55% das *reads* mapearam em regiões de RNAs não codificadores de proteína e cerca de 35% correspondeu a RNAs codificadores de proteína (Figura 5). Os tumores primários de mama e seus tumores metastáticos em cérebro correspondentes apresentaram proporções similares de RNAs codificadores e não codificadores de proteína, tanto entre si quanto em relação ao primeiro grupo. Além disso, a quantidade de transcritos que pertencem a um lócus de pseudogene ou a outras classes menores de RNAs foi bem similar entre os três tipos de tumor.



Figura 5 - Classes de RNAs às quais pertencem as *reads* mapeadas resultantes do RNA-seq das amostras tumorais. Tumores primários de mama (A) pareados com metástases em linfonodos (B) e tumores primários de mama (C) pareados com metástases em cérebro (D).

4.3.3 Análise dos IncRNAs expressos nas amostras tumorais

A partir dos dados do mapeamento, foi observado que os RNAs não codificadores foram os transcritos mais representados dentre todos os expressos e, dentre os ncRNAs, o RNAs longos não codificadores formaram a classe mais abundante. Estes, por sua vez, foram divididos em seis categorias, de acordo com a classificação do GENCODE, sendo elas: antisenso, senso sobreposto, senso intrônico, ncRNA sobreposto a 3', lincRNA e íntron retido. Analisando as classes de lncRNAs, observou-se que elas apresentaram abundâncias de transcritos bem similares entre as amostras tumorais (Figura 6).



Figura 6 - Classes de IncRNAs encontradas nas amostras tumorais após o mapeamento. Tumores primários de mama (A) e seus correspondentes metastáticos em linfonodo (B); tumores primários de mama (C) e seus correspondentes metastáticos em cérebro (D).

Em todos os casos, os íntrons retidos foram os transcritos predominantes, representando quase 60% de todos os lncRNAs expressos nas amostras primárias e metastáticas. Além desses, os transcritos antisenso e intergênicos foram bem representados, com média de 19,5% e 18,5% dos lncRNAs, respectivamente.

A partir da observação de que as amostras tumorais são bastante homogêneas quanto à quantidade de transcritos não codificadores longos, foram construídos diagramas de Venn para verificar se esses lncRNAs eram os mesmos expressos em todas as amostras primárias e nas amostras metastáticas (Figura 7). Observa-se que grande parte dos transcritos, tanto nos pares de mama-linfonodo (Figura 7A) quanto nos pares mama-cérebo (Figura 7B), são compartilhados entre os tumores primários e metastáticos (aproximadamente 16.700 e 17.850, respectivamente), o que mostra mais uma vez que esses dois tipos tumorais são muito semelhantes entre si e o que os diferencia minimamente é uma parcela pequena de transcritos.



Figura 7 - LncRNAs comuns e exclusivos entre os pares de (A) tumores primários de mama e respectivas metástases em linfonodo e (B) tumor primário de mama e respectivas metástases em cérebro.

Também foram construídos diagramas para ilustrar a quantidade de IncRNAs comuns e exclusivos entre os tumores primários e entre os metastáticos (Figura 8). Nas amostras de tumor primário de mama, mais de 16.700 transcritos são comuns entre todos os tumores,

mesmo que os cinco sejam de subtipos diferentes (PM2 e PM18 do subtipo luminal, PM5 e PM20 do Her2 e PM17 do subtipo triplo negativo) (Figura 8A). A amostra PM2 possui o maior número de transcritos exclusivos, enquanto que a PM5 é a que possui a menor quantidade desses transcritos.

Os tumores metastáticos apresentaram um número similar aos tumores primários de transcritos comuns expressos em todos os cinco (Figura 8B). As amostras MC2 e ML12 são as que possuem as maiores quantidades de transcritos exclusivos, com 1.139 e 1.198 respectivamente. O número de IncRNAs comuns aos tumores metastáticos em linfonodo axilar é praticamente o dobro (543) do número de transcritos não codificadores longos comuns aos tumores primários de mama (269) que originaram essas metástases. No caso dos tumores metastáticos em cérebro, o número de IncRNAs compartilhados entre eles é aproximadamente 58% maior (585) que o número de transcritos comuns entre os tumores de mama que os originaram (369).



Figura 8 - LncRNAs comuns e exclusivos entre as amostras de tumor primário de mama (A) e de tumor metastático de mama (B). Cores iguais correspondem a um par tumor primário-tumor metastático.

Esses resultados apontam que as amostras primárias e suas respectivas metástases, assim como as amostras primárias e metastáticas, independente de pareamento, são muito

semelhantes. Tal semelhança não se restringe à representatividade de cada classe de lncRNA, mas também se refere à quantidade de transcritos compartilhados entre os tumores.

Adicionalmente, foi realizada a busca pela função dos IncRNAs expressos nas amostras tumorais em bases de dados (IncRNAdb, LNCipedia, GeneCards, PubMed), porém verificouse que a grande maioria deles ainda não possui nenhuma descrição funcional. Apesar disso, puderam ser encontrados alguns transcritos bem expressos nas amostras tumorais que já estão descritos na literatura como envolvidos no câncer de mama, alguns deles estão contidos na figura 9.







Figura 9 - Expressão de IncRNAs já conhecidos em câncer de mama nas amostras tumorais. À esquerda das colunas de tumores metastáticos (ML e MC) encontram-se as colunas dos seus tumores primários correspondentes.

4.3.4 Análise da expressão diferencial dos IncRNAs

Após a verificação de quais eram os IncRNAs que estavam sendo expressos nas amostras tumorais e de como eles estavam distribuídos em cada uma, foi realizada a análise dos transcritos que se encontravam diferencialmente expressos entre as amostras de tumor primário e as suas respectivas metástases, utilizando o pacote DESeq2 do Bioconductor.

Nos tumores primários de mama e nas suas metástases correspondentes em linfonodo, foi encontrado um número similar de IncRNAs mais expressos – 59 e 54, respectivamente (Anexos A e C). Metade dos transcritos mais expressos nos tumores primários corresponde a íntrons retidos, 24% são transcritos antisenso, 20,4% são lincRNAs e 5,6% são transcritos senso intrônicos, sendo os transcritos RP11-1002K11.1-001 (lincRNA), COL1A2-002 e STEAP2-008 (íntrons retidos) os mais expressos nessas amostras. Nas metástases em linfonodo, a classe de lncRNA mais abundante foi a dos lincRNAs, com 37,3%, seguida dos íntrons retidos, com aproximadamente 34% dos transcritos, dos lncRNAs antisenso com 25,4% e, por fim, dos transcritos sobrepostos a 3' e dos senso intrônicos, que juntos englobam 3,4% dos lncRNAs mais expressos. Os transcritos CD22-011, RP11-297B17.3-001 e LINC00861-001 foram os mais diferencialmente expressos em linfonodos (Figura 10A), quando comparados com suas amostras de mama pareadas, sendo eles das classes íntron retido, antisenso e lincRNA, respectivamente.

O grupo dos pares de tumores de mama pareados com metástases em cérebro apresentou um número menor de lncRNAs diferencialmente expressos, quando comparado ao grupo mama-linfonodo: os tumores primários possuem 18 transcritos altamente expressos e suas respectivas metástases têm 14 (Anexos B e D). Os íntrons retidos representam pouco mais da metade (55,6%) dos lncRNAs mais expressos nas amostras de tumor primário, os transcritos antisenso correspondem a 27,8% e os lincRNAs a 16,6% dos RNAs mais expressos. Dentre eles, o COL6A3-009, o TBX5-AS1-001 e o RP11-40109.4-001 se sobressaem como os mais diferencialmente expressos nas amostras de mama, sendo o primeiro um íntron retido e os dois últimos transcritos antisenso.





log2 fold change

Figura 10 - Expressão diferencial de IncRNAs entre as amostras de tumor de mama primário e seus correspondentes metastáticos. Em azul, IncRNAs mais expressos nas metástases; em vermelho, IncRNAs mais expressos nos tumores primários.

Assim como nas amostras metastáticas em linfonodo, nas amostras de cérebro os lincRNAs são os mais numerosos (43%), já o restante dos transcritos se dividem em 28,6% de íntrons retidos, 21,4% de transcritos senso sobrepostos e 7% de ncRNAs sobrepostos a 3'. Os lncRNAs SOX2-OT-008 e RP11-731J8.2-001 são os mais expressos nessas amostras metastáticas (Figura 10B) e ambos pertencem à classe dos transcritos senso sobrepostos.

É interessante apontar que os tumores primários de mama, independente da via metastática (hematógena ou linfática) que seguiram posteriormente ou dos seus subtipos moleculares, possuem doze transcritos diferencialmente expressos em comum, como o RP11-1002K11.1-001, o TBX5-AS1-002 e o RP11-401O9. Os tumores metastáticos em linfonodo e em cérebro, contrariamente, não possuem nenhum lncRNA diferencialmente expresso comum a ambos.

Adicionalmente, foi realizada a análise de interação entre proteínas e os três IncRNAs mais expressos de cada grupo amostral, utilizando a ferramenta IncRNAtor. Foi observado que a maioria dos IncRNAs interage com proteínas envolvidas com o processamento de mRNA e com *splicing* alternativo (Anexo F). Também foram identificadas interações com proteínas envolvidas em vias de sinalização importantes no contexto do câncer (vias de sinalização de EGFR, de TGF-β e Notch).

Em seguida, os IncRNAs diferencialmente expressos foram utilizados para análise de agrupamento hierárquico não supervisionado dos dados de expressão (Figura 11). As amostras de mama formaram um grupo, as metástases em cérebro formaram outro e as amostras metastáticas em linfonodo consistiram de duas semelhantes entre si e uma que se assemelhou aos tumores primários de mama (ML15). Percebe-se que há três assinaturas de IncRNAs - dos tumores primários, das metástases em linfonodo e das metástases em cérebro – das quais a de linfonodo se sobressai pela homogeneidade de expressão entre ambas as amostras.

As amostras de metástase em cérebro estão mais próximas das amostras de tumor primário do que as amostras metastáticas em linfonodo, com excessão da ML15. No conjunto de tumores primários, curiosamente, as amostras mais semelhantes entre si não pertencem ao mesmo subtipo de câncer de mama: PM17 e PM20 são dos subtipos triplo negativo e Her2, respectivamente; e PM5 e PM18 são provenientes de câncer de mama Her2 e luminal, nesta ordem.



Figura 11 - Agrupamento hierárquico dos IncRNAs com expressão diferencial entre as amostras tumorais.

4.3.5 Validação de IncRNA

Devido aos seus papéis já descritos na literatura e por ser um transcrito recentemente descrito, foi selecionado o IncRNA LINC00472 para validação por PCR semi-quantitativa em seis pares tumorais. O resultado do RNA-Seq demonstrou que, analisando para a par, a

expressão absoluta do LINCO0472 variou bastante nas amostras tumorais, não seguindo um padrão definido (Figura 12A). A PCR semi-quantitativa revelou a mesma falta de padrão nas amostras (Figura 12B), sendo o lncRNA bastante expresso no par tumoral PM1-ML1 e em duas metástases em cérebro (MC4 e MC5), apresentando expressão mesmo na diluição 1:32. Em contrapartida, outros pares tumorais, como PM2-ML2 e PM6-MC6 não apresentaram expressão de LINC00472.





Figura 12 - PCR semi-quantitativa do IncRNA LINC00472. Assim como os valores de expressão obtidos do RNA-Seq (A), a PCR demonstrou variação na expressão desse transcrito nas amostras tumorais (B). TUBB, endógeno (β -tubulina); PM, tumor primário de mama; ML, metástase em linfonodo; MC, metástase em cérebro. Em B, números iguais correspondem a um par tumor primário-metástase.

5 DISCUSSÃO

O câncer de mama é altamente relevante epidemiologicamente, uma vez que sua incidência na população é alta e acomete principalmente as mulheres. Muitos fatores de risco estão associados ao seu desenvolvimento, incluindo fatores reprodutivos, como idade na primeira gravidez e amamentação, fatores hormonais, como o uso de contraceptivos, e fatores genéticos (Ban; Godellas, 2014). Apesar de já haver diversos biomarcadores bem estabelecidos para o prognóstico e tratamento do câncer de mama, como mutações nos genes BRCA1/2 e p53 e outros tantos genes de alta ou moderada penetrância familial (Sana; Malik, 2015; Scalia-Wilbur et al., 2016), eles ainda não são suficientes para explicar por completo as razões e os mecanismos por trás da metástase.

O desenvolvimento de tecnologias de sequenciamento profundo trouxe a possibilidade de sequenciar transcritos pouco expressos e de quantificar a sua abundância na população (Ding et al., 2014). O sequenciamento de RNA (RNA-Seq) é um método de sequenciamento de nova geração (NGS) que permite uma visão mais detalhada e quantitativa da expressão gênica, do *splicing* alternativo e da expressão alelo-específica (Kukurba; Montgomery, 2015). O método do RNA-Seq se baseia na determinação das sequências de pequenos fragmentos de cDNA, possibilitando o sequenciamento de virtualmente todo o transcriptoma e a descoberta de novos *loci* e novos transcritos (Ilott e Ponting, 2013). Tais características tornam o RNA-Seq uma técnica apropriada para o estudo dos lncRNAs, uma vez que os mesmos possuem, de uma forma geral, níveis de expressão mais baixos que os mRNAs (Kornienko et al., 2016 apud Cabili et al., 2011, p. 1923).

Os IncRNAs, dentre todas as moléculas de RNAs não codificadores, são ainda a classe menos conhecida, o que se reflete na grande quantidade de transcritos anotados, porém sem função descrita (Cheetham et al., 2013). Eles têm sido descritos em diversos processos biológicos e patológicos (Zhaoet al., 2013; Herriges, et al., 2014; Reddy et al., 2014) e no câncer, já foi demonstrado que podem atuar tanto como supressores tumorais (Russell et al., 2015) quanto como oncogenes (Colombo et al., 2015). Em câncer de mama já foi demonstrado que um perfil consistindo exclusivamente de IncRNAs está associado a risco de metástase em pacientes com câncer de mama linfonodo negativo, independentemente dos marcadores clínicos tradicionais, tais como tamanho do tumor e status do receptor de

estrogênio (Sørensen et al., 2015). A expressão aberrante no câncer de mama de alguns RNAs, como HOTAIR, MALAT1, XIST e BC200, é indicativo que os lncRNAs podem ser utilizados como fatores prognósticos e de terapia (Song et al., 2015) e com o avanço dos estudos, novos lncRNAs têm sido descobertos e novas funções têm sido atribuídas àqueles já conhecidos. Um bom exemplo é o caso do PANDAR, lncRNA envolvido na regulação da resposta apoptótica em resposta a dano no DNA (Hung et al., 2011) que recentemente foi sugerido como promotor de tumor e regulador do ciclo celular de células do câncer de mama (Sang et al., 2016).

O presente estudo buscou analisar a expressão de IncRNAs em amostras metastáticas pareadas com seus tumores primários correspondentes, utilizando para isso a técnica RNA-Seq. Foi obtido um total de quarenta amostras, correspondentes a vinte pares de tumores primários e metastáticos. De acordo com as informações contidas nos prontuários das pacientes, foi possível classificar as amostras de mama nos seus subtipos moleculares – a maioria (65%) do subtipo HER2, 25% dos subtipos luminal A e B (por falta de informações adicionais sobre o estado proliferativo dos tumores - pela expressão do marcador Ki-67 - os subtipos luminais foram classificados como um único grupo) e apenas uma amostra triplo negativa. No Brasil, o tipo histológico mais frequente, de acordo com Carvalho e cols. (2010), é o luminal A, representando 55% dos casos de carcinoma invasivo de mama. Provavelmnte devido ao pequeno número amostral, as frequências dos subtipos histológicos encontradas no presente estudo divergiu desse dado. O carcinoma ductal invasivo é o subtipo morfológico mais comum, representando 80% dos carcinomas invasivos (Rivenbark; O'Connor; Coleman, 2013), mesma proporção encontrada nas amostras deste estudo.

A análise da qualidade do sequenciamento dos cinco pares tumorais selecionados demonstrou que o resultado é altamente confiável. A qualidade é baseada no escore Phred (valor bruto de qualidade), que é calculado como a função log da probabilidade da chamada de base estar incorreta (Ruffalo et al., 2012). O escore é calculado para cada chamada de base na *read* e varia de zero a quarenta, sendo zero o menos e quarenta o mais confiável. Neste estudo, todas as posições das *reads* geradas apresentaram um escore acima de 30, que corresponde a uma baixa probabilidade (1/1000) da chamada de base estar incorreta (Ewing e Green, 1998).

Após a avaliação da qualidade do sequenciamento, as sequências foram mapeadas contra o genoma de referência GRCh37.p13, o que revelou que aproximadamente 55% das reads de todos os grupos amostrais mapearam em regiões de RNA não codificador de proteína – predominantemente regiões de IncRNAs – enquanto que cerca de 35% foram mapeados em regiões codificadoras. O maior número de IncRNAs em relação aos mRNAs está de acordo com estudo de Iyer e cols. (2015), no qual encontraram que a diversidade genômica de IncRNAs sobrepuja a de transcritos codificadores. Dentre os IncRNAs expressos, aqueles que retêm íntros são os mais abundantes, representando aproximadamente 60% de todos os IncRNAs de todos os grupos amostrais. A retenção de íntrons consiste justamente na manutenção de uma sequência intrônica no transcrito maduro, onde o íntron retido geralmente apresenta sítios de splicing mais fracos que fazem com que o spliceossomo não consiga identificá-los (Ge e Porse, 2013 apud Sakabe; Souza, 2007). Já foi demonstrado, que, juntamente com o éxon skipping, a retenção de íntron é um dos eventos de splicing predominantes em câncer de mama (Eswaran et al., 2013). A maioria dos íntrons retidos é específica do câncer de origem, ou seja, cânceres originados de tecidos similares exibem padrões de retenção intrônica similares (Dvinge e Bradley, 2015), o que pode explicar a semelhança no número de íntrons retidos entre os grupos tumorais. Essa característica parece se estender para as outras classes de IncRNAs, que também apresentaram número de transcritos similares entre as amostras. A análise dos diagramas de Venn também reforça essa ideia, uma vez que a grande maioria dos IncRNAs expressos nas metástases são também expressos nos seus respectivos tumores primários. Além disso, dados da literatura apontam que as mudanças de expressão de uma colônia metastática são sutis e que características biológicas significantes se mantêm similares entre o tumor primário e sua metástase correspondente (Weigelt e van't Veer, 2004). Além disso, observa-se que todos os tumores de mama compartilham um alto número de transcritos comuns, assim como todos os tumores metastáticos, o que indica que, qualitativamente, eles são muito semelhantes entre si.

Procurando saber quais seriam as funções dos IncRNAs que se encontravam comumente expressos nas amostras tumorais, foram realizadas buscas em bases de dados públicas. Verificou-se que a maioria dos genes de IncRNAs encontrados neste estudo ainda não possuem caracterização funcional. De fato, os papéis regulatórios de poucos IncRNAs já

Discussão | 59

foram biologicamente caracterizados, devido, em grande parte, ao rápido acúmulo de dados de estudos de larga escala (Fritah; Niclou; Azuaje, 2014). Alguns IncRNAs já descritos na literatura como envolvidos no desenvolvimento e na progressão do câncer de mama encontravam-se expressos nas amostras tumorais, tanto primárias quanto metastáticas. Um dos que apresentou um dos maiores níveis de expressão em todas as amostras foi o MALAT1 (Metastasis-Associated Lung Adenocarcinoma Transcript 1), primeiramente descrito em associação com metástase em câncer de pulmão (Ji et al., 2003). Ele é altamente expresso em diversos tipos de câncer e promove motilidade celular através de regulação transcricional e pós-transcricional da expressão de genes relacionados a motilidade, como CCT4 e HMMR (Tano et al., 2010). Um estudo de meta-análise demonstrou que a incidência de metástases em linfonodo em pacientes com alta expressão de MALAT1 é maior que a de pacientes com baixa expressão desse IncRNA (Zhai et al., 2015) e, recentemente, foi demonstrado que a diminuição da expressão de MALAT1 resulta em alterações na expressão gênica e mudança nos padrões de *splicing* de genes envolvidos em diferenciação e vias de sinalização pró-tumorigênicas (Arun et al., 2016).Em todas as amostras tumorais, também foi encontrada alta expressão de XIST (X inactive specific transcript), bem conhecido pelo seu papel na inativação do cromossomo X (Cerase et al., 2015). Ele é encontrado altamente expresso em carcinoma de mama e a diminuição da sua expressão tem função supressora tumoral, reduzindo proliferação celular, migração, invasão e induzindo apoptose (Tantai et al., 2015). O BCYRN1 (brain cytoplasmic RNA 1), também conhecido como BC200, foi mais um dos genes de IncRNAs que apresentou alta expressão nas amostras tumorais, o que é corroborado por estudo mostrando que esse IncRNA é altamente expresso em carcinomas de mama invasivos (lacoangeli et al., 2004). O seu possível papel oncogênico está relacionado com sua função de regulador negativo da apoptose (Singh e Mo, 2015).

Outro IncRNA expresso nos tumores deste estudo foi o LINCO0472 (*Long intergenic nonprotein coding RNA 472*). Em estudo recente, no qual ele foi caracterizado, curiosamente, foi observado que o LINCO0472 estava associado com tumores de mama menos agressivos e que sua alta expressão podia suprimir a proliferação e migração das células do câncer de mama (Shen et al., 2015). A expressão desse IncRNA, tanto nas amostras de tumor primário quanto nas amostras metastáticas, não coloca seu papel antitumoral e questão, mas pode demonstrar que o balanço da expressão de oncogenes e

genes supressores tumorais pode determinar a evolução do câncer. O LINCO0472 foi escolhido para validação, devido à sua recente caracterização funcional e expressão nas amostras tumorais. A análise da PCR semi-quantitativa em seis pares tumorais demonstrou que o LINCO0472 apresenta expressão variável entre as amostras, assim como foi observado nos dados de expressão (FPKM) do RNA-Seq, quando avaliados para cada amostra individualmente.

Estudo de Gupta e cols. (2010) demonstrou que HOTAIR (*HOX antisense intergenic RNA*) é altamente expresso em tumores primário de mama e suas metástases. Ele é necessário para o mecanismo de transição epitélio-mesenquimal e para a manutenção do fenótipo de célula-tronco em linhagem celular de câncer de mama (Pádua Alves et al., 2013), o que pode ser consequência da sua capaciade de interagir com o complexo repressivo policombe 2 (*PRC2*), guiando-o para sítios específicos e promovendo silenciamento transcricional e supressão de diferenciação celular (Zhang, J. et al., 2014). Nas amostras do presente estudo, HOTAIR encontrava-se bem expresso nos pares tumorais, porém com um nível de expressão menor se comparado com MALAT1, XIST, BCYRN1 e LINC00472, o que pode sugerir um certo grau de diferenciação das células tumorais. O ZFAS1 (*zinc finger antisense 1*), por sua vez, apresentou níveis de expressão comparativamente menores que os demais lncRNAs acima citados. Uma explicação possível é que esse lncRNA é provavelmente um gene supressor tumoral – estudo de Askarian-Amiri e cols. (2011) mostrou que a diminuição da expressão de Zfas1 em células epiteliais mamárias de camundongos aumentou significativamente a proliferação e a atividade metabólica celular.

Em seguida, foi avaliada a diferença nos níveis de expressão dos IncRNAs comuns entre as amostras de tumor primário e seus tumores metastáticos, para verificar quais os transcritos que poderiam estar envolvidos com o processo metastático. Dentre os genes dos IncRNAs diferencialmente expressos, apenas os codificadores de proteínas (que originam os transcritos com íntron retido) possuem função biológica descrita na literatura, com exceção de cinco IncRNAs: o 7SK, um IncRNA que atua interrompendo o alongamento transcricional (Kugel; Goodrich, 2012); o SOX2-OT, altamente expresso em células embrionárias e excerce função regulatória na progressão do ciclo celular (Shahryari et al., 2015); o DNM3OS, um IncRNA que contém três miRNAs e é indispensável para o desenvolvimento esquelético normal e o crescimento corporal em mamíferos (Watanabe et al., 2008); o TBX-AS1, envolvido na capacidade de crescimento independente de ancoragem de células cancerosas de cólon (Li et al., 2015); e, curiosamente, o MEG3, um IncRNA cujo gene tem a expressão perdida em diversos tumores primários e que atua inibindo a proliferação de células tumorais (Zhou et al., 2012).

Uma vez que a maior parte dos IncRNAs não possui função descrita, foi analisada a interação daqueles transcritos mais expressos com proteínas, a fim de inferir a função desses transcritos (Anexo E). Observou-se que a maioria entá envolvida no processamento de mRNA, principalmente na regulação de splicing alternativo, processo que oferece uma plasticidade importante para as células cancerosas remodelarem seus proteomas e suprirem suas necessidades de crescimento e metástase (David; Manley, 2010). Alguns IncRNAs, como o RP11-400N13.3-001 e o RP11-693J15.5-001, interagem com proteínas envolvidas na via de sinalização do receptor do fator de crescimento epidermal (EGFR). A superfamília EGFR é composta por diversas proteínas, das quais algumas podem atuar como fatores de crescimento pela ligação com receptores de membrana específicos (Normanno et al., 2001). Além disso, os membros dessa família são os principais participantes de uma complexa cascata de sinalização que, além de crescimento, regula diferenciação, migração, adesão e sobrevivência das células cancerosas (Seshacharyulu et al., 2012). Também foi identificada a interação de IncRNAs com proteínas relacionadas com a via de sinalização Notch, cujo papel oncogênico em câncer de mama provavelmente é mediado, em parte, pela sua interação com outras vias de sinalização, como a do estrógeno (Al-Hussaini et al., 2011). De forma geral, a via Notch determina o destino celular e regula a renovação de células-tronco normais e cancerosas, assim como diferenciação (Takebe; Nguyen; Yang, 2014). Além das vias já citadas, houve interação de lncRNAs com proteínas da via do TGF- β , que é uma via contexto-dependente e que excerce efeitos célula-específicos, podendo tanto estimular quanto inibir o crescimento, a apoptose ou a diferenciação (Jakowlew, 2006). A partir desses resultados, observa-se que, apesar de muitos IncRNAs não possuírem função conhecida, suas características funcionais podem ser inferidas a partir das suas interações com outras moléculas.

Em seguida, foi realizado o agrupamento hierárquico não supervisionado, que agrupou as amostras com base nas suas similaridades de IncRNAs diferencialmente expressos. Em 2003, Weigelt e cols. analisaram o perfil de expressão gênica entre oito pares de tumores

primários de mama e suas metástases em sítios distantes, utilizando a técnica de microarranjo para mais de dezoito mil cDNAs. Os pesquisadores observaram que os tumores primários eram mais similares aos seus tumores metastáticos pareados que a outros tumores. Contrariamente ao achado de Weigelt e cols., neste estudo as amostras de tumores primários de mama eram mais similares entre si do que com seus correspondentes metastáticos. Da mesma forma, as amostras de metástase foram agrupadas em linfonodo e em cérebro, com exceção de uma amostra de linfonodo, que assemelhou mais com o grupo de tumores primários. Resultado similar foi encontrado por LaTulippe e cols. (2002), em estudo da diferença na expressão gênica entre tumores de próstata primários e metastáticos. A partir da análise não supervisionada da expressão gênica, o grupo observou forte tendência dos tumores primários e metastáticos apresentarem perfis de expressão bem distintos, sugerindo genes específicos responsáveis pelas diferenças biológicas entre os dois grupos. No presente estudo, foram identificadas três assinaturas de IncRNAs bem distintas entre os grupos amostrais, o que sugere a presença de IncRNAs específicos responsáveis pela distinção biológica de cada tipo de tumor metastático e dos tumores primários. Além disso, a diferença de expressão desses IncRNAs pode indicar possível envolvimento desses transcritos no mecanismo de metástase.

6 CONCLUSÕES

1. Determinação o perfil de expressão de IncRNAs de tumor de mama metastático em cérebro e em linfonodo;

Foram obtidos com sucesso perfis de expressão de IncRNAs de tumores primários de mama e suas respectivas metástases em linfonodo e em cérebro

 Classificação funcionaldos IncRNAs preferencialmente expressos nos tumores metastáticos;

A maioria dos IncRNAs expressos nas amostras deste estudo ainda não são funcionalmente caracterizados

3. Seleção de IncRNAs potencialmente envolvidos com a metástase.

Foram identificados três assinaturas de lncRNAs, uma para cada grupo tumoral.

Em resumo, os resultados indicam que os IncRNAs diferencialmente expressos que compõem as assinaturas podem estar envolvidos no mecanismo de metástase e são potenciais candidatos para estudos funcionais.

REFERÊNCIAS

Al-Hussaini, H. et al. Notch Signaling Pathway as a Therapeutic Target in Breast Cancer. *Molecular Cancer Therapeutics*, v. 10, n. 1, p. 9-15, 2011.

Alitalo, K.; Tammela, T.; Petrova, T. V. Lymphangiogenesis in development and human disease. *Nature*, v. 438, 946-953, 2005.

Aloraifi, F. et al. Detection of novel germline mutations for breast cancer in non-BRCA1/2 families. The FEBS Journal, v. 282, n. 17, p. 3424-3427, 2015.

Antoniou, A. C. et al. Breast-cancer risk in families with mutations in PALB2. *The New England Journal of Medicine*, v. 371, n. 6, p. 497-506, 2014.

Amendola, L. C. B.; Vieira, R. A contribuição dos genes BRCA na predisposição hereditária ao câncer de mama. *Revista Brasileira de Cancerologia*, v. 51, n. 4, p.325-330, 2005.

American Cancer Society. Cancer Facts & Figures, *American Cancer Society*, 2016. Disponível em < <u>http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/index</u> > Acesso em 10 de fevereiro de 2016.

Anderberg, C. et al. Deficiency for endoglin in tumor vasculature weakens the endothelial barrier to metastatic dissemination. *The Journal of Experimental Medicine*, v. 210, n. 3, p. 563-579, 2013.

Anders, S.; Pyl, P. T.; Huber, W. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics*, v. 31, n. 2, p. 166-169, 2014.

Ard, R.; Tong, P.; Allshire, R. C. Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. *Nature Communications*, v. 5, p. 5576, 2013.

Arun, G. et al. Differentiation of mammary tumors and reduction in metastasis upon *Malat1* IncRNA loss. *Genes & Development*, v. 30, p. 34-51, 2016.

Askarian-Amiri, M. E. et al. SNORD-host RNA *Zfas1* is a regulator of mammary development and a potential marker for breast cancer. *RNA*, v. 17, n. 5, p. 878-891, 2011.

Babraham Intitute. FastQC. *Babraham Bioinformatics*, 2014. Disponível em < <u>http://www.bioinformatics.babraham.ac.uk/projects/fastqc/</u>> acesso em 17 de janeiro de 2016.

Badve, S.; Gökmen-Polar, Y. Tumor heterogeneity on breast cancer. *Advances in Anatomic Pathology*, v. 22, n. 5, p. 294-302, 2015.

Baker, M. Long noncoding RNAs: the search for function. *Nature Methods*, v. 8, n. 5, p. 379-383, 2011.

Ban, K. A.; Godellas, C. V. Epidemiology of breast cancer. *Surgical Oncology Clinics of North America*, v. 23, n. 3, p. 409-422, 2014.

Bànkfalvi, A. et al. Immunophenotypic and prognostic analysis of E-cadherin and betacatenin expression during breast carcinogenesis and tumour progression: a comparative study with CD44. *Histopathology*, v. 34, n. 1, p. 25-34, 1999.

Bastos, E. P. et al. MicroRNAs discriminate familial from sporadic non-BRCA1/2 breast carcinoma arising in patients ≤35 years. *PLoS One*, v. 9, n. 7, e101656, 2014.

Bergmann, J. H.; Spector, D. L. Long non-coding RNAs: modulators of nuclear structure and function. *Current Opinion in Cell Biology*, v. 26, p. 10-18, 2014.

Berteaux, N. et al. *H19* mRNA-like noncoding RNA promotes breast cancer cell proliferation through positive control by E2F1. *The Journal of Biological Chemistry*, v. 280, p. 29625-29636, 2005.

Bissell, M. J.; Rizki, A.; Mian, I. S. Tissue architecture: the ultimate regulator of breast epithelial function. *Current Opinion in Cell Biology*, v. 15, n. 6, p. 753-762, 2003.

Brábek, J. et al. The role of the tissue microenvironment in the regulation of cancer cell motility and invasion. *Cell Communication and Signaling*, v. 8, p. 22, 2010.

Brannan, C. I. et al. The product of the H19 gene may function as an RNA. *Mollecular and Cellular Biology*, v. 10, n. 1, p. 28-36, 1990.

Branham, M. T. et al. Epigenetic regulation of ID4 in the determination of the BRCAness phenotype in breast cancer. *Breast Cancer Research and Treatment*, v. 155, n. 1, p. 13-23, 2016.

Braunschweig, U. et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Research*, v. 24, p. 1774-1786, 2014.

Bravo-Cordero, J. J.; Hodgson, L.; Condeelis, J. Directed cell invasion and migration during metastasis. *Current Opinion in Cell Biology*, v. 24, n. 2, p. 277-283, 2012.

Bubien, V. et al. High cumulative risks of cancer in patients with *PTEN* hamartoma tumour syndrome. *Journal of Medical Genetics*, v. 50, p. 255-263, 2013.

Buerguer, H. et al. Genetic relation of lobular carcinoma in situ, ductal carcinoma in situ, and associated invasive carcinoma of the breast. *Journal of Clinical Pathology: Molecular Pathology*, v. 53, p. 118-121, 2000.

Burstein, H. J. et al. Ductal Carcinoma in Situ of the Breast. *The New England Journal of Medicine*, v. 350, p. 1430-1441, 2004.

Cabili, M. N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, v. 25, n. 18, p. 1915–1927, 2011.

Cai, X.; Cullen, B. R. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA*, v. 13, p. 313-316, 2007.

Carvalho, L. V. et al. Molecular characterization of breast cancer in young Brazilian women. *Revista da Associação Médica Brasileira*, v. 56, n. 3, p. 278-287, 2010.

Cerase, A. et al. Xist localization and function: new insights from multiple levels. *Genome biology*, v. 16, n. 1, p. 166, 2015.

Chao, Y. L.; Shepard, C. R.; Wells, A. Research Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition. *Molecular Cancer*, v. 9, p. 179, 2010.

Cheetham, S. W. et al. Long noncoding RNAs and the genetics of cancer. *British Journal of Cancer*, v. 108, p. 2419-2425, 2013.

CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancercases and 9,065 controls from 10 studies. *American Journal of Human Genetics*, v. 74, n. 6, p. 1175-1182, 2004.

Chou, J. et al. MALAT1 induced migration and invasion of human breast cancer cells by competitively binding miR-1 with cdc42. *Biochemical and Biophysical Research Communications*, v. 472, n. 1, p. 262-269, 2016.

Colombo, T. et al. PVT1: A rising star among oncogenic long noncoding RNAs. *BioMed Research International*, v. 2015, artigo 304208, 10 p., 2015.

Cooke, V. G. et al. Pericyte depletion results in hypoxia-associated epithelial-tomesenchymal transition and metastasis mediated by Met signaling pathway. *Cancer Cell*, v. 21, n. 1, p. 66-81, 2012.

Curtit, E. et al. First description of a sporadic breast cancer in a woman with BRCA1 germline mutation. *Oncotarget*, v. 6, n. 34, p. 35616-35624, 2015.

Danza, K. et al. Combined microRNA and ER expression: a new classifier for familial and sporadic breast cancer patients. *Journal of Translational Medicine*, v. 12, p. 319, 2014.

David, C. J.; Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & Development*, v. 24, p.2343-2364, 2010.

de Boer, M. et al. Breast cancer prognosis and occult lymph node metastases, isolated tumor cells, and micrometastases. *Journl of the National Cancer Institute*, v. 102, n. 6, p. 410-425, 2010.

Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, v. 22, p. 1775-1789, 2012.

Ding, X. et al. Long intergenic non-coding RNAs (LincRNAs) identified by RNA-Seq in breast cancer. *PLoS One*, v. 9, n. 8, e103270, 2014.

Domchek, S. M. Evolution of Genetic Testing for Inherited Susceptibility to Breast Cancer. *Journal of Clinical Oncology*, v. 33, n. 4, p. 295-296, 2015. Dvinge, H.; Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Medicine*, v. 7, n. 1, p. 45, 2015.

Economopoulou, P.; Dimitriades, G.; Psyrri, A. Beyond *BRCA*: New hereditary breast cancer susceptibility genes. *Cancer Treatment Reviews*, v. 41, n. 1, p. 1-8, 2015.

Eswaran, J. et al. RNA sequencing of cancer reveals novel splicing alterations. *Scientific Reports*, v. 3, artigo 1689, 2013.

Ewing, B.; Green, P. Base-calling of automated sequencer traces using *Phred*. II. Error probabilities. *Genome Research*, v. 8, p. 186-194, 1998.

Fritah, S.; Niclou, S. P.; Azuaje, F. Databases for IncRNAs: a comparative evaluation of emerging tools. *RNA*, v. 20, n. 11, p. 1655–1665, 2014.

Gallego Romero, I. et al. RNA-seq: impact of RNA degradation on transcript quantification. BMC Biology, v. 12, p. 42, 2014.

Ge, Y.; Porse, B. T. The functional consequences of intron retention: Alternative splicing coupled to NMD as a regulator of gene expression. *Bioessays*, v. 36, p. 236–243, 2013.

Gialeli, C.; Theocharis, A. D.; Karamanos, N. K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *FEBS Journal*, v. 278, n. 1, p. 16-27, 2011.

Gil-Barnabé, A. M.; Lucotti, S.; Muschel, R. J. Coagulation and metastasis: what does the experimental literature tell us? *British Journal of Haematology*, v. 162, n. 4, p. 433-41, 2013.

Guil, S.; Esteller, M. *Cis*-acting noncoding RNAs: friends and foes. *Nature Structural & Molecular Biology*, v. 19, p. 1068–1075, 2012.

Gupta, R. A. et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, v. 464, p. 1071-1078, 2010.

Gutschner, T.; Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biology*, v. 9, n. 6, p. 703-719, 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, 2015.

Han, P.; Chang, C. P. Long non-coding RNA and chromatin remodeling. *RNA Biology*, v. 12, n. 10, p. 1094-1098, 2015.

Hangauer, M. J.; Vaughn, I. W.; McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genetics*, v. 9, n. 6, e1003569, 2013.

Hansji, H. et al. Keeping abreast with long non-coding RNAs in mammary gland development and breast cancer. *Frontiers in Genetics*, v. 5, 379, 2014.

Henry, N. L.; Haynes, D. F. Cancer biomarkers. *Molecular Oncology*, v. 6, n. 2, p. 140-146, 2012.

Herriges, M. J. et al. Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes & Development*, v. 28, p. 1363-1379, 2014.

Herschkowitz, J. I. et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology*, v. 8, n. 5, R76, 2007.

Hiemer, S. E.; Szymaniak, A. D.; Varelas, X. The transcriptional regulators TAZ and YAP direct transforming growth factor β -induced tumorigenic phenotypes in breast cancer cells. *The Journal of Biological Chemistry*, v. 289, n. 19, p. 13461-13474, 2014.

Hu, M. et al. Regulation of in situ to invasive breast carcinoma transition. *Cancer Cell*, v. 13, n. 5, p. 394-406, 2008.

Hung, T. et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genetics*, v. 43, p. 621–629, 2011.

Iacoangeli, A. et al. BC200 RNA in invasive and preinvasive breast cancer. *Carcinogenesis*, v. 25, n. 11, p. 2125-2133, 2004.

Ilott, N. E.; Ponting, C. P. Predicting long non-coding RNAs using RNA sequencing. *Methods*, v. 63, n. 1, p. 50-59, 2013.

Instituto Nacional do Câncer. Tipos de câncer – mama. *INCA*, 2016. Disponível em < <u>http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama</u> > acesso em 10 de fevereiro de 2016.

Iyer, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. *Nature* Genetics, v. 47, p. 199–208, 2015.

Jakowlew, S. B. Transforming growth factor-β in cancer and metastasis. *Cancer Metastasis Reviews*, v. 25, p. 3, p. 435-457, 2006.

Ji, P. et al. MALAT-1, a novel noncoding RNA, and thymosin β 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, v. 22, p. 8031–8041, 2003.

Jones, P. A.; Baylin, S. B. The epigenomics of cancer. Cell, v. 128, p. 683-692, 2007.

Kallen, A. N. et al. The imprinted H19 IncRNA antagonizes let-7 microRNAs. *Molecular Cell*, v. 52, n. 1, p. 101-112, 2013.

Kamińska, M. et al. Breast cancer risk factors. Prz Menopauzalny, v. 14, n. 3, p. 196-202, 2015.

Kanada, M. et al. Endothelial cell-initiated extravasation of cancer cells visualized in zebrafish. *PeerJ*, v. 2, e688, 2014.

Karaman, S.; Detmar, M. Mechanisms of lymphatic metastasis. *Journal of Clinical Investigation*, v. 124, n. 3, p. 922-928, 2014.

Khachane, A. N.; Harrison, P. M. Mining mammalian transcript data for functional long noncoding RNAs. *PLoS One*, v. 5, n. 4, e10316, 2010.

Khuon, S. et al. Myosin light chain kinase mediates transcellular intravasation of breast cancer cells through the underlying endothelial cells: a three-dimensional FRET study. *Journal of Cell Science*, v. 123, p. 431-440, 2010.

Knowles, L. M. et al. Fibronectin matrix formation is a prerequisite for colonization of kidney tumor cells in fibrin. *Journal of Cancer*, v. 6, n. 2, p. 98-104, 2015.

Konstantopoulos, K.; Thomas, S. N. Cancer cells in transit: the vascular interactions of tumor cells. *Annual Review of Biomedical Engineering*, v. 11, p. 177-202, 2009.

Kornblihtt, A. R. et al. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Reviews Molecular Cell Biology*, v. 14, p. 153-165, 2013.

Koren, S.; Betires-Alj, M. Breast tumor heterogeneity: source of fitness, hurdle for therapy. *Molecular Cell*, v. 60, n. 4, p. 537-546, 2015.

Kornienko, A. E. et al. Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. *Genome Biology*, v. 17, n. 1, p. 14, 2016.

Koujan, S. E.; Gargarib, B. P.; Khalili, M. Matrix metalloproteinases and breast cancer. *Thrita*, v. 4, n. 1, e21959, 2015.

Kugel, J. F.; Goodrich, J. A. Non-coding RNAs: key regulators of mammalian transcription. *Trends in Biochemical Sciences*, v. 37, n. 4, p. 144-151, 2012.

Kukurba, K. R.; Montgomery, S. B. RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 2015.

Labelle, M.; Hynes, R. O. The initial hours of metastasis: the importance of cooperative hosttumor cell interactions during hematogenous dissemination. *Cancer Discovery*, v. 2, n. 12, p. 1091-1099, 2012.

Lamouille, S.; Xu, J.; Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, v. 15, n. 3, p. 178-196, 2014.

LaTulippe, E. et al. Comprehensive Gene Expression Analysis of Prostate Cancer Reveals Distinct Transcriptional Programs Associated with Metastatic Disease. *Cancer Research*, v. 62, n. 15, p. 4499-4506, 2002.

Leber, M. F.; Efferth, T. Molecular principles of cancer invasion and metastasis (Review). *International Journal of Oncology*, v. 34, p. 881-895, 2009.

Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science*, v. 338, n. 6113, p. 1435-1439, 2012.

Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of Clinical Investigation*, v. 121, n. 7, p. 2750-2767, 2011.

Leong, H. S. et al. Invadopodia are required for cancer cell extravasation and are a therapeutic target for metastasis. *Cell Reports*, v. 8, n. 5, p. 1558-1570, 2014.

Lerner-Ellis, J. et al. Genetic risk assessment and prevention: the role of genetic testing panels in breast cancer. *Expert Review of Anticancer Therapy*, v. 15, n. 11, p. 1315-1326, 2015.

Li, H. et al. Circulating Prostaglandin Biosynthesis in Colorectal Cancer and Potential Clinical Significance. *EBioMedicine*, v. 2, n. 2, p. 165-171, 2015.

Li, L. et al. Long non-coding RNA MALAT1 promote triple-negative breast cancer progression by regulating miR-204 expression. *International Journal of Clinical and Experimental Pathology*, v. 9, n. 2, p. 969-977, 2016.

Li, Y. M. et al. Upregulation of CXCR4 is essential for HER2-mediated tumor metastasis. *Cancer Cell*, v. 6, n. 5, p. 459-469, 2004.

Liu, J. C. et al. Common and distinct features of mammary tumors driven by Pten-deletion or activating Pik3ca mutation. *Oncotarget*, v. 7, n. 8, p. 9060-9068, 2016.

Liu, Z.; Zhang, X. S.; Zhang, S. Breast tumor subgroups reveal diverse clinical prognostic power. *Scientific Reports*, v. 4, p. 4002, 2014.

Lottin, S. et al. Overexpression of an ectopic *H19* gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis*, v. 23, n. 11, p. 1885-1895, 2002.

Love, M. I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, v. 15, p. 550, 2014.

Malih, S.; Saidijam, M.; Malih, N. A brief review on long noncoding RNAs: a new paradigm in breast cancer pathogenesis, diagnosis and therapy. *Tumour Biology*, 2015. No prelo. Disponível em: http://link.springer.com/article/10.1007%2Fs13277-015-4572-y Acesso em 9 jan. 2016.
Mangone, F. R. et al. *ATM* gene mutations in sporadic breast cancer patients from Brazil. *Springerplus*, v. 4, p. 23, 2015.

Massink, M. P. G. et al. Genomic profiling of *CHEK2**1100delC-mutated breast carcinomas. *BMC Cancer*, v. 15, p. 877, 2015.

Mattick, J. S. The functional genomics of noncoding RNA. *Science*, v. 309, n. 5740, p. 1527-1528, 2005.

Melchor, L. et al. Identification of cellular and genetic drivers of breast cancer heterogeneity in genetically engineered mouse tumour models. *The Journal of Pathology*, v. 233, n. 2, p. 124-137, 2014.

Michailidou, K. et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature Genetics*, v. 47, n. 4, p. 373-380, 2015.

Mueller, B. M. et al. Expression of tissue factor by melanoma cells promotes efficient hematogenous metastasis. *PNAS*, v. 89, p. 11832-11836, 1992.

Müller, A. et al. Involvement of chemokine receptors in breast cancer metastasis. *Nature*, v. 410, p. 50-56, 2001.

Nathanson, S. D. Insights into the mechanisms of lymph node metastasis. *Cancer*, v. 98, n. 2, p. 413-423, 2003.

Nazário, A. C. P.; Facina, G.; Filassi, J. R. Breast cancer: news in diagnosis and treatment. *Revista da Associação Médica Brasileira*, v. 61, n. 6, p. 543-552, 2015.

Nguyen, D. X.; Bos, P. D.; Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nature Reviews Cancer*, v. 9, p. 274-284, 2009.

Nickels, S. et al. Evidence of gene-environment interactions between common breast câncer susceptibility loci and established environmental risk factors. *Plos Genetics*, v. 9, n. 3, p. e1003284, 2013.

Nie, N. et al. Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. *American Journal of Translational Research*, v. 4, n. 2, p. 127-150, 2012.

Nisticò, P.; Bissel, M. J.; Radisky, D. C. Epithelial-mesenchymal transition: General principles and pathological relevance with special emphasis on the role of matrix metalloproteinases. *Cold Spring Harbor Perspectives in Biology*, v. 4, p. a011908, 2012.

Normanno, N. et al. The role of EGF-related peptides in tumor growth. *Frontiers in Bioscience*, v. 6, p. D685-707, 2001.

Pádua Alves, C. et al. Brief Report: The lincRNA Hotair is required for epithelial-tomesenchymal transition and stemness maintenance of cancer cell Lines. *Stem Cells*, v. 31, n. 12, p. 2827-2832, 2013.

Palumbo, J. S. et al. Platelets and fibrin(ogen) increase metastatic potential by impeding natural killer cell–mediated elimination of tumor cells. *Blood*, v. 105, n. 1, p. 178-185, 2005.

Pecorino, L. *Molecular Biology of Cancer: Mechanisms, Targets, and Therapeutics*. 3^a ed. Oxford: Oxford University Press, 2012.

Perou, C. M. et al. Molecular portraits of human breast tumours. *Nature*, v. 406, p. 747-752, 2000.

Pignatelli, J. et al. Invasive breast carcinoma cells from patients exhibit Mena^{INV}- and macrophage-dependent transendothelial migration. *Science Signaling*, v. 7, n. 253, ra112, 2014.

Polyak, K. Breast cancer: origins and evolution. *The Journal of Clinical Investigation*, v. 117, n. 11, p. 3155-3163, 2007.

Polyak, K. Heterogeneity in breast cancer. *The Journal of Clinical Investigation*, v. 121, n. 10, p. 3786-3788, 2011.

Poumpouridou, N. et al. Development of a novel PTT assay for mutation detection in PALB2 large exons and PALB2 screening in medullary breast cancer. *Familial Cancer*, v. 15, n. 2, p. 183-191, 2015.

Quan, M.; Chen, J.; Zhang, D. Exploring the Secrets of Long Noncoding RNAs. *International Journal of Melecular Sciences*, v. 16, n. 3, p. 5467-5496, 2015.

R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. 2012.

Rahman, M.; Mohammed, S. Breast cancer metastasis and the lymphatic system (Review). *Oncology Letters*, v. 10, p. 1233-1239, 2015.

Ravasi, T. et al. Experimental validation of the regulated expression of large numbers of noncoding RNAs from the mouse genome. *Genome Research*, v. 16, n. 1, p. 11-19, 2006.

Reddy, M. et al. Regulation of Inflammatory Phenotype in Macrophages by a Diabetes-Induced Long Noncoding RNA. *Diabetes*, v. 63, n. 12, p. 4249–4261, 2014.

Reymond, N.; d'Água, B. B.; Ridley, A. J. Crossing the endothelial barrier during metastasis. *Nature Reviews Cancer*, v. 13, n. 12, p. 858-870, 2013.

Rinn, J. L. et al. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Non-Coding RNAs. *Cell*, v. 129, n. 7, p. 1311–1323, 2007.

Rivenbark, A. G.; O'Connor, S. M.; Coleman, W. B. Molecular and cellular heterogeneity in breast cancer: Challenges for personalized medicine. *The American Journal of Pathology*, v. 183, n. 4, p. 1113-1124, 2013.

Rizwan, A. et al. Metastatic breast cancer cells in lymph nodes increase nodal collagen density. *Scientific Reports*, v. 5, p. 10002, 2015.

Ruffalo, M. et al. Accurate estimation of short read mapping quality for next-generation genome sequencing. *Bioinformatics*, v. 28, n. 18, p. i349–i355, 2012.

Russell, M. R. et al. CASC15-S is a tumor suppressor IncRNA at the 6p22 neuroblastoma susceptibility locus. *Tumor and Stem Cell Biology*, v. 75, n. 15, p. 3155-3166, 2015.

Sabatier, R. et al. Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Molecular Cancer*, v. 13, p. 228, 2014.

Sakabe, N. J.; Souza, S. J. Sequence features responsible for intron retention in human. *BMC Genomics*, v. 8, p. 59, 2007.

Samatov, T. R. et al. Epithelial-mesenchymal transition: focus on metastatic cascade, alternative splicing, non-coding RNAs and modulating compounds. *Molecular Cancer*, v. 12, p. 107, 2013.

Sana, M.; Malik, H. J. Current and emerging breast cancer biomarkers. *Journal of Cancer Research and Therapeutics*, v. 11, n. 3, p. 508-513, 2015.

Sang, Y. et al. LncRNA PANDAR regulates the G1/S transition of breast cancer cells by suppressing p16^{INK4A} expression. *Scientific Reports*, v. 6, artigo 22366, 2016.

Scalia-Wilbur, J. et al. Breast cancer risk assessment: moving beyond BRCA 1 and 2. *Seminars in Radiation Biology*, v. 26, n. 1, p. 3-8, 2016.

Seshacharyulu, P. et al. Targeting the EGFR signaling pathway in cancer therapy. *Expert Opinion on Therapeutic Targets*, v. 16, n. 1, p. 15-31, 2012.

Shahryari, A. et al. Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis. *Frontiers in Genetics*, v. 6, p. 196, 2015.

Sheikh, A. et al. The spectrum of genetic mutations in breast cancer. *Asian Pacifi Journal of Cancer Prevention*, v. 16, n. 6, p. 2177-2185, 2015.

Shen, Y. et al. Prognostic and predictive values of long non-coding RNA *LINCO0472* in breast cancer. *Oncotarget*, v. 6, n. 11, p. 8579–8592, 2015.

Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nature Biotechnology*, v. 26, p. 1135-1145, 2008.

Shi, X. et al. Long non-coding RNAs: A new frontier in the study of human diseases. *Cancer Letters*, v. 339, p. 159-166, 2013.

Shibue, T.; Weinberg, R. A. Metastatic colonization: Settlement, adaptation and propagation of tumor cells in a foreign tissue environment. *Seminars in Cancer Biology*, v. 21, n. 2, p. 99-106, 2011.

Shore, A. N.; Rosen, J. M. Regulation of mammary epithelial cell homeostasis by IncRNAs. *The International Journal of Biochemistry & Cell Biology*, v. 54, p. 318-330, 2014.

Sibbering, M.; Courtney, C. A. Management of breast cancer: basic principles. Surgery, 2015.

Sigurgeirsson, B. et al. Sequencing degraded RNA addressed by 3' tag counting. *PLoS One*, v. 9, n. 3, e91851, 2014.

Silva, J. M. et al. Identification of Long stress-induced non-coding transcripts that have altered expression in cancer. *Genomics*, v. 95, n. 6, p. 355-362, 2010.

Silva, J. M. et al. LSINCT5 is over expressed in breast and ovarian cancer and affects cellular proliferation. *RNA Biology*, v. 8, n. 3, p. 496-505, 2011.

Singh, R.; Mo, Y. Y. LncRNA BC200 is induced by estradiol and regulates apoptosis in MCF-7 breast cancer cells. *Cancer Research*, v. 75, p. 156, 2015.

Smigal, C. et al. Trends in breast cancer by race and ethnicity: update 2006. *CA Cancer Journal for Clinicians*, v. 56, p. 168-183, 2006.

Song, J. et al. The research progression of correlation between long non-coding RNA and breast cancer. *Computacional Molecular Biology*, v. 5, n. 2, p. 1-4, 2015.

Sørensen, K. P. et al. Long non-coding RNA expression profiles predict metastasis in lymph node-negative breast cancer independently of traditional prognostic markers. *Breast Cancer Research*, v. 17, p. 55, 2015.

Sotiriou, C.; Pusztai, L. Gene-expression signatures in breast cancer. *The New England Journal of Medicine*, v. 360, p. 790-800, 2009.

St. Laurent, G. et al. The Landscape of long noncoding RNA classification. *Trends in Genetics*, v. 31, n. 5, p. 239-251, 2015.

Stoletov, K. et al. Visualizing extravasation dynamics of metastatic tumor cells. *Journal of Cell Science*, v. 123, p. 2332-2341, 2010.

Su, X. et al. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. *Oncotarget*, v. 5, n. 20, p. 9864-9876, 2014.

Takebe, N.; Nguyen, D.; Yang, S. X. Targeting notch signaling pathway in cancer: clinical development advances and challenges. *Pharmacology and* Therapeutics, v. 141, n. 2, p. 140-149, 2014.

Tano, K. et al. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Letters*, v. 584, n. 22, p. 4575-4580, 2010.

Tantai, J. et al. Combined identification of long non-coding RNA XIST and HIF1A-AS1 in serum as an effective screening for non-small cell lung cancer. *International Journal of Clinical and Experimental Pathology*, v. 8, n. 7, p. 7887–7895, 2015.

Thiery, J. P. Epithelial-mesenchymal transitions in tumor progression. *Nature Reviews Cancer*, v. 2, p. 6, p. 442-454, 2002.

Thiery, J. P.; Sleeman, J. P. Complex networks orchestrate epithelial–mesenchymal transitions. *Nature Reviews. Molecular Cell Biology*, v. 7, n. 2, p. 131-142, 2006.

Thiery, J. P. et al. Epithelial-mesenchymal transitions in development and disease. *Cell*, v. 139, p. 871-890, 2009.

Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, v. 28, n. 5, p. 511-515, 2010.

Tuo, Y. L. et al. Long noncoding RNA UCA1 modulates breast cancer cell growth and apoptosis through decreasing tumor suppressive miR-143. *European Review for Medical and Pharmacological Sciences*, v. 9, n. 18, p. 3403-3411, 2015.

Vara, J. A. F. et al. PI3K/Akt signalling pathway and cancer. *Cancer Treatments Reviews*, v. 30, n. 2, p. 193-204, 2004.

Visvader, J. E.; Stingl, J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes & Development*, v. 28, p. 1143-1158, 2014.

Vikram, R. et al. Functional significance of long non-coding RNAs in breast cancer. *Breast Cancer*, v. 21, p. 515-521, 2014.

Vymetalkova, V. et al. Genotype and haplotype analyses of TP53 gene in breast cancer patients: association with risk and clinical outcomes. *Plos One*, v. 10, n. 7, e0134463, 2015.

Watanabe, T. et al. Dnm3os, a non-coding RNA, is required for normal growth and skeletal development in mice. *Developmental Dynamics*, v. 237, n. 12, p. 3738-3748, 2008.

Wculek, S. K.; Malanchi, I. Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature*, v. 528, p. 413-417, 2015.

Weigelt, B. et al. Gene expression profiles of primary breast tumors maintained in distant metástases. *PNAS*, v. 100, n. 26, p 15901-15905, 2003.

Weigelt, B.; van't Veer, L. J. Hard-wired genotype in metastatic breast cancer. *Cell Cycle*, v. 3, n. 6, p. 756-757, 2004.

Wong, S. Y.; Hynes, R. O. Lymphatic or hematogenous dissemination: How does a metastatic tumor cell decide? *Cell Cycle*, v. 5, n. 8, p. 812-817, 2006.

Yadav, P. et al. Clinical significance of TP53 (R72P) and MDM2 (T309G) polymorphisms in breast cancer patients. *Clinical and Translational Oncology*, 2015. [*Epub ahead of print*]

Yao, D. et al. Mechanism of the mesenchymal–epithelial transition and Its relationship with metastatic tumor formation. *Molecular Cancer Research*, v. 9, n. 12, p. 1608-1620, 2011.

Yarden, Y. Biology of HER2 and its importance in breast cancer. *Oncology*, v. 61, supl. 2, p. 1-13, 2001.

Zervantonakis, I. K. et al. Three-dimensional microfluidic model for tumor cell intravasation and endothelial barrier function. *PNAS*, v. 109, n. 34, p. 13515-13520, 2012.

Zhai, H. et al. Long noncoding RNA MALAT1 as a putative biomarker of lymph node metastasis: a meta-analysis. *International Journal of Clinical and Experimental Meddicine*, v. 8, n. 5, p. 7648-7654, 2015.

Zhang, J. et al. Long non-coding RNA HOTAIR in carcinogenesis and metastasis. *Acta Biochimica et Biophysica Sinica*, v. 46, n. 1, p. 1-5, 2014.

Zhang, L.; Long, X. Association of BRCA1 promoter methylation with sporadic breast cancers: Evidence from 40 studies. *Scientific Reports*, v. 5, p. 17869, 2015.

Zhang, M. et al. Intratumoral heterogeneity in a p53 null mouse model of human breast cancer. *Cancer Discovery*, v. 5, n. 5, p. 520-533, 2015.

Zhao, X. et al. A long noncoding RNA contributes to neuropathic pain by silencing Kcna2 in primary afferent neurons. *Nature Neuroscience*, v. 16, n. 8, p. 1024-1031, 2013.

Zhou, Y. et al. MEG3 noncoding RNA: a tumor suppressor. *Journal of Molecular Endocrinology*, v. 48, n. 3, p. R45-R53, 2012.

Zhu, X. et al. Hypermethylation of BRCA1 gene: implication for prognostic biomarker and therapeutic target in sporadicprimary triple-negative breast cancer. *Breast Cancer Research and Treatment*, v. 150, n. 3, p. 479-486, 2015.

ANEXOS

Anexo A - IncRNAs mais expressos nos tumores primários de mama que deram origem a metástases em linfonodo.

IncRNA	Тіро	log2Fold change	-log10padj
RP11-400N13.3-001	lincRNA	5,1545	3,8446
RP11-1002K11.1-001	lincRNA	4,7936	5,5971
STEAP2-008	íntron retido	4,7915	4,0258
RP11-40109.4-001	antisenso	4,7032	3,5113
RP3-333A15.2-001	senso intrônico	4,6637	2,5782
CTC-327F10.5-001	antisenso	4,3190	2,3502
RP11-680C21.1-001	lincRNA	4,2305	2,3940
RP5-1112F19.2-001	lincRNA	4,0882	2,6741
COL12A1-005	íntron retido	4,0711	2,1380
RP11-53019.1-002	antisenso	4,0100	2,0146
CA12-006	íntron retido	3,9923	2,4011
CTC-327F10.4-001	antisenso	3,7799	2,0664
GABRE-007	íntron retido	3,7584	2,1901
CORIN-006	íntron retido	3,7309	2,5782
MLPH-006	íntron retido	3,6362	2,9814
RP11-417E7.2-001	lincRNA	3,6254	2,2651
AC099684.1-001	antisenso	3,5935	2,6711
SCUBE2-004	íntron retido	3,4480	2,6784
RP11-62I21.1-001	antisenso	3,4129	3,1041
COL1A2-005	íntron retido	3,3705	3,5991
FAP-003	íntron retido	3,3574	2,6395
COL1A2-012	íntron retido	3,3304	3,6959
C7orf63-006	íntron retido	3,2892	2,9651
RP5-1027G4.3-001	antisenso	3,2791	2,3182
COL1A2-006	íntron retido	3,2747	3,5027
COL1A2-007	íntron retido	3,2726	2,8875
RP11-275F13.1-002	antisenso	3,2352	2,0084
PPP2R2C-006	íntron retido	3,2322	2,0510
COL1A2-009	íntron retido	3,2276	2,4920
COL1A2-004	íntron retido	3,1810	2,8728
RP11-709B3.2-001	lincRNA	3,1759	2,8984
COL1A2-002	íntron retido	3,1674	4,1515

IncRNA	Тіро	log2Fold change	-log10padj
FAP-008	íntron retido	3,1416	2,0755
RP11-27M24.3-001	lincRNA	3,1306	2,4629
AC005162.5-001	Antisenso	3,1129	2,8926
MEG3-010	lincRNA	3,0941	2,5706
RP1-60N8.1-001	senso intrônico	3,0588	2,0746
RP11-86H7.7-001	lincRNA	3,0454	2,3304
LINC00702-003	lincRNA	2,9595	2,1075
RP11-5407.3-003	íntron retido	2,9445	2,3900
CTD-2127H9.1-002	lincRNA	2,9373	3,1113
AC004538.3-005	Antisenso	2,8811	2,0606
COL3A1-002	íntron retido	2,8567	2,9997
TBX5-AS1-001	Antisenso	2,6099	3,1113
FARP1-017	íntron retido	2,6000	2,5187
RP11-282K24.3-001	senso intrônico	2,4165	2,3316
MICAL2-003	íntron retido	2,3122	2,3957
AC092171.2-001	Antisenso	2,3067	2,1188
FN1-018	íntron retido	2,3044	2,2651
MICAL2-004	íntron retido	2,2461	2,1235
RP3-439F8.1-001	Antisenso	2,1898	2,8390
WWC2-010	íntron retido	2,1534	2,0118
PARVA-002	íntron retido	2,1202	2,4489
COL6A2-007	íntron retido	2,1149	2,8728

Anexo A - IncRNAs mais expressos nos tumores primários de mama que deram origem a metástases em linfonodo.

IncRNA	Тіро	log2Fold change	-log10padj
RP11-40109.4-001	antisenso	5,8546	4,7356
TBX5-AS1-001	antisenso	4,7860	5,9774
CORIN-006	íntron retido	4,6557	2,4248
RP11-62I21.1-001	antisenso	3,9247	2,9017
COL6A3-009	íntron retido	3,8251	7,4754
COL1A2-004	íntron retido	3,6172	2,5137
COL1A2-005	íntron retido	3,4980	2,7186
COL1A2-002	íntron retido	3,3056	3,6430
COL3A1-002	íntron retido	3,2436	2,7570
COL1A2-006	íntron retido	3,2351	2,0085
COL6A3-007	íntron retido	3,1903	2,1289
COL1A2-012	íntron retido	3,1779	2,2894
RP11-1002K11.1-001	lincRNA	3,0786	2,0879
LINC00861-001	lincRNA	3,0443	2,1658
RP11-426C22.4-001	lincRNA	3,0336	2,3791
DNM3OS-001	antisenso	3,0147	2,5248
COL6A2-007	íntron retido	2,6669	2,7774
RP11-815J21.4-001	antisenso	2,0050	2,1254

Anexo B - IncRNAs mais expressos nos tumores primários de mama que deram origem a metástases em cérebro.

IncRNA	Тіро	log2Fold change	-log10padj
RP11-297B17.3-001	antisenso	5,8813	3,8676
RP11-164H13.1-001	antisenso	5,6246	3,4619
RP11-148021.6-001	lincRNA	5,2166	3,4262
CD22-011	íntron retido	5,0077	4,1515
MS4A1-009	íntron retido	4,8229	2,8875
LINC00494-004	lincRNA	4,8165	3,1041
RP11-138I18.2-001	lincRNA	4,6246	3,4103
RP11-344B23.2-001	antisenso	4,4494	2,1316
RP11-148021.4-001	antisenso	4,4285	3,0339
LRMP-004	íntron retido	4,3664	2,1947
RP11-693J15.5-001	lincRNA	4,3627	3,4009
U62631.5-001	senso intrônico	4,3556	3,1041
RP11-148021.2-001	antisenso	4,3334	2,9434
BLK-006	íntron retido	4,2860	2,8079
CHRM3-AS2-001	antisenso	4,2138	2,9434
TCL6-013	íntron retido	4,1232	2,4489
HVCN1-005	íntron retido	4,0361	2,6540
RP4-555D20.2-001	lincRNA	4,0311	2,7332
CTA-250D10.23-001	lincRNA	4,0209	3,2969
RP11-348F1.3-001	antisenso	4,0010	2,5398
FCRL2-002	íntron retido	3,9851	2,9882
LINC00402-001	lincRNA	3,9207	2,3870
CD79B-005	íntron retido	3,8682	2,1188
LINC00926-001	lincRNA	3,7919	3,1041
RP11-498M5.2-001	antisenso	3,7795	2,4384
RP11-428G5.5-001	lincRNA	3,7131	2,5740
RP11-554l8.1-001	lincRNA	3,6621	2,3597
RP11-16E12.1-002	lincRNA	3,6345	2,0975
FAM65B-004	íntron retido	3,5549	2,1222
AL928768.3-001	lincRNA	3,5333	2,1118
HLA-DOB-002	íntron retido	3,5221	2,4946
CD69-004	íntron retido	3,4560	2,4747
RP11-861A13.4-001	lincRNA	3,4554	2,3494
LINC00926-002	lincRNA	3,4264	2,3498
RP11-796E2.4-001	antisenso	3,4228	2,3502

Anexo C - IncRNAs mais expressos nos tumores metástaticos de mama em linfonodo.

IncRNA	Тіро	log2Fold change	-log10padj
CD69-002	íntron retido	3,3125	2,1245
LRMP-017	íntron retido	3,2681	2,4929
LY9-006	íntron retido	3,2611	2,4273
LINC00861-001	lincRNA	3,2347	3,8446
P2RX5-003	íntron retido	3,2321	2,1571
CYTIP-005	íntron retido	3,2058	2,6335
CTC-457L16.2-001	antisenso	3,1525	2,8390
IL16-015	íntron retido	3,1142	2,1125
RP11-761I4.4-001	ncRNA sobreposto a 3'	3,1078	2,7127
RP3-455J7.4-001	antisenso	3,0820	2,3502
RP11-77C3.3-002	antisenso	2,9986	2,1118
RP11-730K11.1-001	antisenso	2,9639	2,4629
AC006129.2-001	lincRNA	2,9609	2,5118
AC006129.2-003	lincRNA	2,8875	2,8592
SNX29P2-001	lincRNA	2,8346	2,6294
TMEM163-002	íntron retido	2,7138	2,4920
7SK.4-001	lincRNA	2,6110	2,6193
CTD-2306M10.1-001	lincRNA	2,5495	2,1705
AC007386.4-002	antisenso	2,4658	2,3663
RP3-323N1.2-001	antisenso	2,4415	2,2372
TRAF3IP3-004	íntron retido	2,3763	2,0996
FGD2-002	íntron retido	2,3245	2,5258
RP11-94L15.2-001	lincRNA	2,2698	2,1237
RP11-16E12.1-001	lincRNA	2,2376	2,1791

Anexo C - IncRNAs mais expressos nos tumores metástaticos de mama em linfonodo.

	Tine	log2Fold shange	log10nodi
INCKINA	про	logzfold change	-logropadj
SOX2-OT-008	senso sobreposto	5,4457	7,2162
MIR137HG-001	lincRNA	4,9659	3,0307
RP11-731J8.2-001	senso sobreposto	4,9327	4,3548
KRT23-006	íntron retido	4,9128	3,0199
RP11-215H22.1-001	lincRNA	4,8938	2,1800
GRIA2-003	íntron retido	4,6603	2,3900
RP1-240B8.3-001	ncRNA sobreposto a 3'	4,3557	2,0830
RP11-389G6.3-001	senso sobreposto	4,3420	2,6009
RP11-286B14.1-001	lincRNA	4,1462	2,0820
RP11-2L8.1-001	lincRNA	4,1103	2,2973
RP11-297M9.2-001	lincRNA	4,0002	2,3104
TMEM132B-001	íntron retido	3,8079	2,1289
RCAN1-011	íntron retido	3,3055	2,3656
RP11-379F4.6-001	lincRNA	3,2115	3,0064

Anexo D - IncRNAs mais expressos nos tumores metástaticos de mama em cérebro.

Amostra	IncRNA	Proteínas de interação	Função molecular	Processo biológico						
_	RP11- 400N13.3- 001	CPSF100	Ligação de RNA	Expressão gênica, processamento de extremidade 3' de mRNA, clivagem de mRNA						
		MOV10	Ligação de ATP, atividade de helicase, ligação de RNA poli(A)	Via de sinalização do EGFR, via de sinalização Notch, via de sinalização do FGFR						
		eIF4AIII	Ligação de ATP, ligação de mRNA, ligação de poli(A)	Via de sinalização mediada por citocina, splicing de mRNA, regulação negativa da tradução						
	RP5-	Fip1	Ligação de RNA poli(A)	Expressão gênica, processamento de extremidade 3' de mRNA, splicing de mRNA						
Mama (linfonodo)	001	hnRNPA1	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de DNA dupla-fita	Splicing alternativo de mRNA, exportação e importação nuclear, regulação da manutenção telomérica via telomerase						
			LIN28A	Ligação de DNA, ligação de miRNA, ligação de mRNA	Processamento de pre-miRNA, manutenção de população de células- tronco somáticas, diferenciação de células-tronco					
		DGCR8	Ligação de RNA dupla fita, ligação de miRNA primário, atividade de homodimerização de proteína	Silenciamento gênico por RNA, regulação de proliferação de célula- tronco, processamento de miRNA primário						
	AC099684. 1-001	elF4AIII	Ligação de ATP, ligação de mRNA, ligação de poli(A)	Via de sinalização mediada por citocina, splicing de mRNA, regulação negativa da tradução						
		hnRNPA1	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de DNA dupla-fita	Splicing alternativo de mRNA, exportação e importação nuclear, regulação da manutenção telomérica via telomerase						
		HuR	Ligação de RNA dupla fita, ligação de RNA, ligação de nucleotídeo	Estabilização de mRNA, regulação positiva da tradução, regulação da manutenção de população de células-tronco						
								NUDT21	Ligação de histona deacetilase, atividade de hidrolase, ligação de mRNA	Splicing de mRNA, poliadenilação de mRNA, expressão gênica
			RTCB	Ligação de ATP, ligação de RNA poli(A), ligação de vinculina	Desenvolvimento de placenta, processamento de tRNA, desenvolvimento embrionário					
	0011	AGO2	Atividade endonucleásica, ligação de miRNA e de siRNA	Via de sinalização do EGFR, via de sinalização Notch, silenciamento gênico por RNA						
Linfonodo	693J15.5-	hnRNPA1	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de DNA dupla-fita	Splicing alternativo de mRNA, exportação e importação nuclear, regulação da manutenção telomérica via telomerase						
	001	РТВ	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de pre-mRNA	Splicing alternativo de mRNA, regulação negativa de diferenciação de célula muscular, processamento de mRNA						

ATP, adenosina tri-fosfato; mRNA, RNA mensageiro; EGFR, receptor de fator de crescimento epidermal; miRNA, micro RNA; tRNA, RNA transportador.

Amostra	IncRNA	Proteínas de interação	Função molecular	Processo biológico	
	RP11- 693J15.5-	AGO2	Atividade endonucleásica, ligação de miRNA e de siRNA	Via de sinalização do EGFR, via de sinalização Notch, silenciamento gênico por RNA	
		hnRNPA1	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de DNA dupla-fita	Splicing alternativo de mRNA, exportação e importação nuclear, regulação da manutenção telomérica via telomerase	
	001	РТВ	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de pre-mRNA	Splicing alternativo de mRNA, regulação negativa de diferenciação de célula muscular, processamento de mRNA	
	CHRM3-	CPSF100	Ligação de RNA	Expressão gênica, processamento de extremidade 3' de mRNA, clivagem de mRNA	
	AS2-001	hnRNPF	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de RNA fita simples	Splicing de mRNA, processamento de RNA, expressão gênica	
Linfonodo		AGO2	Atividade endonucleásica, ligação de miRNA e de siRNA	Via de sinalização do EGFR, via de sinalização Notch, silenciamento gênico por RNA	
	TCL6-013	ALKBH5	Ligação de íon metal, ligação de RNA poli(A), atividade de oxidorredutase	Diferenciação celular, resposta à hipóxia, processamento de mRNA	
			Fip1	Ligação de RNA poli(A)	Expressão gênica, processamento de extremidade 3' de mRNA, splicing de mRNA
		FMR1	Ligação de mRNA, ligação de RNA poli(A), ligação de RNA	Desenvolvimento do sistema nervoso central, transporte de mRNA, regulação negativa da tradução	
		hnRNPA1	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de DNA dupla-fita	Splicing alternativo de mRNA, exportação e importação nuclear, regulação da manutenção telomérica via telomerase	
				hnRNPU	Ligação de ATP, ligação a <i>enhancer</i> , ligação a RNA telomerase
	TBX5-AS1-	AGO2	Atividade endonucleásica, ligação de miRNA e de siRNA	Via de sinalização do EGFR, via de sinalização Notch, silenciamento gênico por RNA	
Mama	001	HuR	Ligação de RNA dupla fita, ligação de RNA, ligação de nucleotídeo	Estabilização de mRNA, regulação positiva da tradução, regulação da manutenção de população de células-tronco	
(cérebro)		CPSF100	Ligação de RNA	Expressão gênica, processamento de extremidade 3' de mRNA, clivagem de mRNA	
	LINCUU861	elF4AIII	Ligação de ATP, ligação de mRNA, ligação de poli(A)	Via de sinalização mediada por citocina, splicing de mRNA, regulação negativa da tradução	

ATP, adenosina tri-fosfato; mRNA, RNA mensageiro; EGFR, receptor de fator de crescimento epidermal; miRNA, micro RNA; tRNA, RNA transportador.

LINC00861 hnRNPA1 Ligação de nucleotídeo, ligação de RNA poli(A), Splicing alternativo de mRNA, exportação e importação nuclear, r ligação de DNA dupla-fita da manutenção telomérica via telomerase	gulação
CPSF100 Ligação de RNA Expressão gênica, processamento de extremidade 3' de mRNA, c de mRNA de mRNA	ivagem
Ligação de RNA dupla fita, ligação de miRNA primário, atividade de homodimerização de proteínaSilenciamento gênico por RNA, regulação de proliferação de célula processamento de miRNA primário	-tronco,
(cérebro) DNM3OS FMR1 Ligação de mRNA, ligação de RNA poli(A), ligação de RNA Desenvolvimento do sistema nervoso central, transporte de m regulação negativa da tradução	RNA,
hnRNPA2/B1 Ligação de miRNA, ligação de 3'-UTR de mRNA, Transporte de miRNA, processamento de mRNA, regulação nega ligação de RNA poli(A) splicing de RNA	tiva do
IGF2BP1 Atividade reguladora da tradução, ligação de mRNA, ligação de nucleotídeo Manutenção de população de células-tronco neuronais, reguladora da tradução, ligação de mRNA envolvido em resposta ao estresse, transporta a	ão da orte de
ALKBH5 Ligação de íon metal, ligação de RNA poli(A), atividade de oxidorredutase Diferenciação celular, resposta à hipóxia, processamento de m	RNA
CCNT1 Ligação à cromatina, ligação a DNA, ligação a SnRNA Divisão celular, ciclo celular, via de sinalização do receptor de T	GF-β
CPSF100 Ligação de RNA Expressão gênica, processamento de extremidade 3' de mRNA, c de mRNA de mRNA	ivagem
CérebroSOX2-OTLigação de RNA dupla fita, ligação de miRNA primário, atividade de homodimerização de proteínaSilenciamento gênico por RNA, regulação de proliferação de célula processamento de miRNA primário	-tronco,
elF4AIII Ligação de ATP, ligação de mRNA, ligação de Via de sinalização mediada por citocina, splicing de mRNA, regu poli(A) negativa da tradução	lação
FMR1Ligação de mRNA, ligação de RNA poli(A), ligação de RNADesenvolvimento do sistema nervoso central, transporte de m regulação negativa da tradução	RNA,

ATP, adenosina tri-fosfato; mRNA, RNA mensageiro; EGFR, receptor de fator de crescimento epidermal; miRNA, micro RNA; tRNA, RNA transportador; TGF-β, fator de crescimento tumoral beta.

Amostra	IncRNA	Proteínas de interação	Função molecular	Processo biológico	
		ALKBH5	Ligação de íon metal, ligação de RNA poli(A), atividade de oxidorredutase	Diferenciação celular, resposta à hipóxia, processamento de mRNA	
		CCNT1	Ligação à cromatina, ligação a DNA, ligação a snRNA	Divisão celular, ciclo celular, via de sinalização do receptor de TGF- β	
		CPSF100	Ligação de RNA	Expressão gênica, processamento de extremidade 3' de mRNA, clivagem de mRNA	
SOX2-OT Cérebro		DGCR8	Ligação de RNA dupla fita, ligação de miRNA primário, atividade de homodimerização de proteína	Silenciamento gênico por RNA, regulação de proliferação de célula- tronco, processamento de miRNA primário	
	SOX2-OT	elF4AIII	Ligação de ATP, ligação de mRNA, ligação de poli(A)	Via de sinalização mediada por citocina, splicing de mRNA, regulação negativa da tradução	
		FMR1	Ligação de mRNA, ligação de RNA poli(A), ligação de RNA	Desenvolvimento do sistema nervoso central, transporte de mRNA, regulação negativa da tradução	
		hnRNPA1	Ligação de nucleotídeo, ligação de RNA poli(A), ligação de DNA dupla-fita	Splicing alternativo de mRNA, exportação e importação nuclear, regulação da manutenção telomérica via telomerase	
		hnRNPH	Ligação de nucleotídeo, ligação de RNA poli(A), splicing de RNA	Regulação de splicing de RNA, expressão gênica, processamento de RNA	
		HuR	Ligação de RNA dupla fita, ligação de RNA, ligação de nucleotídeo	Estabilização de mRNA, regulação positiva da tradução, regulação da manutenção de população de células-tronco	
	RP11- 286B14.1	CPSF100	Ligação de RNA	Expressão gênica, processamento de extremidade 3' de mRNA, clivagem de mRNA	
RP11-2L8		CCNT1	Ligação à cromatina, ligação a DNA, ligação a snRNA	Divisão celular, ciclo celular, via de sinalização do receptor de TGF- β	
			elF4AIII	Ligação de ATP, ligação de mRNA, ligação de poli(A)	Via de sinalização mediada por citocina, splicing de mRNA, regulação negativa da tradução
	NP11-2L8.1	Fip1	Ligação de RNA poli(A)	Expressão gênica, processamento de extremidade 3' de mRNA, splicing de mRNA	
		hnRNPH	Ligação de nucleotídeo, ligação de RNA poli(A), splicing de RNA	Regulação de splicing de RNA, expressão gênica, processamento de RNA	

ATP, adenosina tri-fosfato; mRNA, RNA mensageiro; EGFR, receptor de fator de crescimento epidermal; miRNA, micro RNA; tRNA, RNA transportador; TGF-β, fator de crescimento tumoral beta.

Anexo F - Aprovação do Comitê de Ética em Pesquisa do HCFMRP-USP.



HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA USP -

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Análise qualitativa e quantitativa de IncRNA de tumores metastáticos de mama no sistema nervoso central e osso Pesquisador: Isabela Ichihara de Barros

Área Temática: Versão: 1 CAAE: 35281214.1.0000.5440 Instituição Proponente: Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da USP -Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 829.151 Data da Relatoria: 27/08/2014

Apresentação do Projeto:

Trata-se de um estudo que será coordenado pelo Prof. Dr. Wilson Araújo da Silva Jr e pela aluna de mestrado Isabela Ichihara de Barros do programa de pós-graduação em Genética da FMRP-USP. O estudo será realizado por meio da utilização amostras que serão fornecidas pelo Banco de Tumores (aprovação junto ao Comitê de Ética no 7645/99) e pelo Banco de Tumores de Mama (aprovação junto ao Comitê de ética no 2467/2009) do HCFMRP/USP. As amostras de tumores do sistema nervoso central serão obtidas em colaboração com o Prof. Dr. Carlos Gilberto Carlotti Junior e as de tumores ósseo em colaboração com o Prof. Dr. Helton Luiz Aparecido Defino. Serão analisadas 20 amostras de tumor de mama, sendo 5 tumores metastáticos no sistema nervoso central e 5 de tumores metastáticos ósseos. Adicionalmente serão analisados 5 tumores primários, pareados com os respectivos tumores metastáticos

Objetivo da Pesquisa:

O objetivo Geral do estudo é obter o perfil de expressão de IncRNAs e identificar novos IncRNA associados com tumores metastáticos de sistema nervoso central e osso e os específicos: determinar o perfil de expressão de IncRNAs de tumor de mama metastático no sistema nervoso central e ósseo quanto à: Classificação funcional dos IncRNAs preferencialmente expressos nos tumores metastáticos; e Seleção de IncRNAs potencialmente envolvidos com a metástase.

Endereço: CAMPUS UNIVERSITÁRIO					
Bairro: MONTE ALEGRE CEP:			14.048-900		
UF: SP	Município:	RIBEIRAO PRETO			
Telefone:	(16)3602-2228	Fax: (16)3633-1144	E-mail:	cep@hcrp.usp.br	

Página 01 de 03

Anexo F - Aprovação do Comitê de Ética em Pesquisa do HCFMRP-USP.



```
HOSPITAL DAS CLÍNICAS DA
FACULDADE DE MEDICINA DE
RIBEIRÃO PRETO DA USP -
```

Continuação do Parecer: 829.151

Avaliação dos Riscos e Benefícios:

Os pesquisadores mencionam que com os resultados obtidos, será possível estabelecer uma assinatura de expressão de IncRNAs, o qual será útil para uma melhor compreensão do

papel efetivo desses RNAs na regulação do mecanismo de metástase, além de possibilitar a identificação biomarcadores de prognóstico e terapêuticos. E descrevem que não há riscos uma vez que não haverá novas coletas para este estudo.

Comentários e Considerações sobre a Pesquisa:

O projeto está bem descrito contendo todas as etapas com clareza, descrevem critérios de inclusão das amostras para o estudo e a forma de análise de dados. Os pesquisadores solicitam a dispensa do Termo de Consentimento Esclarecido por se tratar de estudo que será realizado com amostras de de tumores de mama já coletadas.

Considerações sobre os Termos de apresentação obrigatória:

Apresentou folha de rosto devidamente assinada pelo pesquisador e Orçamento que foi aprovado pela Unidade de Pesquisa Clínica (UPC) do HCFMRP-USP; Solicitam dispensa do TCLE visto que os pacientes são graves e que seriam de difícil acesso. Os pesquisadores mencionam que poderão consultar prontuário para confirmação de dados.

Recomendações:

Considerando que os pesquisadores irão consultar prontuário os pesquisadores resposnáveis devem garantir o sigilo e anonimato das informações pessoais.

Conclusões ou Pendências e Lista de Inadequações:

Diante do exposto e à luz da Resolução CNS 466/2012, o projeto de pesquisa, assim como a solicitação de dispensa de aplicação do Termo de Consentimento Livre e Esclarecido, podem ser enquadrados na categoria APROVADO.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

Projeto Aprovado: Tendo em vista a legislação vigente, devem ser encaminhados ao CEP, relatórios parciais anuais referentes ao andamento da pesquisa e relatório final ao término do trabalho. Qualquer modificação do projeto original deve ser apresentada a este CEP em nova versão, de forma objetiva e com justificativas, para nova apreciação.

```
        Endereço:
        CAMPUS UNIVERSITÁRIO

        Bairro:
        MONTE ALEGRE
        CEP:
        14.048-900

        UF:
        SP
        Município:
        RIBEIRAO PRETO

        Telefone:
        (16)3602-2228
        Fax:
        (16)3633-1144
        E-mail:
        cep@hcrp.usp.br
```