

**ESTUDO DE VESÍCULAS EXTRACELULARES  
EM LINHAGENS CELULARES E EM PACIENTES  
COM CÂNCER DE MAMA POSITIVAS PARA  
A AMPLIFICAÇÃO DE ERBB2/HER2**

**MARIA GALLI DE AMORIM**

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

**Área de Concentração: Oncologia**

**Orientador: Dr. Emmanuel Dias-Neto**

**Co-Orientadora: Dra. Diana Noronha Nunes**

**São Paulo**

**2015**

## FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Amorim, Maria Galli de

**Estudo de vesículas extracelulares em linhagens celulares e em pacientes com câncer de mama positivas para a amplificação de ERBB2/HER2** / Maria Galli de Amorim - São Paulo, 2015.

165p.

Tese (Doutorado)-Fundação Antônio Prudente.

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

Orientador: Emmanuel Dias-Neto

Descritores: 1. RECEPTOR ERBB-2/genética. 2. NEOPLASIAS DA MAMA/genética. 3. MICROPARTÍCULAS DERIVADAS DE CÉLULAS. 4. EXOSSOMOS.

*“Just when the caterpillar thought the world was over, it became a butterfly.”*

*- anônimo*

## DEDICATÓRIA

*Aos meus queridos pais, Vânia e Lucio, por apesar da distância geográfica,  
sempre estarem muito presentes em tudo, me dando amor e incentivo  
incondicionais.*

*Aos meus irmãos, Tomás e Gustavo, pelo carinho.*

*Ao Marcos Vinicius, por todo o seu amor, paciência, companheirismo e por  
me fazer muito feliz.*

*Amo vocês!*

## **AGRADECIMENTOS**

**Ao meu orientador, Dr. Emmanuel Dias-Neto**, pela confiança, ensinamentos e apoio diário durante todo o desenvolvimento deste trabalho. Sou muita grata pela oportunidade profissional que me proporcionou.

**À minha co-orientadora, Dra. Diana Noronha Nunes**, mãe deste projeto, por todo o seu entusiasmo, auxílio e acompanhamento constante.

**Aos colegas e amigos do Laboratório de Genômica Médica:** Dra. Ana Costa, Andrew Thomas, Frederico Netto, Gabriela Branco, Dr. Gustavo Ribeiro, Dra. Juliana Laino, Jéssica Takarada, Melissa Pizzi, e Sheila Garcia, pelos aprendizados que compartilhamos e os muitos momentos divertidos que passamos juntos.

**Aos colegas da Bioinformática,** Dr. Jorge Estafano de Souza, Matheus Bürger e em especial ao Renan Valieris, que foi fundamental para a análise de dados, pela sua imensa paciência e dedicação.

**A todos que me auxiliaram na coleta de amostras:** Louise Danielle de Carvalho Mota, Dr. Diogo Patrão, Dr. Rubens Chojniak, Dr. Eduardo Nóbrega Lima, enfermeiras Leia, Téo e Rita, anestesistas do centro cirúrgico, Ana Carolina Ferreira, Dra. Maria do Socorro Maciel, Dra. Maria Auxiliadora Bernardi, e em especial ao José Ivanildo Neves, que foi fundamental para a identificação das pacientes elegíveis.

**Aos colaboradores do Hospital de Câncer de Barretos:** Dr. José Humberto Tavares Guerreiro Fregnani e Viviane Andrade por terem tornado possível termos completado todos os pacientes inicialmente planejados.

**Ao Dr. Daniel Martins-de-Souza**, que realizou a análise proteômica publicada e sempre foi muito atencioso com este trabalho.

**Ao Dr. Paulo Sergio Lopes de Oliveira**, pela ajuda com a análise dos dados de proteômica com o programa MetaCore.

**À Dra. Vilma Martins**, pelo empréstimo de diversos anticorpos e por manter o seu laboratório sempre aberto.

**A todos os membros do Laboratório de Biologia Celular e Molecular**, e em especial ao Dr. Martín Roffé e a Dra. Glaucia Hajj, pela troca de experiências e discussões.

**Ao Dr. Marcos Salles, e aos técnicos do ICB/USP Gaspar Ferreira de Lima, e Edson Rocha de Oliveira** pelo auxílio com os experimentos de microscopia eletrônica.

**À Dra. Dirce Carraro**, pelo empréstimo das linhagens celulares e dos reagentes para sequenciamento Sanger.

**À Dra. Bruna Barros**, por inúmeras ajudas com sequenciamento.

**À Dra. Elisa Napolitano e Ferreira**, pelos seus ensinamentos e frequentes consultorias científicas.

**A todos os membros do Laboratório de Genômica e Biologia Molecular**, por estarem sempre dispostos a ajudar.

**Aos membros da banca de qualificação, Dr. Sergio Verjovski-Almeida e Dr. Vladimir Cláudio Cordeiro de Lima**, pelo acompanhamento e sugestões ao longo da realização deste trabalho.

**Aos funcionários da pós-graduação e biblioteca**, em especial a Ana Kuninari, Vanuza Barros, Luciana Costa, e Suely Francisco por toda a atenção ao longo desses 4 anos.

**À Carla Regina, Gilmara Silva, Kelly Cristina, Patricia Santos e Vanessa Dantas** pelo apoio administrativo.

**À infraestrutura do Centro Internacional de Pesquisa e Ensino (CIPE)** do AC Camargo Cancer Center, essencial para a realização deste trabalho.

**Ao Dr. Wagner Gattaz**, pelo empréstimo de ultracentrífuga fundamental para a conclusão de certas análises desta tese.

**Ao Prof. Ricardo Brentani**, em memória, que tinha um carinho especial por este projeto desde o seu início.

**Ao suporte financeiro indispensável da FAPESP (2011/09172-3).**

**Às pacientes**, que por voluntariamente cederem seu material biológico possibilitaram a realização deste trabalho.

## RESUMO

Amorim MG. **Estudo de vesículas extracelulares em linhagens celulares e em pacientes com câncer de mama positivas para a amplificação de ERBB2/HER2.** São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente]

As Vesículas Extracelulares (EVs) são organelas essenciais para processos fisiológicos e aparentemente contribuem para o desenvolvimento de diversas doenças através da transferência entre células de ácidos nucléicos e proteínas contidos em seu interior. Deste modo, EVs são promissoras como fonte de biomarcadores. Neste trabalho, avaliamos o conteúdo de EVs secretadas *in vitro* e *in vivo*, com o intuito de identificar moléculas relacionadas com a amplificação do oncogene ERBB2/HER2 e com estádios de progressão tumoral de câncer de mama HER2+. Utilizando um modelo de duas linhagens celulares - HB4a derivada de mama normal e C5.2, um clone de HB4a que superexpressa o oncogene ERBB2/HER2 - isolamos duas populações de EVs provenientes do meio de cultura celular condicionado: 20K, enriquecida para microvesículas e 100K, enriquecida para exossomos. Demonstramos que a superexpressão de um único oncogene (ERBB2/HER2) altera o perfil proteômico das EVs secretadas pela células C5.2. Proteínas capazes de induzir a transformação maligna foram encontradas superexpressas nas duas populações de EVs da linhagem C5.2, e interessante observamos uma maior expressão de HER2 nas EVs em relação a esta mesma linhagem celular. Uma análise por sequenciamento de nova geração do transcriptoma completo (RNAs pequenos e longos) das EVs e células secretoras revelou centenas de transcritos diferencialmente representados entre as duas linhagens, que podem estar direta ou indiretamente sendo afetados pela superexpressão de HER2, e possibilitou uma investigação das moléculas preferencialmente incorporadas nas EVs. Avaliamos também EVs circulantes no plasma de



trinta pacientes com câncer de mama ductal invasivo HER2+, divididas em três grupos de estadiamento progressivo da doença. Caracterizamos o perfil transcricional destas EVs, validamos experimentalmente um miRNA (miR-223-3p) que pode estar envolvido com o desenvolvimento de metástase e identificamos 1.808 transcritos únicos diferencialmente representados entre os três grupos, que podem estar correlacionados com a progressão tumoral. De maneira inédita, identificamos em EVs derivadas tanto de linhagens celulares quanto de pacientes centenas de RNAs circulares, que são moléculas capazes de atuar como esponjas de miRNAs. O estudo apresentado traz novas oportunidades para a descoberta de biomarcadores em câncer de mama HER2+ e contribui com o melhor entendimento do papel biológico das moléculas carregadas em EVs.

## SUMMARY

Amorim MG. **[Study of extracellular vesicles in cell lines and breast cancer patients positive for ERBB2/HER2 amplification]**. São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente]

Extracellular vesicles (EVs) are essential organelles involved in physiological processes and apparently contribute to the development of several pathological conditions by the transfer between cells of nucleic acid and protein cargo. Therefore, EVs are promising sources of biomarkers. In the study presented here, we evaluated the content of EVs secreted *in vitro* and *in vivo*, aiming to identify molecules related to ERBB2/HER2 oncogene amplification and with progressing tumor stages of HER2+ breast cancer. Using a model of two cell lines – HB4a derived from normal breast and C5.2, a clone of HB4a that overexpresses ERBB2/HER2 – we isolated two populations of EVs derived from the conditioned cell culture medium: 20K, enriched for microvesicles and 100K, enriched for exosomes. We demonstrate that the overexpression of a single oncogene (ERBB2/HER2) alters the proteomic landscape of EVs secreted by C5.2 cells. Proteins capable of inducing malignant transformation were found overexpressed in both EVs populations secreted by C5.2 cells, and interestingly, we observed an increased expression of HER2 in the EVs compared to the secreting cells. Next generation sequencing analysis of the entire transcriptome (long and short RNAs) of EVs and secreting cells revealed hundreds of transcripts differentially expressed between the two cell lines, which may be directly or indirectly affected by HER2 overexpression, and enabled the investigation of molecules that are preferentially shuttled in EVs. We also evaluated circulating EVs in the plasma of thirty patients with invasive HER2+ ductal breast cancer, divided into three groups of progressive disease staging. We characterized the transcriptional profile of these EVs, experimentally validated a miRNA (miR-223-3p) that may be involved with metastasis and

identified 1,808 unique transcripts that are differentially represented between the three groups and that may be correlated to tumor progression. We identified, for the first time in EVs derived from both cell lines and patients, hundreds of circular RNAs, which are molecules capable of acting as miRNAs sponges. Our study creates new opportunities for the discovery of new biomarkers in HER2+ breast cancers and contributes to the better understanding of the biological role of molecules shuttled by EVs.

## LISTA DE FIGURAS

|                  |  |    |
|------------------|--|----|
| <b>Figura 1</b>  | Biogênese das duas principais populações de EVs – microvesículas e exossomos.....  | 3  |
| <b>Figura 2</b>  | Mecanismo de formação das vesículas intraluminais nos endossomos multivesiculares.....   | 4  |
| <b>Figura 3</b>  | Representação esquemática do conteúdo proteico de EVs...   | 7  |
| <b>Figura 4</b>  | Dois principais mecanismos de biogênese dos miRNAs: via canônica e via não-canônica.....   | 21 |
| <b>Figura 5</b>  | Desenho esquemático da PCR em emulsão, onde ocorre a amplificação clonal das moléculas de DNA nas <i>beads</i> .....                 | 45 |
| <b>Figura 6</b>  | Desenho esquemático do controle de qualidade feito após a PCR em emulsão.....  | 46 |
| <b>Figura 7</b>  | Desenho esquemático da química de sequenciamento da plataforma Ion Proton.....   | 47 |
| <b>Figura 8</b>  | Desenho esquemático dos pares de iniciadores utilizados para amplificar as isoformas linear (L) e circular (C) de <i>UBXN7</i> ..... | 52 |
| <b>Figura 9</b>  | Caracterização das EVs 20K e 100K.....   | 57 |
| <b>Figura 10</b> | Análise das proteínas exclusivas das EVs 20K e 100K de C5.2 e HB4a.....  | 64 |

|                  |   |    |
|------------------|---|----|
| <b>Figura 11</b> | Perfil representativo do <i>Bioanalyzer</i> de biblioteca (C5.2 20K) para sequenciamento.....                                 | 67 |
| <b>Figura 12</b> | Distribuição dos transcritos identificados em EVs e células nas diversas categorias de RNAs.....                              | 73 |
| <b>Figura 13</b> | Distribuição dos transcritos identificados em EVs e células na categoria short noncoding RNAs.....                            | 74 |
| <b>Figura 14</b> | <i>Heat map</i> dos 100 miRNAs diferencialmente representados entre as EVs (20K e 100K) e células C5.2 e HB4a.....            | 77 |
| <b>Figura 15</b> | <i>Heat map</i> dos 3.190 outros RNAs diferencialmente representados entre as EVs (20K e 100K) e células das C5.2 e HB4a..... | 79 |
| <b>Figura 16</b> | Categorias dos 3.190 outros RNAs diferencialmente representados entre as EVs (20K e 100K) e células C5.2 e HB4a.....          | 80 |
| <b>Figura 17</b> | <i>Heat map</i> dos 16 miRNAs diferencialmente representados entre as EVs (20K e 100K) das células C5.2 e HB4a.....           | 82 |
| <b>Figura 18</b> | <i>Heat map</i> dos 644 miRNAs diferencialmente representados entre as EVs (20K e 100K) das células C5.2 e HB4a.....          | 83 |
| <b>Figura 19</b> | Categorias dos 644 outros RNAs diferencialmente representados entre as EVs (20K e 100K) das células C5.2 e HB4a.....          | 84 |

|                  |  |     |
|------------------|--|-----|
| <b>Figura 20</b> | Categorias dos 592 outros RNAs diferencialmente representados entre as células C5.2 e HB4a.....  | 88  |
| <b>Figura 21</b> | Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em C5.2 20K.....   | 90  |
| <b>Figura 22</b> | Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em C5.2 100K.....  | 91  |
| <b>Figura 23</b> | Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em HB4a 20K.....   | 92  |
| <b>Figura 24</b> | Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em HB4a 100K.....  | 93  |
| <b>Figura 25</b> | Comparação do percentual de AT dos miRNAs contidos em EVs e células.....   | 94  |
| <b>Figura 26</b> | Diagrama de Venn dos motivos enriquecidos em EVs.....  | 95  |
| <b>Figura 27</b> | Distribuição posicional de todos os transcritos.....   | 98  |
| <b>Figura 28</b> | Distribuição posicional de oncogenes e genes supressores tumorais.....   | 99  |
| <b>Figura 29</b> | Distribuição posicional de oncogenes e genes supressores tumorais de relevância clínica para o câncer de mama.....                                       | 100 |
| <b>Figura 30</b> | Avaliação por Nanotracking Particle Analysis (NTA) da concentração de EVs isoladas do plasma de pacientes dos três grupos de agressividade tumoral ..... | 103 |

|                  |   |     |
|------------------|---|-----|
| <b>Figura 31</b> | Imunodeteccção por Western Blot dos marcadores de EVs RAB7 e Flotilina em EVs derivadas de pools de plasma de pacientes.....                | 104 |
| <b>Figura 32</b> | Perfil representativo do Bioanalyzer de biblioteca para sequenciamento, obtido a partir de RNAs em EVs derivadas do plasma de paciente..... | 105 |
| <b>Figura 33</b> | <i>Heat map</i> dos 7 miRNAs diferencialmente representados entre os grupos de pacientes N0 e N.....  | 111 |
| <b>Figura 34</b> | <i>Heat map</i> dos 593 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e N.....                                 | 112 |
| <b>Figura 35</b> | Categorias dos 593 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e N.....                                      | 113 |
| <b>Figura 36</b> | <i>Heat map</i> dos 19 miRNAs diferencialmente representados entre os grupos de pacientes N0 e M.....                                       | 116 |
| <b>Figura 37</b> | <i>Heat map</i> dos 1.016 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e M.....                               | 117 |
| <b>Figura 38</b> | Categorias dos 1.016 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e M.....                                    | 118 |
| <b>Figura 39</b> | <i>Heat map</i> dos 20 miRNAs diferencialmente representados entre os grupos de pacientes N e M.....  | 120 |
| <b>Figura 40</b> | <i>Heat map</i> dos 708 outros RNAs diferencialmente representados entre os grupos de pacientes N e M.....                                  | 121 |

|                  |   |     |
|------------------|---|-----|
| <b>Figura 41</b> | Categorias dos 708 outros RNAs diferencialmente representados entre os grupos de pacientes N e M.....   | 122 |
| <b>Figura 42</b> | Expressão do mRNA de ERBB2 entre os grupos de pacientes.....  | 123 |
| <b>Figura 43</b> | Análise de miRNAs diferencialmente representados entre os grupos de pacientes N e N0.....   | 124 |
| <b>Figura 44</b> | Análise de miRNAs diferencialmente representados entre os grupos de pacientes M e N.....  | 125 |
| <b>Figura 45</b> | Análise de miRNAs diferencialmente representados entre os grupos de pacientes M e N0.....   | 125 |
| <b>Figura 46</b> | Boxplots dos miRNAs diferencialmente representados entre os grupos de pacientes.....  | 126 |
| <b>Figura 47</b> | Boxplots dos miRNAs selecionados como controles endógenos a partir dos dados de NGS.....  | 127 |
| <b>Figura 48</b> | Boxplots dos valores de Ct dos miRNAs inicialmente selecionados como controles endógenos.....   | 128 |
| <b>Figura 49</b> | Boxplots dos miRNAs avaliados por qRT-PCR.....  | 129 |
| <b>Figura 50</b> | Estrutura predita e número de reads de um dos potenciais precursores de novos miRNAs, mostrando alinhamento das sequências aqui obtidas para os miRNAs 5p e 3p a partir do precursor predito..... | 136 |



|                  |  |     |
|------------------|--|-----|
| <b>Figura 51</b> | Visualização em gel de poliacrilamida dos produtos de amplificação do transcrito UBXN7 em cDNA de célula HB4a.....                                   | 139 |
| <b>Figura 52</b> | Visualização em gel de poliacrilamida dos produtos de amplificação de UBXN7 (isoforma circular) em cDNA de EVs derivadas do plasma de pacientes..... | 140 |
| <b>Figura 53</b> | Visualização em gel de poliacrilamida dos produtos de amplificação de UBXN7 (isoforma linear) em cDNA de EVs derivadas do plasma de pacientes.....   | 140 |
| <b>Figura 54</b> | Análise de um dos reads do circRNA UBXN7 no UCSC Genome Browser.....   | 141 |
| <b>Figura 55</b> | Resultado da análise por sequenciamento Sanger do RNA circular do transcrito UBXN7 em cDNA de célula HB4a.....                                       | 142 |

## LISTA DE TABELAS

|                 |   |    |
|-----------------|---|----|
| <b>Tabela 1</b> | Principais tipos de RNAs descritos em células de mamíferos.....   | 19 |
| <b>Tabela 2</b> | Dez proteínas mais superexpressas nas EVs 20K e 100K das linhagens C5.2 (cinza claro) e HB4a (cinza escuro).....                  | 60 |
| <b>Tabela 3</b> | Análise por MetaCore de enriquecimento de vias de sinalização nas EVs derivadas de HB4a e C5.2.....                               | 63 |
| <b>Tabela 4</b> | Análise dos dados de sequenciamento em larga escala de miRNAs de EVs derivadas das linhagens C5.2 e HB4a.....                     | 69 |
| <b>Tabela 5</b> | Comparação do número de precursores de miRNAs identificados nas EVs em relação as células.....                                    | 71 |
| <b>Tabela 6</b> | Análise dos dados de sequenciamento em larga escala de outros RNAs de EVs derivadas das linhagens C5.2 e HB4a.....                | 72 |
| <b>Tabela 7</b> | Análise estatística da distribuição dos transcritos identificados em EVs e células nas diversas categorias de RNAs.....           | 73 |
| <b>Tabela 8</b> | Análise estatística da distribuição dos transcritos identificados em EVs e células na categoria <i>short noncoding RNAs</i> ..... | 75 |

|                  |  |     |
|------------------|--|-----|
| <b>Tabela 9</b>  | Análise dos <i>reads</i> de C5.2 e HB4a que não anotaram após alinhamento contra miRBase e hg19/Ensembl.....           | 76  |
| <b>Tabela 10</b> | Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 20K da linhagem C5.2.....  | 90  |
| <b>Tabela 11</b> | Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 100K da linhagem C5.2..... | 91  |
| <b>Tabela 12</b> | Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 20K da linhagem HB4a.....  | 92  |
| <b>Tabela 13</b> | Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 100K da linhagem HB4a..... | 93  |
| <b>Tabela 14</b> | Resumo dos dados gerais das pacientes do estudo incluídas no sequenciamento em larga escala.....                       | 102 |
| <b>Tabela 15</b> | Pacientes incluídas no sequenciamento em larga escala...   | 102 |
| <b>Tabela 16</b> | Análise dos dados de sequenciamento em larga escala de miRNAs de EVs derivadas do plasma de pacientes.....             | 107 |
| <b>Tabela 17</b> | Análise dos dados de sequenciamento em larga escala de outros RNAs de EVs derivadas do plasma de pacientes.....        | 108 |

|                  |   |     |
|------------------|---|-----|
| <b>Tabela 18</b> | Análise dos <i>reads</i> de EVs derivadas do plasma de pacientes que não anotaram após alinhamento contra miRBase e hg19/Ensembl.....   | 109 |
| <b>Tabela 19</b> | Comparação dos <i>fold-changes</i> obtidos por sequenciamento (NGS) e na validação por qRT-PCR.....                                     | 130 |
| <b>Tabela 20</b> | Potenciais novos miRNAs preditos pelo programa miRDeep2 em EVs derivadas das linhagens C5.2 e HB4a.                                     | 135 |
| <b>Tabela 21</b> | Potenciais novos miRNAs preditos pelo programa miRDeep2 em EVs derivadas do plasma de pacientes.....                                    | 135 |
| <b>Tabela 22</b> | Distribuição do número de <i>reads</i> dos 513 RNAs circulares (circRNAs) identificados nas EVs derivadas do plasma de pacientes.....   | 137 |
| <b>Tabela 23</b> | Distribuição dos <i>reads</i> dos 365 RNAs circulares (circRNAs) identificados em C5.2 e HB4a e EVs derivadas do meio condicionado..... | 138 |

## LISTA DE QUADROS

|                 |   |     |
|-----------------|---|-----|
| <b>Quadro 1</b> | Classificação molecular do câncer de mama.....  | 29  |
| <b>Quadro 2</b> | Concentração molar e fator de diluição das bibliotecas<br>construídas a partir de RNA das EVs e das linhagens<br>celulares que as originaram..... | 44  |
| <b>Quadro 3</b> | Concentração molar e fator de diluição das bibliotecas<br>construídas a partir de RNA obtidos das EVs derivadas do<br>plasma de pacientes.....    | 44  |
| <b>Quadro 4</b> | Ensaio de qRT-PCR usados na avaliação dos miRNAs...   | 50  |
| <b>Quadro 5</b> | Sequências dos iniciadores utilizados para amplificação<br>do transcrito <i>UBXN7</i> isoforma linear (L) e circular (C).....                     | 52  |
| <b>Quadro 6</b> | Distribuição dos seis RNAs circulares (circRNAs) mais<br>abundantes no sequenciamento das EVs derivadas do<br>plasma de pacientes.....            | 137 |
| <b>Quadro 7</b> | Distribuição dos quinze RNAs circulares (circRNAs) mais<br>abundantes em C5.2 e HB4a e nas EVs do meio<br>condicionado.....                       | 138 |

## LISTA DE SIGLAS E ABREVIATURAS

|                                     |  |
|-------------------------------------|--|
| <b><math>\Delta\Delta Ct</math></b> | Delta delta Ct   |
| <b>20K</b>                          | População de EVs obtida por ultracentrifugação a 20.000 xg                                       |
| <b>100K</b>                         | População de EVs obtida por ultracentrifugação a 100.000 xg                                      |
| <b>3'-UTR</b>                       | <i>3'-untranslated region</i> - Região 3' não traduzida  |
| <b>ACD</b>                          | Ácido cítrico, citrato, dextrose   |
| <b>ADAPT</b>                        | <i>Adaptive dialysis-like affinity platform technology</i>                                       |
| <b>ADCC</b>                         | <i>Antibody-dependent cellular cytotoxicity</i> - citotoxicidade celular dependente de anticorpo |
| <b>C</b>                            | Isoformar circular   |
| <b>cDNA</b>                         | DNA complementar   |
| <b>CEP</b>                          | Comitê de Ética em Pesquisa  |
| <b>circRNA</b>                      | RNA circular   |
| <b>CO<sub>2</sub></b>               | Dióxido de carbono   |
| <b>CodProt</b>                      | Transcritos codificadores de proteína  |
| <b>CONEP</b>                        | Comissão Nacional de Ética em Pesquisa   |
| <b>Ct</b>                           | <i>Cycle threshold</i>   |
| <b>ddPCR</b>                        | <i>Droplet Digital PCR</i>   |
| <b>DTT</b>                          | Ditiotreitol   |
| <b>DEPC</b>                         | Dietilpirocarbonato  |
| <b>DNA</b>                          | <i>Deoxyribonucleic Acid</i> - Ácido Desoxirribonucléico   |
| <b>dNTP</b>                         | <i>Deoxyribonucleotide triphosphate</i> - Desoxirribonucleotídeos Fosfatados                     |
| <b>EMT</b>                          | <i>Epithelial-mesenchymal transition</i> - Transição epitélio-mesênquima                         |
| <b>ENCODE</b>                       | <i>Encyclopedia of DNA Elements</i>  |
| <b>ER</b>                           | Receptor de estrogênio   |
| <b>ERBB2/HER2</b>                   | Receptor do fator de crescimento epidermal 2   |
| <b>ESCRT</b>                        | <i>Endosomal sorting complex required for transport</i>  |
| <b>ESI</b>                          | <i>Electrospray ionization</i> - Ionização por electrospray                                      |
| <b>EVs</b>                          | <i>Extracellular vesicles</i> – Vesículas extracelulares   |

|                       |   |
|-----------------------|---|
| <b>F</b>              | Iniciador <i>forward</i>  |
| <b>FBS</b>            | <i>Fetal bovine serum</i> - Soro fetal bovino   |
| <b>FDA</b>            | <i>Food and Drug Administration</i>   |
| <b>FDR</b>            | <i>False-discovery rate</i>   |
| <b>FISH</b>           | <i>Fluorescent in situ hybridization</i> - Hibridização fluorescente <i>in situ</i>                                 |
| <b>FPKM</b>           | <i>Fragments Per Kilobase of transcript per Million mapped reads</i>  |
| <b>h</b>              | Hora  |
| <b>HDL</b>            | <i>High-density lipoprotein</i> - Lipoproteína de alta densidade  |
| <b>HMEC</b>           | <i>Human Mammary Epithelial Cells</i>   |
| <b>HRP</b>            | <i>Horseradish peroxidase</i> - Peroxidase de rábano  |
| <b>HUVEC</b>          | <i>Human Umbilical Vein Endothelial Cells</i>   |
| <b>ICPL</b>           | <i>Isotope-coded protein label</i>  |
| <b>INCA</b>           | Instituto Nacional do Câncer  |
| <b>IPA</b>            | <i>Ingenuity Pathway Analysis</i>   |
| <b>L</b>              | Isoforma linear   |
| <b>LC-MS/MS</b>       | <i>Liquid chromatography-tandem mass spectrometry</i> -<br>Cromatografia líquida-espectrometria de massas em tandem |
| <b>lincRNA</b>        | <i>Long intergenic noncoding RNA</i>  |
| <b>LNC</b>            | <i>Long noncoding</i> - transcritos não-codificadores de proteínas longos   |
| <b>lncRNA</b>         | <i>Long noncoding RNA</i>   |
| <b>M</b>              | Doença Metastática  |
| <b>M</b>              | Molar   |
| <b>mg</b>             | Miligrama   |
| <b>miRNA</b>          | microRNA  |
| <b>miscRNA</b>        | <i>miscellaneous RNA</i>  |
| <b>ml</b>             | Mililitro   |
| <b>mm</b>             | Milimetro   |
| <b>mM</b>             | Mili-molar  |
| <b>mm<sup>3</sup></b> | Milimetro cúbico  |
| <b>mRNA</b>           | RNA mensageiro  |
| <b>MVE</b>            | <i>Multivesicular endosomes</i> – Endossomos multivesiculares   |
| <b>N</b>              | Linfonodo   |

|                  |  |
|------------------|--|
| <b>n</b>         | Número   |
| <b>N0</b>        | Linfonodo Negativo   |
| <b>ncRNA</b>     | <i>Noncoding RNA</i> - RNA não-codificador de proteína                               |
| <b>ND</b>        | Não definido   |
| <b>ng</b>        | Nanograma  |
| <b>NGS</b>       | <i>Next generation sequencing</i> - Sequenciamento de nova geração                   |
| <b>nm</b>        | Nanometro  |
| <b>nt</b>        | Nucleotídeo  |
| <b>NTA</b>       | <i>Nanotracking Particle Analysis</i>  |
| <b>PAR-CLIP</b>  | <i>Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation</i> |
| <b>pb</b>        | Pares de base  |
| <b>PBS</b>       | <i>Phosphate Buffered Saline</i>   |
| <b>PCR</b>       | <i>Polymerase Chain Reaction</i> - Reação em cadeia da polimerase                    |
| <b>pg</b>        | Picograma  |
| <b>piRNA</b>     | <i>PIWI-interacting RNA</i>  |
| <b>pM</b>        | Pico-molar   |
| <b>pmol</b>      | Pico-mole  |
| <b>PR</b>        | Receptor de progesterona   |
| <b>pre-miRNA</b> | miRNA precursor  |
| <b>pri-miRNA</b> | miRNA primário   |
| <b>Psdg</b>      | Pseudogene   |
| <b>Q20</b>       | Escore de qualidade 20   |
| <b>qRT-PCR</b>   | <i>Quantitative reverse transcription-PCR</i> - PCR quantitativa em tempo real       |
| <b>R</b>         | Iniciador reverso  |
| <b>RefSeq</b>    | <i>Reference Sequence</i>  |
| <b>RISC</b>      | <i>RNA-induced silencing complex</i>   |
| <b>RNA</b>       | <i>Ribonucleic Acid</i> - Ácido Ribonucléico   |
| <b>RNA-Seq</b>   | Sequenciamento de RNAs em larga escala   |
| <b>rRNA</b>      | RNA ribossomal   |
| <b>SAP</b>       | <i>Shrimp Alkaline Phosphatase</i>   |
| <b>scaRNA</b>    | <i>small Cajal body-associated RNA</i>   |



|                 |  |
|-----------------|--|
| <b>SDS-PAGE</b> | <i>Sodium dodecyl sulfate-polyacrylamide gel electrophoresis</i>           |
| <b>siRNA</b>    | <i>small interfering RNA</i>   |
| <b>SNC</b>      | <i>short noncoding</i> - Transcritos não-codificadores de proteínas curtos |
| <b>snoRNA</b>   | <i>small nucleolar RNA</i>   |
| <b>snRNA</b>    | <i>small nuclear RNA</i>   |
| <b>SRP-RNA</b>  | <i>signal recognition particle-RNA</i>                                     |
| <b>TA</b>       | Temperatura ambiente   |
| <b>TCLE</b>     | Termo de consentimento livre e esclarecido                                 |
| <b>tRNA</b>     | RNA transportador  |
| <b>U</b>        | Unidade  |
| <b>vRNA</b>     | <i>vault RNA</i>   |

# ÍNDICE

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b> .....  | <b>1</b>  |
| 1.1      | Vesículas Extracelulares (EVs) .....   | 1         |
| 1.1.1    | Histórico.....   | 1         |
| 1.1.2    | Biogênese e Secreção .....   | 2         |
| 1.1.3    | Composição e Conteúdo .....  | 5         |
| 1.1.4    | Métodos de Estudo.....   | 8         |
| 1.1.5    | Funções de EVs Derivadas de Células Tumorais .....   | 10        |
| 1.1.6    | EVs em outras Doenças Humanas.....   | 16        |
| 1.2      | RNAs não-Codificadores de Proteínas (ncRNAs) .....   | 17        |
| 1.2.1    | miRNAs .....   | 20        |
| 1.2.2    | <i>Long noncoding</i> RNAs (lncRNAs) .....   | 24        |
| 1.2.3    | RNAs Circulares (circRNAs).....  | 25        |
| 1.3      | Câncer de Mama .....   | 27        |
| 1.3.1    | Incidência e Mortalidade.....  | 27        |
| 1.3.2    | Classificação Histopatológica .....  | 28        |
| 1.3.3    | Classificação Molecular .....  | 28        |
| 1.3.4    | HER2.....  | 29        |
| <b>2</b> | <b>OBJETIVOS</b> .....   | <b>32</b> |
| <b>3</b> | <b>MATERIAIS E MÉTODOS</b> .....   | <b>33</b> |
| 3.1      | Seleção de Pacientes e Coleta de Amostras.....   | 33        |
| 3.2      | Cultura Celular.....   | 34        |
| 3.3      | Isolamento das EVs Derivadas do Plasma.....  | 34        |
| 3.4      | Isolamento das EVs Derivadas do Meio de Cultura Celular<br>Condicionado.....                     | 35        |
| 3.5      | Caracterização das EVs Derivadas do Plasma por <i>NTA</i> .....                                  | 36        |
| 3.6      | Caracterização das EVS Derivadas do meio de Cultura Celular<br>condicionado por <i>NTA</i> ..... | 36        |

|          |   |           |
|----------|---|-----------|
| 3.7      | Caracterização das EVs Derivadas do Plasma por <i>Western Blot</i> .....                                    | 36        |
| 3.8      | Caracterização das EVs Derivadas do Meio de Cultura Celular<br>Condicionado por <i>Western Blot</i> .....   | 37        |
| 3.9      | Caracterização das EVs Derivadas do Meio de Cultura Celular<br>Condicionado por Microscopia Eletrônica..... | 37        |
| 3.10     | Análise Proteômica de EVs e Células das Linhagens HB4a e C5.2.....  | 38        |
| 3.10.1   | Extração de Proteínas .....   | 38        |
| 3.10.2   | Marcação ICPL e Pré-Fracionamento dos Proteomas Celulares .....   | 38        |
| 3.10.3   | Digestão em Solução do Proteoma de EVs.....   | 39        |
| 3.10.4   | LC-MS/MS .....  | 39        |
| 3.10.5   | Análise de Dados .....  | 39        |
| 3.11     | Extração de RNA das EVs.....  | 40        |
| 3.12     | Extração de RNA das Células, Tratamento com DNase e Depleção de<br>rRNA.....                                | 41        |
| 3.13     | Sequenciamento de Nova Geração.....   | 41        |
| 3.13.1   | Construção das Bibliotecas .....  | 41        |
| 3.13.2   | PCR em Emulsão .....  | 43        |
| 3.13.3   | Enriquecimento.....   | 45        |
| 3.13.4   | Avaliação da Eficiência da PCR em Emulsão .....   | 46        |
| 3.13.5   | Sequenciamento.....   | 47        |
| 3.13.6   | Análises de Bioinformática e Estatística.....   | 48        |
| 3.14     | Validação dos miRNAs Diferencialmente Representados por PCR<br>Quantitativo em Tempo Real (qRT-PCR) .....   | 49        |
| 3.14.1   | Síntese do cDNA por Transcrição Reversa .....   | 49        |
| 3.14.2   | Pré-Amplificação do cDNA .....  | 50        |
| 3.14.3   | Avaliação da Expressão dos miRNAs por qRT-PCR.....  | 51        |
| 3.15     | Validação dos RNAs Circulares (circRNAs) .....  | 51        |
| 3.15.1   | Síntese do cDNA por Transcrição Reversa .....   | 51        |
| 3.15.2   | Desenho dos Iniciadores e PCR.....  | 52        |
| 3.15.3   | Sequenciamento pelo Método de Sanger .....  | 53        |
| <b>4</b> | <b>RESULTADOS E DISCUSSÃO .....</b>   | <b>55</b> |

|          |   |            |
|----------|---|------------|
| 4.1      | Caracterização de EVs Derivadas de HB4a e C5.2 .....  | 55         |
| 4.1.1    | Caracterização do Proteoma de EVs Derivadas de HB4a e C5.2 .....                                      | 55         |
| 4.1.2    | Caracterização do Conjunto de RNAs de EVs Derivadas de HB4a e<br>C5.2 por NGS .....                   | 65         |
| 4.1.3    | Análise de Enriquecimento de Motivos.....   | 89         |
| 4.1.4    | Análise de Distribuição Posicional.....   | 95         |
| 4.2      | Caracterização de EVs Derivadas do Plasma de Pacientes .....  | 101        |
| 4.2.1    | Pacientes e Amostras.....   | 101        |
| 4.2.2    | Caracterização das EVs Derivadas do Plasma de Pacientes por <i>NTA</i><br>e <i>Western Blot</i> ..... | 103        |
| 4.2.3    | Caracterização do Conjunto de RNAs de EVs Derivadas do Plasma<br>de Pacientes por NGS.....            | 104        |
| 4.2.4    | Validação dos miRNAs Diferencialmente Representados em EVs<br>Derivadas do Plasma de Pacientes.....   | 124        |
| 4.3      | Predição de Novos miRNAs .....  | 134        |
| 4.4      | Identificação de RNAs Circulares em EVs.....  | 136        |
| <b>5</b> | <b>CONCLUSÕES .....</b>   | <b>143</b> |
| <b>6</b> | <b>PERSPECTIVAS FUTURAS.....</b>  | <b>145</b> |
| <b>7</b> | <b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>   | <b>147</b> |

## **ANEXOS**

**Anexo 1** Carta de aprovação do Comitê de Ética e Pesquisa do Hospital A.C. Camargo

**Anexo 2** Carta de aprovação do Comitê de Ética e Pesquisa do Hospital de Câncer de Barretos

**Anexo 3** Termo de consentimento assinado pelas pacientes incluídas neste estudo.

**Anexo 4** Artigo publicado na revista Proteomics.

**Anexo 5** Perfil do Bioanalyzer de bibliotecas para sequenciamento, obtido a partir de RNAs de EVs e células das linhagens C5.2 e HB4a.

**Anexo 6** Dados clínicos de todas as pacientes recrutadas para este estudo

**Anexo 7** Perfil do Bioanalyzer de bibliotecas para sequenciamento, obtido a partir de RNAs de EVs derivadas do plasma de pacientes

# 1 INTRODUÇÃO

## 1.1 VESÍCULAS EXTRACELULARES (EVS)

### 1.1.1 Histórico

Sabe-se desde 1967 que o plasma depletado de plaquetas intactas contém material particulado (originalmente chamado de “*platelet dust*”) que pode ser separado por ultracentrifugação (WOLF 1967). As partículas observadas WOLF (1967) foram descritas como sendo ricas em fosfolípidos e com propriedades coagulantes, porém suas funções fisiológicas ainda eram desconhecidas. Foi na década de 1980 que as vesículas ganharam a atenção da comunidade científica. Estudos de microscopia eletrônica de dois grupos independentes demonstraram o envolvimento de vesículas de ~50nm na reciclagem do receptor de transferina durante a maturação de reticulócitos de ratos e ovelhas (HARDING et al. 1983, 1984; PAN e JOHNSTONE 1983; PAN et al. 1985). O nome de “exossomo” foi dado a essas vesículas de origem endossomal em 1987 (JOHNSTONE et al. 1987), porém pouco se sabia então de suas funções além de exocitose de certas moléculas.

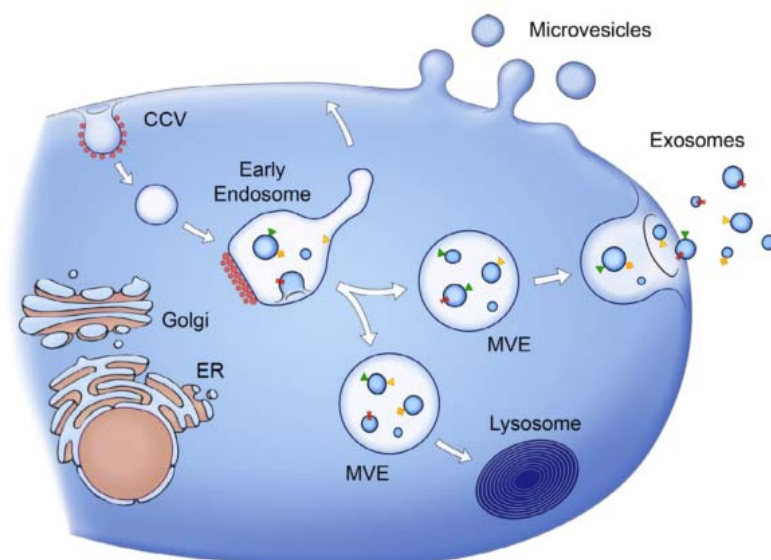
Em 1980, TAYLOR et al. descreveram a presença de fragmentos de membrana de células tumorais em fluídos (cisto e ascite) de pacientes com câncer de ovário. Logo depois, em 1981, outro grupo demonstrou que o líquido ascítico de animais inoculados com hepatocarcinoma (porcos da índia) e carcinoma de mama (camundongos) centrifugados a 100.000xg continham vesículas com atividade pró-coagulante (DVORAK et al. 1981). A significância biológica desta atividade pró-coagulante vesicular, sugeriram os autores, poderia estar relacionada com os extensos depósitos de fibrina encontrados na cavidade peritoneal dos animais, que funcionaria como uma barreira física isolando as células tumorais do sistema imune do hospedeiro (DVORAK et al. 1981; LIPINSKI e EGYUD 2000).

Em 1996, com a descoberta de que linfócitos do tipo B secretavam vesículas contendo moléculas de MHC classe II e que eram capazes de ativar linfócitos T (RAPOSO et al. 1996), foi atribuído aos exossomos um papel de apresentação de antígenos. A capacidade das vesículas desencadear respostas imunes também foi comprovada em estudo de 1998 com exossomos isolados do meio condicionado de células dendríticas murinas estimuladas *in vitro* com peptídeos tumorais. Quando administrados em camundongos com tumores já estabelecidos (carcinoma mamário e mastocitoma), esses exossomos levaram à supressão do crescimento tumoral de maneira dependente de linfócitos T (ZITVOGEL et al. 1998). Em 2002, ANDRE et al. demonstraram que líquido ascítico de pacientes com diversos tipos de câncer acumula grandes quantidades de exossomos contendo moléculas relacionadas ao sistema imune (MHC I) e antígenos tumorais, como MART1 em pacientes com melanoma e Her2/Neu em pacientes com câncer de mama. A estimulação de células dendríticas autólogas com os exossomos isolados do líquido ascítico do próprio paciente foi capaz de induzir a produção de linfócitos T direcionados contra os antígenos tumorais contidos nos exossomos. Neste sentido, o uso de vacinas derivadas de vesículas, e portanto acelulares, é uma área promissora na imunoterapia do câncer, com diversos estudos clínicos em andamento (CHAPUT e THÉRY 2011).

### **1.1.2 Biogênese e secreção**

A classificação dos dois principais tipos de vesículas, os exossomos e as microvesículas, veio em 1999, com o estudo de plaquetas ativadas (HEIJNEN et al. 1999). Enquanto os exossomos já eram sabidamente derivados da membrana interna do endossomo, e de tamanho de 40-100nm, as microvesículas foram descritas como provenientes da membrana plasmática, e de maior tamanho (100-1000nm). As duas populações de vesículas são secretadas pelas células, porém possuem biogênese e composições proteicas distintas, sugerindo funções diferenciadas (HEIJNEN et al. 1999).

Conforme ilustrado na **Figura 1**, as microvesículas são geradas diretamente da membrana plasmática, enquanto que os exossomos são derivados de invaginações da membrana que formam os endossomos e MVEs (do Inglês *multivesicular endosomes*), que eventualmente se fundem novamente com a membrana celular liberando o seu conteúdo (exossomos) no meio extracelular (RAPOSO e STOORVOGEL 2013). Alternativamente, os endossomos também podem ser destinados ao lisossomo para degradação. Atualmente o conjunto de vesículas secretadas pelas células tem sido denominado Vesículas Extracelulares (ou EVs, do Inglês, *Extracellular Vesicles*).



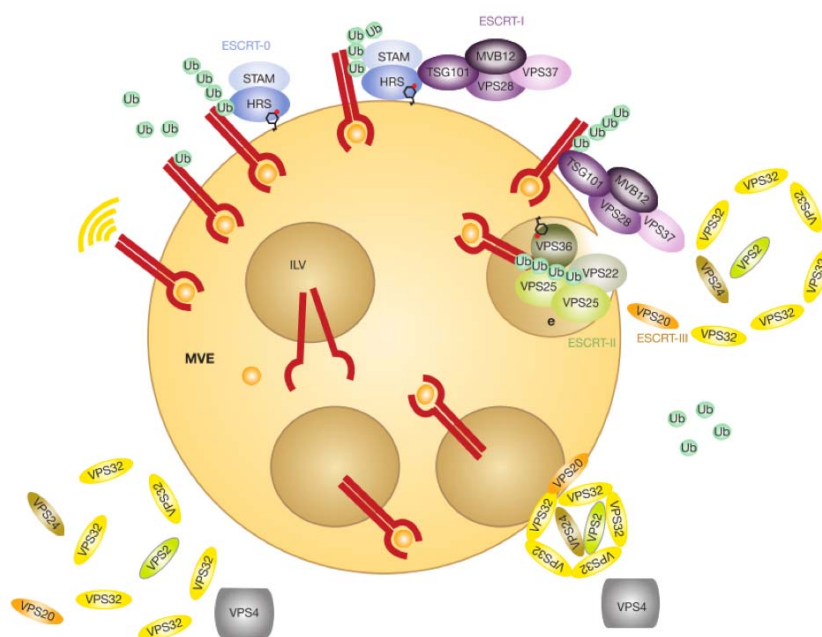
**Fonte:** RAPOSO e STOORVOGEL (2013).

**Figura 1** - Biogênese das duas principais populações de EVs – microvesículas e exossomos.

Alguns dos mecanismos envolvidos na formação dos exossomos já foram bem caracterizados. Vesículas intraluminais se formam por invaginações da membrana dos endossomos gerando os chamados endossomos multivesiculares. Este processo é coordenado por proteínas da maquinaria ESCRT (do Inglês *endosomal sorting complex required for transport*), que consiste de aproximadamente trinta proteínas organizadas em quatro complexos (ESCRT-0, -I, -II, e -III) em conjunto com proteínas



associadas (ALIX e VPS4, dentre outras) (COLOMBO et al. 2014). A **Figura 2** ilustra os principais passos do processo de formação das vesículas intraluminais que eventualmente são secretadas como exossomos. Primeiramente a proteína HRS do complexo ESCRT-0 reconhece a ubiquitina presente em determinadas proteínas transmembranas, e recruta a proteína TSG101 do complexo ESCRT-I, que por sua vez recruta o complexo ESCRT-III por meio de ESCRT-II ou ALIX. Os complexos ESCRT-I e -II são responsáveis pela deformação da membrana, e o complexo ESCRT-III subsequentemente realiza a excisão da membrana, por meio de oligômeros em forma de espiral que constroem para “quebrar o pescoço” da vesícula em formação. Finalmente, a ATPase VPS4 é recrutada para dissociar os oligômeros, que são então reciclados (RUSTEN et al. 2012).



**Fonte:** RUSTEN et al. (2012).

**Figura 2** - Mecanismo de formação das vesículas intraluminais nos endossomos multivesiculares. ILV – do Inglês *intraluminal vesicle*; MVE – do Inglês *multivesicular endosome*.

Uma via independente de ESCRT foi proposta, mediada por microdomínios na membrana (*lipid-rafts*) que contêm altas concentrações do esfingolípídeo ceramida (TRAJKOVIC et al. 2008). Foi demonstrado que as esfingomielinases (SMases), enzimas responsáveis por processar os esfingolípídeos, são necessárias para a formação dos exossomos (BIANCO et al. 2009; KOSAKA et al. 2010). Várias proteínas da família Rab (que em humanos inclui aproximadamente 70 proteínas) participam do processo de tráfico intracelular, por exemplo, as proteínas Rab27a e Rab27b são importantes para a fusão dos MVEs com a membrana plasmática (OSTROWSKI et al. 2010). Sabe-se que a secreção de exossomos é dependente de  $Ca^{2+}$  intracelular (SAVINA et al. 2003), e que outros fatores fisiológicos como pH e hipóxia também afetam a produção de EVs (KING et al. 2012; PAROLINI et al. 2009). Proteínas como TP53 e TSAP6 também regulam a secreção de exossomos (YU et al. 2006). Como TP53 é o gene mais frequentemente mutado em câncer (LEVINE e OREN 2009), podemos especular que existam alterações na secreção de exossomos em câncer.

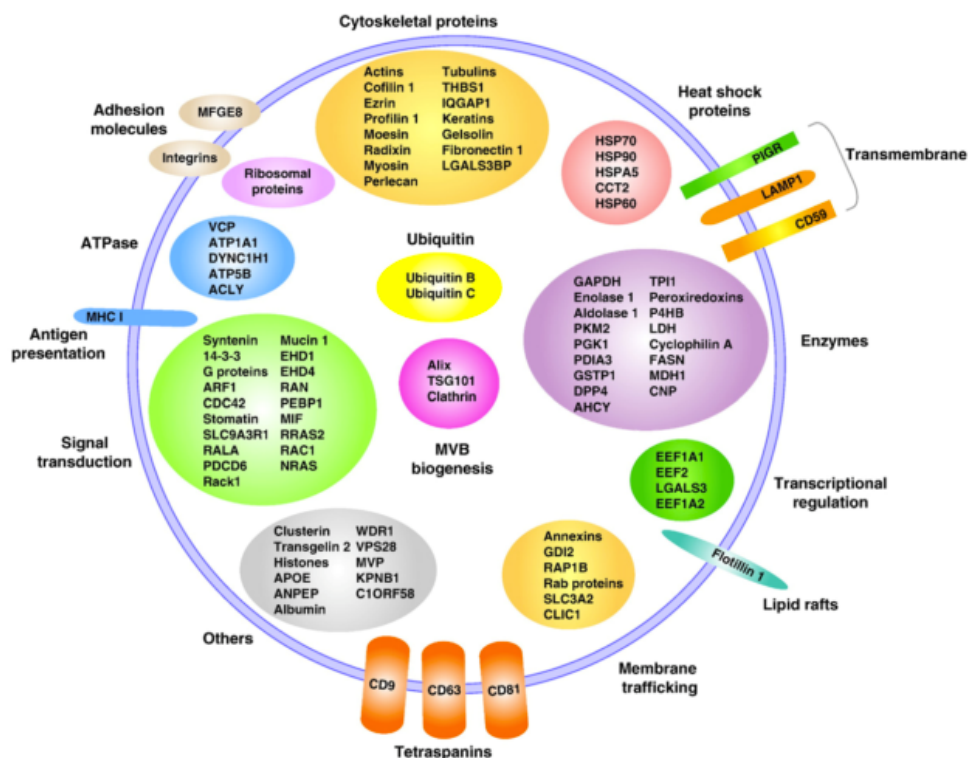
### 1.1.3 Composição e conteúdo

As EVs são compostas por uma bicamada lipídica, e contêm em seu interior proteínas e ácidos nucléicos. Já que as EVs aparentemente são importantes carreadores de moléculas bioativas, recentemente diversos grupos têm buscado determinar a sua composição e o seu conteúdo. Neste sentido, foram criados repositórios públicos que buscam organizar os achados dos grupos que avaliaram o conteúdo de EVs obtidas em diferentes contextos e a partir de diferentes fontes. O primeiro destes bancos de dados foi o ExoCarta (<http://exocarta.org>) (MATHIVANAN e SIMPSON 2009), que cataloga o conteúdo de exossomos de mamíferos. Subsequentemente o ExoCarta foi incorporado a um banco denominado Vesiclepedia (<http://microvesicles.org>) (KALRA et al. 2012), que reúne dados obtidos de todos os tipos de EVs de eucariotos e também de procariotos. Recentemente foi criado um outro repositório denominado EVpedia (<http://evpedia.info>) (KIM et al. 2013b), que em sua atualização de Setembro

de 2014 apresentava o conteúdo de 263 estudos em larga-escala, com mais de 172 mil moléculas catalogadas.

Em termos de composição do conteúdo lipídico de suas membranas, uma comparação entre membrana celular e membrana vesicular mostra um enriquecimento nas EVs de ceramida, colesterol, fosfatidilserina, e esfingomiéline (COLOMBO et al. 2014). Diferenças na composição lipídica entre duas populações (de tamanho médio 50 e 100nm) de EVs secretadas em sêmen (prostassomos) também foram encontradas, sugerindo que as duas populações podem ter biogêneses e funções distintas (BROUWERS et al. 2013). Os lipídeos expostos em EVs derivadas de células tumorais podem ter um papel funcional na indução da angiogênese, pois foi demonstrado que a esfingomiéline presente em EVs derivadas de linhagens celulares tumorais (HT1080 fibrosarcoma e DU-145 carcinoma de próstata) é capaz de induzir a migração, invasão e formação de tubos em células endoteliais humanas (HMEC-1) (KIM et al. 2002).

Quando o conteúdo carregado pelas EVs é avaliado quantitativa e qualitativamente, observamos uma ampla diversidade protéica que reflete as células de origem, estados fisiológicos, patológicos ou de desenvolvimento. A **Figura 3** representa as principais proteínas identificadas em EVs, agrupadas de acordo com a sua atividade (MATHIVANAN et al. 2010). Várias destas proteínas são utilizadas como marcadores de EVs para separação por imunoafinidade ou para caracterização do material isolado por *Western Blot*, por exemplo. Destacam-se as proteínas envolvidas com adesão celular, como integrinas; as proteínas de citoesqueleto como actina e cofilina; as proteínas de *lipid-rafts* como flotilina; as proteínas transmembrana da família de tetraespaninas, como CD9, CD63 e CD81; as proteínas envolvidas com a biogênese das EVs, como Alix, TSG101 e clatrina; e as proteínas de apresentação de antígeno, como MHC I e II.



Fonte: MATHIVANAN et al. (2010).

**Figura 3** - Representação esquemática do conteúdo proteico de EVs. As proteínas estão agrupadas de acordo com a função. Várias das proteínas aqui representadas, tais como HSP70, Flotillin1, CD9, CD63 e CD81, são frequentemente utilizadas como marcadores de EVs em imunodeteção por *Western Blot*.

Os primeiros ácidos nucleicos identificados em EVs foram os miRNAs e os mRNAs, em 2007 (VALADI et al. 2007). Neste estudo os autores demonstraram que os mRNAs presentes nos exossomos eram funcionais, pois quando um destes mRNAs foi transferido de células de camundongo para células humanas, este foi passível de ser traduzido *in vitro*, produzindo uma proteína murina na célula receptora. Diferenças no conteúdo transcricional entre os exossomos e as células das quais eles se originaram foram encontradas, sendo que 270/1.300 transcritos foram detectados exclusivamente nos exossomos e não nas células que os originaram. Em um outro estudo, com microvesículas derivadas de culturas primárias de glioblastoma, 4.700/27.000 transcritos foram encontrados exclusivamente nas EVs e não nas células secretoras (SKOG et al. 2008). Além de miRNAs e mRNAs, as EVs também contêm grandes quantidades de proteínas e de

outros RNA não-codificantes de proteínas. Em um trabalho com EVs derivadas de células dendríticas e linfócitos T, foi demonstrado um enriquecimento de RNAs estruturais como SRP-RNAs, vRNAs e Y-RNAs nas EVs, em comparação com o transcriptoma celular (NOLTE-'T HOEN et al. 2012). Tais achados sugerem que algumas moléculas são preferencialmente destinadas para os exossomos, e que a composição das EVs não é uma simples amostragem aleatória do conteúdo das células.

A presença de DNA genômico em EVs também já foi reportada. O primeiro estudo identificou DNA simples fita em microvesículas derivadas de linhagens tumorais de meduloblastoma e demonstrou a amplificação do oncogene c-Myc (BALAJ et al. 2011). Já um estudo mais recente avaliou um amplo painel de linhagens tumorais e demonstrou que o DNA contido em EVs é predominantemente dupla fita, representa sem viés todo o genoma e reflete o perfil mutacional da célula secretora, sendo possível a detecção de mutações em *BRAF* e *EGFR* nas EVs tumorais de melanoma e pulmão, respectivamente (THAKUR et al. 2014). Um outro estudo também detectou a presença de DNA dupla fita em EVs representando todos os cromossomos. Neste estudo mutações em *KRAS* e *TP53* foram identificadas no DNA de EVs derivadas do soro de pacientes com adenocarcinoma de pâncreas (KAHLERT et al. 2014). Se confirmado, o achado de DNA genômico em EVs representa um importante avanço pois a detecção de mutações câncer-específicas em EVs tumorais circulantes pode ser usada para diagnóstico e monitoramento de resposta a tratamento dentro do conceito de biópsias líquidas.

#### **1.1.4 Métodos de estudo**

O isolamento das EVs é geralmente feito a partir de meio condicionado de cultura de células ou de biofluidos como o plasma, saliva, líquido cerebrospinal, e outros. Independente da fonte, o protocolo clássico de isolamento das EVs utiliza uma série de centrifugações para primeiro eliminar células, células mortas e debris celulares, e depois baixar as EVs que são lavadas com PBS e então resuspendidas em um volume

reduzido (THÉRY et al. 2006). Para uma maior separação de contaminantes, as EVs podem ser isoladas por um gradiente de densidade (tipicamente preparado utilizando sacarose ou iodixanol). Também existem métodos alternativos de isolamento, como a resina ExoQuick, desenvolvida comercialmente pela empresa SBI (System Biosciences, California, EUA) e que precipita as EVs tendo como base a afinidade entre estas e um polímero presente na resina. Métodos baseados em filtração, ou ainda em imunoseparação utilizando um conjugado de *beads* magnéticas e anticorpos que reconhecem proteínas específicas presentes na superfície das EVs, como EpCAM para a separação de EVs de origem epitelial, também são disponíveis comercialmente. Cada método de isolamento tem vantagens e desvantagens, mas nenhum é capaz de separar completamente as diferentes populações de EVs (exossomos e microvesículas), permitindo um enriquecimento das mesmas, mas não uma purificação absoluta (RAPOSO e STOORVOGEL 2013).

Após o isolamento, as EVs são caracterizadas quanto a sua morfologia, tamanho e conteúdo. Devido ao seu tamanho, as EVs em geral são caracterizadas morfologicamente por microscopia eletrônica ou microscopia de força atômica. Para uma análise quantitativa da concentração e tamanho das EVs, foram desenvolvidas metodologias específicas baseadas em espalhamento da luz. O NanoSight (Malvern Instruments, Worcestershire, Reino Unido) é uma delas e consiste em um microscópio com feixe de laser que incide na amostra contendo as EVs em suspensão, em um compartimento apropriado. O movimento Browniano das vesículas em suspensão difrata a luz do laser proporcionalmente ao tamanho das partículas. O equipamento grava um vídeo das EVs em movimento e calcula sua concentração e tamanho baseado no padrão de dispersão do laser em um dado intervalo de tempo. Lasers com comprimentos de onda diferentes podem ser utilizados para a detecção de populações específicas de EVs que foram previamente marcadas com anticorpos conjugados com fluoróforos. Para a quantificação da concentração de EVs baseada em marcadores de interesse pode-se

também utilizar a citometria de fluxo, porém os citômetros convencionais permitem a detecção direta somente de partículas com tamanho superior a 300nm, fazendo-se necessário o conjugamento prévio das EVs a esferas de látex (BARTENEVA et al. 2013). A análise do conteúdo protéico das EVs geralmente é feito em larga escala por espectrometria de massas e quando proteínas específicas de interesse são avaliadas utiliza-se a técnica de *Western Blot*. A análise do conteúdo de ácidos nucleicos (miRNAs, mRNAs, e DNA) pode ser feita por microarranjos ou por sequenciamento de nova geração.

### 1.1.5 Funções de EVs derivadas de células tumorais

- *Microambiente tumoral*

Trabalhos recentes têm demonstrado que EVs secretadas por células tumorais podem afetar o microambiente tumoral ao induzirem a transformação de células normais do estroma. Esta transformação parece contribuir para a progressão do câncer (expansão da massa tumoral) de uma maneira que não depende somente da proliferação das células tumorais em si (ANTONYAK et al. 2011).

EVs derivadas de células tumorais de mama (MDA-MB-231) e glioblastoma (U87) foram capazes de induzir *in vitro* a transformação de fibroblastos normais (NIH 3T3) em cultura (ANTONYAK et al. 2011). O tratamento de fibroblastos com EVs derivadas das células tumorais ativou as vias de sinalização AKT/ERK, fazendo com que os fibroblastos adquirissem a capacidade de sobreviver em baixas concentrações de soro e de crescer independente de ancoragem. A transferência da enzima transglutaminase tecidual e de fibronectina entre EVs e fibroblastos foi essencial para a aquisição do fenótipo transformado.

A comunicação entre células tumorais e células do microambiente tumoral é uma via de mão-dupla. Foi demonstrado que exossomos CD81-positivos secretados por fibroblastos primários, derivados de pacientes com câncer de mama, estimulam a motilidade e a capacidade metastática de células MDA-MB-231 através da ativação autócrina de vias de sinalização

Wnt-PCP (LUGA et al. 2012). Neste sentido, a expressão de CD81 foi encontrada significativamente aumentada no estroma associado a tumores de mama do tipo ductal invasivo.

- *Angiogênese*

A angiogênese, ou formação de novos vasos sanguíneos a partir de vasos já existentes, é um processo crucial no crescimento tumoral, pois possibilita aporte de oxigênio e nutricional permitindo que os tumores cresçam além do limite de 1-2 mm<sup>3</sup> (MCDUGALL et al. 2006). Seu papel no crescimento tumoral, assim como a sugestão de sua importância como possível alvo terapêutico, foram inicialmente sugeridos por Judah Folkman no início da década de 1970 (FOLKMAN et al. 1971). EVs secretadas por progenitores de células endoteliais estimulam a angiogênese em células endoteliais (dos tipos HMECs e HUVECs), que formam estruturas capilares *in vitro* em matrigel ao internalizarem as EVs (DEREGIBUS et al. 2007). As células endoteliais tratadas com EVs também formam vasos funcionais *in vivo* quando implantadas subcutaneamente juntamente com matrigel em camundongos, sendo visivelmente conectados à vasculatura murina pela presença de eritrócitos em seu lúmen. A transferência de RNAs é necessária para a estimulação das células endoteliais, pois o tratamento com RNase inibiu o efeito pro-angiogênico das EVs.

EVs secretadas por células tumorais também estimulam a angiogênese pela transferência do receptor oncogênico EGFR, ativando nas células endoteliais vias de sinalização MAKT e AKT que levam à expressão autócrina do fator pró-angiogênico VEGF (AL-NEDAWI et al. 2009). Este efeito angiogênico é por sua vez inibido pelo bloqueio da interação EV-célula pela anexina-V, demonstrando que a transferência de EVs é responsável pela ativação das células endoteliais.

- *Evasão do sistema imune*

Apesar de expressarem moléculas como MHC I e II, e estarem envolvidas em apresentação de antígeno e ativação de resposta imune, as



EVs também são capazes de imunossupressão. A evasão do sistema imune é o processo pelo qual as células tumorais escapam da destruição por células imunológicas do hospedeiro. EVs secretadas por células tumorais (da linhagem de carcinoma de próstata LNCaP) contribuem para a evasão do sistema imune pela expressão do ligante FasL na membrana, que por sua vez se liga ao receptor Fas de linfócitos T citotóxicos CD8, induzindo sua apoptose (ABUSAMRA et al. 2005). Este efeito é inibido pelo tratamento das EVs com anticorpos que bloqueiam FasL, sugerindo que uma interação direta entre os ligantes das EVs e o receptores nos linfócitos T seja necessária para a indução da apoptose. A presença de altas concentrações de EVs FasL+ circulantes foi associada com um maior tamanho tumoral em pacientes com carcinoma oral (KIM et al. 2005).

Também já foi demonstrado que EVs obtidas a partir das linhagens OVCAR3, SKOV3 and AD10 de carcinoma de ovário estimulam a expansão e a atividade de linfócitos T regulatórios, que promovem a tolerância imunológica pela secreção de citocinas imunossupressoras, como TGF $\beta$  e IL-10 (SZAJNIK et al. 2010). Um outro estudo demonstrou que EVs derivadas do plasma de pacientes com melanoma avançado alteram a diferenciação normal de monócitos (que em indivíduos saudáveis se transformam em células dendríticas) para células mielóides sem capacidade fagocítica e com atividade de imunossupressão em células-T mediada por TGF $\beta$  (VALENTI et al. 2006).

- *Invasão e metástase*

O potencial invasivo *in vitro* de linhagens tumorais de mama aparentemente se correlaciona com a quantidade de EVs liberadas, tendo sido observado que linhagens mais agressivas secretam uma quantidade maior de vesículas com maior atividade proteolítica (GINESTRA et al. 1998). Pacientes com câncer de mama têm maior quantidade de EVs circulantes (que expressam anexina-V e fator tecidual) em comparação com mulheres saudáveis ou com tumores benignos (LIEBHARDT et al. 2010; TESSELAAR et al. 2007). A concentração de EVs também se correlaciona com o

estadiamento, aumentando com o tamanho tumoral (T1 vs T2) e sendo aparentemente maior em pacientes com doença metastática (TOTH et al. 2008).

Exossomos secretados *in vitro* por linhagens tumorais de mama contêm a forma ativa de metaloproteinase (MMP-9), capaz de degradar a matriz extracelular e promover a migração e invasão celular (DOLO et al. 1998). Neste estudo também foi detectada a presença de  $\beta_1$  integrina na membrana dos exossomos, que por interagir com diversas macromoléculas como fibronectina, colágeno e laminina, é capaz de atuar como uma “âncora”, mediando a degradação focal da matriz extracelular pelas enzimas proteolíticas presentes nos exossomos. As EVs também podem carrear proteínas como Hsp90 $\alpha$  e Hsp70, que aumentam a motilidade celular ao ativarem outras proteínas efetoras, como a metaloproteinase MMP-2 e a protease plasmina (MCCREADY et al. 2010; SIMS et al. 2011).

Segundo a hipótese de “seed and soil” proposta pelo cirurgião Stephen Paget em 1889, a formação de metástases depende das células tumorais (as sementes) encontrarem condições favoráveis no sítio metastático (o solo) (FIDLER e POSTE 2008). Neste sentido, foi demonstrado que EVs podem contribuir para a preparação do microambiente pré-metastático da medula óssea para receber as células tumorais de melanoma (PEINADO et al. 2012). Camundongos tratados sistemicamente com exossomos derivados da linhagem agressiva de melanoma B16F10, antes da injeção subcutânea de células tumorais de melanoma, desenvolvem mais metástases (240X), tanto no pulmão quanto na medula óssea, em comparação com camundongos tratados com exossomos derivados da linhagem menos agressiva B16F1. O efeito pró-metastático dos exossomos foi atribuído à transferência da oncoproteína MET para as células progenitoras da medula óssea, pois após o silenciamento desta proteína os animais tratados com os exossomos desenvolverem menos metástases. Enfatizando o potencial clínico destes achados, um aumento da proteína MET em exossomos foi observado em pacientes com melanoma em estadios mais avançados (3 e 4), em

comparação com indivíduos saudáveis. Também foi detectada, por citometria de fluxo uma maior quantidade de células progenitoras da médula óssea positivas para MET no sangue dos pacientes com estadios mais avançados.

- *Aquisição de fenótipo agressivo*

As EVs parecem contribuir para a aquisição de fenótipo agressivo pela transferência horizontal de RNAs e proteínas derivadas das células tumorais (VALADI et al. 2007). No caso de gliomas, foi demonstrado *in vitro* a transferência do receptor mutante EGFRvIII para células U373 que foram tratadas com EVs derivadas da linhagem U373vIII, positiva para EGFR-vIII (AL-NEDAWI et al. 2008). A incorporação do EGFRvIII nas células U373 tratadas com as EVs levou à ativação das vias de sinalização MAPK e Akt, efeito que foi inibido pelo tratamento com anexina-V. As células que incorporaram EGFRvIII adquiriram um fenótipo transformado, com morfologia fusiforme e capacidade de crescimento independente de ancoragem. A transferência de EGFRvIII via EVs pode contribuir para a transformação de células *in vivo* e a progressão tumoral, uma vez que o EGFRvIII foi detectado em EVs derivadas do plasma de pacientes com glioblastoma (SKOG et al. 2008).

- *Resistência à terapia*

EVs secretadas por células tumorais podem contribuir com a resistência a quimioterápicos, radioterapia e terapias-alvo. Estudos demonstraram que as EVs atuam fazendo um efluxo de quimioterápicos. Após o tratamento *in vitro* com cisplatina, células de carcinoma de ovário resistentes à cisplatina exportaram, via exossomos, 4,9 vezes mais cisplatina do que as células sensíveis (SAFAEI et al. 2005). O acúmulo de doxorubicina em EVs também foi observado em células resistentes a este quimioterápico (SHEDDEN et al. 2003). Utilizando um painel de expressão gênica de 60 linhagens tumorais da ATCC, os autores deste estudo mostraram uma relação inversa entre a expressão de genes envolvidos na secreção de EVs e o perfil de quimiosensibilidade das células.

Uma interação mediada por exossomos entre células do estroma e células tumorais de mama do subtipo *basal-like* já foi descrita como importante para a indução de resistência ao tratamento com quimio- e radio-terapia (BOELENIS et al. 2014). Exossomos secretados por fibroblastos do estroma são enriquecidos em RNAs não-codificantes que contêm 5' trifosfatos. Tipicamente encontrados em RNAs virais, os 5' trifosfatos ativam o receptor RIG-I e desencadeiam a resposta imune inata, levando à produção de interferon e à ativação de diversos genes, dentre eles fatores de transcrição da família STAT. Também presente nas células tumorais de mama, o RIG-1 ativado pelo 5' trifosfato dos exossomos estromais ativa STAT1, que regula a expressão de genes da via NOTCH de sinalização. O resultado é a expansão de uma população de células com características de células tronco tumorais, que são mais resistentes à terapia (CD44<sup>+</sup>CD24<sup>low+</sup>). Em pacientes com tumores do tipo *basal-like* foi observada uma correlação entre a expressão aumentada de uma assinatura composta por NOTCH3 e STAT1, dentre outros, e uma menor sobrevida após tratamento com quimio- e radio-terapia (BOELENIS et al. 2014).

EVs derivadas de células que superexpressam o HER2 contêm este receptor na membrana (KOGA et al. 2005), mostrando que a composição protéica de EVs tumorais reflete a célula de origem. Assim, terapias cujos alvos sejam receptores específicos das células tumorais podem ser menos eficazes na presença de EVs que também expressam este mesmo alvo. Portadores de tumores que superexpressam HER2 podem ser tratados com trastuzumabe, anticorpo monoclonal humanizado que se liga especificamente ao domínio extracelular do receptor HER2. No entanto, já foi demonstrado *in vitro* que o trastuzumabe se liga ao receptor HER2 exposto na superfície das EVs, inibindo o efeito anti-proliferativo desta terapia (CIRAVOLO et al. 2012). Um dos princípios de ação do trastuzumabe é a indução de apoptose por ADCC (do inglês *antibody-dependent cellular cytotoxicity*) mediada por células efectoras com receptores Fc. Além de reduzirem o efeito ADCC, as EVs HER2<sup>+</sup> diminuem a

concentração ativa de trastuzumabe disponível para se ligar nas células tumorais (BATTKE et al. 2011).

EVs HER2+ secretadas por células tumorais e circulantes em pacientes com câncer de mama podem contribuir para a resistência a terapias-alvo como o trastuzumabe. Com o intuito de eliminar EVs tumorais e diminuir o seu efeito quelante, a empresa *Aethlon Medical* desenvolveu uma tecnologia chamada ADAPT (do inglês *adaptive dialysis-like affinity platform technology*) para hemofiltração (MARLEAU et al. 2012). O sistema consiste basicamente de cartuchos, que podem ser incorporados em máquinas de diálise já existentes na maioria dos hospitais, contendo agentes de afinidade (tais como anticorpos contra proteínas presentes em EVs tumorais) imobilizados. Componentes do plasma menores que 200nm de tamanho passam por filtros porosos e então interagem com os agentes de afinidade. Moléculas com afinidade pelos agentes são seletivamente retidas no filtro, enquanto os demais componentes do sangue passam livremente e podem ser re-introduzidos no paciente por acesso vascular cirurgicamente implementado. Espera-se que o uso desta tecnologia possa diminuir especificamente os níveis de EVs tumorais circulantes, contribuindo com a diminuição não somente da resistência ao tratamento, mas também de todos os outros efeitos pró-tumorigênicos mediados por EVs discutidos acima.

#### **1.1.6 EVs em outras doenças humanas**

Todos os tipos celulares secretam EVs, que são encontradas em diversos biofluidos como plasma (CABY et al. 2005), urina (PISITKUN et al. 2004), saliva (OGAWA et al. 2008) e leite materno (ADMYRE et al. 2007). Apesar de serem mais amplamente estudadas no câncer, as EVs também participam de vários outros processos patológicos, dentre eles: doenças neurodegenerativas como a doença de Alzheimer (RAJENDRAN et al. 2006) e a doença de Parkinson (EMMANOUILIDOU et al. 2010); doenças infecciosas como HIV (NGUYEN et al. 2003) e doenças de Prion (FEVRIER et al. 2004); doenças cardiovasculares (BERNAL-MIZRACHI et al. 2003); e

doenças metabólicas como diabetes (SABATIER et al. 2002). Em todas estas situações, o conteúdo carregado pelas EVs parece ter um papel importante no estabelecimento do quadro patológico. Além de proteínas, as EVs carregam diversas classes de RNAs, incluindo as que são apresentadas e discutidas a seguir.

## 1.2 RNAs NÃO-CODIFICADORES DE PROTEÍNAS (ncRNAs)

O genoma humano contém menos de 2% de sequências codificadoras de proteínas, correspondendo a cerca de 20,000-25,000 genes (International Human Genome Sequencing Consortium 2004; LANDER et al. 2001; VENTER et al. 2001). Os 98% de sequências restantes, não codificadoras de proteínas, eram consideradas lixo genômico (“*junk DNA*”) até que o projeto ENCODE (do Inglês *Encyclopedia of DNA Elements*) determinou que mais de 90% do genoma humano é ativamente transcrito (BIRNEY et al. 2007) e atribuiu alguma função bioquímica a cerca de 80% do genoma humano (The Encode Project Consortium 2012). Hoje sabe-se que os RNAs não-codificadores de proteína (ncRNAs) desempenham funções estruturais e regulatórias fundamentais para diversos processos celulares, e que a sua expressão desregulada pode levar a processos patológicos (ESTELLER 2011).

Diversas classes de ncRNAs pequenos (<200nt) e longos (>200nt) já são conhecidas, e com o avanço das tecnologias de sequenciamento de nova geração, novas moléculas tem sido continuamente descritas. Recentemente o projeto ENCODE, por exemplo, anotou 8,801 ncRNAs pequenos e 9,640 ncRNAs longos (THE ENCODE PROJECT CONSORTIUM 2012). Dentre outras, as principais classes de ncRNAs são os miRNAs (*microRNAs*), snoRNAs (*small nucleolar RNAs*), snRNAs (*small nuclear RNAs*), piRNAs (*PIWI-interacting RNAs*), e lncRNAs (*long noncoding RNAs*). A classe dos miRNAs, a mais estudada entre os ncRNAs, é detalhadamente descrita abaixo. Os snRNAs desempenham um papel central no *splicing* de RNA, processo pelo qual os introns são removidos de

pre-mRNAs, sendo o U6 um dos membros mais conhecidos desta classe, e frequentemente utilizado como controle endógeno na quantificação de miRNAs (MORRIS e MATTICK 2014). Os snoRNAs são responsáveis por modificações químicas - metilação e pseudouridilação - de nucleotídeos dos rRNAs, tRNAs e snRNAs, que são essenciais para a maturação e função de moléculas destas classes de ncRNAs (MORRIS e MATTICK 2014). Os piRNAs atuam epigeneticamente no silenciamento pós-transcricional de transposons presentes nas células germinativas (MORRIS e MATTICK 2014). Os lncRNAs são descritos em mais detalhe abaixo.

A **Tabela 1** resume o conteúdo de RNAs de uma típica célula de mamífero, indicando o percentual por massa, o número de moléculas e o tamanho médio das principais classes de RNAs (PALAZZO e LEE 2015). É sabido que os RNAs ribossomais e transportadores são os mais abundantes, mas dentre as demais categorias, é importante ressaltar que alguns RNAs não-codificadores de proteínas, dentre eles os snRNAs, snoRNAs e miRNAs, são bastante abundantes, e são expressos na mesma ordem de grandeza em termos de número de moléculas ( $10^5$ ) que os RNAs mensageiros.

**Tabela 1** – Principais tipos de RNAs descritos em células de mamíferos.

| <b>Type</b>         | <b>Percent of total RNA by mass</b> | <b>Molecules per cell</b>         | <b>Average size (kb)</b> |
|---------------------|-------------------------------------|-----------------------------------|--------------------------|
| rRNAs               | 80 to 90                            | $3-10 \times 10^6$<br>(ribosomes) | 6.9                      |
| tRNA                | 10 to 15                            | $3-10 \times 10^7$                | <0.1                     |
| mRNA                | 3 to 7                              | $3-10 \times 10^5$                | 1.7                      |
| hnRNA<br>(pre-mRNA) | 0.06 to 0.2                         | $1-10 \times 10^3$                | 10*                      |
| Circular RNA        | 0.002 to 0.03                       | $3-20 \times 10^3$                | ~0.5                     |
| snRNA               | 0.02 to 0.3                         | $1-5 \times 10^5$                 | 0.1-0.2                  |
| snoRNA              | 0.04 to 0.2                         | $2-3 \times 10^5$                 | 0.2                      |
| miRNA               | 0.003 to 0.02                       | $1-3 \times 10^5$                 | 0.02                     |
| 7SL                 | 0.01 to 0.2                         | $3-20 \times 10^4$                | 0.3                      |
| Xist                | 0.0003 to 0.02                      | $0.1-2 \times 10^3$               | 2.8                      |
| Other lncRNA        | 0.03 to 0.2                         | $3-50 \times 10^3$                | 1                        |

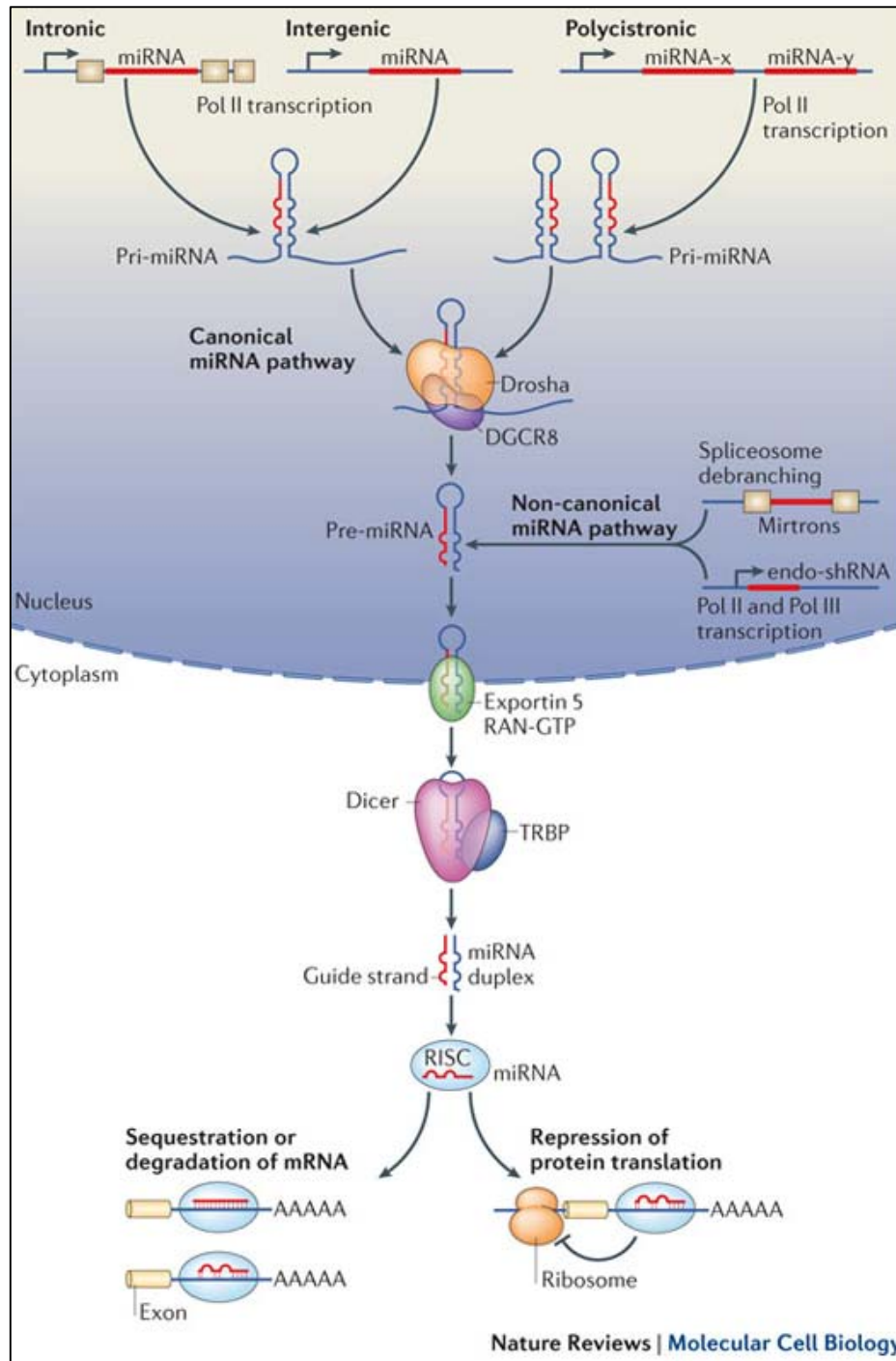
**Fonte:** PALAZZO e LEE (2015)



### 1.2.1 miRNAs

Os miRNAs são pequenos RNAs (~22 nt) que regulam pós-transcricionalmente a expressão gênica. Em 1993 o miRNA lin-4 de *Caenorhabditis elegans* foi o primeiro desta classe a ser descrito e foi demonstrado que ele era complementar a uma região no 3'-UTR do mRNA lin-14 e que regulava negativamente o nível da proteína homônima (LEE et al. 1993). Em 2000, o segundo miRNA descoberto, let-7, também foi descrito em *C. elegans* (REINHART et al. 2000) e observou-se que seu tamanho, sequência e expressão temporal eram altamente conservados entre diversas espécies (PASQUINELLI et al. 2000). Um ano depois, com a descoberta de novos miRNAs em *Drosophila melanogaster* e em humanos, os miRNAs foram oficialmente denominados como uma nova classe de pequenos RNAs não-codificantes de proteína, com ampla função regulatória (LAGOS-QUINTANA et al. 2001; LAU et al. 2001; LEE e AMBROS 2001). Atualmente já são conhecidos 2588 miRNAs maduros humanos, segundo a mais recente atualização (*release 21*) do banco de dados miRBase (GRIFFITHS-JONES et al. 2006).

Os miRNAs são transcritos pela enzima RNA polimerase II a partir de regiões intrônicas ou intergênicas do genoma (LEE et al. 2004). Eles também podem ser policistrônicos, sendo transcritos em *clusters* contendo mais de um miRNA. Conforme ilustrado na **Figura 4**, os transcritos formam uma estrutura secundária em grampo (*hairpin*) característica desta classe de miRNAs (ROTTIERS e NÄÄR 2012).



Fonte: ROTTIERS et al. (2012)

**Figura 4** - Dois principais mecanismos de biogênese dos miRNAs: via canônica e via não-canônica.

A biogênese dos miRNAs tem como molécula inicial um miRNA-primário que, segundo a via canônica de biogênese, é clivado no núcleo pela Drosha, enzima do tipo RNaseIII, gerando os miRNAs precursores com cerca de 70 nt (LEE et al. 2003). Pela via não-canônica de biogênese, independente de Drosha, os precursores de miRNAs são formados a partir de *splicing* de introns e então denominados mirtrons (RUBY et al. 2007). Os miRNAs precursores são exportados do núcleo pela proteína Exportina 5 (LUND et al. 2004), e são clivados no citoplasma pela enzima Dicer, gerando duas fitas (5p e 3p) de miRNAs maduros com cerca de 22 nt (BERNSTEIN et al. 2001). Uma destas fitas é incorporada a um complexo de proteínas denominado RISC (do inglês *RNA-induced silencing complex*), que contém proteínas da família Argonata. Este complexo guia o miRNA, que se liga então ao seu mRNA-alvo pela complementaridade de sequência das bases 2 a 8 (chamada *seed sequence*) do miRNA aos sítios de ligação localizados principalmente na região 3'-UTR do mRNA. Em geral, a perfeita complementaridade de sequência do miRNA ao seu mRNA-alvo leva à degradação do mRNA, mecanismo comum em plantas (VOINNET 2009). Em animais a complementariedade é parcial e ocorre a redução da estabilidade do mRNA por deadenilação (redução da cauda poli-A) e a inibição da tradução do mRNA para proteína, sendo que a contribuição de cada um desses processos ainda não está totalmente esclarecida. Alguns trabalhos mostram que ocorre predominantemente a desestabilização do mRNA (EICHHORN et al. 2014; GUO et al. 2010), já outros determinam que o bloqueio da tradução é o mecanismo inicial de repressão (BAZZINI et al. 2012; DJURANOVIC et al. 2012; JANAS e NOVINA 2012).

Estima-se que mais de 60% dos mRNAs humanos sejam regulados por miRNAs (FRIEDMAN et al. 2009). Como um único miRNA tem o potencial de regular centenas de mRNAs-alvo, alterações no nível de expressão de miRNAs podem ter um grande efeito na regulação global de expressão gênica (KREK et al. 2005; LIM et al. 2005). Aberrações na expressão de miRNAs estão envolvidas com diversas doenças, sendo a leucemia mielóide crônica a primeira doença humana associada à perda de

expressão de dois miRNAs (miR-15 e miR-16) (CALIN et al. 2002). A baixa expressão global de miRNAs está associada a vários tipos tumorais, e também ao maior grau de indiferenciação celular (LU et al. 2005). Sabe-se que mais da metade dos genes codificadores para miRNAs se encontram em regiões cromossômicas mais susceptíveis à deleções e quebras (sítios frágeis), e regiões genômicas que são frequentemente envolvidas com o câncer (CALIN et al. 2004). Portanto, a avaliação do perfil de expressão de miRNAs em diversos biofluidos, tecidos, e também nas EVs, pode ter aplicação diagnóstica e prognóstica muito importante.

Aplicações prognósticas da detecção de miRNAs derivam principalmente da correlação entre sua expressão e a agressividade tumoral. Por exemplo, o nível de expressão de miR-10b em câncer de mama se correlaciona com a progressão da doença, ocorrendo um aumento da expressão deste miRNA nos tumores primários de pacientes com metástase comparado com sua expressão em tecido mamário normal (MA et al. 2007). A superexpressão de miR-21 também está associada com estadiamento clínico avançado, comprometimento linfonodal e pior sobrevida em pacientes com câncer de mama (YAN et al. 2008) e também em outros tipos tumorais (WANG et al. 2014b). Já a baixa expressão de miR-335 e miR-126 em tumores primários de mama foi correlacionada com maior chance de desenvolvimento de metástases e pior sobrevida (TAVAZOIE et al. 2008). Similarmente, em outro estudo, a baixa expressão de miR-31 foi também associada com o desenvolvimento de metástases, independentemente do grau histológico ou subtipo molecular do tumor primário de mama (VALASTYAN et al. 2009b). Neste estudo, a expressão de miR-31 foi diminuída *in vitro* pela transfecção estável de “esponjas” de miRNA, que são construtos capazes de adsorver miRNAs específicos por conterem vários sítios de ligação para este dado miRNA, em uma linhagem celular não-invasiva (MCF7-Ras). Quando injetada ortotopicamente em camundongos imunodeficientes, esta linhagem transformada teve uma capacidade 10X maior de formar metástases no pulmão dos animais em comparação com a linhagem com expressão endógena de miR-31. Em

estudo subsequente os autores concluíram que o miR-31 inibe a formação de metástases *in vivo* pela supressão concomitante da expressão de somente três genes alvos efetores, de um repertório com mais de 200 alvos preditos deste miRNA (VALASTYAN et al. 2009a).

A expressão de miRNAs também é correlacionada com resposta ao tratamento. A expressão aumentada de miRNAs miR-221 e 222 foi encontrada em pacientes com tumores de mama HER2<sup>+</sup> resistentes ao tratamento com tamoxifeno (MILLER et al. 2008). Em 2012, a alta expressão de miR-210 no plasma de pacientes com câncer de mama HER2<sup>+</sup> foi associada à resistência ao tratamento com trastuzumabe (JUNG et al. 2012). A elevada expressão deste miRNA foi antes correlacionada com maior grau de hipóxia tumoral e pior sobrevida em pacientes com câncer de mama (CAMPS et al. 2008).

Além de estarem presentes no interior das EVs, miRNAs extracelulares circulantes foram recentemente descritos associados a um complexo contendo a proteína Argonauta 2 (ARROYO et al. 2011), ou associados à lipoproteína de alta densidade (HDL) (VICKERS et al. 2011). Nestes três cenários, os miRNAs estariam protegidos da degradação pelas RNases presentes no sangue, possibilitando seu uso como biomarcadores para diagnóstico e prognóstico.

### **1.2.2 Long noncoding RNAs (lncRNAs)**

Os lncRNAs participam da regulação de expressão gênica por diversos mecanismos, dentre eles o recrutamento de complexos remodeladores de cromatina. Um dos primeiros lncRNAs descobertos, o Xist, é responsável pela inativação de um dos dois cromossomos X em fêmeas mamíferas, pelo recrutamento do complexo proteico PRC2 que metila as histonas levando à repressão da transcrição (BROWN et al. 1992; ZHAO et al. 2008).

Os lncRNAs mais estudados são aqueles transcritos de regiões intergênicas (lincRNAs). Em câncer eles podem atuar como supressores de tumor ou oncogenes (REIS e VERJOVSKI-ALMEIDA 2012). Um exemplo de

lincRNA supressor de tumor é o *GAS5*, que se liga ao receptor glucocorticoide, e é menos expresso em células tumorais de mama em comparação com o tecido normal adjacente, e cuja superexpressão em células MCF-7 leva à apoptose (MOURTADA-MAARABOUNI et al. 2009). Por outro lado, o lincRNA *HOTAIR*, que regula a expressão dos genes homeobox, atua como um oncogene e sua superexpressão *in vitro* em linhagens tumorais de mama foi correlacionada com uma maior capacidade de invasão *in vivo*. Sua alta expressão no tumor primário é um fator preditivo para o desenvolvimento de metástase e pior sobrevida dos pacientes (GUPTA et al. 2010).

Estudos apontam a importância dos lncRNAs derivados de regiões intrônicas ao sugerirem que para 81% dos genes do banco RefSeq ocorre a transcrição de pelo menos um lncRNA totalmente intrônico (LOURO et al. 2008). Alguns dos lncRNAs intrônicos antisensos são conservados, têm expressão tecido-específica e foram correlacionados com o grau de diferenciação tumoral em câncer de próstata (REIS et al. 2004).

### 1.2.3 RNAs circulares (circRNAs)

RNAs circulares (circRNAs) são transcritos cujos exons encontram-se em ordem não-sequencial em relação à isoforma linear (mRNA canônico), devido à circularização. Apesar dos circRNAs serem transcritos a partir de genes codificadores de proteínas, ainda não existem relatos na literatura de circRNAs que codifiquem proteínas. Desde o final da década de 1970 experimentos de microscopia eletrônica foram capazes de demonstrar que os RNAs eucarióticos podem existir em forma circular (HSU e COCA-PRADOS 1979). A partir de 2012, graças ao avanço das tecnologias de sequenciamento de RNAs em larga escala (RNA-Seq), descobriu-se que circRNAs são abundantes, estáveis, conservados evolutivamente, e constituem a isoforma predominante para centenas de genes humanos (SALZMAN et al. 2012; WANG et al. 2014a). A molécula de RNA circular, ao contrário da linear, é resistente à digestão pela enzima RNA exonuclease RNase R. Deste modo, é possível enriquecer o RNA inicial para uma

população de moléculas circulares com um tratamento enzimático antes da construção de bibliotecas (JECK e SHARPLESS 2014). Ao comparar uma amostra de RNA celular total pré-enriquecida com RNase R com a mesma amostra não-enriquecida por este tratamento, foi possível identificar mais de 25 mil circRNAs distintos em fibroblastos humanos, sendo que do total de genes expressos nestas células, 14,4% dos genes produziram um circRNA. Para alguns transcritos detectou-se que as isoformas circulares eram pelo menos 10 vezes mais abundantes que as lineares (JECK et al. 2013). Os autores também investigaram se havia características específicas da sequência genômica que poderiam estar associadas com a formação de circRNAs e encontraram um enriquecimento em circRNAs de exons longos, flanqueados por introns longos contendo elementos repetitivos complementares do tipo ALU (JECK et al. 2013). Dois outros trabalhos também descreveram a importância de sequências reverso-complementares nos introns flanqueadores (IVANOV et al. 2014) e a presença de sítios de *splicing* canônicos para o processo de circularização (STARKE et al. 2014). Apesar destes avanços, o entendimento da biogênese e da função dos circRNAs ainda é muito incompleto.

Recentemente foi descrito que RNAs circulares podem atuar como esponjas de miRNAs, impedindo que estes se liguem aos seus mRNAs alvos. Neste sentido, os circRNAs regulariam a expressão da proteína codificada pelo mRNA-alvo do miRNA, pois várias repetições dos sítios de ligação presentes no circRNA competiriam com o mRNA-alvo para a ligação dos miRNAs. No circRNA *ciRS-7/CDR1as* foram encontrados 73 repetições do sítio de ligação (regiões complementares à *seed sequence* do miRNA) para o miR-7, enquanto que no circRNA *Sry* foram encontrados 16 sítios de ligação para o miR-138 (HANSEN et al. 2013). Neste mesmo trabalho foi demonstrado que a regulação negativa *in vitro* da expressão de três alvos endógenos (SNCA, EGFR and IRS2) do miR-7 foi menos eficiente em células HeLa estavelmente transfectadas com construtos do circRNA *ciRS-7/CDR1as* e quantidades crescentes do miR-7, em comparação com células não-transfectadas. Outros confirmaram esta mesma associação circRNA:

miRNAs por imunoprecipitação da proteína Argonauta, seguida de sequenciamento dos RNAs associados (PAR-CLIP) (MEMCZAK et al. 2013).

Apesar de já terem sido detectados em um biofluido corporal (saliva livre de células) (BAHN et al. 2015), RNAs circulares ainda não foram descritos especificamente associados a vesículas extracelulares. Se comprovado o papel funcional destes RNAs como esponjas de miRNAs, podemos imaginar a transferência de circRNAs por EVs como um mecanismo regulatório à distância de modulação da expressão gênica de células que incorporam o conteúdo molecular das EVs.

Deste modo, é fascinante a possibilidade de encontrarmos moléculas regulatórias carregadas por EVs – sejam elas RNAs, DNAs, proteínas ou outras – associadas a certas doenças, na medida que estudos deste tipo abrem a possibilidade de avançarmos no entendimento de mecanismos moleculares alternativos com potencial impacto no diagnóstico, prognóstico e na compreensão das bases moleculares destas doenças.

## **1.3 CÂNCER DE MAMA**

### **1.3.1 Incidência e mortalidade**

O câncer de mama é o tipo de tumor mais comum entre as mulheres em todo o mundo. No Brasil foram estimados 57,120 novos casos para 2014, segundo estimativa do INCA (MINISTÉRIO DA SAÚDE 2013). Apesar de grandes avanços no diagnóstico precoce e no tratamento das neoplasias da mama, incluindo terapias-alvo, a mortalidade continua alta. Mesmo em regiões mais desenvolvidas, o câncer de mama tem a segunda maior taxa de mortalidade entre as mulheres (15% do total de mortes por câncer), atrás somente do câncer de pulmão, segundo estimativa do GLOBOCAN 2012 (FERLAY et al. 2013).



### 1.3.2 Classificação Histopatológica

O câncer de mama é uma doença heterogênea, classificada morfológicamente pelo padrão de crescimento e aspectos citológicos das células tumorais em dois principais tipos: tumores *in situ*, onde as células tumorais não invadiram a membrana basal, e tumores invasivos, onde ocorre a infiltração da membrana e a disseminação para o estroma e outros tecidos. O tipo histológico mais comum no carcinoma da mama é o tipo carcinoma ductal invasivo não-especificado (~80% dos casos), cujo diagnóstico é feito por exclusão quando o tumor não pode ser classificado em um tipo especial de carcinoma invasivo, como lobular, tubular, mucinoso, medular, entre outros (BURSTEIN et al. 2011). Os tumores são também subclassificados de acordo com a diferenciação celular pelo grau histológico que leva em consideração três características: formação tubular, pleomorfismo nuclear, e atividade mitótica (ELSTON e ELLIS 1991). O grau histológico é um dos fatores prognósticos reconhecidos como sendo importante para prever a evolução clínica da doença, além do tamanho do tumor, status do linfonodo axilar e subtipo molecular (FITZGIBBONS et al. 2000).

### 1.3.3 Classificação Molecular

A classificação molecular baseada na avaliação de biomarcadores no tumor complementa a classificação histopatológica dos carcinomas de mama. Estudos clássicos realizados pelos grupos de Perou e Sorlie caracterizaram os tumores como tendo diferentes perfis moleculares, que os agrupam em subclasses com importantes implicações clínicas (PEROU et al. 2000; SØRLIE et al. 2001). Os achados destes estudos foram traduzidos em marcadores protéicos que são avaliados por imunohistoquímica na rotina diagnóstica, sendo eles: os receptores hormonais de estrogênio (ER) e progesterona (PR), além do receptor do fator de crescimento epidermal 2 (HER2 ou ERBB2). O status de *HER2* também pode ser avaliado pelo método de *fluorescent in situ hybridization* (FISH), capaz de determinar o grau de amplificação gênica deste oncogene. A positividade para o marcador HER2 é definida, segundo a mais atual recomendação de 2014 da *American*

*Society of Clinical Oncology/College of American Pathologists*, por imunistoquímica com >10% das células com marcação intensa e circunferencial de membrana; ou por *dual-probe FISH* com uma razão de HER2/CEP17  $\geq 2.0$  (WOLFF et al. 2014). De acordo com a positividade para os receptores hormonais, o HER2, o marcador de proliferação celular ki-67, as citoqueratinas epiteliais 5/6, e o receptor do fator de crescimento epidermal (EGFR), os tumores são subclassificados em 4 principais subtipos: luminal A, luminal B, HER2, e *basal-like*, conforme **Quadro 1** (HOFSTATTER et al. 2011)

**Quadro 1** – Classificação molecular do câncer de mama.

| <b>Subtipo</b>    | <b>Marcadores</b>  |
|-------------------|--|
| Luminal A         | ER- e/ou PR-positivo/HER2-negativo   |
| Luminal B         | ER- e/ou PR-positivo/HER2-positivo ou ki-67 $\geq 20\%$ ;                  |
| HER2-positivo     | ER- e PR-negativo/HER2-positivo  |
| <i>basal-like</i> | ER-, PR- e HER2-negativo e/ou citokeratina 5/6-positivo e/ou EGFR-positivo |

A determinação do subtipo molecular é de suma importância para definir a conduta terapêutica adjuvante sistêmica: somente pacientes com tumores positivos para o receptor de estrógeno são tratadas com Tamoxifeno, enquanto somente pacientes com tumores positivos para o HER2 são tratadas com Trastuzumabe (UNTCH et al. 2013). Do mesmo modo, o estudo de tumores de mama sem um agrupamento por subtipo molecular pode gerar dados conflitantes e que podem contribuir muito pouco para o avanço no conhecimento desta neoplasia.

#### 1.3.4 HER2

O receptor do fator de crescimento epidermal 2 (*HER2/ERBB2*), cujo gene se localiza no braço longo cromossomo 17 (17q12), é um membro da família dos receptores tirosina quinase que inclui o *EGFR (HER1/ErbB1)*, o *HER3/ErbB3* e o *HER4/ErbB4*. A proteína codificada por este gene possui

três domínios principais: a região extracelular onde ocorre a interação com ligantes, a região transmembrana e o domínio citoplasmático com atividade tirosina quinase que ativa vias intracelulares de sinalização por cascatas de fosforilação de proteínas. O HER2 heterodimeriza preferencialmente com o HER3, mas ele também pode formar homodímeros. A ligação do ligante ao domínio extracelular induz a dimerização de dois receptores, levando à ativação dos domínios quinases que se trans-fosforilam. A fosforilação dos dímeros do receptor HER2 leva à ativação de duas principais vias de sinalização: *ras/raf/MEK/MAPK* e *PI3K/AKT*, ambas levando à transcrição de genes envolvidos com proliferação celular (HOFSTATTER et al. 2011).

A superexpressão de HER2, seja por amplificação gênica ou superexpressão da proteína, leva à ativação constitutiva das vias de sinalização promovendo o crescimento tumoral por diversos processos. Um deles é pela desregulação do controle da fase G<sub>1</sub>/S do ciclo celular, com a superexpressão das ciclinas D1, E e cdk6, além da degradação de p27 (TIMMS et al. 2002). A angiogênese também é estimulada, pois a superexpressão de HER2 leva a um aumento da expressão de HIF-1 $\alpha$  e de VEGF (LI et al. 2005).

Desde 1987 é sabido que cerca de 30% dos carcinomas invasivos de mama superexpressam HER2, e que estes tumores têm uma pior sobrevida na ausência de tratamento (SLAMON et al. 1987). Os tumores tipicamente tem entre 25-50 cópias do gene *HER2*, e um aumento de 40-100 vezes na expressão da proteína HER2, levando a mais de 2 milhões de receptores expressos na membrana de uma única célula (MOASSER 2007).

O trastuzumabe, um anticorpo monoclonal humanizado que se liga a HER2, é usado para tratar pacientes com tumores que superexpressam HER2. Ele foi inicialmente aprovado pelo FDA em 1998 para tratar pacientes com tumores metastáticos, mas atualmente é utilizado também em tumores iniciais, após a divulgação em 2005 do resultado de cinco ensaios clínicos randomizados que o uso adjuvante de trastuzumabe diminuiu a taxa de recorrência pela metade e reduziu a mortalidade em 30% (BRAGA et al. 2006).

A eficácia do trastuzumabe se deve a vários mecanismos: 1) o bloqueio da heterodimerização do receptor e, portanto, de sua ativação; 2) a diminuição da angiogênese; 3) maior citotoxicidade celular dependente de anticorpo (ADCC) mediada pelo sistema imune. O tratamento com trastuzumabe tem efeitos colaterais, principalmente a cardiotoxicidade, e somente um terço das pacientes tem resposta imediata. A resistência ao tratamento é comum e a maioria das pacientes com resposta inicial progride após um ano. Vários possíveis mecanismos de resistência ao trastuzumabe já foram descritos, dentre eles a superexpressão de IGF-1R e a perda da fosfatase PTEN, ambos eventos que levam à ativação constitutiva de vias de sinalização independentemente do bloqueio do receptor HER2 (BARTSCH et al. 2007).

Nesta tese buscamos caracterizar o conteúdo molecular de EVs circulantes no plasma de pacientes com cancer de mama HER2+, a fim de avaliarmos a sua composição durante a evolução da doença. Os RNAs e proteínas contidos em EVs são protegidos de degradação pela camada lípidica e deste modo podem ser excelentes biomarcadores. A caracterização destas moléculas apresentada aqui é inédita nos grupos de pacientes avaliados e esperamos que este trabalho contribuia para o melhor entendimento dos processos biológicos envolvidos com a progressão da doença.

## 2 OBJETIVOS

- Avaliar, nas linhagens celulares HB4a e C5.2, a existência de alterações no conteúdo de vesículas extracelulares (EVs) relacionadas à expressão aumentada do oncogene ERBB2/HER2.
- Avaliar, nas mesmas linhagens celulares citadas acima, diferenças no conteúdo de EVs em relação às células secretoras.
- Determinar a quantidade de EVs, assim como o seu conteúdo de RNAs, em amostras de plasma de mulheres portadoras de câncer de mama ductal invasivo HER2<sup>+</sup>, alocadas em três grupos de estadiamento progressivo da doença.
- Avaliar, nestas mesmas amostras, a variabilidade intra- e inter-grupo da quantidade de EVs e do conteúdo de RNAs derivados de EVs.
- Buscar validar a expressão de moléculas de interesse utilizando uma segunda abordagem experimental.

### 3 MATERIAIS E MÉTODOS

#### 3.1 SELEÇÃO DE PACIENTES E COLETA DE AMOSTRAS

Este estudo prospectivo foi aprovado pelos Comitês de Ética em Pesquisa do AC Camargo Cancer Center (nº 1554/11) e do Hospital de Câncer de Barretos (nº 689/2013) e também pelo CONEP (nº 16580). As cartas de aprovação constam no **Anexos 1 e 2**. Recebemos ainda o apoio financeiro da FAPESP (Projeto Regular 2011/09172-3). Todas as pacientes foram provenientes do AC Camargo Cancer Center ou do Hospital de Câncer de Barretos. Os critérios de inclusão das pacientes foram: diagnóstico de câncer de mama ductal HER2<sup>+</sup> (*in situ* e invasivo), ausência de qualquer tratamento prévio à coleta da amostra (cirurgia, quimio- ou radio-terapia) e assinatura do termo de consentimento livre e esclarecido (TCLE, **Anexo 3**). A positividade para o marcador HER2 foi considerada por escore 3<sup>+</sup> em imunohistoquímica ou, nos casos com escore 2<sup>+</sup>, pela razão HER2/CEP17 (centrômero cromossomo 17)  $\geq 2.0$  em FISH, indicado no laudo anatomopatológico da biópsia. O sangue das pacientes (15ml) foi coletado em tubos BD contendo ácido cítrico, citrato, dextrose (solução ACD) (GYÖRGY et al. 2014). Imediatamente após a coleta, o plasma foi separado por duas centrifugações a 2500xg por 15 minutos à temperatura ambiente (LACROIX et al. 2012). O plasma foi alíquotado em tubos de 1,5ml e congelado a -80°C até a sua utilização. As pacientes foram divididas em três grupos, de modo a mimetizar a progressão natural da doença: a) doença inicialmente restrita ao tumor primário (ou simplesmente N0, para Linfonodo Negativo); b) doença com acometimento linfonodal (simplesmente N, para Linfonodo) e c) doença com lesões metastáticas à distancia (ou M, para doença Metastática). Cada grupo foi composto por 10 pacientes.

## 3.2 CULTURA CELULAR

As linhagens celulares HB4a e C5.2, originalmente cedidas pelo Dr. Michael O'Hare (Ludwig Institute for Cancer Research, UK) e a nós transferidas pela Dra. Dirce M. Carraro (AC Camargo Cancer Center), foram cultivadas a 37°C e 5% CO<sub>2</sub> em meio RPMI 1640 suplementado com 10% soro fetal bovino (Invitrogen, California, EUA), 1% antibiótico-antimicótico - penicilina, estreptomicina, anfotericina B (Invitrogen, California, EUA), e 5 µg/ml hidrocortisona (Sigma-Aldrich, Missouri, EUA). Ambas linhagens se mostraram negativas para teste por PCR que avaliou contaminação por micoplasma. O meio completo foi depletado de EVs provenientes do soro fetal bovino, por centrifugação a 100.000 x g por >16 horas, antes de sua utilização no cultivo das células.

## 3.3 ISOLAMENTO DAS EVs DERIVADAS DO PLASMA

Uma alíquota de plasma de cada paciente (2,5ml) foi diluída com 9,5ml de PBS para completar o volume de 12ml necessário para preencher o tubo da ultracentrífuga (Beckman modelo 331372). Para eliminar células e debris celulares, foram realizadas centrifugações a 2.000xg por 30 minutos e 12.000xg por 45 minutos, descartando os *pellets* (centrífuga Eppendorf 5810R, rotor de ângulo fixo F-34-6-38). O sobrenadante foi então centrifugado em ultracentrífuga Optima L-90K (Beckman Coulter, California, EUA), rotor *swinging* SW-41Ti a 110.000xg por 120 minutos. O *pellet* de EVs foi lavado em PBS e re-centrifugado (110.000xg por 70 minutos). Todas as centrifugações foram realizadas a 4°C. Diferentemente do protocolo empregado para os meios de cultura celulares condicionados, diante da quantidade limitante de sangue/plasma das pacientes, para estas foi feito um único *pellet* (conteúdo equivalente a 20K + 100K). O *pellet* final de EVs foi ressuspendido em solução de lise do kit de extração de RNA, conforme detalhado abaixo.

Para a etapa de validação, as EVs foram isoladas a partir de um

volume menor de plasma (1,5ml) diluído em 2,5ml de PBS para completar o volume de 4ml requerido para a ultracentrifugação (Hitachi tubo S404332A). Para eliminar células e debris celulares, foram realizadas duas centrifugações (Eppendorf 5810R, rotor de ângulo fixo F-34-6-38) a 2.000xg por 30 minutos e 12.000xg por 45 minutos, descartando os *pellets*. O sobrenadante foi então ultracentrifugado em ultracentrífuga CS-150GX (Hitachi, Tokyo, Japão) com rotor de ângulo fixo S100AT6 a 110.000xg por 120 minutos. O *pellet* de EVs foi lavado com PBS e re-centrifugado a 110.000xg por 70 minutos. Todas as centrifugações foram realizadas a 4°C. O *pellet* final de EVs foi ressuspensionado em solução de lise do kit de extração de RNA, conforme detalhado abaixo.

### **3.4 ISOLAMENTO DAS EVs DERIVADAS DO MEIO DE CULTURA CELULAR CONDICIONADO**

Inicialmente, o meio condicionado de cultura de células (120ml) de cada linhagem (HB4a e C5.2) foi centrifugado (centrífuga Eppendorf 5810R, rotor de ângulo fixo F-34-6-38) a 300xg por 10 minutos, 2.000xg por 20 minutos e 10.000xg por 30 minutos para eliminar células e debris celulares. Após cada uma destas etapas de centrifugação os *pellets* obtidos foram descartados. O sobrenadante final foi centrifugado em ultracentrífuga Optima L-90K (Beckman Coulter, California, EUA) com rotor *swinging* SW-28 a 20.000xg por 60 minutos para obtenção de população de EVs enriquecida para microvesículas (denominada 20K), e a 100.000 x g por 60 minutos para população de EVs enriquecida para exossomos (denominada 100K). Cada um dos *pellets* de EVs (20K e 100K) foi lavado com PBS e re-centrifugado a 100.000xg por 60 minutos. Posteriormente, estes *pellets* foram ressuspensionados em solução de lise do kit de extração de RNA, ou tampão de lise para extração proteica, conforme detalhado abaixo. Todas as centrifugações foram realizadas a 4°C.



### **3.5 CARACTERIZAÇÃO DAS EVs DERIVADAS DO PLASMA POR NTA**

As EVs foram isoladas por ultracentrifugação, conforme descrito acima, a partir de 100 $\mu$ l de plasma e o *pellet* foi ressuspendido em 450 $\mu$ l de PBS filtrado (filtro de 0,2 $\mu$ m). As amostras foram avaliadas no equipamento *NanoSight LM10* (Malvern Instruments, Worcestershire, Reino Unido), utilizando seringas de 1ml descartáveis, sendo cada amostra lida em triplicata. Entre cada amostra o leitor foi limpo com PBS filtrado e cada amostra seguinte só foi inserida após ter sido verificada a ausência de partículas. Os parâmetros utilizados foram: *camera shutter* – 1495, *camera gain* – 512, *detection threshold* – 10, duração de captura do vídeo – 60 segundos.

### **3.6 CARACTERIZAÇÃO DAS EVs DERIVADAS DO MEIO DE CULTURA CELULAR CONDICIONADO POR NTA**

Igual descrito acima, sendo que o *pellet* foi ressuspendido em 1000 $\mu$ l de PBS filtrado e os seguintes parâmetros foram utilizados: *camera shutter* – 414, *camera gain* – 196, *detection threshold* – 10.

### **3.7 CARACTERIZAÇÃO DAS EVs DERIVADAS DO PLASMA POR WESTERN BLOT**

As EVs foram isoladas por ultracentrifugação conforme descrito acima, a partir de cerca de 4ml de plasma e ao final o *pellet* foi ressuspendido com 100 $\mu$ l de tampão de lise contendo 300mM NaCl, 50mM Tris pH 7,4, 0,5% NP-40, e coquetel anti-protease (cOmplete Mini - Roche, Indiana, EUA). A concentração de proteínas foi determinada pelo método colorimétrico de Bradford (Bio-Rad, California, EUA), e 1 $\mu$ g de proteínas foi submetido a SDS-PAGE (gel 10% poliacrilamida). Após transferência para membrana de

nitrocelulose (GE Healthcare, Pennsylvania, EUA), foi feito o bloqueio por 1 hora com 5% leite, incubação com anticorpo primário (anti-FLOTILLIN produzido em coelho 1:500 – ab41927 e anti-RAB7 produzido em camundongo 1:1,000 – ab50533, Abcam, Massachusetts, EUA) *overnight* a 4°C, e incubação por 1 hora com anticorpo secundário conjugado com HRP 1:5,000 (GE Healthcare, Pennsylvania, EUA), as membranas foram reveladas com *ECL Prime* (GE Healthcare, Pennsylvania, EUA) em filme.

### **3.8 CARACTERIZAÇÃO DAS EVs DERIVADAS DO MEIO DE CULTURA CELULAR CONDICIONADO POR WESTERN BLOT**

Igual à descrição do item 3.7 acima, sendo que os seguintes anticorpos primários foram utilizados: anti-HER2 produzido em coelho 1:10.000 (2165, Cell Signaling, Massachusetts, EUA), anti-HSP70 produzido em coelho 1:500 (EXOAB-Hsp70A-1, System Biosciences, California, EUA), anti-Flotillin produzido em coelho 1:1.000 (ab41927, Abcam, Massachusetts, EUA), e anti-RAB27B produzido em coelho 1:250 (HPA019849, Sigma-Aldrich, Missouri, EUA).

### **3.9 CARACTERIZAÇÃO DAS EVs DERIVADAS DO MEIO DE CULTURA CELULAR CONDICIONADO POR MICROSCOPIA ELETRÔNICA**

O meio condicionado (~240ml) proveniente da cultura de células foi centrifugado conforme descrito acima. Os pellets finais (20K e 100K) foram fixados *overnight* a 4°C com glutaraldeído e paraformaldeído 2%, lavados com PBS, seguido de tetróxido de ósmio 1% por 10 minutos, e uranil acetato 1% por 10 minutos. As amostras foram então desidratadas por 1 minuto em soluções de etanol a 50%, 70%, 90% e 100% (2X). Em seguida as amostras foram incluídas em resina 812 (EMS, Pennsylvania, EUA) no próprio tubo e polimerizadas com incubação *overnight* a 60°C. Cortes ultra-finos de 65-75 nm feitos em ultramicrótomo com lâmina de diamante foram depositados em

grades de cobre 200-mesh (EMS, Pennsylvania, EUA), contrastados com uranil acetato 0,5% por 5 minutos e citrato de chumbo por 4 minutos, e observados ao microscópio eletrônico de transmissão JEM -1010 (JEOL, Tokyo, Japão) operado a 80kV (Depto de Biologia Celular e do Desenvolvimento, ICB/USP).

### **3.10 ANÁLISE PROTEÔMICA DE EVs E CÉLULAS DAS LINHAGENS HB4a E C5.2**

#### **3.10.1 Extração de proteínas**

As proteínas foram extraídas após ressuspensão dos pellets 20K e 100K de EVs derivadas do meio de cultura celular condicionado em 100µl de tampão de lise contendo 300mM NaCl, 50mM Tris pH 7,4, 0,5% NP-40 e coquetel anti-proteases (Roche, Indiana, EUA). Extratos celulares foram preparados incubando os pellets de células em 200µl de tampão de lise por 20 minutos no gelo, e coletando o sobrenadante de centrifugação a 20.000xg por 2 minutos. As proteínas foram liofilizadas e enviadas ao laboratório do Dr. Daniel Martins-de-Souza na Ludwig-Maximilians-University Munich, Alemanha, para a realização da análise proteômica.

#### **3.10.2 Marcação ICPL e pré-fracionamento dos proteomas celulares**

Na Alemanha, os proteomas das células foram marcados por *isotope-coded protein label* (ICPL) 4-plex (SERVA Electrophoresis, Heidelberg, Alemanha), conforme previamente descrito (MACCARRONE et al. 2010). *Pools* de 100µl de proteína total de cada um dos quatro grupos amostrais (duas linhagens em duplicata) foram reduzidos com ditioneitol (DTT) por 30 minutos a 60°C e resfriados à temperatura ambiente (TA). Em seguida, grupos tiol livres foram alquilados no escuro por 30 minutos a TA com 1ml de 0,4M iodoacetamida, seguido de 1ml de 0,5M N-acetilcisteína. Para marcação das proteínas, isótopos estáveis – ICPL 0, ICPL 4, ICPL 6, e ICPL 10 – foram adicionados em um excesso molar de 10X em relação as amostras e incubados por 2 horas a TA. Em seguida, 4µl de 1,5M

hidroxilamina foi adicionada a cada amostra para inativação da marcação. Grupos éster remanescentes foram hidrolisados aumentando o pH a 11-12 por 20 minutos. As quatro amostras (duas replicatas de cada linhagem) foram unidas e pré-fraccionadas em um mini-gel SDS-PAGE 12%. As proteínas foram visualizadas com o corante Coomassie brilliant blue. Cada fileira corada no gel foi cortada em 20 fatias, que foram submetidas a digestão *in-gel* com tripsina e liofilizadas ao final.

### **3.10.3 Digestão em solução do proteoma de EVs**

O proteoma das EVs foi digerido em solução usando uma razão 1:50 de tripsina para amostra. As proteínas digeridas foram liofilizadas ao final.

### **3.10.4 LC-MS/MS**

Cada fatia digerida de amostra das células assim como o proteoma digerido das EVs foi dissolvido em solução aquosa de 0,1% ácido fórmico e injetadas no sistema nano-LC (Eksigent, California, EUA) em uma coluna Zorbax-300 SB-C18 (5µm, 5×0.3mm id, Agilent Technologies, California, EUA) e separadas em uma nano-coluna, 75 µm id × 15 cm empacotada com partículas de 3µm revestidas por C18. Os peptídeos eluídos foram analisados online em um espectrômetro de massas LTQ-Orbitrap XL (Thermo Scientific, Bremen, Alemanha) equipado com uma fonte nano-ESI em modo positivo (MACCARRONE et al. 2013). A separação cromatográfica e a aquisição de espectra foram automaticamente realizados pelo programa Xcalibur (versão 2.0.7, Thermo Scientific, California, EUA).

### **3.10.5 Análise de dados**

Os dados brutos de espectrometria de massas foram processados diretamente no programa *MASCOT Distiller* (Matrix Sciences, Londres, Reino Unido) e buscados contra a database Uniprot de proteínas humanas (*release 2013\_09*). Os parâmetros de busca foram: acurácia da massa iônica dos peptídeos e fragmentos: 10 ppm e 0,6 DA, respectivamente; tripsina como enzima permitindo no máximo duas clivagens perdidas;

carbamidometilação de cisteína como modificação fixa e oxidação de metionina como modificação variável. Os experimentos foram realizados em duplicada e só foram consideradas proteínas identificadas por pelo menos dois peptídeos distintos com >95% de confiança determinadas pelo algoritmo do MASCOT. Quantificações também foram realizadas no mesmo programa. Por empregarmos marcação ICPL para a quantificação do proteoma das células, a intensidade relativa de cada peptídeo precursor na sua forma marcada foi obtida dos cromatogramas iônicos e razões foram calculadas por comparações dois a dois. Para as EVs, foi realizada uma análise *spectral counting label-free* usando o algoritmo do *MASCOT Distiller*. As análises de enriquecimento de vias foram realizadas com o programa MetaCore em colaboração com o Dr. Paulo Oliveira do Laboratório Nacional de Biociências (LNBio) em Campinas.

### 3.11 EXTRAÇÃO DE RNA DAS EVs

Os RNAs das EVs, obtidas a partir do plasma (2,5ml para NGS e 1,5ml para validação) ou do meio de cultivo celular condicionado (120ml), foram extraídos com o *Total RNA Purification Kit* (Norgen, Ontario, Canada). Este kit utiliza uma coluna contendo uma resina que permite a purificação de RNA de todos os tamanhos, incluindo os pequenos RNAs, com tamanho inferior a 200nt. O *pellet* de EVs após a ultracentrifugação final foi ressuspenso em 350 $\mu$ l de *Lysis Solution* contendo 1%  $\beta$ -mercaptoetanol, e em seguida foi então adicionado 200 $\mu$ l de 100% etanol. Todo o lisado foi transferido para a coluna, que foi centrifugada a 14.000xg por 1 minuto à TA e lavada 3 vezes com 400 $\mu$ l de *Wash Solution*. O RNA foi eluído em 50 $\mu$ l de *Elution Solution*.

### 3.12 EXTRAÇÃO DE RNA DAS CÉLULAS, TRATAMENTO COM DNASE E DEPLEÇÃO DE rRNA

Os RNAs obtidos das linhagens celulares HB4a e C5.2 foram extraídos a partir de  $1,2 \times 10^7$  células de cada linhagem (divididas em 4 alíquotas), usando o *Total RNA Purification Kit* (Norgen, Ontario, Canada) e eluídos em 50 $\mu$ l conforme descrito acima. Para a quantificação foi usado o aparelho NanoDrop (Thermo Scientific, Massachusetts, EUA) e as amostras foram diluídas a 10 $\mu$ g/50 $\mu$ l para tratamento com DNase (*TURBO DNA-free kit* - Ambion, California, EUA). Em seguida, 3 $\mu$ g de RNA livre de DNA foram depletados de rRNA com o *Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat)* (Epicentre, Wisconsin, EUA) que utiliza sondas biotiniladas complementares ao rRNA, possibilitando a sua captura e remoção com *beads* magnéticas acopladas à estreptavidina. A depleção do rRNA foi verificada pelo equipamento Bioanalyzer (Agilent Technologies, California, EUA), que realiza uma eletroforese capilar das amostras utilizando reagentes do *Agilent RNA 6000 Pico Kit* (Agilent Technologies, California, EUA) e 1 $\mu$ l de RNA.

### 3.13 SEQUENCIAMENTO DE NOVA GERAÇÃO

#### 3.13.1 Construção das bibliotecas

As bibliotecas foram construídas utilizando reagentes do *Ion Total RNA-Seq Kit v2* (Life Technologies, California, EUA) nas etapas detalhadas abaixo.

##### 1) Fragmentação e purificação do RNA

O volume total de RNA eluído da coluna (50 $\mu$ l) foi concentrado no SpeedVac para um volume de 16 $\mu$ l. Todo o RNA obtido foi fragmentado com a enzima *RNase III* (2 $\mu$ l) e *10X RNase III Reaction Buffer* (2 $\mu$ l) a 37°C por 3 horas, em um volume final de 20 $\mu$ l. Em seguida este RNA foi purificado com *Nucleic Acid Binding Beads*, que ligam somente RNA e permitem que restos

de tampão e enzima sejam removidos por lavagem utilizando estante magnética. O RNA foi eluído das *beads* com 12 $\mu$ l de água livre de nucleases.

## 2) Hibridização e ligação do RNA aos adaptadores

Os adaptadores aqui usados são oligonucleotídeos de sequência degenerada simples fita em uma extremidade, que hibridiza diretamente com as moléculas de RNA, seguida de uma parte dupla fita contendo sequência complementar aos iniciadores usados para a amplificação da biblioteca e posteriormente para a reação de sequenciamento *in situ*. Utilizamos um adaptador específico para cada extremidade da molécula de RNA, de modo a preservarmos a orientação da fita transcrita. O sequenciamento foi feito sempre a partir da extremidade 5' da fita senso de RNA.

Após a purificação do RNA fragmentado, o volume total eluído das *beads* (12 $\mu$ l) foi concentrado no SpeedVac para um volume de 3 $\mu$ l, ao qual foi adicionado 3 $\mu$ l de *Hybridization Solution* e 2 $\mu$ l de *Ion Adaptor Mix v2* para um volume final de 8 $\mu$ l, que foi incubado a 65°C por 30 minutos. Em seguida foi adicionado 10 $\mu$ l de *2X Ligation Buffer* e 2 $\mu$ l de *Ligation Enzyme Mix*, para um volume final de 20 $\mu$ l, que foi incubado a 16°C por 16 horas.

## 3) Síntese e purificação do cDNA

Os RNAs ligados aos adaptadores foram convertidos em cDNA por transcrição reversa. Aos 20 $\mu$ l de reação de ligação foram adicionados 2 $\mu$ l de água livre de nucleases, 4 $\mu$ l de *10X RT Buffer*, 2 $\mu$ l de *dNTP Mix*, e 8 $\mu$ l de *Ion RT Primer*, para um volume final de 36 $\mu$ l que foi incubado a 75°C por 15 minutos, e transferido para o gelo. Em seguida foram adicionados 4 $\mu$ l de *10X SuperScript Enzyme Mix*, para um volume final de 40 $\mu$ l que foi incubado a 42°C por 30 minutos. O cDNA assim produzido foi purificado com as *Nucleic Acid Binding Beads* utilizando estante magnética. O cDNA purificado foi eluído das *beads* com 12 $\mu$ l de água livre de nucleases.

#### 4) Amplificação do cDNA e purificação dos *amplicons*

O cDNA purificado foi amplificado por PCR, etapa na qual também são inseridos os *barcodes* que permitem a atribuição da origem das amostras após o sequenciamento. A PCR consistiu de 6µl de cDNA purificado, 45µl de *Platinum PCR SuperMix High Fidelity*, 1µl de *Ion Xpress RNA 3' Barcode Primer* e 1µl de *Ion Xpress RNA-Seq Barcode BC primer*, conforme **Quadros 2 e 3**, para um volume final de 53µl. O programa utilizado para a amplificação por PCR foi: 94°C por 2 minutos, 2 ciclos de 94°C por 30 segundos e 50°C por 30 segundos, 18 ciclos de 94°C por 30 segundos e 62°C por 30 segundos, e 68°C por 5 minutos. O cDNA amplificado foi purificado como anteriormente descrito e eluído das *beads* com 10µl de água livre de nucleases.

#### 5) Avaliação do cDNA amplificado (biblioteca final)

O controle de qualidade das bibliotecas foi feito no equipamento Bioanalyzer, que realiza uma eletroforese capilar das amostras utilizando reagentes do *High Sensitivity DNA Kit* (Agilent Technologies, California, EUA). Um eletroferograma é gerado a partir de 1µl de amostra, permitindo a determinação tanto do tamanho (em pares de bases) das moléculas presentes na biblioteca, assim como de sua concentração (em pM). O tamanho esperado para miRNAs ligados à adaptadores é de 94-114bp, segundo protocolo do fabricante (Life Technologies) do kit de preparo de biblioteca.

### 3.13.2 PCR em emulsão

A partir da molaridade da biblioteca, determinada pelo Bioanalyzer, foi feita uma diluição de cada biblioteca para uma concentração final de 11pM (para as bibliotecas das pacientes, utilizado versão v2 do kit de preparo de emulsão; Life Technologies, California, EUA) ou 100pM (para as bibliotecas das linhagens celulares, utilizado *kit* versão v3 do kit de preparo de emulsão; Life Technologies, California, EUA), seguindo recomendação do fabricante para a realização da PCR em emulsão, conforme **Quadros 2 e 3**.



**Quadro 2** - Concentração molar e fator de diluição das bibliotecas construídas a partir de RNA das EVs e das linhagens celulares que as originaram.

| Biblioteca         | Barcode | Molaridade da biblioteca (pM) | Fator de diluição da biblioteca para 100pM |
|--------------------|---------|-------------------------------|--|
| C5.2 EVs 20K 1:10  | 1       | 10.990,3                      | 109,9                                      |
| C5.2 EVs 100K 1:10 | 2       | 13.614,5                      | 136,1                                      |
| HB4a EVs 20K 1:10  | 3       | 6.042,3                       | 60,4                                       |
| HB4a EVs 100K 1:10 | 4       | 15.302,3                      | 153,0                                      |
| C5.2 célula 1:10   | 5       | 17.564,0                      | 175,6                                      |
| HB4a célula 1:10   | 6       | 15.691,5                      | 156,9                                      |

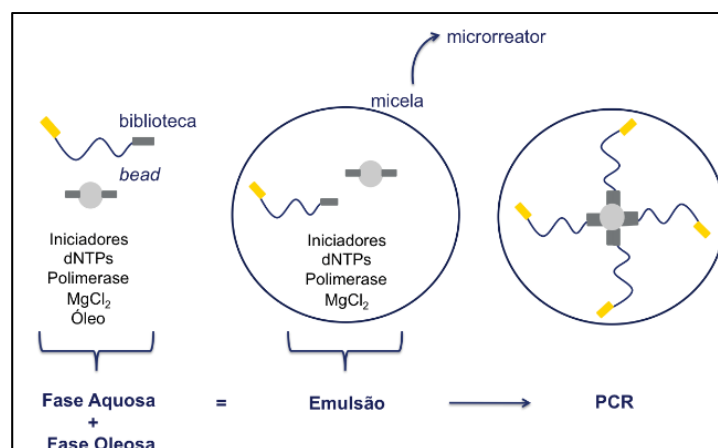
**Legenda:** O *barcode* é uma sequência de 10pb específica para cada amostra, de forma que elas podem ser misturadas para o sequenciamento, e os resultados separados posteriormente por análise bioinformática.

**Quadro 3** - Concentração molar e fator de diluição das bibliotecas construídas a partir de RNA obtidos das EVs derivadas do plasma de pacientes.

| Grupo         | Biblioteca    | Barcode | Chip    | Molaridade da biblioteca (pM) | Fator de diluição da biblioteca para 11pM |
|---------------|---------------|---------|---------|-------------------------------|---|
| NO            | Mama 19 1:100 | 1       | 1       | 3.221,0                       | 292,8                                     |
|               | Mama 22 1:100 | 2       | 1       | 3.256,1                       | 296,0                                     |
|               | Mama 30 1:10  | 3       | 1       | 1.668,4                       | 151,7                                     |
|               | Mama 32 1:100 | 4       | 1       | 5.673,3                       | 515,8                                     |
|               | Mama 33 1:100 | 5       | 1       | 3.362,3                       | 305,7                                     |
|               | Mama 44 1:100 | 6       | 1       | 4.000,0                       | 363,6                                     |
|               | Mama 26 1:100 | 7       | 1       | 4.277,2                       | 388,8                                     |
|               | Mama 34 1:100 | 8       | 1       | 5.658,1                       | 514,4                                     |
|               | Mama 36 1:100 | 9       | 1       | 1.848,8                       | 168,1                                     |
| N             | Mama 31 1:100 | 14      | 1       | 8.895,6                       | 808,7                                     |
|               | Mama 18 1:100 | 10      | 1       | 4.315,6                       | 392,3                                     |
|               | Mama 25 1:100 | 11      | 1       | 1.497,7                       | 136,2                                     |
|               | Mama 38 1:10  | 12      | 1       | 8.528,7                       | 775,3                                     |
|               | Mama 6 1:100  | 15      | 1       | 3.081,4                       | 280,1                                     |
|               | Mama 24 1:10  | 16      | 2       | 6.129,9                       | 557,3                                     |
|               | Mama 28 1:100 | 1       | 2       | 1.078,0                       | 98,0                                      |
|               | Mama 29 1:10  | 2       | 2       | 10.734,5                      | 975,9                                     |
|               | Mama 41       | 3       | 2       | 7.463,4                       | 678,5                                     |
| M             | Mama 42 1:100 | 4       | 2       | 3.091,8                       | 281,1                                     |
|               | Mama 48 1:100 | 5       | 2       | 3.397,1                       | 308,8                                     |
|               | Mama 21 1:100 | 6       | 2       | 4.109,0                       | 373,5                                     |
|               | Mama 37 1:100 | 7       | 2       | 4.012,6                       | 364,8                                     |
|               | Mama 46 1:100 | 8       | 2       | 6.506,0                       | 591,5                                     |
|               | Mama 50 1:100 | 13      | 1       | 6.987,9                       | 635,3                                     |
|               | Mama 51 1:100 | 9       | 2       | 1.146,7                       | 104,2                                     |
|               | Mama 52 1:100 | 10      | 2       | 4.430,2                       | 402,7                                     |
|               | Mama 53 1:100 | 11      | 2       | 3.230,7                       | 293,7                                     |
| Mama 54 1:100 | 12            | 2       | 3.278,5 | 298,0                         |   |
| Mama 55 1:100 | 13            | 2       | 4.141,1 | 376,5                         |   |
| Mama 56 1:100 | 14            | 2       | 7.022,8 | 638,4                         |   |

**Legenda:** O *barcode* é uma sequência de 10pb específica para cada amostra, de forma que elas podem ser misturadas para o sequenciamento, e os resultados separados posteriormente por análise bioinformática. Como só existem 16 *barcodes* para RNA, as amostras foram analisadas em corridas separadas (indicadas pela coluna *chip*).

A PCR em emulsão tem como objetivo a amplificação clonal da biblioteca em *beads*, que posteriormente são depositadas em poços do *chip* de sequenciamento. Conforme ilustrado na **Figura 5**, a reação ocorre em um micro-reator, ou micela, formada por moléculas de água em meio oleoso. As condições de diluição devem favorecer a distribuição em cada *bead* de apenas uma molécula de DNA da biblioteca, que será então amplificada por PCR. A ciclagem é feita no equipamento *OneTouch2*, que possui também uma centrífuga que ao final da reação permite a separação das *beads* do óleo.



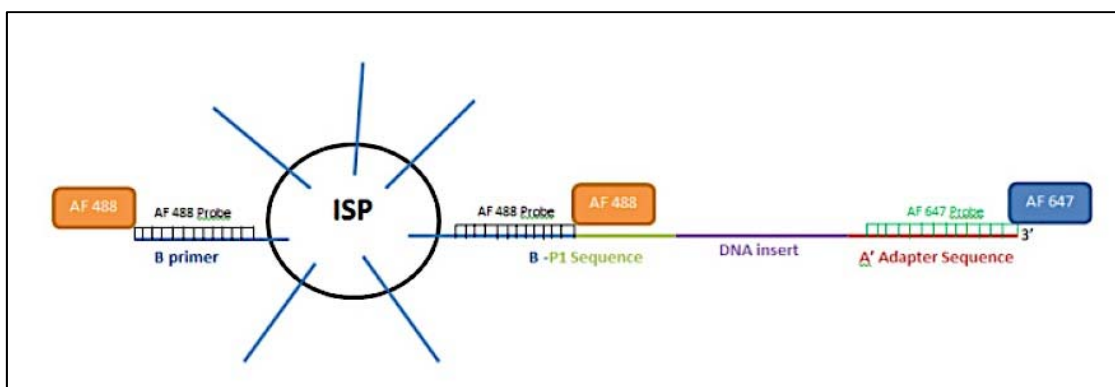
**Figura 5** - Desenho esquemático da PCR em emulsão, onde ocorre a amplificação clonal das moléculas de DNA nas *beads*.

### 3.13.3 Enriquecimento

Após a PCR em emulsão, é feito um enriquecimento para *beads* contendo moléculas de DNA amplificado das bibliotecas, permitindo eliminar *beads* vazias, onde não houve incorporação de moléculas. Este enriquecimento é possível pois os adaptadores utilizados na construção das bibliotecas são biotinizados, permitindo que apenas as *beads* contendo moléculas de biblioteca sejam capturadas com estreptavidina acoplada às *beads* magnéticas.

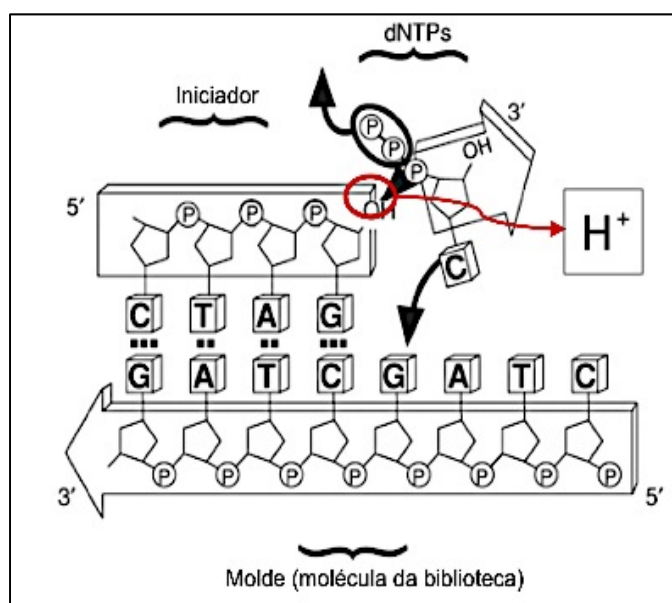
### 3.13.4 Avaliação da eficiência da PCR em emulsão

Um controle de qualidade desta etapa é feito antes da etapa final de sequenciamento, para assegurar o prosseguimento apenas dos casos onde a PCR em emulsão tenha sido eficiente. Esta etapa é possível pois os adaptadores possuem sítios de ligação de sondas contendo fluoróforos específicos. Conforme ilustrado na **Figura 6**, a sonda que se liga ao adaptador A (que está na ponta livre da molécula de biblioteca) fluoresce a 647nm, enquanto que a sonda que se liga ao adaptador P1 (que está na extremidade da molécula ligada à *bead*), e ao oligo B que é complementar ao adaptador P1 e está presente em toda a superfície da *bead*, fluoresce a 488nm. A leitura das fluorescências das duas sondas é feita no equipamento Qubit 2.0 (Life Technologies, California, EUA) e a relação entre ambas permite a determinação da quantidade relativa de *beads* recobertas por moléculas das bibliotecas, versus *beads* vazias. O fabricante determina que a quantidade ideal de *beads* contendo biblioteca esteja entre 10 e 25%. Quantidades superiores a isso podem indicar alta frequência de *beads* policlonais, onde mais de uma molécula diferente de DNA foi incorporada em cada uma das *beads*. *Beads* policlonais geram sinais conflitantes no sequenciamento, e devem ser evitadas.



### 3.13.5 Sequenciamento

As amostras foram depositadas em *chips* semicondutores P1 para sequenciamento no equipamento *Ion Proton* (Life Technologies, California, EUA). O *chip* é formado por poços, cujo tamanho permite a deposição de uma única *bead* por poço. A quantidade de poços em cada *chip* determina a quantidade máxima de seqüências que se é possível obter, que é em torno de 60-80 milhões para este *chip* (utilizado em todos os sequenciamentos descritos nesta tese). Para a determinação das seqüências, o sequenciador acrescenta sucessivamente na amostra cada um dos 4 nucleotídeos individualmente. Conforme ilustrado na **Figura 7**, se o nucleotídeo oferecido é complementar à seqüência da molécula de DNA molde, a sua incorporação é feita pela enzima DNA polimerase, havendo a liberação de um íon de hidrogênio que muda o pH. O *chip* detecta o sinal químico da mudança de pH e o converte em sinal digital que é interpretado como uma ou mais bases nitrogenadas.



**Figura 7** - Desenho esquemático da química de sequenciamento da plataforma Ion Proton. O sequenciador detecta a mudança de pH gerada pela liberação de um íon de hidrogênio quando ocorre a incorporação pela polimerase de um nucleotídeo, ou dNTP, complementar à fita molde.

### 3.13.6 Análises de Bioinformática e Estatística

As análises aqui apresentadas foram realizadas com o auxílio do grupo de Bioinformática do AC Camargo Cancer Center. O *pipeline* estabelecido incluiu um filtro inicial pelo programa *FastQ Screen* ([http://www.bioinformatics.babraham.ac.uk/projects/fastq\\_screen/](http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/)), para a remoção de RNA ribossomal (rRNA), RNA transportador (tRNA), sequências de elementos repetitivos abundantes no genoma humano (satélites, ALU, LINE) e contaminantes de outras espécies (bactérias, fungos, vírus, *C. elegans*, etc.). As sequências com tamanho entre 12 e 25 bases foram mapeadas com o programa *miRDeep2* (MACKOWIAK 2011) para a anotação dos miRNAs humanos descritos no banco de dados *miRBase v20* (KOZOMARA e GRIFFITHS-JONES 2014). As demais sequências, assim como aquelas que não foram anotadas no passo anterior, foram alinhadas no genoma humano (montagem Hg19) (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/>) utilizando o software de mapeamento *TopHat2* (KIM et al. 2013a) seguido de *Bowtie2* (LANGMEAD e SALZBERG 2012). Para a anotação dos transcritos foi utilizado a base de dados do *Ensembl* - Homo\_sapiens.GRCh37.73 (<http://grch37.ensembl.org/info/data/ftp/index.html>). A abundância relativa dos transcritos foi calculada utilizando FPKM (do Inglês, *Fragments Per Kilobase of transcript per Million mapped reads*) utilizando o software *Cufflinks* (TRAPNELL et al. 2010). Para gerar o número total de reads por gene foi utilizado o software HTSeq (ANDERS et al. 2014). A identificação dos transcritos com expressão diferencial foi realizada com o pacote estatístico *DESeq2* (LOVE et al. 2014) do *Bioconductor* (<http://www.bioconductor.org/>). Foram realizadas comparações da expressão de cada miRNA entre os três grupos de pacientes, dois a dois (N vs N0, M vs N, M vs N0) e foram considerados estatisticamente significativos aqueles miRNAs com p-valor menor ou igual a 0,01. Os *heat maps* foram gerados com o pacote *gplots* (<http://cran.fhcrc.org/web/packages/gplots/index.html>). A predição de potenciais novos miRNAs foi feita com o programa *miRDeep2* (MACKOWIAK 2011), e só foram considerados aqueles candidatos com

FDR=0. A identificação de RNAs circulares foi feita utilizando *scripts* desenvolvidos por MEMCZAK et al. (2013) (<http://circbase.org/cgi-bin/downloads.cgi>).

### **3.14 VALIDAÇÃO DOS miRNAs DIFERENCIALMENTE REPRESENTADOS POR PCR QUANTITATIVO EM TEMPO REAL (qRT-PCR)**

Os miRNAs com expressão diferencial nas EVs derivadas dos três grupos de pacientes são potenciais candidatos a biomarcadores de progressão tumoral, podendo estar envolvidos com a invasão linfonodal ou o surgimento de lesão metastática à distância. Neste sentido, todos os transcritos identificados como diferencialmente representados são potenciais biomarcadores de interesse. Focamos inicialmente nos miRNAs por eles serem importantes reguladores de expressão gênica, e por já haverem diversos trabalhos na literatura demonstrando que EVs são enriquecidas em miRNAs que podem ser transferidos entre células (CHENG et al. 2014; MONTECALVO et al. 2012; MORELLO et al. 2013; VALADI et al. 2007). Para a confirmação da expressão diferencial dos miRNAs utilizamos ensaios individuais para cada um dos miRNAs selecionados, que foram então avaliados por qRT-PCR utilizando as mesmas amostras utilizadas no sequenciamento. Na ausência de controles endógenos descritos para EVs, utilizamos na normalização ensaios específicos para 5 miRNAs (miR-26a-5p, miR-30c-5p, let-7g-5p, miR-101-3p, e miR-20a-5p) que mostraram expressão homogênea entre os três grupos por NGS, e que foram considerados como os mais estáveis segundo análise no programa *geNorm* (VANDESOMPELE et al. 2002).

#### **3.14.1 Síntese do cDNA por transcrição reversa**

Após a extração do RNA (*Total RNA Purification Kit* - Norgen, Ontario, Canada) conforme descrito acima, obteve-se 12 $\mu$ l do total de 50 $\mu$ L para a síntese do cDNA utilizando o *miScript II RT* (Qiagen, California, EUA), em

reação de 20µl contendo 4µl de 5X *miScript HiSpec Buffer*, 2µl de 10X *miScript Nucleics Mix* e 2µl de *miScript Reverse Transcriptase Mix*, que foi incubada a 37°C por 60 minutos e 95°C por 5 minutos.

### 3.14.2 Pré-amplificação do cDNA

Como a quantidade de RNA inicial derivada das EVs das amostras de pacientes era limitante, realizamos um passo de pré-amplificação do cDNA utilizando um *mix* equimolar contendo todos os 16 ensaios de miRNAs (11 candidatos e 5 endógenos, listados na **Quadro 4**) e o *miScript PreAMP PCR Kit* (Qiagen, California, EUA). Para a pré-amplificação foi preparada para cada amostra uma reação de 25µl contendo 5µl de 5X *miScript PreAMP Buffer*, 2µl de *HotStarTaq DNA Polymerase*, 5µl de mix customizado com os 16 ensaios de interesse, 7µl de água livre de nucleases, 1µl de *miScript PreAMP Universal Primer* e 5µl de cDNA diluído 1:5. As reações foram incubadas no termociclador a 95°C por 15 minutos para ativação da polimerase, seguido de 12 ciclos de: desnaturação a 94°C por 30 segundos e anelamento/extensão a 60°C por 3 minutos.

**Quadro 4** - Ensaios de qRT-PCR usados na avaliação dos miRNAs.

| miRNA           | Nº de catálogo Qiagen | Sequência                 |
|-----------------|-----------------------|---------------------------|
| hsa-miR-130a-3p | MS00003444            | 5'CAGUGCAAUGUAAAAGGGCAU   |
| hsa-miR-128-3p  | MS00008582            | 5'UCACAGUGAACCGGUCUCUUU   |
| hsa-miR-223-3p  | MS00003871            | 5'UGUCAGUUUGUCAAAUACCCCA  |
| hsa-miR-142-3p  | MS00031451            | 5'UGUAGUGUUUCCUACUUUAUGGA |
| hsa-miR-144-5p  | MS00008701            | 5'GGAUAUCAUCAUAUACUGUAAG  |
| hsa-miR-191-5p  | MS00003682            | 5'CAACGGAAUCCCAAAGCAGCUG  |
| hsa-miR-23a-3p  | MS00031633            | 5'AUCACAUUGCCAGGGAUUUCC   |
| hsa-miR-23b-3p  | MS00031647            | 5'AUCACAUUGCCAGGGAUUACC   |
| hsa-miR-301a-3p | MS00009317            | 5'CAGUGCAAUAGUAUUGUCAAGC  |
| hsa-miR-584-5p  | MS00037198            | 5'UUAUGGUUUGCCUGGGACUGAG  |
| hsa-miR-92a-3p  | MS00006594            | 5'UAUUGCACUUGUCCCGGCCUGU  |
| hsa-miR-26a-5p  | MS00029239            | 5'UUCAAGUAAUCCAGGAUAGGCU  |
| hsa-miR-30c-5p  | MS00009366            | 5'UGUAAACAUCUACACUCUCAGC  |
| hsa-let-7g-5p   | MS00008337            | 5'UGAGGUAGUAGUUUGUACAGUU  |
| hsa-miR-101-3p  | MS00008372            | 5'UACAGUACUGUGAUAACUGAA   |
| hsa-miR-20a-5p  | MS00003199            | 5'UAAAGUGCUUAUAGUGCAGGUAG |

### 3.14.3 Avaliação da expressão dos miRNAs por qRT-PCR

A avaliação da expressão dos miRNAs foi realizada por PCR quantitativa em tempo real (qRT-PCR) no equipamento *7500 Fast Real-Time PCR System* (Life Technologies, California, EUA) utilizando o *miScript SYBR Green PCR Kit* (Qiagen, California, EUA), em reação de 10 $\mu$ l contendo 5 $\mu$ l de *2X QuantiTect SYBR Green PCR Master Mix*, 1 $\mu$ l de *10X miScript Universal Primer*, 1 $\mu$ l de *10X miScript Primer Assay* (para cada um dos ensaios de interesse), 2 $\mu$ l de água livre de nucleases, e 1 $\mu$ l de cDNA pré-amplificado diluído 1:20. As condições de ciclagem foram: 95°C por 15 minutos para ativação da polimerase e 40 ciclos de: desnaturação a 94°C por 15 segundos, anelamento a 55°C por 30 segundos e extensão a 70°C por 30 segundos. As curvas de desnaturação foram geradas para a verificação da especificidade da reação por incubação a: 95°C por 15 segundos, 60°C por 1 minuto, 95°C por 30 segundos, e 60°C por 15 segundos. Todas as reações foram realizadas em triplicata. A análise foi realizada pelo método de  $\Delta\Delta C_t$ , considerando a média dos dois endógenos mais estáveis (miR-26a-5p e miR-30c-5p) e a normalização foi feita adotando o grupo N0 como referência.

## 3.15 VALIDAÇÃO DOS RNAs CIRCULARES (circRNAs)

### 3.15.1 Síntese do cDNA por transcrição reversa

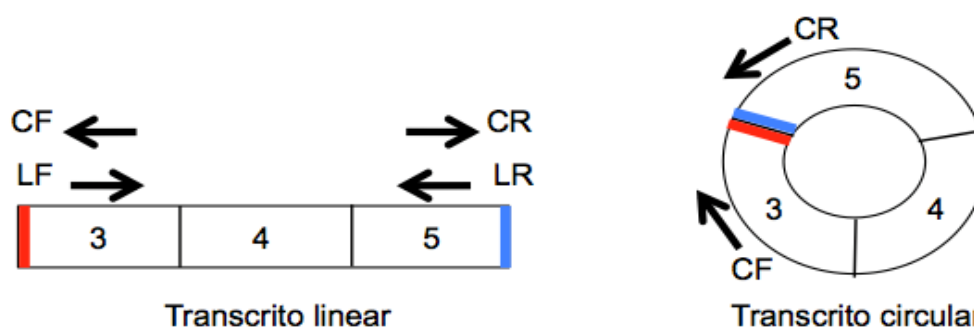
Após a extração do RNA (*Total RNA Purification Kit* - Norgen, Ontario, Canada) conforme descrito acima, utilizamos 8 $\mu$ l (pacientes) ou 1 $\mu$ l (linhagem) do total de 50 $\mu$ L para a síntese do cDNA utilizando o *SuperScript III First-Strand Synthesis System for RT-PCR* (Invitrogen, California, EUA). O cDNA foi feito em reação de 20 $\mu$ l contendo 1 $\mu$ l 50ng/ $\mu$ l *Random hexamers*, 1 $\mu$ l *10mM dNTP mix*, 7 $\mu$ l *água-DEPC* (somente linhagem), 2 $\mu$ l *10X RT buffer*, 4 $\mu$ l *25mM MgCl<sub>2</sub>*, 2 $\mu$ l *0.1M DTT*, 1 $\mu$ l *RNaseOUT (40U/ $\mu$ l)*, 1 $\mu$ l *SuperScript III RT (200U/ $\mu$ l)* que foi incubada a 65°C por 5 minutos, 4°C por 1 minuto, 25°C por 10 minutos, 50°C por 50 minutos, 85°C por 5 minutos, 4°C por 1 minuto.



Foi então adicionado 1 $\mu$ l de RNase H à reação, que foi incubada a 37°C por 20 minutos.

### 3.15.2 Desenho dos iniciadores e PCR

Os pares de iniciadores foram desenhados de forma a amplificar a isoforma linear (L) ou circular (C) do transcrito *UBXN7*, conforme ilustrado na **Figura 8**. As sequências dos iniciadores estão listadas no **Quadro 5**.



**Figura 8** – Desenho esquemático dos pares de iniciadores utilizados para amplificar as isoformas linear (L) e circular (C) de *UBXN7*. Os pares de iniciadores estão representados como setas.

**Quadro 5** - Sequências dos iniciadores utilizados para amplificação do transcrito *UBXN7* isoforma linear (L) e circular (C).

| Iniciador | Sequência (5' – 3')  |
|-----------|----------------------|
| CF        | GAATTGGGGCACGAACTTC  |
| CR        | GCCACCCATTGATTTGATGC |
| LF        | CAAGAAGAAGTTCGTGCC   |
| LR        | GCATCAAATCAATGGGTGGC |

A PCR foi realizada em volume final de 15 $\mu$ l contendo: 7.5 $\mu$ l de 2X *GoTaq Green Master Mix* (Promega, Wisconsin, EUA), 1 $\mu$ l de iniciador F a 2,5 $\mu$ mol/ $\mu$ l, 1 $\mu$ l de iniciador R a 2,5 $\mu$ mol/ $\mu$ l, 4,5 $\mu$ l de água livre de nucleases, e 1 $\mu$ l de cDNA (sem diluir para pacientes, e 1:10 para linhagem). As condições de ciclagem foram: 95°C por 2 minutos para ativação da

polimerase e 40 ciclos (35 ciclos para linhagem) de: desnaturação a 95°C por 1 minuto, anelamento a 58°C por 30 segundos e extensão a 72°C por 30 segundos, seguido de uma extensão final a 72°C por 5 minutos. Posteriormente 10µl de cada produto de PCR foram avaliados em géis de poli-acrilamida 8% corados com prata (SANGUINETTI et al. 1994).

### 3.15.3 Sequenciamento pelo método de Sanger

Os produtos de PCR foram purificados por degradação e desfosforilação do excesso de iniciadores e dNTPs em reação de 10µl contendo 5µl de *amplicon*, 0,3µl de enzima Exonuclease I (Affymetrix, California, EUA) 1µl de enzima *Shrimp Alkaline Phosphatase* (SAP) (Affymetrix, California, EUA), 1µl de *10X SAP Reaction Buffer* (Affymetrix, California, EUA), e 2,7µl de água, que foi incubada a 37°C por 30 minutos e 80°C por 15 minutos.

O sequenciamento dos *amplicons* foi realizado pelo método de terminação de cadeia (método de Sanger). A incorporação dos dideoxynucleotídeos fluorescentes foi feita em reação de 10µl contendo 2,5µl de produtos de PCR purificados, 1,6µl de iniciador F ou R a 2,5pmol/µl, 0,7µl de *BigDye Terminator v3.1 Ready Reaction Mix* (Life Technologies, California, EUA), 2µl de *5X Sequencing Buffer* (Life Technologies, California, EUA), e 3,2µl de água livre de nucleases, que foi incubada a 95 °C por 2 minutos para ativação da polimerase e 40 ciclos de: desnaturação a 95°C por 18 segundos, anelamento a 55°C por 18 segundos e extensão a 60°C por 4 minutos.

Após a reação de sequenciamento, os produtos foram precipitados com 1µl de acetato de sódio 3M pH 5,2, 1µl de EDTA 125mM pH 8,0 e 25µl de etanol 100%, e incubação a temperatura ambiente por 15 minutos. Após centrifugação a 4000rpm por 30 minutos a 4°C (centrífuga Eppendorf 5810R, rotor A-4-62), o sobrenadante foi descartado (placa foi centrifugada invertida a 650 rpm por 1 minuto), e o *pellet* foi lavado com 35µl de etanol 70% gelado. Após centrifugação a 4000 rpm por 20 minutos, o sobrenadante foi descartado (placa foi centrifugada invertida a 650 rpm por 4 minutos), e o *pellet* foi seco a 95°C por 20 minutos. Após a secagem, o *pellet* foi

ressuspendido com 13 $\mu$ l de formamida Hi-Di (Life Technologies, California, EUA) e desnaturado a 95°C por 3 minutos, seguido de 3 minutos no gelo.

A detecção dos produtos marcados com fluorescência foi realizada por eletroforese em sequenciador capilar automático *3130xl Genetic Analyzer* (Life Technologies, California, EUA). As sequências geradas foram verificadas no programa *CLC Genomics Workbench 6.0.3*, após alinhamento contra a sequência referência do transcrito *UBXN7* (NCBI NM\_015562).

## 4 RESULTADOS E DISCUSSÃO

### 4.1 CARACTERIZAÇÃO DE EVs DERIVADAS DE HB4a E C5.2

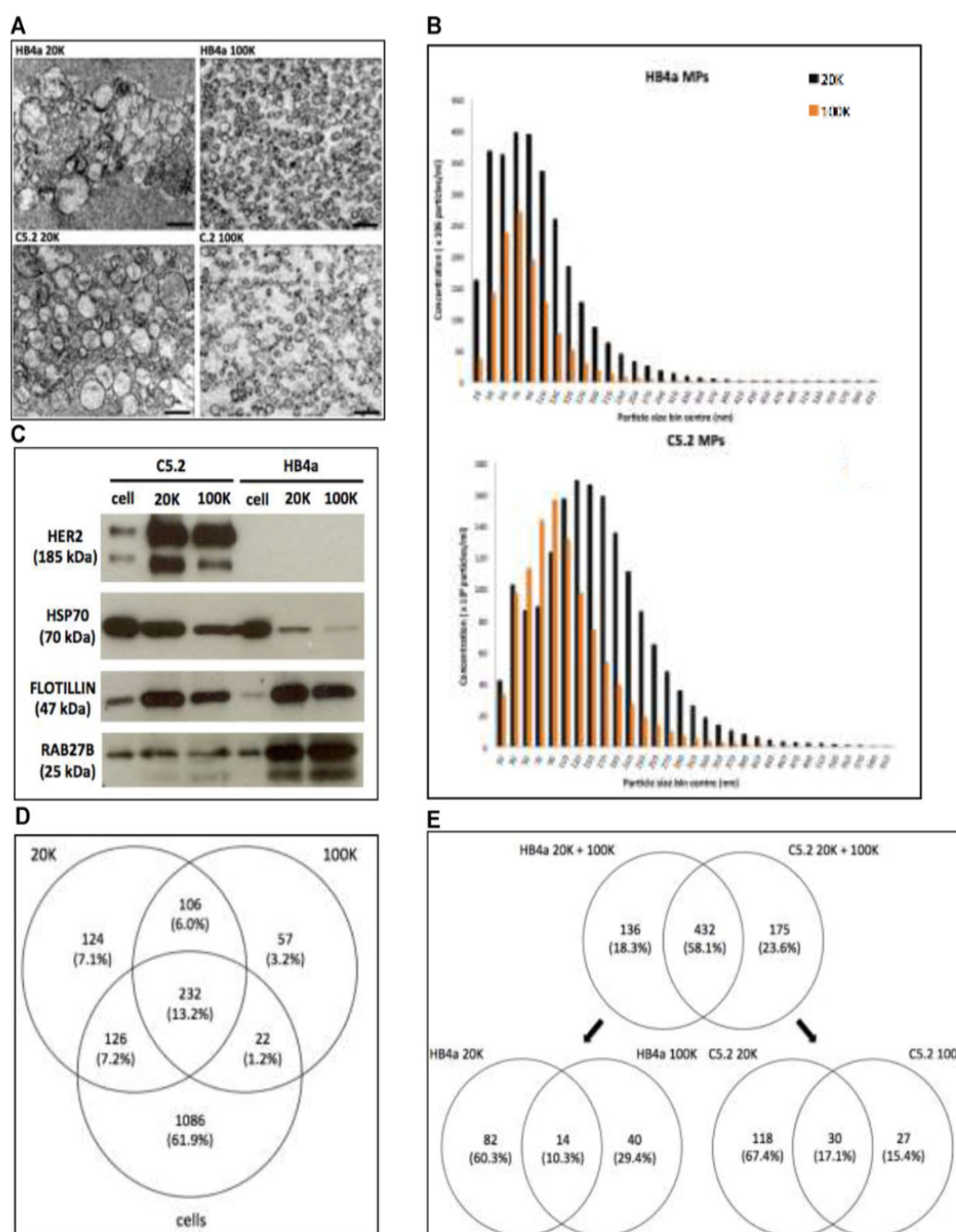
#### 4.1.1 Caracterização do proteoma de EVs derivadas de HB4a e C5.2

A caracterização do proteoma de EVs derivadas das linhagens celulares HB4a e C5.2 foi realizada durante a fase inicial de avaliação dos protocolos de isolamento e caracterização de EVs. Por dispormos de quantidade limitada de plasma de pacientes, optamos por realizar as padronizações iniciais com meio de cultura celular condicionado, que é de fácil obtenção. O modelo celular escolhido permite avaliar os efeitos da superexpressão de um único oncogene, usando as linhagens celulares HB4a – derivada de tecido luminal epitelial de mama normal humana imortalizada com o vírus SV-40 e C5.2 – um clone de HB4a transformado que superexpressa ERBB2/HER2 (HARRIS et al. 1999; STAMPS et al. 1994). Os resultados iniciais se mostraram promissores e por fim se tornaram um projeto paralelo de caracterização completa das EVs secretadas por estas duas linhagens celulares, que serão apresentados nesta tese em conjunto com os dados das amostras de pacientes. Os resultados das análises proteômicas, feitos em colaboração com o Dr. Daniel Martins-de-Souza, foram publicados na revista *Proteomics* (**Anexo 4**) onde demonstramos que a superexpressão do oncogene ERBB2/HER2 na linhagem C5.2 altera o perfil proteômico das EVs (AMORIM et al. 2014).

Isolamos por ultracentrifugação duas populações de EVs (20K e 100K) a partir do meio de cultura celular condicionado. A microscopia eletrônica (**Figura 9A**) revelou que ambas populações de EVs são morfologicamente distintas; a fração 20K é enriquecida em vesículas maiores e mais heterogêneas enquanto as EVs de 100K demonstram a morfologia *cup-shaped* típica de exossomos (THÉRY et al. 2006). Estes achados foram confirmados por *nanoparticle tracking analysis* (**Figura 9B**) que mostrou que as EVs 20K são maiores que as EVs 100K para ambas linhagens (teste Wilcoxon rank: W: 627, p-valor: 0,04 para HB4a; W: 618, p-valor: 0,05 para C5.2).

Análises por *Western Blot* (**Figura 9C**) usando marcadores associados a vesículas demonstraram a presença de Flotilina nas duas populações de ambas linhagens, enquanto HSP70 pareceu ser mais abundante em C5.2 (células e ambos sets de vesículas), quando comparada com HB4a. Por outro lado, RAB27B estava superexpressa em ambas 20K e 100K subpopulações de HB4a em comparação com C5.2. Como esperado, observamos alta expressão de HER2 em C5.2 e ausente ou muito pouco expressa em HB4a e as suas EVs derivadas, achado também confirmado por qRT-PCR (dado não mostrado) e por proteômica. *Western Blot* (**Figura 9C**) usando a mesma massa de proteína mostrou HER2 muito mais abundante nas EVs de C5.2, em relação às células C5.2.

As análises proteômicas (**Figura 9D**) permitiram a identificação de 1.753 proteínas únicas; 83,6% (1466) foram encontradas nas células, 33,5% (588) foram identificadas em EVs 20K e 23,8% (417) foram encontradas nas EVs 100K. O componente exclusivo das células era composto por 1.086 (61,9%) proteínas, enquanto 287 (16,4%) foram encontradas somente em EVs e 232 (13,2%) estavam presentes em ambas populações de EVs e nas células. Os processos biológicos preditos e as localizações subcelulares de proteínas exclusivas de cada população de EVs de ambas linhagens encontram-se na **Figura 10**. Das 667 proteínas únicas identificadas por pelo menos dois peptídeos nos dois subsets de EVs de ambas linhagens, 650 (97%) já foram previamente descritas em EVs na EVpedia (KIM et al. 2013b), sugerindo o sucesso do nosso protocolo de isolamento das EVs. A maioria das proteínas encontradas e não descritas na EVpedia (12/17) eram exclusivas das EVs 20K, achado que provavelmente reflete a maior diversidade de partículas desta fração, assim como a exclusão de vesículas maiores pelos protocolos de isolamento de exossomos baseados em filtração (RAPOSO e STOORVOGEL 2013).



Fonte: AMORIM et al. (2014)

**Figura 9** - Caracterização das EVs 20K e 100K. **(A)** Microscopia eletrônica de transmissão mostrando a heterogeneidade das populações de EVs: as EVs 20K são maiores e mais heterogêneas que as EVs 100K, que demonstram a morfologia *cup-shaped* típica de exossomos. **(B)** Análise por Nanosight mostrando o deslocamento de tamanho entre as frações 20K (barras pretas) e 100K (barras laranjas) de ambas linhagens. **(C)** *Western Blots* mostrando a superexpressão de HER2 na linhagem C5.2 e subsets de EVs, assim como os marcadores de vesículas Flotillin, RAB27B, e HSP-70. **(D)** Diagrama de Venn mostrando o *overlap* de proteínas derivadas das linhagens e suas frações de EVs. **(E)** Diagramas de Venn demonstrando a distribuição de proteínas exclusivas às EVs de cada linhagem.

Para minimizar o stress celular conhecidamente causado pela carência de soro fetal bovino (FBS) (WITWER et al. 2013) e focar o nosso estudo nos efeitos induzidos por HER2, as células HB4a e C5.2 foram crescidas em meio contendo FBS previamente depletado de EVs. Porém, como proteínas derivadas do FBS podem potencialmente ter sido co-purificadas junto com as EVs nós colocamos um *flag* nas proteínas mais abundantes do FBS (BUNKENBORG et al. 2010) que também foram identificadas aqui. Queratinas, abundantemente encontradas no FBS, também foram encontradas em ambos subsets de EVs. Interessantemente, alguns estudos apontam para o papel funcional destas proteínas no remodelamento de citoesqueleto, invasão celular e metástase (OBERMAJER et al. 2009), assim é possível que a presença de queratinas em EVs indique mais que contaminação da amostra.

A análise proteômica de HB4a e C5.2 indicou 1.466 proteínas compartilhadas entre as duas linhagens (**Figura 9D**). Interessantemente, 74% destas proteínas não foram encontradas nas EVs. Por outro lado, das 667 proteínas consistentemente encontradas em todas as EVs, 43% não foram detectadas nas células secretoras. Ambos achados reforçam o conceito que EVs servem como shuttles de sets específicos de proteínas (NOLTE-'T HOEN et al. 2012; SKOG et al. 2008; VALADI et al. 2007).

Nós comparamos os proteomas de EVs derivadas das duas linhagens para investigar proteínas direta ou indiretamente afetadas pelo superexpressão de HER2. Um total de 136 (18,3%) de proteínas foram exclusivamente encontradas nas EVs derivadas de HB4a, enquanto 175 (23,6%) proteínas foram encontradas somente nas EVs derivadas de C5.2, considerando ambas populações de EVs (**Figura 9E**). Alterações proteômicas direta ou indiretamente correlacionadas com a superexpressão de HER2 estavam presentes em ambas populações de EVs, e suas atividades biológicas estavam principalmente relacionadas a processos como migração celular, transformação maligna, angiogênese, metástase e remodelamento de citoesqueleto. Notavelmente, apesar do agrupamento das 10 top proteínas das subpopulações 20K e 100K de EVs nos mesmos

processos biológicos, elas não se sobrepõem (**Tabela 2**). Este achado reforça que o *shuttling* de proteínas não é arbitrariamente definido e sugere que as proteínas carregadas pelos sets distintos de vesículas podem ter papéis complementares ao menos para algumas vias biológicas.

Proteínas capazes de induzir transformação maligna foram encontradas super-representadas em EVs da linhagem HER2+ (C5.2), incluindo as *top* proteínas das subpopulações 20K e 100K, ambas aumentadas >90X: cofilina-1 e CD44 (**Tabela 2**), as duas envolvidas em adesão celular, motilidade e invasão, e remodelamento de citoesqueleto. HER2 regula a via de cofilina por ativar PLC $\gamma$  para hidrolisar PIP2 que mantém cofilina presa a membrana celular. Cofilina livre leva ao remodelamento de actina, essencial para a motilidade e invasão celular (WANG et al. 2007). A inibição de cofilina foi descrita como envolvida na formação de EVs, quando a proteína quinase associada a Rho (ROCK) leva a fosforilação da quinase LIM (LIMK), que por fim fosforila cofilina que é então impedida de clivar actina, resultando em fibras de actina estendidas (LI et al. 2012). A ligação de ácido hialurônico a CD44 modula a interação CD44-HER2 (BOURGUIGNON 1997), que afeta o crescimento, a sobrevivência e a diferenciação celular (GÜNTHERT et al. 1991; PONTA et al. 2003). CD44 clivada por ADAM10 também foi reportada em exossomos, como uma via de secreção da forma solúvel deste receptor (STOECK et al. 2006). Também notável foi a superexpressão de fibronectina-1 no subset 20K das células C5.2, já que foi demonstrado que quando essa proteína é carregada em EVs derivadas de células tumorais ela é capaz de transformar células não-tumorais (ANTONYAK et al. 2011). Devemos mencionar que fibronectina-1 é abundante em FBS (BUNKENBORG et al. 2010). Porém a sua descrição prévia em EVs de células carenciadas (ANTONYAK et al. 2011) e o seu papel no remodelamento de citoesqueleto, migração celular e metástase (RUOSLAHTI 1984) podem ser indicativos do seu papel biológico em EVs e não uma simples contaminação do soro. Dentre as *top* 10 proteínas superexpressas em EVs de C5.2 (**Tabela 2**), algumas foram descritas por interagirem entre si, como a proteína 14-3-3 zeta/delta e



cofilina-1 (BIRKENFELD et al. 2003), ou CAP-1 e cofilina-1 (NORMOYLE e BRIEHER 2012), ambas associações sendo relacionadas à reciclagem e reorganização de actina.

**Tabela 2** - Dez proteínas mais superexpressas nas EVs 20K e 100K das linhagens C5.2 (cinza claro) e HB4a (cinza escuro).

| Accession                 | Ratio C5.2/Hb4a | Number of Peptides | Description  | EVs subset |
|---------------------------|-----------------|--------------------|--|------------|
| COF1_HUMAN <sup>a)</sup>  | 95.96           | 3                  | Cofilin-1  | 20 K       |
| EF1A1_HUMAN               | 33.63           | 6                  | Elongation factor 1 alpha 1                          |            |
| FINC_HUMAN <sup>a)</sup>  | 26.83           | 71                 | Fibronectin  |            |
| EPS8_HUMAN                | 21.98           | 2                  | Epidermal growth factor receptor kinase substrate 8  |            |
| 1433Z_HUMAN <sup>a)</sup> | 20.38           | 6                  | 14-3-3 protein zeta/delta                            |            |
| RL31_HUMAN                | 19.62           | 2                  | 60S ribosomal protein L31                            |            |
| NUCL_HUMAN                | 18.83           | 4                  | Nucleolin  |            |
| THIL_HUMAN                | 18.30           | 7                  | Acetyl-CoA acetyltransferase, mitochondrial          |            |
| EF1D_HUMAN                | 16.11           | 5                  | Elongation factor 1 delta                            |            |
| CAP1_HUMAN <sup>a)</sup>  | 15.89           | 4                  | Adenylyl cyclase associated protein 1                |            |
| CD44_HUMAN                | 90.09           | 2                  | CD44 antigen   | 100 K      |
| RRAS_HUMAN                | 31.51           | 2                  | Ras-related protein R-Ras                            |            |
| IF4A3_HUMAN               | 25.81           | 2                  | Eukaryotic initiation factor 4A-III                  |            |
| AKA12_HUMAN               | 22.26           | 2                  | A-kinase anchor protein 12                           |            |
| CYBP_HUMAN                | 21.48           | 2                  | Calcyclin-binding protein                            |            |
| HIBCH_HUMAN               | 19.76           | 2                  | 3-hydroxyisobutyryl-CoA hydrolase, mitochondrial     |            |
| ERBB2_HUMAN               | 18.33           | 18                 | Receptor tyrosine-protein kinase erbB-2              |            |
| H3L_HUMAN                 | 18.14           | 2                  | Histone H3-like                                      |            |
| RL5_HUMAN                 | 16.60           | 2                  | 60S ribosomal protein L5                             |            |
| YBOX1_HUMAN               | 16.51           | 3                  | Nuclease-sensitive element-binding protein 1         |            |
| IF2A_HUMAN                | 0.50            | 2                  | Eukaryotic translation initiation factor 2 subunit 1 | 20 K       |
| NPM_HUMAN                 | 0.46            | 6                  | Nucleophosmin  |            |
| PARP1_HUMAN               | 0.45            | 2                  | Poly (ADP-ribose) polymerase 1                       |            |
| TCPG_HUMAN                | 0.44            | 11                 | T-complex protein 1 subunit gamma                    |            |
| TFR1_HUMAN                | 0.44            | 10                 | Transferrin receptor protein 1                       |            |
| K1109_HUMAN               | 0.41            | 2                  | Uncharacterized protein KIAA1109                     |            |
| RL12_HUMAN                | 0.41            | 2                  | 60S ribosomal protein L12                            |            |
| RHOA_HUMAN                | 0.40            | 2                  | Transforming protein RhoA                            |            |
| G6PD_HUMAN                | 0.40            | 7                  | Glucose-6-phosphate 1-dehydrogenase                  |            |
| SFPO_HUMAN                | 0.38            | 8                  | Splicing factor, proline- and glutamine rich         |            |
| FUMH_HUMAN                | 0.50            | 5                  | Fumarate hydratase, mitochondrial                    | 100 K      |
| AHNAK_HUMAN               | 0.48            | 5                  | Neuroblast differentiation associated protein AHNAK  |            |
| RAB35_HUMAN               | 0.46            | 4                  | Ras-related protein Rab-35                           |            |
| MOT4_HUMAN                | 0.45            | 4                  | Monocarboxylate transporter 4                        |            |
| TGM2_HUMAN                | 0.42            | 3                  | Protein-glutamine gamma-glutamyltransferase 2        |            |
| TFR1_HUMAN                | 0.42            | 5                  | Transferrin receptor protein 1                       |            |
| RL19_HUMAN                | 0.40            | 2                  | 60S ribosomal protein L19                            |            |
| RS14_HUMAN                | 0.39            | 2                  | 40S ribosomal protein S14                            |            |
| RAB3A_HUMAN               | 0.32            | 2                  | Ras-related protein Rab-3A                           |            |
| TCPB_HUMAN                | 0.31            | 7                  | T-complex protein 1 subunit beta OS = Homo sapiens   |            |

a) This protein was also found in FBS [21].

**Fonte:** AMORIM et al. (2014)

Ao avaliarmos as proteínas derivadas de HB4a e C5.2 por MetaCore, foi confirmado o enriquecimento de proteínas envolvidas com adesão celular e remodelamento de citoesqueleto, assim como proteínas agrupadas em vias distintas relacionadas ao desenvolvimento. Estas vias corresponderam a 32 de 80 vias metabólicas e de sinalização encontradas enriquecidas pelo MetaCore ( $p < 0,01$ ) (**Tabela 3**). A via mais enriquecida, LRRK2 em neurônios, foi consistente em todas as frações de EVs ( $p < 1,7e-12$ ) e pode ser sugestivo do seu envolvimento na via de biogênese de EVs. A **Tabela 3** também demonstra vias enriquecidas exclusivamente em EVs 20K ou 100K, assim como vias que aparentam ser induzidas pela superexpressão de HER2, tais como processos relacionados a invasão tumoral e migração celular, que podem ser relevante para a biologia dos tumores HER2+. Um exemplo é o enriquecimento de sinalização por *sphingosine-1-phosphate* (S1P) em ambos sets de EVs derivadas de C5.2. Interessantemente, dentre outras vias relacionadas ao câncer afetadas por S1P, ele parece regular a angiogênese e a função de células endoteliais vasculares pelo seu receptor SIP3 (HLA et al. 2000). Junto com HER2 e o seu receptor S1P4, S1P é capaz de ativar a via de ERK-1/2 na linhagem de mama MDA-MB-453, um evento que é provavelmente relacionado a progressão tumoral (LONG et al. 2010). Curiosamente, proteínas relacionadas a via de S1P3 foram encontradas somente nas EVs 20K, enquanto proteínas relacionadas a via de S1P4 foram encontradas exclusivamente nas EVs 100K, reforçando de novo a natureza complementar do sets de proteínas das EVs 20K e 100K.

Nossos achados indicam que a superexpressão de HER2 modifica o conteúdo proteico da linhagem C5.2 e de suas EVs. O conteúdo proteico alterado das EVs inclui sets de proteínas complementares, que é reafirmado pela presença de proteínas distintas da mesma via, relacionadas a fenótipos tumor-relevantes principalmente associados à invasão tumoral. A investigação destas proteínas em EVs derivadas do plasma de pacientes com tumores HER2+ pode revelar potenciais biomarcadores que podem ser úteis para diagnóstico ou prognóstico de forma não-invasiva. Deve ser ressaltado também que os níveis aumentados de HER2 vistos em C5.2 não

só são refletidos nas EVs correspondentes, mas as EVs derivadas de C5.2 carregam uma quantidade ainda mais aumentada deste oncogene. Se este achado for confirmado em tumores, as EVs podem potencialmente agir como agentes de *shuttling* de HER2, transferindo-o para sítios distantes, ou mesmo para outras células tumorais não-HER2+ — uma questão relevante dado que >20% dos pacientes com tumores de mama HER2+ demonstram heterogeneidade de HER2, com entre 5 e 50% das células tumorais tendo amplificação deste oncogene (ALLISON et al. 2011). Um excesso de HER2 em vesículas poderia também atuar sequestrando a droga trastuzumab de seu alvo original (as células tumorais) (CIRAVOLO et al. 2012), de modo que uma menor quantidade desta droga estivesse disponível para atuar nas células tumorais. Cabe notar, de modo relevante, que estes modos de atuação não são mutuamente exclusivos e sua ocorrência poderia contribuir de modo muito relevante para a malignidade destes tumores e refratariedade ao tratamento.

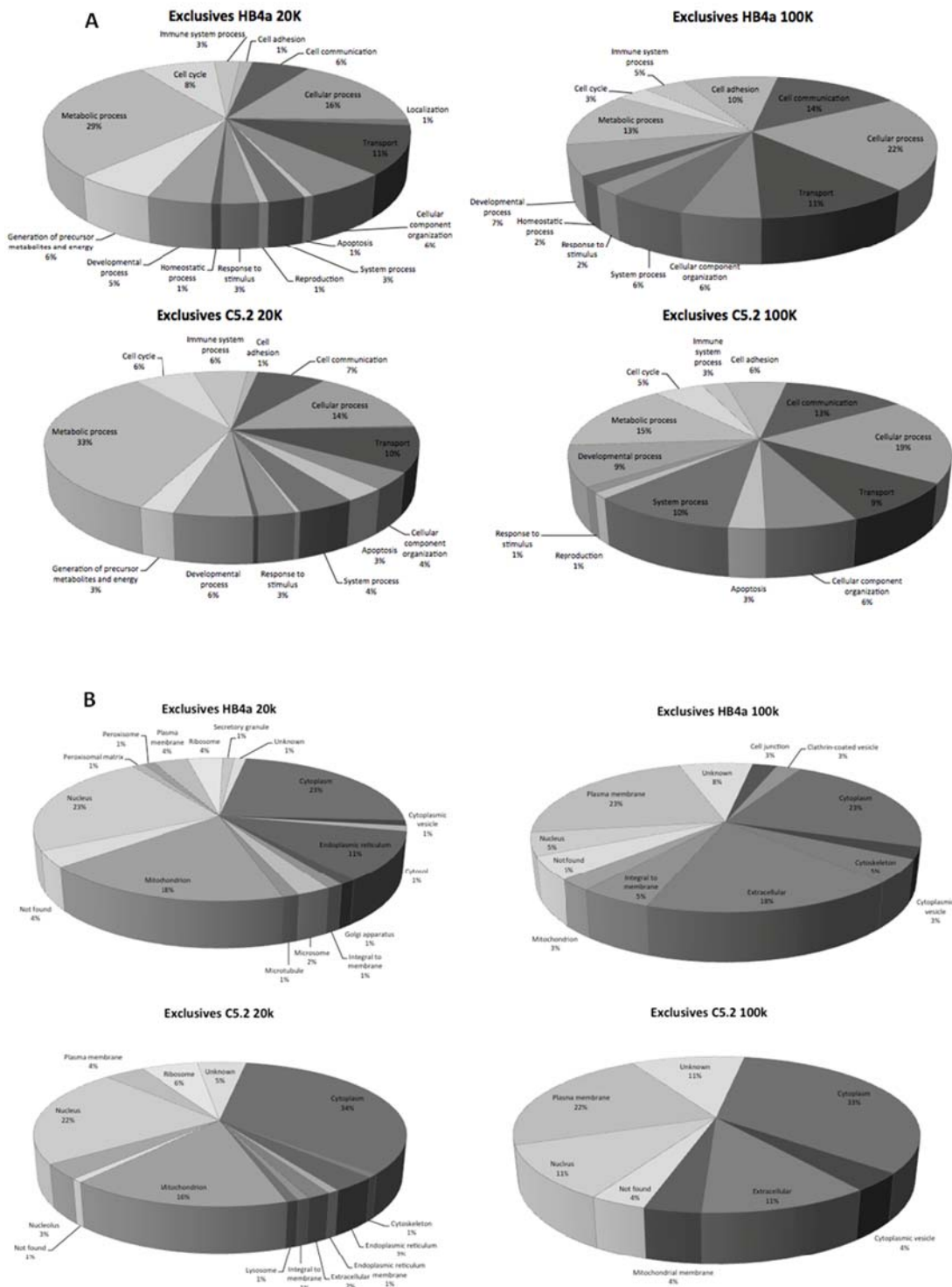
**Tabela 3** - Análise por MetaCore de enriquecimento de vias de sinalização nas EVs derivadas de HB4a e C5.2.

| Signaling and metabolic pathways  | 20 K     |        |          |        | 100 K    |        |          |        |
|---|----------|--------|----------|--------|----------|--------|----------|--------|
|   | HB4a     |        | C5.2     |        | HB4a     |        | C5.2     |        |
|   | p-value  | PP     | p-value  | PP     | p-value  | PP     | p-value  | PP     |
| Cell adhesion_Cadherin-mediated cell adhesion   | 1.61e-04 | 6/26   | 3.67e-07 | 9/26   | –        | 0/26   | –        | 0/26   |
| Cell adhesion_CHEMOKINES and adhesion   | 2.25e-05 | 13/100 | 6.36e-15 | 26/100 | –        | 0/100  | –        | 0/100  |
| Cell adhesion_ECM remodeling  | p > 0.01 | 5/52   | 2.95e-05 | 10/52  | p > 0.01 | 5/52   | 5.71e-04 | 7/52   |
| Cell adhesion_Endothelial cell contacts by non-junctional mechanisms                        | 3.44e-08 | 9/24   | 7.79e-13 | 13/24  | 4.01e-13 | 12/24  | 1.22e-11 | 11/24  |
| Cell adhesion_Ephrin signaling  | p > 0.01 | 2/45   | 8.11e-03 | 6/45   | p > 0.01 | 3/45   | 8.16e-03 | 5/45   |
| Cell adhesion_Histamine H1 receptor signaling in the interruption of cell barrier integrity | 6.10e-04 | 7/45   | 7.61e-06 | 10/45  | –        | 0/45   | –        | 0/45   |
| Cell adhesion_Integrin-mediated cell adhesion and migration                                 | 1.52e-04 | 8/48   | 2.23e-10 | 15/48  | 6.40e-06 | 9/48   | 6.09e-08 | 11/48  |
| Cell adhesion_Role of CDK5 in cell adhesion   | p > 0.01 | 2/9    | 1.04e-05 | 5/9    | 1.66e-03 | 3/9    | 6.81e-05 | 4/9    |
| Cell adhesion_Role of tetraspanins in the integrin-mediated cell adhesion                   | 2.25e-06 | 9/37   | 2.84e-12 | 15/37  | 3.30e-09 | 11/37  | 1.80e-10 | 12/37  |
| Cytoskeleton remodeling_Cytoskeleton remodeling   | 3.07e-08 | 17/102 | 7.90e-13 | 24/102 | 1.50e-07 | 15/102 | 3.40e-10 | 18/102 |
| Cytoskeleton remodeling_FAK signaling   | p > 0.01 | 6/57   | 6.75e-05 | 10/57  | –        | 0/57   | –        | 0/57   |
| Cytoskeleton remodeling_Fibronectin-binding integrins in cell motility                      | 4.33e-07 | 9/31   | 2.54e-12 | 14/31  | 1.64e-06 | 8/31   | 3.48e-10 | 11/31  |
| Cytoskeleton remodeling_Integrin outside-in signaling                                       | 1.76e-04 | 8/49   | 3.10e-10 | 15/49  | –        | 0/49   | –        | 0/49   |
| Cytoskeleton remodeling_Keratin filaments   | –        | 0/36   | –        | 0/36   | 5.50e-05 | 7/36   | 4.50e-07 | 9/36   |
| Cytoskeleton remodeling_Neurofilaments  | 1.16e-03 | 5/25   | 2.49e-07 | 9/25   | –        | 0/25   | –        | 0/25   |
| Cytoskeleton remodeling_RalA regulation pathway   | p > 0.01 | 4/30   | 1.32e-04 | 7/30   | 9.60e-03 | 4/30   | 1.54e-04 | 6/30   |
| Cytoskeleton remodeling_Regulation of actin cytoskeleton by Rho GTPases                     | 4.04e-07 | 8/23   | 5.94e-09 | 10/23  | 2.21e-06 | 7/23   | 1.16e-07 | 8/23   |
| Cytoskeleton remodeling_TGF $\beta$ WNT and cytoskeletal remodeling                         | 1.55e-05 | 14/111 | 2.97e-07 | 18/111 | 1.15e-03 | 10/111 | 1.26e-05 | 13/111 |
| Development_Beta-adrenergic receptors regulation of ERK                                     | p > 0.01 | 4/47   | 4.49e-04 | 8/47   | –        | 0/47   | –        | 0/47   |
| Development_Ligand-independent activation of ESR1 and ESR2                                  | –        | 0/45   | –        | 0/45   | p > 0.01 | 3/45   | 8.16e-03 | 5/45   |
| Development_Lipoxin inhibitory action on PDGF, EGF and LTD4 signaling                       | p > 0.01 | 3/36   | 2.61e-03 | 6/36   | –        | 0/36   | –        | 0/36   |
| Development_Osteopontin signaling in osteoclasts  | 2.73e-03 | 5/30   | 1.46e-06 | 9/30   | 1.39e-03 | 5/30   | 1.17e-06 | 8/30   |
| Development_Regulation of telomere length and cellular immortalization                      | –        | 0/35   | –        | 0/35   | p > 0.01 | 3/35   | 2.72e-03 | 5/35   |
| Development_S1P1 receptor signaling via beta-arrestin                                       | p > 0.01 | 3/34   | 1.93e-03 | 6/34   | –        | 0/34   | –        | 0/34   |
| Development_S1P1 signaling pathway  | p > 0.01 | 4/44   | 4.43e-05 | 9/44   | p > 0.01 | 3/44   | 1.32e-03 | 6/44   |
| Development_S1P3 receptor signaling pathway   | p > 0.01 | 2/43   | 6.49e-03 | 6/43   | –        | 0/43   | –        | 0/43   |
| Development_S1P4 receptor signaling pathway   | –        | 0/22   | –        | 0/22   | p > 0.01 | 2/22   | 2.97e-03 | 4/22   |
| Development_Slit-Robo signaling   | 4.18e-05 | 7/30   | 1.22e-07 | 10/30  | 1.62e-04 | 6/30   | 7.93e-08 | 9/30   |
| Development_TGF $\beta$ -beta-dependent induction of EMT via MAPK                           | p > 0.01 | 4/47   | 4.49e-04 | 8/47   | p > 0.01 | 4/47   | 1.88e-03 | 6/47   |
| Development_TGF $\beta$ -beta-dependent induction of EMT via RhoA, PI3K & ILK.              | p > 0.01 | 4/46   | 2.02e-03 | 7/46   | 9.26e-03 | 5/46   | 3.54e-05 | 8/46   |
| Development_VEGF signaling via VEGFR2 – generic cascades                                    | p > 0.01 | 5/84   | 1.68e-03 | 10/84  | p > 0.01 | 6/84   | 5.46e-04 | 9/84   |
| Development_WNT signaling pathway. Part 2   | p > 0.01 | 3/53   | 4.58e-03 | 7/53   | –        | 0/53   | –        | 0/53   |
| Glycolysis and gluconeogenesis (short map)  | –        | 0/66   | –        | 0/66   | 2.86e-10 | 15/66  | 2.32e-07 | 12/66  |
| Glycolysis and gluconeogenesis p.3 / Human version  | 9.56e-04 | 5/24   | 2.75e-05 | 7/24   | –        | 0/24   | –        | 0/24   |
| LRRK2 in neurons in Parkinson's disease   | 6.11e-13 | 14/33  | 6.27e-16 | 17/33  | 7.80e-17 | 16/33  | 1.68e-12 | 13/33  |

Note: PP indicates the proteins identified over the total proteins in the pathway. Rows of the same colors were used to group the main MetaCore™ pathways.

**Fonte:** AMORIM et al. (2014)

**Legenda:** PP indica o número de proteínas identificadas em relação ao total de proteínas presentes na via. Fileiras da mesma cor foram usadas para agrupar as principais vias.



Fonte: AMORIM et al. (2014)

**Figura 10** – Análise das proteínas exclusivas das EVs 20K e 100K de C5.2 e HB4a. **A)** Processos biológicos GO das proteínas identificadas conforme classificação pela database PANTHER. **B)** Localização subcelular primária das proteínas identificadas conforme classificação pelo banco HPRD.

#### **4.1.2 Caracterização do conjunto de RNAs de EVs derivadas de HB4a e C5.2 por NGS**

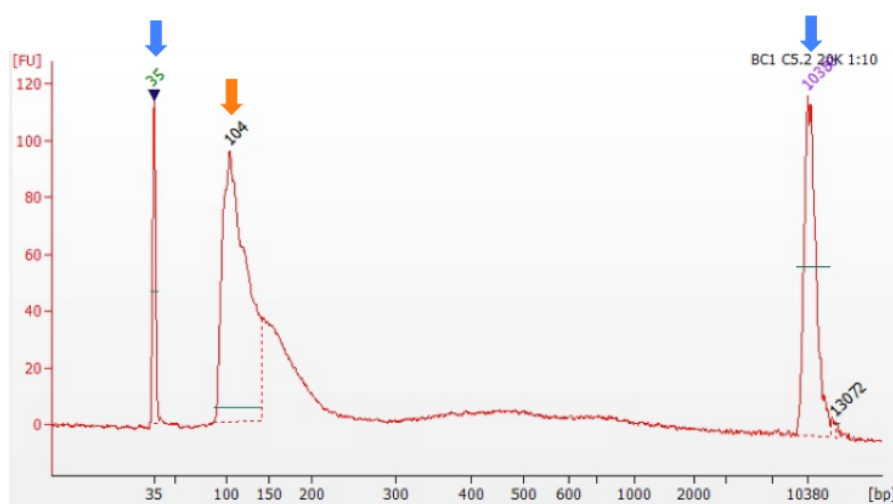
De modo similar às análises proteômicas, utilizamos meio de cultura celular condicionado das linhagens C5.2 e HB4a para padronizar a caracterização do conjunto de RNAs contidos em EVs de 20K e 100K. Além de possibilitar o estudo dos efeitos transcricionais da superexpressão de *ERBB2/HER2*, este modelo também permite uma maior investigação das diferenças entre os RNAs carregados pelas EVs em relação aqueles RNAs contidos nas células secretoras, pois não contem EVs derivadas de outras fontes, como ocorre nas EVs derivadas de plasma de pacientes. Tais análises contribuem para o entendimento da diversidade de conteúdo das EVs e também dos mecanismos que determinam a secreção vesicular de determinadas moléculas.

Logo que iniciamos estes estudos nos deparamos com a baixíssima quantidade de RNAs que poderia ser obtida das vesículas, algo muito desfavorável principalmente para o estudo de EVs obtidas do plasma de pacientes. Como exemplo, enquanto os protocolos usuais de construção de bibliotecas para o sequenciamento de nova geração (NGS) requerem 100-500ng de RNA, a quantidade de RNA que conseguíamos obter a partir de 120ml de meio de cultivo era indetectável por fluorometria (método de quantificação mais específico que o baseado em absorbância). Durante meses de padronização a pequena quantidade de RNA levava a um excesso de dímeros de adaptadores formados (chegando a 82% das sequências geradas), devido à pequena quantidade de RNA total obtido das EVs. Apesar da grande diminuição dos dímeros alcançada com o protocolo otimizado, mesmo assim ainda perdemos de 21 a 35% das sequências geradas das amostras de pacientes por serem dímeros de adaptadores.

Diante disto, partimos então para um protocolo de fragmentação do RNA total com a intenção de aumentar o número de moléculas disponíveis para ligação aos adaptadores, sem haver o aumento da massa total do RNA, de modo que estas moléculas competissem pela ligação dos adaptadores. Este procedimento é normalmente realizado no protocolo de construção de

biblioteca para transcriptoma (que utiliza RNAs longos), com intuito diferente, mas não é usado na construção de bibliotecas de pequenos RNAs. Ao fragmentar o RNA total derivado das EVs conseguimos reduzir significativamente a quantidade de dímeros de adaptadores formados e também possibilitamos a análise simultânea de todos os transcritos incluindo RNAs pequenos e longos, a partir de uma única biblioteca. Essa otimização foi muito importante nas etapas posteriores quando tínhamos que lidar com quantidades ínfimas de RNA obtidas a partir de EVs extraídas de uma quantidade limitada de plasma.

Após uma série de padronizações do preparo de biblioteca, obtivemos sucesso na caracterização do conjunto de RNAs (todas as classes de pequenos RNAs e RNAs longos, incluindo classes de RNAs não codificantes e mRNAs) de EVs utilizando a plataforma *Ion Proton* (Life Technologies). Nesta etapa trabalhamos com 120ml de meio condicionado obtido de células das linhagens HB4a e C5.2 cultivadas em meio complementado com soro fetal bovino previamente depletado de EVs. Isolamos as EVs por ultracentrifugação, extraímos o RNA total, e construímos bibliotecas para a avaliação de todo o transcriptoma. Extraímos RNA também a partir do *pellet* celular para podermos fazer a caracterização transcricional de todos os componentes, incluindo as duas células e os dois grupos de vesículas (20K e 100K). A verificação do tamanho final das bibliotecas foi feita no equipamento *Bioanalyzer* utilizando o *High Sensitivity DNA Kit* (Agilent Technologies, California, EUA), confirmando que o maior pico tinha tamanho de 104pb, que está dentro da faixa de tamanho esperado de 94-114pb para fragmentos do tamanho de miRNAs ligados aos adaptadores, segundo o fabricante do kit de preparo de biblioteca (exemplo apresentado na **Figura 11** e os demais no **Anexo 5**, onde vemos que o tamanho do maior pico variou de 97 a 104pb entre as 6 bibliotecas).



**Figura 11** - Perfil representativo do *Bioanalyzer* de biblioteca (C5.2 20K) para sequenciamento. As setas azuis correspondem a marcadores de tamanho (conforme eixo X), e a seta laranja indica o pico obtido que corresponde ao tamanho esperado (104pb) após a ligação de adaptadores aos RNAs, seguido de PCR com a adição dos *barcodes*.

No sequenciamento geramos no total 94,796,201 *reads*, com distribuição de tamanhos com média 32,5nt, desvio padrão 20,6nt e variando entre 8-371nt considerando as 4 populações de EVs e de tamanho médio 27,7nt, desvio padrão 16,9nt e variando entre 8-370nt considerando as 2 células. As sequências geradas foram inicialmente filtradas com o programa FastQ Screen para a remoção de RNAs ribossomais (rRNA), RNAs transportadores (tRNA), sequências de elementos repetitivos abundantes no genoma humano (satélites, ALU, LINE) e contaminantes de outras espécies (bactérias, fungos, vírus, *C. elegans*, etc.) (**Tabela 4**). Neste passo em média 82% das sequências geradas para as quatro bibliotecas de EVs (C5.2 20K e 100K; HB4a 20K e 100K) e em média 55% das sequências geradas para as duas bibliotecas de células (C5.2 e HB4a) foram filtradas. Dentre todos os *reads* filtrados, as duas categorias mais representadas foram: 1) rRNA para as bibliotecas de EVs - com em média 72,9% de *reads* filtrados, enquanto que nas bibliotecas de células, que passaram por depleção de rRNA, somente 4,0% em média foi filtrado nesta categoria e 2) elementos repetitivos para as bibliotecas de células - com em média 28,8% de *reads* filtrados, enquanto que nas bibliotecas de EVs somente 2,7% foi filtrado



nesta categoria. O resultado que apontou uma média de 72,9% de rRNA nas bibliotecas de EVs foi de certo modo inesperado. A literatura é controversa em relação a este aspecto, com os primeiros trabalhos que caracterizaram RNAs de EVs relatando a ausência de picos característicos de rRNA em eletroferogramas gerados pelo Bioanalyzer (SKOG et al. 2008; VALADI et al. 2007). Os trabalhos mais recentes que utilizaram NGS para caracterizar os RNAs de EVs mostram resultados discrepantes: em alguns trabalhos é encontrado relativamente pouco rRNA, cerca de 9% (HUANG et al. 2013), já em outros trabalhos é identificada uma quantidade de rRNA ainda superior à que detectamos, chegando a 97% de rRNA do total de sequências geradas (JENJAROENPUN et al. 2013). Deste modo, nossos dados encontram respaldo na literatura e de maneira importante, acreditamos que a metodologia utilizada na construção das bibliotecas nos permitiu descrever precisamente o real conteúdo de todo o transcriptoma das EVs. Devido à pequena quantidade de RNA obtida das EVs, não foi possível realizarmos uma depleção de rRNA como fizemos para o RNA derivado do extrato celular, pois era necessário 3µg de RNA para a depleção com o *Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat)* (Epicentre, Wisconsin, EUA), e mesmo usando 120ml de meio condicionado, não conseguimos obter sequer uma quantidade mínima que permitisse a dosagem dos RNAs derivados das EVs, pois a concentração de RNA encontrava-se abaixo do limite mínimo de detecção do fluorímetro Qubit (Life Technologies, California, EUA), que é de 20pg/µl. Apesar da maioria das sequências terem sido filtradas, ainda obtivemos em média 947,944 sequências maiores que 12nt para cada biblioteca das EVs. Para a anotação destas sequências usamos inicialmente o programa *miRDeep2* (MACKOWIAK 2011) para a anotação dos miRNAs humanos descritos no banco de dados *miRBase v20*, e em seguida as sequências que não anotaram foram alinhadas no genoma humano para a anotação dos transcritos contidos no banco *Ensembl*. Conseguimos identificar um grande número de miRNAs e outras categorias de RNAs nas EVs e nas células, conforme **Tabelas 4 e 6**.

**Tabela 4** – Análise dos dados de sequenciamento em larga escala de miRNAs de EVs derivadas das linhagens C5.2 e HB4a.

| Amostra     | Total reads                        | # reads após filtro FastQ Screen (%) | # reads usados miRDeep2 ( $\geq 12nt$ ) | # reads mapeados miRBase            | # miRNAs identificados              |
|-------------|------------------------------------|--------------------------------------|---|-------------------------------------|-------------------------------------|
| C5.2 20K    | 9.167.123                          | 2.026.337 (22,1%)                    | 1.252.427                               | 683                                 | 108                                 |
| C5.2 100K   | 6.741.573                          | 1.232.651 (18,3%)                    | 654.397                                 | 1.350                               | 125                                 |
| C5.2 célula | 25.114.667                         | 10.929.299 (43,5%)                   | 8.806.324                               | 10.575                              | 338                                 |
| HB4a 20K    | 6.857.816                          | 1.097.932 (16,0%)                    | 820.854                                 | 1.202                               | 128                                 |
| HB4a 100K   | 9.996.981                          | 1.490.507 (14,9%)                    | 1.064.096                               | 3.896                               | 209                                 |
| HB4a célula | 36.918.041                         | 17.276.874 (46,8%)                   | 14.160.581                              | 10.689                              | 466                                 |
| Amostra     | # miRNAs maduros $\geq 1$ read     | # miRNAs maduros $\geq 2$ reads      | # miRNAs maduros $\geq 3$ reads         | # miRNAs maduros $\geq 4$ reads     | # miRNAs maduros $\geq 5$ reads     |
| C5.2 20K    | 76                                 | 47                                   | 35                                      | 32                                  | 23                                  |
| C5.2 100K   | 98                                 | 66                                   | 48                                      | 42                                  | 37                                  |
| C5.2 célula | 132                                | 88                                   | 64                                      | 59                                  | 56                                  |
| HB4a 20K    | 108                                | 71                                   | 58                                      | 44                                  | 36                                  |
| HB4a 100K   | 172                                | 117                                  | 100                                     | 84                                  | 74                                  |
| HB4a célula | 194                                | 123                                  | 98                                      | 87                                  | 74                                  |
| Amostra     | # miRNAs precursores $\geq 1$ read | # miRNAs precursores $\geq 2$ reads  | # miRNAs precursores $\geq 3$ reads     | # miRNAs precursores $\geq 4$ reads | # miRNAs precursores $\geq 5$ reads |
| C5.2 20K    | 32                                 | 14                                   | 7                                       | 3                                   | 2                                   |
| C5.2 100K   | 27                                 | 5                                    | 3                                       | 2                                   | 2                                   |
| C5.2 célula | 206                                | 112                                  | 86                                      | 72                                  | 56                                  |
| HB4a 20K    | 20                                 | 11                                   | 6                                       | 3                                   | 3                                   |
| HB4a 100K   | 37                                 | 15                                   | 11                                      | 8                                   | 5                                   |
| HB4a célula | 272                                | 153                                  | 106                                     | 89                                  | 76                                  |

A quantidade de miRNAs aqui identificados nas EVs está na ordem de grandeza de outros estudos (CHENG et al. 2014; HUANG et al. 2013; JENJAROENPUN et al. 2013; JI et al. 2014; LI et al. 2013). JI et al. (2014) por exemplo, utilizaram a linhagem celular LIM1863 de carcinoma de cólon humano, que formam esferas multicelulares (organóides) *in vitro* que se assemelham as criptas do cólon onde as células são altamente polarizadas, e caracterizaram três populações diferentes de EVs secretadas em cultura: microvesículas, exossomos EpCAM-positivos que são secretados por células na superfície apical (voltada para o lúmen), e exossomos A33-positivos que são secretados por células na superfície basolateral. Eles identificaram um total de 254 miRNAs com pelo menos 5 reads em cada população de EVs. Vale aqui ressaltar dois aspectos metodológicos importantes que diferem entre o nosso trabalho e este mencionado: eles utilizaram a plataforma *Illumina HiSeq* de maior *throughput*, obtendo mais de 57 milhões de sequências de boa qualidade para as bibliotecas de EVs,

enquanto nós obtivemos ~1 milhão de sequências por biblioteca/EV. Eles também utilizaram um sistema de cultura de células contínua em bioreatores que os possibilitou a obtenção de um volume inicial de 2L de meio condicionado para o isolamento das EVs, enquanto nós partimos de 120ml (~16X menos). Em nosso trabalho optamos por fragmentar o RNA total a fim de caracterizar o transcriptoma total das EVs, incluindo miRNAs e outros RNAs longos como mRNAs e lincRNAs. A maioria dos estudos foca somente na população de pequenos RNAs – no estudo acima mencionado, por exemplo, eles fracionaram o RNA total em gel de poliacrilamida e selecionaram somente os fragmentos de 18 a 30nt para a construção das bibliotecas – nós fizemos uma análise sem viés de todos os RNAs contidos nas EVs, e acreditamos que esta seja uma representação mais fiel do transcriptoma vesicular.

Uma comparação do número de miRNAs maduros em relação ao número de precursores de miRNAs identificados revelou uma diferença interessante entre EVs e células. Conforme **Tabela 5**, quando comparamos a porcentagem de precursores em relação ao total de miRNAs mapeados, vemos que as EVs tem de 15,6% (na HB4a 20K) a 29,6% (na C5.2 20K) de precursores, enquanto as células tem de 58,4% (HB4a célula) a 60,9% (C5.2 célula) de precursores. Se ordenamos os 50 miRNAs mais abundantes em termos de quantidade de *reads*, vemos que as EVs tem apenas entre 6,0% (C5.2 100K, HB4a 20K e HB4a 100K) a 22,0% (na C5.2 20K) de precursores, enquanto as células tem de 48,0% (HB4a célula) a 54,0% (C5.2 célula) de precursores dentre os 50 miRNAs mais abundantes. Estes achados sugerem que as EVs carregam preferencialmente miRNAs já maduros, que de fato são as moléculas funcionais com maior potencial regulatório ao serem transferidas para a célula receptora, ao invés da transferência de moléculas precursoras que ainda teriam que ser processadas. Recentemente, um estudo analisou exossomos secretados pela linhagem tumoral de mama MDA-MB-231 ao longo de 7 diferentes pontos de coleta (os exossomos foram mantidos em meio de cultura livre de células por 6, 12, 24, 36, 48, 72 e 96h), e foi observado uma relação inversa entre a expressão

de miRNAs maduros e precursores ao longo do tempo: após 24h houve uma queda significativa da expressão dos precursores e um aumento proporcional da expressão dos miRNAs maduros, atingindo um *plateau* após 72h (MELO et al. 2014). Os autores demonstram que os exossomos contêm proteínas do complexo RISC (DICER e AGO2), que são fundamentais para o processamento dos miRNAs precursores, e propõem que a biogênese de miRNAs pode ocorrer em exossomos independentemente de células. Apesar de não termos avaliado EVs coletadas em outros pontos de tempo, já que isolamos EVs do meio de cultura celular condicionado apenas após aproximadamente 24 horas de cultivo das células, o nosso achado da baixa quantidade de miRNAs precursores nas EVs em relação as células secretoras é concordante com o trabalho acima mencionado. Também é importante ressaltar que não identificamos por proteômica as proteínas DROSHA, DICER, ou AGO2, importantes na biogênese de miRNAs, em nenhuma das EVs de C5.2 ou de HB4a.

**Tabela 5** - Comparação do número de precursores de miRNAs identificados nas EVs em relação as células.

| Amostra     | # precursores (%) | # precursores top-50 (%) |
|-------------|-------------------|--------------------------|
| C5.2 20K    | 32 (29,6%)        | 11 (22,0%)               |
| C5.2 100K   | 27 (21,6%)        | 3 (6,0%)                 |
| C5.2 célula | 206 (60,9%)       | 27 (54,0%)               |
| HB4a 20K    | 20 (15,6%)        | 3 (6,0%)                 |
| HB4a 100K   | 37 (17,7%)        | 3 (6,0%)                 |
| HB4a célula | 272 (58,4%)       | 24 (48,0%)               |

**Legenda:** *Top-50* refere-se aos 50 miRNAs (precursores e maduros) mais abundantes (maior número de *reads*) identificados em cada amostra.

Em relação aos demais RNAs anotados, a **Tabela 6** indica a quantidade de transcritos que foram identificados nas principais categorias. Em média nas quatro bibliotecas de EVs identificamos 10,120 RNAs codificadores de proteínas (mRNAs), 149 snoRNAs, 60 snRNAs, 318

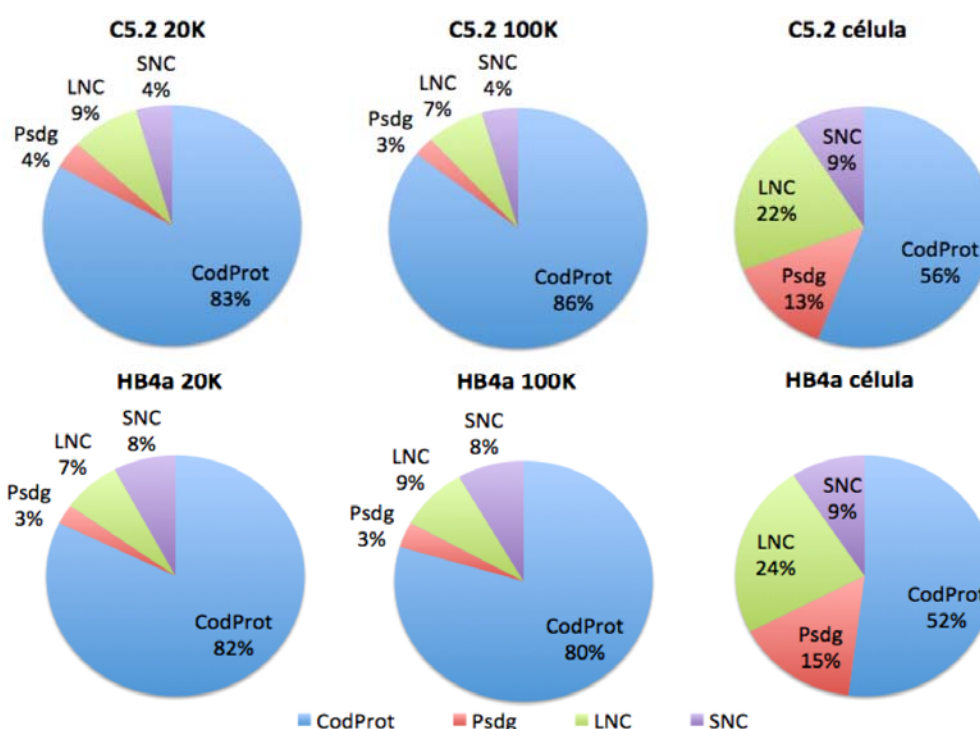
lincRNAs, e 268 outros transcritos classificados na categoria miscelânea (misc), considerando todos os transcritos com FPKM >0. Os RNAs não-codificadores de proteínas agrupados na categoria misc (uma das categorias do Ensembl) incluem dentre outros os SRP-RNAs, Vault RNAs, e Y-RNAs, que são RNAs que tem associação com proteínas citoplasmáticas, e que foram descritos como sendo enriquecidos em EVs, em relação ao RNA celular (NOLTE-'T HOEN et al. 2012). Apesar de nós não termos encontrado um enriquecimento desta classe de RNAs em EVs em relação às células quando consideramos o número total de transcritos identificados (**Tabela 8**), nas análises de expressão diferencial vemos muitos miscRNAs com expressão significativamente aumentada nas EVs em relação às células secretoras (**Figura 16**). Em média nas duas bibliotecas de células foram identificados 17,411 mRNAs, 463 snoRNAs, 544 snRNAs, 3,154 lincRNAs, e 1,219 miscRNAs, considerando todos os transcritos com FPKM >0.

**Tabela 6** - Análise dos dados de sequenciamento em larga escala de outros RNAs de EVs derivadas das linhagens C5.2 e HB4a.

| Amostra     | # reads não anotados no miRDeep2 | # reads usados no Ensembl (≥20nt) | # reads mapeados no Ensembl | # transcritos CodProt | # transcritos snoRNA | # transcritos snRNA | # transcritos lincRNA | # transcritos misc_RNA |
|-------------|----------------------------------|-----------------------------------|-----------------------------|-----------------------|----------------------|---------------------|-----------------------|------------------------|
| C5.2 20K    | 2.025.654                        | 1.227.032                         | 471.350                     | 11.198                | 105                  | 37                  | 370                   | 203                    |
| C5.2 100K   | 1.231.301                        | 637.557                           | 259.019                     | 9.977                 | 71                   | 27                  | 265                   | 198                    |
| C5.2 célula | 10.918.724                       | 8.522.184                         | 7.209.540                   | 16.915                | 438                  | 531                 | 2.802                 | 1.140                  |
| HB4a 20K    | 1.096.730                        | 806.559                           | 498.817                     | 8.485                 | 206                  | 77                  | 241                   | 254                    |
| HB4a 100K   | 1.486.611                        | 1.040.703                         | 467.821                     | 10.820                | 214                  | 101                 | 397                   | 417                    |
| HB4a célula | 17.266.185                       | 13.743.431                        | 11.952.924                  | 17.907                | 488                  | 557                 | 3.507                 | 1.298                  |

Uma análise global das categorias de RNAs contidos nas EVs em relação aos RNAs presentes nas células revelou diferenças marcantes entre o transcriptoma vesicular e o transcriptoma celular (**Figuras 12 e 13**). Considerando a média das quatro populações de EVs em relação à média das duas populações de células para teste t não-pareado (**Tabelas 7 e 8**), vemos um aumento significativo nas EVs de transcritos codificantes de proteína (com 82,64% de RNAs codificantes em EVs versus 53,91% em células). O enriquecimento de transcritos codificantes de proteínas em EVs

ainda não foi reportado na literatura, possivelmente porque a maioria dos estudos de NGS de EVs analisa somente pequenos RNAs, mas deve estar relacionado a alguma função biológica ainda não compreendida. Dentre outras diferenças, observamos um aumento significativo em EVs de miRNAs maduros (com 14,93% em EVs versus 5,45% em células) e uma diminuição significativa de miRNAs precursores (com 4,04% em EVs versus 8,03% em células).

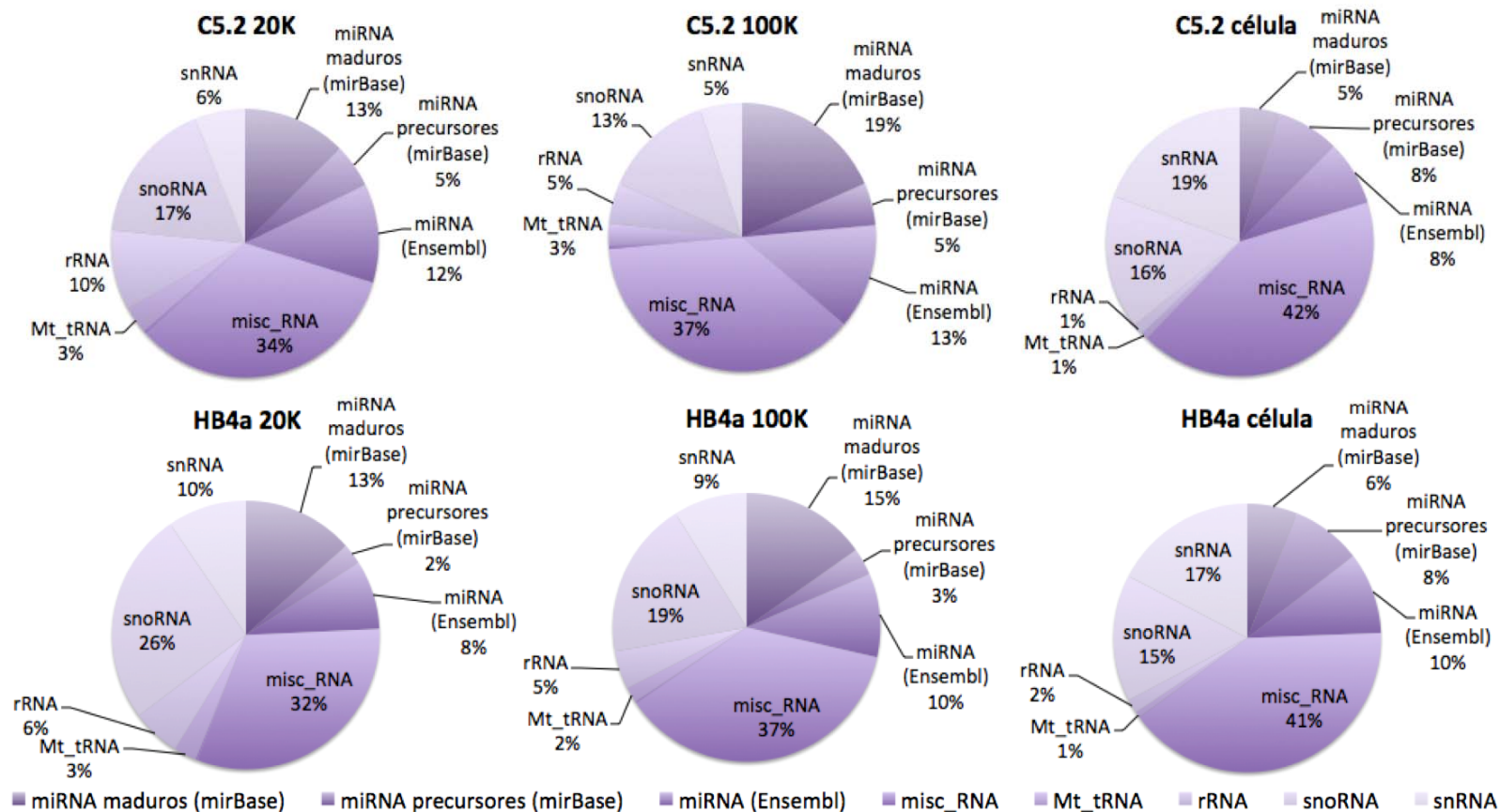


**Figura 12** - Distribuição dos transcritos identificados em EVs e células nas diversas categorias de RNAs. CodProt = transcritos codificadores de proteína; Psdg = Pseudogenes; LNC = transcritos não-codificadores de proteínas longos (*long noncoding*); SNC = transcritos não-codificadores de proteínas curtos (*short noncoding*).

**Tabela 7** - Análise estatística da distribuição dos transcritos identificados em EVs e células nas diversas categorias de RNAs.

| Categoria              | Média EVs | Média Células | p-valor |
|------------------------|-----------|---------------|---------|
| CodProt                | 82,64%    | 53,91%        | 0,0074  |
| <i>Long noncoding</i>  | 7,99%     | 22,63%        | 0,0094  |
| <i>Short noncoding</i> | 6,28%     | 9,12%         | 0,0680  |
| Pseudogene             | 3,09%     | 14,34%        | 0,0404  |

**Legenda** – Média EVs = média de distribuições das 4 bibliotecas de EVs; Média Células = média das 2 bibliotecas de células; p-valor = resultado de teste t não-pareado comparando as médias de EVs e células. CodProt = transcritos codificadores de proteína.



**Figura 13** - Distribuição dos transcritos identificados em EVs e células na categoria *short noncoding RNAs*. miRNA (Ensembl) = miRNAs que escaparam da anotação feita pelo miRBase mas foram anotados pelo Ensembl, ou possivelmente a transcritos que estão indevidamente anotados como miRNAs no Ensembl.

**Tabela 8** - Análise estatística da distribuição dos transcritos identificados em EVs e células na categoria *short noncoding RNAs*.

| <b>Categoria</b>             | <b>Média EVs</b> | <b>Média Células</b> | <b>p-valor</b> |
|------------------------------|------------------|----------------------|----------------|
| miRNAs maduros (miRBase)     | 14,93%           | 5,45%                | 0,0031         |
| miRNAs precursores (miRBase) | 4,04%            | 8,03%                | 0,0089         |
| miRNAs (Ensembl)             | 10,74%           | 8,85%                | 0,2655         |
| misc_RNA                     | 34,86%           | 41,22%               | 0,0147         |
| snoRNA                       | 18,84%           | 15,67%               | 0,3046         |
| snRNA                        | 7,43%            | 18,46%               | 0,0050         |
| rRNA                         | 6,24%            | 1,50%                | 0,0244         |
| Mt_tRNA                      | 2,64%            | 0,75%                | 0,0067         |
| Mt_rRNA                      | 0,28%            | 0,07%                | 0,0158         |

**Legenda** – Média EVs = média de distribuições das 4 bibliotecas de EVs; Média Células = média das 2 bibliotecas de células; p-valor = resultado de teste t não-pareado comparando as médias de EVs e células. miRNA (Ensembl) = miRNAs que escaparam da anotação feita pelo miRBase mas foram anotados pelo Ensembl, ou possivelmente a transcritos que estão indevidamente anotados como miRNAs no Ensembl.

Após a anotação em todos os bancos, ainda restaram em média 70% de *reads* não anotados nas quatro bibliotecas de EVs, e 32% de *reads* não anotados nas duas bibliotecas de células, em relação ao total de *reads* úteis após o filtro FastQ Screen (**Tabela 9**). Ao avaliarmos especificamente as sequências não anotadas, obtivemos indícios por análise manual de BLAST de que as mesmas poderiam ser derivadas de rRNAs, e resolvemos re-mapear contra as sequências de rRNA do FastQ Screen com o mesmo alinhador que utilizado inicialmente (*Bowtie2*), mas reduzindo a estrigência. Com esta modificação, conseguimos mapear uma grande quantidade de *reads* que eram de fato rRNA mas que haviam passado pelo filtro inicial. Após verificarmos do total de *reads* que ainda assim não anotaram quantos eram de boa qualidade ( $\geq Q20$ ), chegamos a uma porcentagem final de em média 15,2% de *reads* que não anotam e tem qualidade nas quatro bibliotecas de EVs e de em média 22,4% nas duas bibliotecas de células. Estas quantidades estão dentro do esperado para NGS, de acordo com a nossa experiência para outras bibliotecas sequenciadas no laboratório.



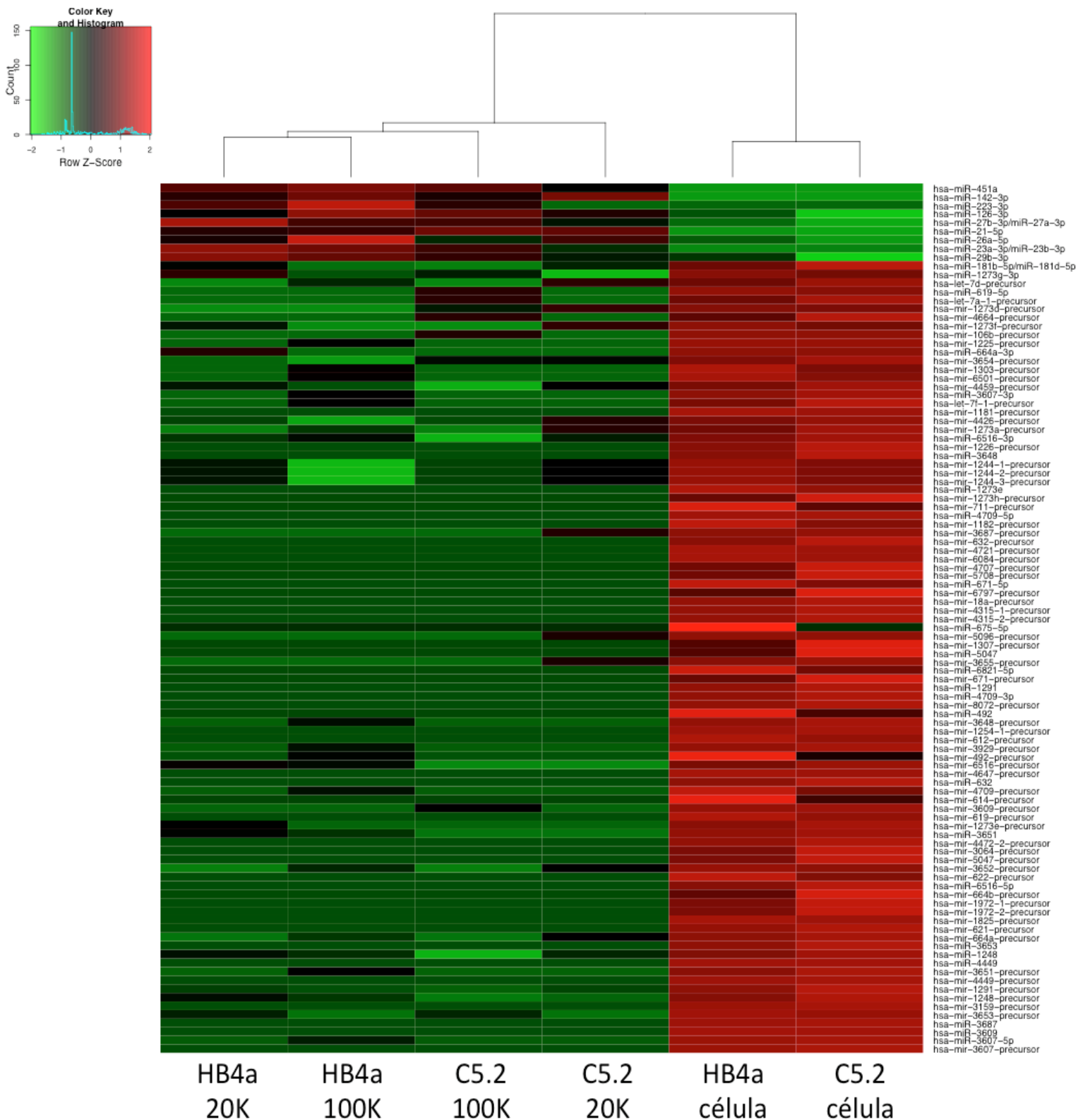
**Tabela 9** - Análise dos *reads* de C5.2 e HB4a que não anotaram após alinhamento contra miRBase e hg19/Ensembl.

| Amostra     | # <i>reads</i> não anotados no Ensembl | % <i>reads</i> não anotados no Ensembl | # <i>reads</i> mapeados rRNA | # <i>reads</i> que não anotam | % <i>reads</i> que não anotam | # <i>reads</i> que não anotam e tem qualidade ( $\geq$ Q20) | % <i>reads</i> que não anotam e tem qualidade ( $\geq$ Q20) |
|-------------|--|--|------------------------------|-------------------------------|-------------------------------|---|---|
| C5.2 20K    | 1.554.304                              | 76,7                                   | 786.819                      | 767.485                       | 49,4                          | 214.711   | 13,8  |
| C5.2 100K   | 972.282                                | 78,9                                   | 617.442                      | 354.840                       | 36,5                          | 138.293   | 14,2  |
| C5.2 célula | 3.709.184                              | 33,9                                   | 1.276.464                    | 2.432.720                     | 65,6                          | 821.334   | 22,1  |
| HB4a 20K    | 597.913                                | 54,5                                   | 316.331                      | 281.582                       | 47,1                          | 103.854   | 17,4  |
| HB4a 100K   | 1.018.790                              | 68,4                                   | 470.837                      | 547.953                       | 53,8                          | 156.380   | 15,3  |
| HB4a célula | 5.313.261                              | 30,8                                   | 1.892.543                    | 3.420.718                     | 64,4                          | 1.205.177   | 22,7  |

**Legenda:** Q20 é um escore de qualidade baseado em *Phred* que indica uma acurácia de chamada de base de 99% (ou seja, a probabilidade de erro é de 1 em 100).

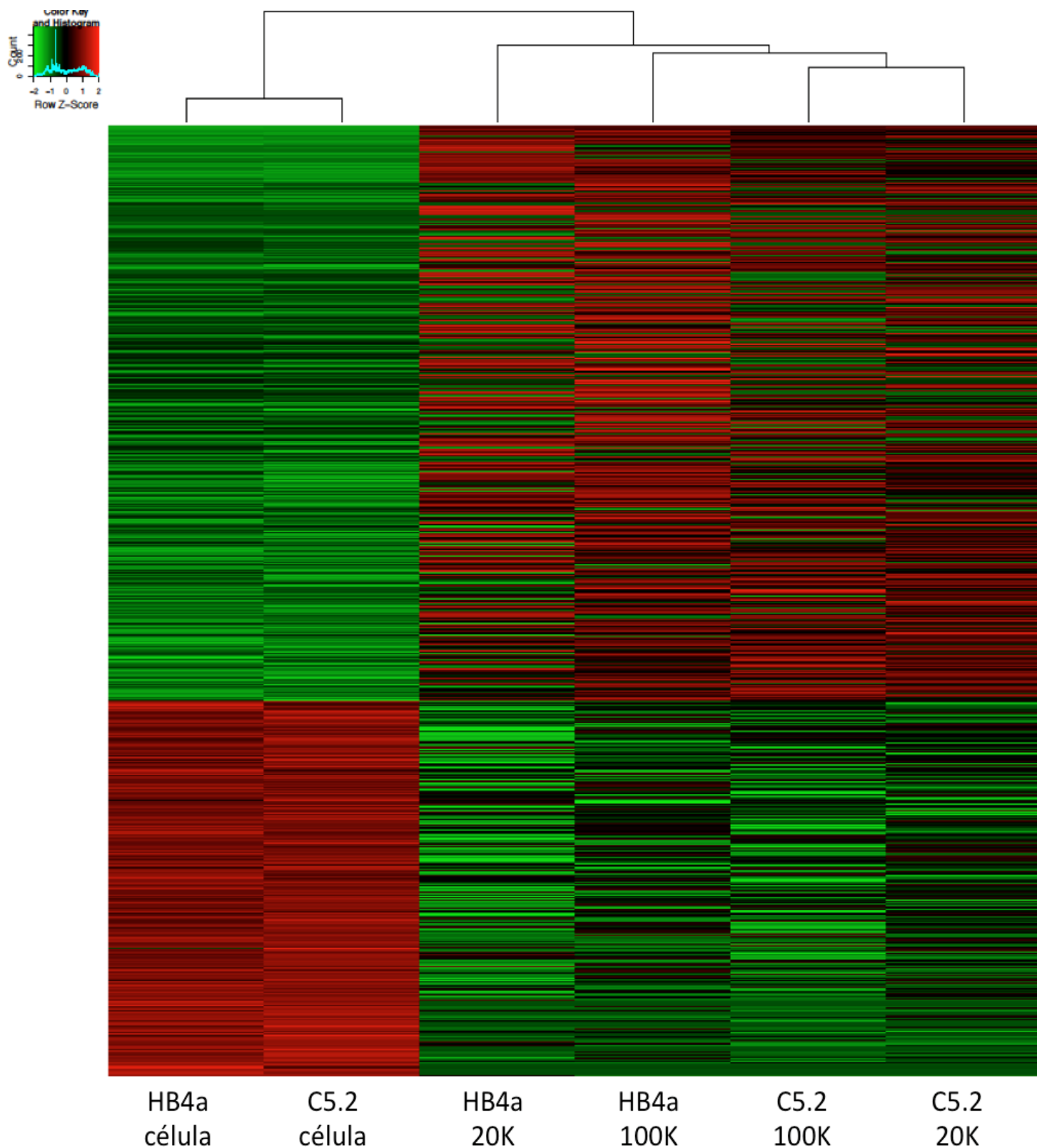
## A Expressão diferencial dos transcritos entre células e EVs

A análise de expressão diferencial dos miRNAs e outros RNAs identificados foi realizada com o pacote *DESeq2* do *Bioconductor* (LOVE et al. 2014). Inicialmente avaliamos os transcritos contidos nas EVs em relação às células secretoras, que poderiam indicar moléculas que são preferencialmente exportadas via EVs. Ao agruparmos as quatro populações de EVs (C5.2 20K, C5.2 100K, HB4a 20K, e HB4a 100K) e compararmos com as duas células (C5.2 e HB4a), encontramos 100 miRNAs e 3.190 outros RNAs diferencialmente representados com p-valor  $\leq 0,05$ . Dos 100 miRNAs diferencialmente representados, 9 (9,0%) estão enriquecidos nas EVs e 91 (91,0%) estão enriquecidos nas células (**Figura 14**). Todos os 9 miRNAs (miR-451a, miR-142-3p, miR-223-3p, miR-126-3p, miR-27b-3p/miR-27a-3p, miR-21-5p, miR-26a-5p, miR-23a-3p/miR-23b-3p, miR-29b-3p) preferencialmente representados nas EVs já foram descritos no banco EVpedia (<http://evpedia.info>) em outros estudos que analisaram o transcriptoma de EVs de eucariotos, sendo que o miR-23a-3p/miR-23b-3p foi descrito em 29 datasets diferentes (número máximo de estudos compilados nesta mesma database), e o hsa-miR-142-3p (menos prevalente dos 9) em pelo menos 6 estudos.



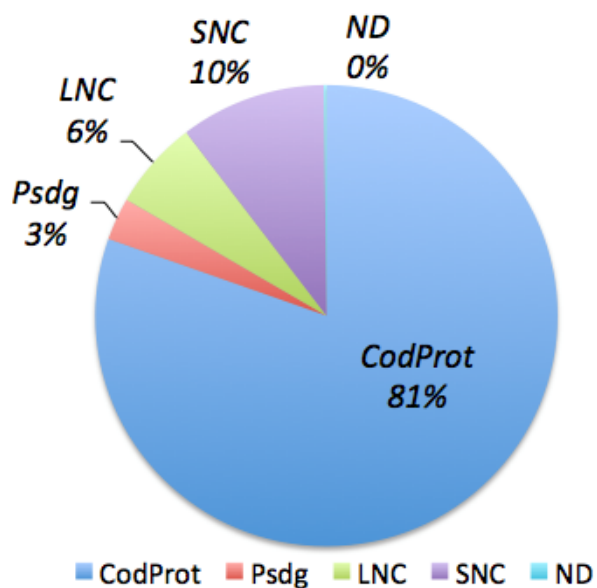
**Figura 14** - Heat map dos 100 miRNAs diferencialmente representados entre as EVs (20K e 100K) e células C5.2 e HB4a.

Dos 3.190 outros RNAs diferencialmente representados, 1.928 (60,4%) estão enriquecidos nas EVs e 1.262 (39,6%) estão enriquecidos nas células (**Figura 15**). A distribuição de categorias destes 3.190 transcritos diferencialmente representados está representada na **Figura 16**: 81% são *codificadores de proteína*, 10% são *short noncoding*, 6% são *long noncoding* e 3% foram anotados como *pseudogenes*. Dentre as classes que compõem a categoria *short noncoding*, a mais representada é miscRNA, com 39% dos transcritos. Interessantemente, dos 128 miscRNAs diferencialmente representados, 114 (89,0%) estão enriquecidos nas EVs e somente 14 (10,9%) estão enriquecidos nas células, corroborando o dado da literatura que miscRNAs são enriquecidos em EVs em comparação com o RNA celular (NOLTE-'T HOEN et al. 2012). Interessantemente, entre os miscRNAs diferencialmente representados, 71,1% são Y-RNAs. Os Y-RNAs também estão significativamente representados dentre os transcritos encontrados diferencialmente representados nas EVs derivadas do plasma de pacientes, e serão discutidos em mais detalhe abaixo.

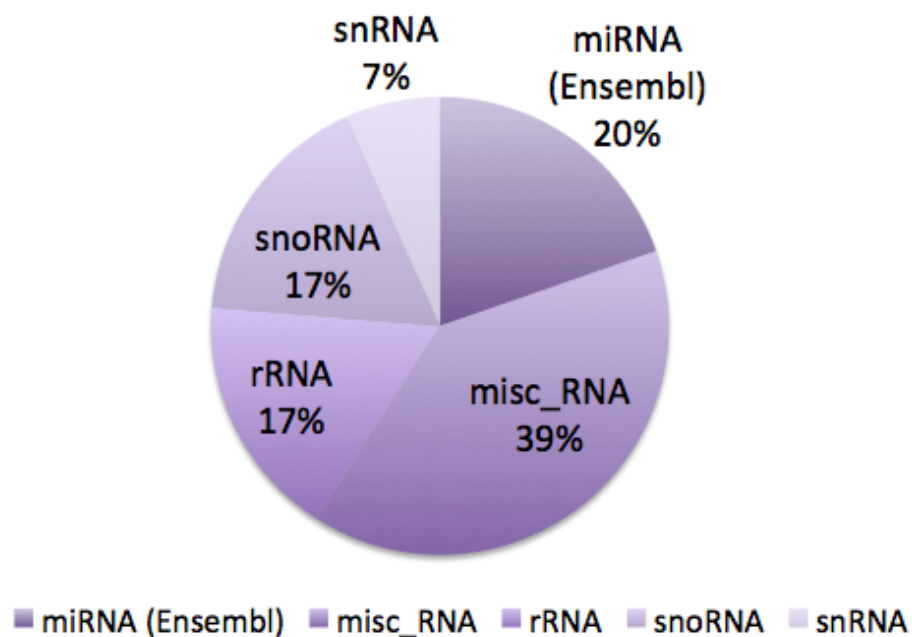


**Figura 15** - Heat map dos 3.190 outros RNAs diferencialmente representados entre as EVs (20K e 100K) e células das C5.2 e HB4a.

A



B

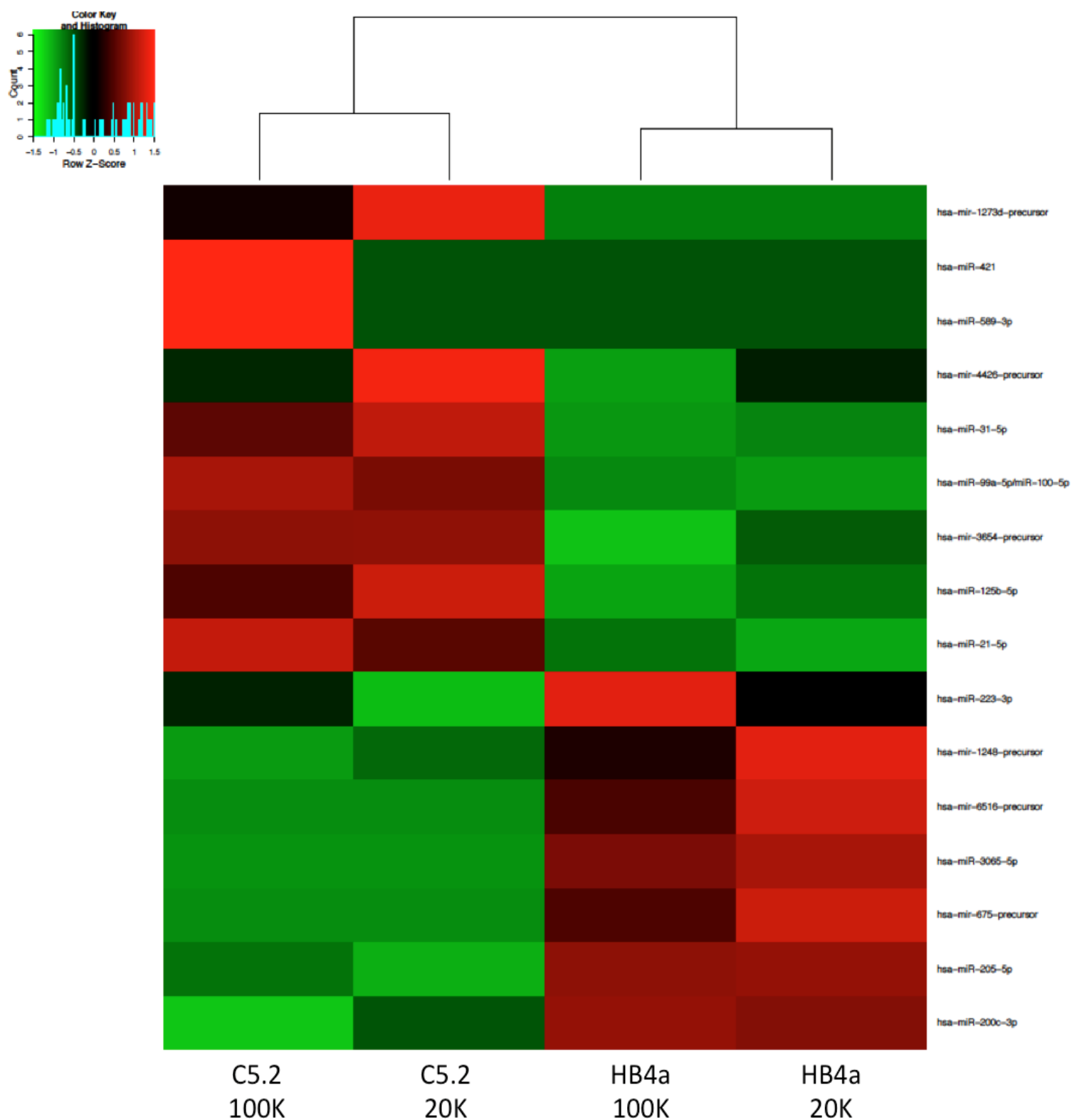


**Figura 16** - Categorias dos 3.190 outros RNAs diferencialmente representados entre as EVs (20K e 100K) e células C5.2 e HB4a. **A)** Distribuição de todos os transcritos diferencialmente representados. CodProt = transcritos codificadores de proteína; Psdg = Pseudogenes; LNC = transcritos não-codificadores de proteínas longos (*long noncoding*); SNC = transcritos não-codificadores de proteínas curtos (*short noncoding*); ND = não definido. **B)** Distribuição dos transcritos na categoria *short noncoding*.

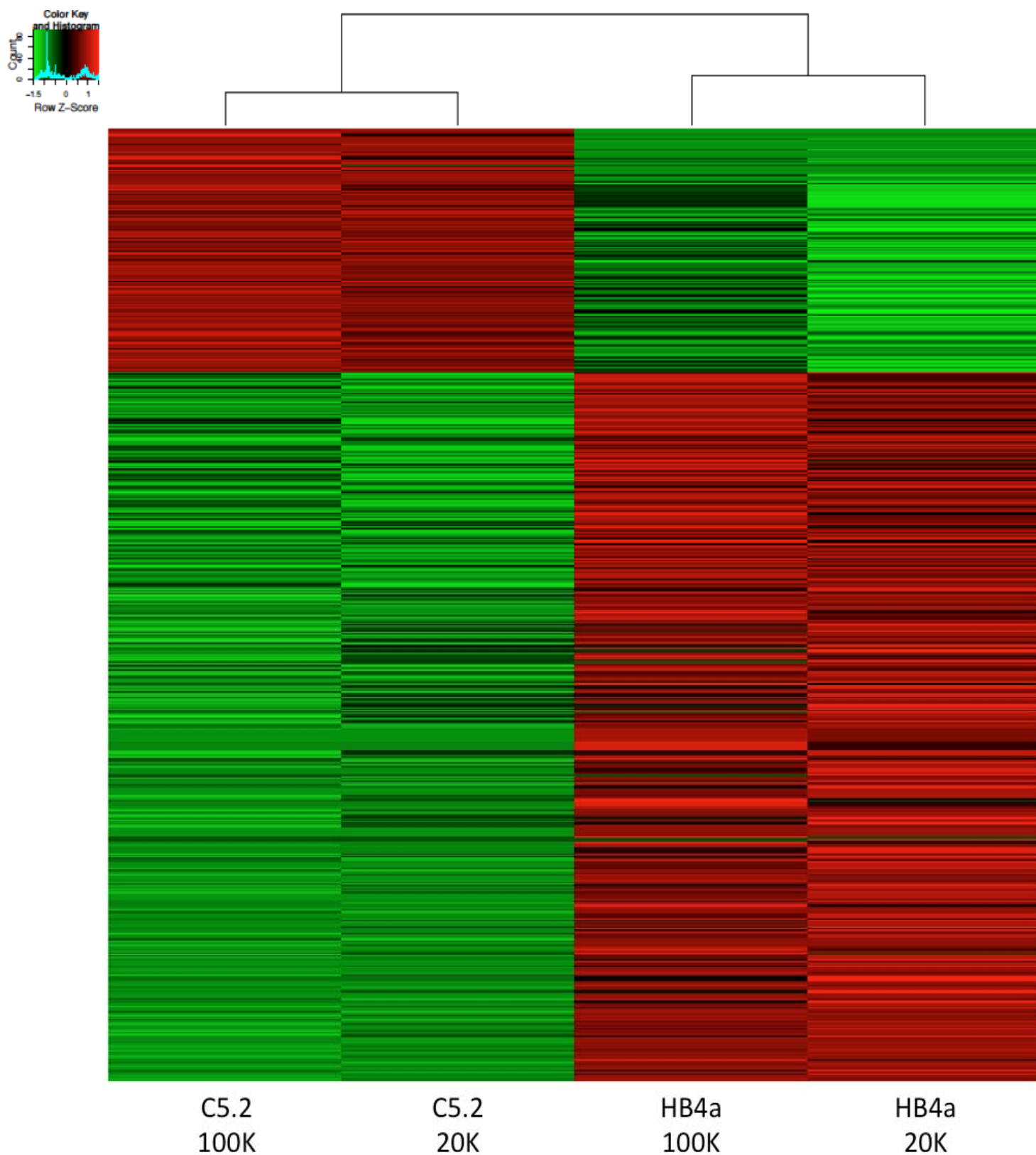
## **B Expressão diferencial do transcriptoma de vesículas e células entre C5.2 e HB4a**

Em seguida comparamos o conteúdo de RNAs nas EVs (20K + 100K) de C5.2 em relação a HB4a, buscando possíveis transcritos diferencialmente representados associados com a superexpressão de *ERBB2* e que poderiam contribuir para o fenótipo das células. No que se refere ao conjunto de miRNAs, dentre os 16 miRNAs diferencialmente representados com ( $p \leq 0,05$ ), 9 (56,2%) estão enriquecidos nas EVs de C5.2 e 7 (43,8%) estão enriquecidos nas EVs de HB4a (**Figura 17**). Dentre os miRNAs aumentados nas EVs de C5.2, destaca-se o miR-31-5p que já foi descrito com expressão reduzida *in vitro* em células invasivas em relação a células não-invasivas, e *in vivo* a expressão de miR-31 em tumores primários de mama é inversamente correlacionada com o desenvolvimento de metástases (VALASTYAN et al. 2009b). A alta expressão de miR-31 em EVs da linhagem C5.2 pode ser um mecanismo de redução da concentração intracelular deste miRNA que é exportado através das EVs, e assim talvez contribuía com o fenótipo mais agressivo das células.

Para os demais 644 RNAs diferencialmente representados ( $p \leq 0,05$ ), 164 (25,4%) encontram-se aumentados e 480 (74,6%) reduzidos em EVs de C5.2 quando comparadas às EVs de HB4a (**Figura 18**). A distribuição de categorias destes 644 transcritos diferencialmente representados encontra-se na **Figura 19**: 64% são codificadores de proteína, 28% são *short noncoding*, 6% *long noncoding* e 2% foram anotados como pseudogenes. Assim observamos que também dentre os transcritos diferencialmente representados temos a predominância de RNAs codificadores de proteína nas EVs, quando comparadas com as células de origem.



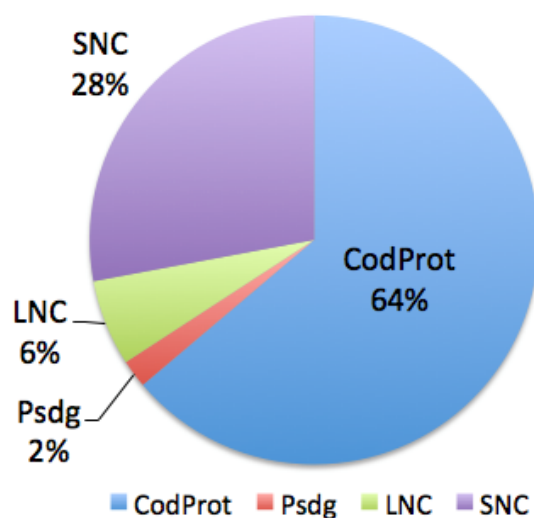
**Figura 17** - Heat map dos 16 miRNAs diferencialmente representados entre as EVs (20K e 100K) das células C5.2 e HB4a.



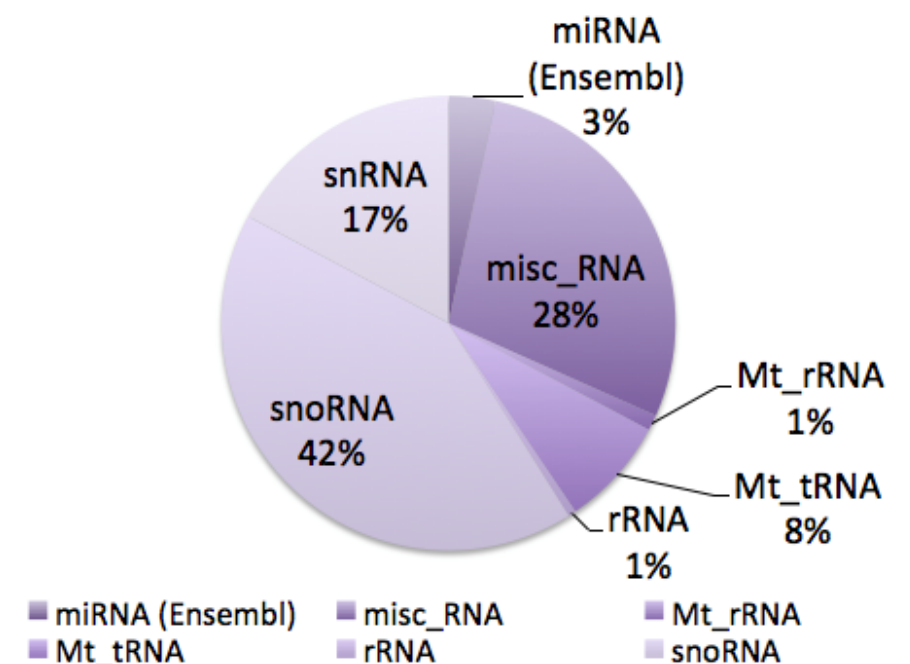
**Figura 18** - *Heat map* dos 644 miRNAs diferencialmente representados entre as EVs (20K e 100K) das células C5.2 e HB4a.



A



B



**Figura 19** - Categorias dos 644 outros RNAs diferencialmente representados entre as EVs (20K e 100K) das células C5.2 e HB4a. **A)** Distribuição de todos os transcritos diferencialmente representados. CodProt = transcritos codificadores de proteína; Psdg = Pseudogenes; LNC = transcritos não-codificadores de proteínas longos (*long noncoding*); SNC = transcritos não-codificadores de proteínas curtos (*short noncoding*). **B)** Distribuição dos transcritos na categoria *short noncoding*.

Dentre as classes que compõem a categoria *short noncoding*, a mais representada foi a snoRNAs, com 42% dos transcritos. Dos 75 snoRNAs diferencialmente representados, todos (100%) estavam reduzidos nas EVs de C5.2 em relação às EVs de HB4a.

Os pequenos RNAs nucleolares (snoRNAs) atuam principalmente na modificação pós-transcricional de rRNAs e snRNAs. Estas modificações são de dois tipos: metilação 2'-O-ribose e pseudouridilação, e são importantes para a formação correta dos ribossomos, onde ocorre a tradução de proteínas. Além de seu conhecido papel no processo de síntese proteica, função *housekeeping* essencial para o funcionamento celular, recentemente os snoRNAs foram implicados na tumorigênese (MANNOOR et al. 2012; WILLIAMS e FARZANEH 2012).

A maioria dos snoRNAs são transcritos a partir de introns de “genes hospedeiros” codificadores de proteínas e não-codificadores de proteínas. O lncRNA GAS5, por exemplo, codifica 10 snoRNAs diferentes que estão localizados dentro de seus introns (SMITH e STEITZ 1998). A expressão de GAS5 encontra-se frequentemente reduzida em tumores primários de mama em relação ao tecido normal adjacente (MOURTADA-MAARABOUNI et al. 2009), provavelmente levando a uma menor expressão dos snoRNAs contidos em seus introns; sendo que a expressão diminuída de um deles (o SNORD44) no tumor já foi correlacionada com pior sobrevida em uma série de 219 pacientes com câncer de mama (GEE et al. 2011). Um outro snoRNA independente, o SNORD50, também foi encontrado com expressão reduzida em câncer de mama (DONG et al. 2009). Similarmente, uma redução global no perfil de expressão de snoRNAs foi detectada em pacientes com mieloma múltiplo (RONCHETTI et al. 2012). Recentemente, a alta expressão de uma assinatura de 6 snoRNAs (SNORA47, SNORA68, SNORA78, SNORA21, SNORD28 e SNORD66) foi associada com uma pior sobrevida em pacientes com câncer de pulmão de células não-pequenas estágio I (GAO et al. 2015).

Pouco ainda se sabe sobre o significado funcional de snoRNAs em EVs. JENJAROENPUN et al. (2013) analisaram por NGS o transcriptoma de duas linhagens celulares de câncer de mama, MDA-MB-231 e MDA-MB-436,

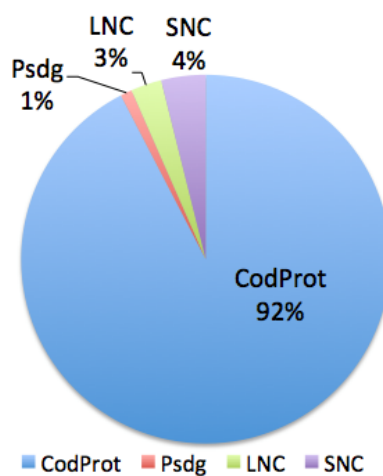
e encontraram os snoRNAs como a classe mais abundante de transcritos não-codificadores de proteína em EVs secretadas por ambas linhagens. O nosso achado de 75 snoRNAs diferencialmente representados entre as EVs da linhagem C5.2 em relação às EVs de HB4a, sendo todos reduzidos nas EVs de C5.2, pode estar relacionado ao processo de transformação maligna das células secretoras, induzida pela super-expressão de HER2.

Comparamos ainda o conteúdo de miRNAs entre as duas linhagens. Somente dois miRNAs foram encontrados diferencialmente expressos com  $p$ -valor  $\leq 0,05$  entre as células C5.2 e HB4a, sendo que ambos estão superexpressos em HB4a. Interessantemente, um deles, o miR-205-5p, foi descrito como sendo reduzido por ERBB2 e pouco expresso em células que passam por transição epitélio-mesênquima (EMT) (ADACHI et al. 2011; GREGORY et al. 2008). A análise dos demais RNAs indicou outros 592 RNAs diferencialmente expressos com  $p$ -valor  $\leq 0,05$  entre as duas linhagens, sendo 184 (31,1%) deles superexpressos em C5.2 e 408 (68,9%) superexpressos em HB4a. A distribuição de categorias destes 592 transcritos diferencialmente expressos encontra-se na **Figura 20**: 92% são genes codificadores de proteína, 4% são *short noncoding*, 3% são *long noncoding* e 1% são *pseudogenes*. Dentre as classes que compõem a categoria *short noncoding*, a mais representada foi a de snRNAs, com 52% dos transcritos. Dos 12 snRNAs diferencialmente expressos, todos (100%) se mostraram superexpressos na célula C5.2 em relação a célula HB4a.

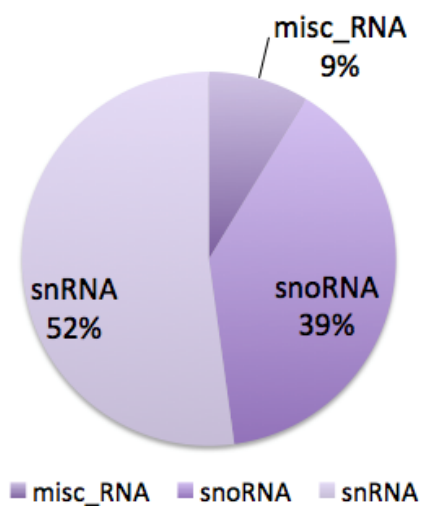
Os pequenos RNAs nucleares (snRNAs) formam parte do complexo denominado spliceossoma que reconhece e remove os introns de precursores de mRNAs, durante o processo de *splicing* (MATERA e WANG 2014). O spliceossoma é formado por mais de 200 proteínas e cinco snRNAs conhecidos como U1, U2, U4, U5 e U6 (WRIGHT e BRUFORD 2011). Sabe-se que o *splicing* alternativo de transcritos pode estar envolvido com a carcinogênese pois determinadas isoformas de vários transcritos são associadas com a progressão tumoral (OLTEAN e BATES 2014). Porém pouco foi investigado sobre o perfil de expressão dos snRNAs em células tumorais. Neste sentido, um estudo detectou um aumento significativo do

snRNA U6 em soro de pacientes com câncer de mama metastático, ou pacientes que tiveram câncer de mama mas que no momento da coleta não tinham doença clinicamente detectável, em relação a indivíduos saudáveis (APPAIAH et al. 2011). Os autores concluem que os níveis elevados de U6 no soro de pacientes que não tinham doença ativa pode refletir mudanças sistêmicas persistentes causadas pelo tumor. Interessantemente, nós detectamos 12 snRNAs diferencialmente expressos entre as células C5.2 e HB4a, sendo todos aumentados na célula C5.2. O snRNA U6, por exemplo, foi encontrado mais de 5X aumentado na célula C5.2 em relação a HB4a, sugerindo que ele pode estar envolvido com o processo de transformação destas células via HER2.

A



B



**Figura 20** - Categorias dos 592 outros RNAs diferencialmente representados entre as células C5.2 e HB4a. **A)** Distribuição de todos os transcritos diferencialmente representados. CodProt = transcritos codificadores de proteína; Psdg = Pseudogenes; LNC = transcritos não-codificadores de proteínas longos (*long noncoding*); SNC = transcritos não-codificadores de proteínas curtos (*short noncoding*). **B)** Distribuição dos transcritos na categoria *short noncoding*.

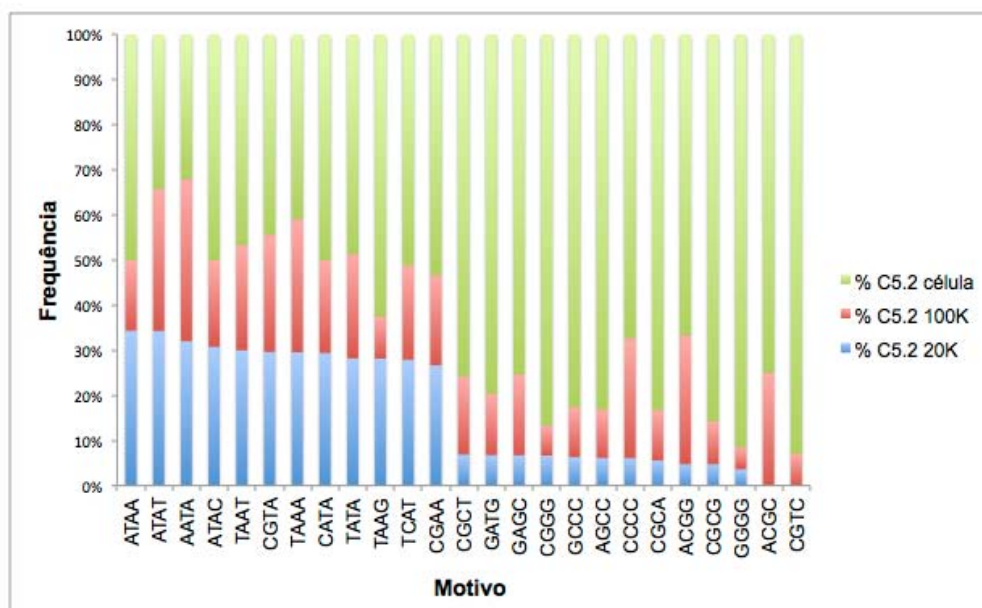
### 4.1.3 Análise de enriquecimento de motivos

Estudos recentes tem buscado identificar possíveis motivos de sequências presentes nas moléculas que são enriquecidas em EVs em relação às células secretoras (BATAGOV et al. 2011; BOLUKBASI et al. 2012; KOPPERS-LALIC et al. 2014; SZOSTAK et al. 2014; VILLARROYA-BELTRI et al. 2013). Tais motivos podem funcionar como “códigos postais” que endereçam as moléculas que os contêm para serem exportadas da célula via EVs, possivelmente por favorecerem que estes RNAs se liguem a proteínas ligantes de RNA específicas. VILLARROYA-BELTRI et al. (2013) descreveram o enriquecimento em miRNAs de motivos de 4 bases que são reconhecidos pela ribonucleoproteína hnRNPA2B1 sumoilada, que controla o carreamento destes miRNAs para os exossomos. Em relação ao trabalho citado, buscamos os motivos de sequências descritos e não observamos os mesmos enriquecidos nos miRNAs derivados de nenhuma das populações de EVs, nem mesmo na fração 100K que é enriquecida em exossomos. Buscamos então a ocorrência de algum motivo predominante em nossos dados, que pudesse indicar outros possíveis motivos envolvidos no mecanismo de seleção preferencial dos miRNAs para as duas populações distintas de EVs (20K e 100K) secretadas pelas células C5.2 e HB4a. Para isto, investigamos a frequência de todos os 256 ( $4^4$ ) possíveis motivos de 4 bases em cada população (20K, 100K e células).

Após verificarmos para cada uma das quatro populações de EVs analisadas (C5.2 20K, C5.2 100K, HB4a 20K, HB4a 100K) que as frequências dos motivos seguem distribuições aproximadamente Gaussianas (normais), nós selecionamos para cada população os motivos cujas frequências nas EVs estão acima/abaixo da média  $\pm 1,64$ (desvio padrão), e que portanto correspondem aos motivos 5% mais/menos frequentes, respectivamente. Interessantemente, os motivos mais/menos frequentes nas EVs se mostraram com tendência oposta nas células. Conforme **Tabela 10 e Figura 21**, na população C5.2 20K, os motivos 5% mais frequentes ( $n=12$ ) tem frequências acima de 26,6%, e os motivos 5% menos frequentes ( $n=13$ ) tem frequências abaixo de 7,64%.

**Tabela 10** – Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 20K da linhagem C5.2.

| Motivo | % C5.2 20K | % C5.2 100K | % C5.2 célula |
|--------|------------|-------------|---------------|
| ATAA   | 34,38      | 15,63       | 50,00         |
| ATAT   | 34,29      | 31,43       | 34,29         |
| AATA   | 32,00      | 36,00       | 32,00         |
| ATAC   | 30,77      | 19,23       | 50,00         |
| TAAT   | 30,00      | 23,33       | 46,67         |
| CGTA   | 29,63      | 25,93       | 44,44         |
| TAAA   | 29,55      | 29,55       | 40,91         |
| CATA   | 29,41      | 20,59       | 50,00         |
| TATA   | 28,21      | 23,08       | 48,72         |
| TAAG   | 28,13      | 9,38        | 62,50         |
| TCAT   | 27,91      | 20,93       | 51,16         |
| CGAA   | 26,67      | 20,00       | 53,33         |
| CGCT   | 6,90       | 17,24       | 75,86         |
| GATG   | 6,78       | 13,56       | 79,66         |
| GAGC   | 6,74       | 17,98       | 75,28         |
| CGGG   | 6,67       | 6,67        | 86,67         |
| GCCC   | 6,35       | 11,11       | 82,54         |
| AGCC   | 6,15       | 10,77       | 83,08         |
| CCCC   | 6,12       | 26,53       | 67,35         |
| CGCA   | 5,56       | 11,11       | 83,33         |
| ACGG   | 4,76       | 28,57       | 66,67         |
| CGCG   | 4,76       | 9,52        | 85,71         |
| GGGG   | 3,66       | 4,88        | 91,46         |
| ACGC   | 0,00       | 25,00       | 75,00         |
| CGTC   | 0,00       | 7,14        | 92,86         |

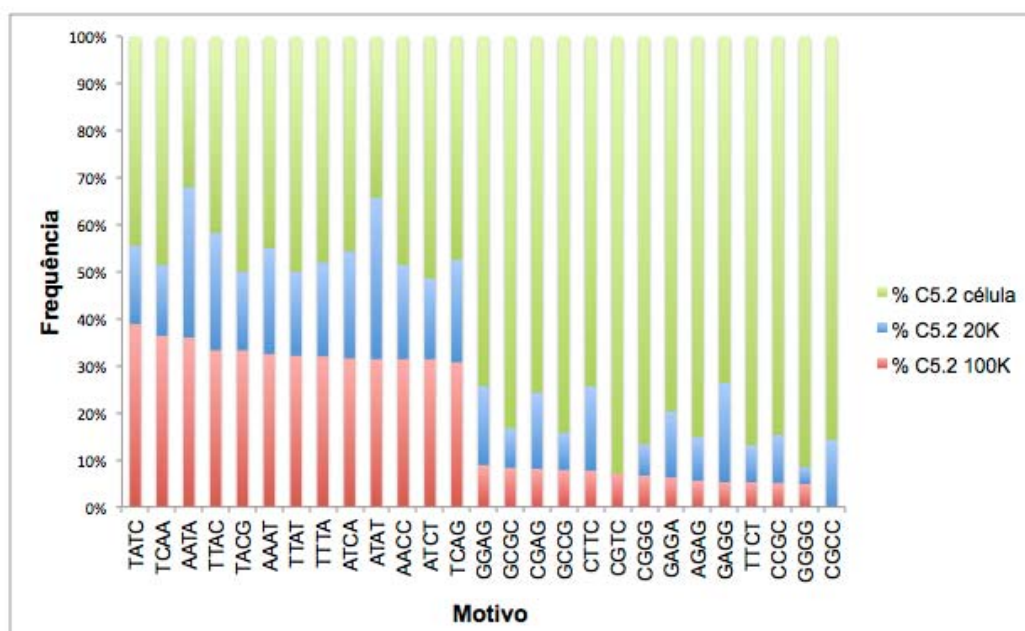


**Figura 21** - Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em C5.2 20K.

Conforme **Tabela 11** e **Figura 22**, na população C5.2 100K os motivos 5% mais frequentes (n=13) tem frequências acima de 30,3%, e os motivos 5% menos frequentes (n=14) tem frequências abaixo de 9.1%.

**Tabela 11** - Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 100K da linhagem C5.2.

| Motivo | % C5.2 20K | % C5.2 100K | % C5.2 célula |
|--------|------------|-------------|---------------|
| TATC   | 16,67      | 38,89       | 44,44         |
| TCAA   | 15,15      | 36,36       | 48,48         |
| AATA   | 32,00      | 36,00       | 32,00         |
| TTAC   | 25,00      | 33,33       | 41,67         |
| TACG   | 16,67      | 33,33       | 50,00         |
| AAAT   | 22,50      | 32,50       | 45,00         |
| TTAT   | 17,86      | 32,14       | 50,00         |
| TTTA   | 20,00      | 32,00       | 48,00         |
| ATCA   | 22,81      | 31,58       | 45,61         |
| ATAT   | 34,29      | 31,43       | 34,29         |
| AACC   | 20,00      | 31,43       | 48,57         |
| ATCT   | 17,14      | 31,43       | 51,43         |
| TCAG   | 21,79      | 30,77       | 47,44         |
| GGAG   | 16,83      | 8,91        | 74,26         |
| GCGC   | 8,33       | 8,33        | 83,33         |
| CGAG   | 16,22      | 8,11        | 75,68         |
| GCCG   | 7,89       | 7,89        | 84,21         |
| CTTC   | 17,95      | 7,69        | 74,36         |
| CGTC   | 0,00       | 7,14        | 92,86         |
| CGGG   | 6,67       | 6,67        | 86,67         |
| GAGA   | 14,06      | 6,25        | 79,69         |
| AGAG   | 9,26       | 5,56        | 85,19         |
| GAGG   | 21,05      | 5,26        | 73,68         |
| TTCT   | 7,89       | 5,26        | 86,84         |
| CCGC   | 10,26      | 5,13        | 84,62         |
| GGGG   | 3,66       | 4,88        | 91,46         |
| CGCC   | 14,29      | 0,00        | 85,71         |



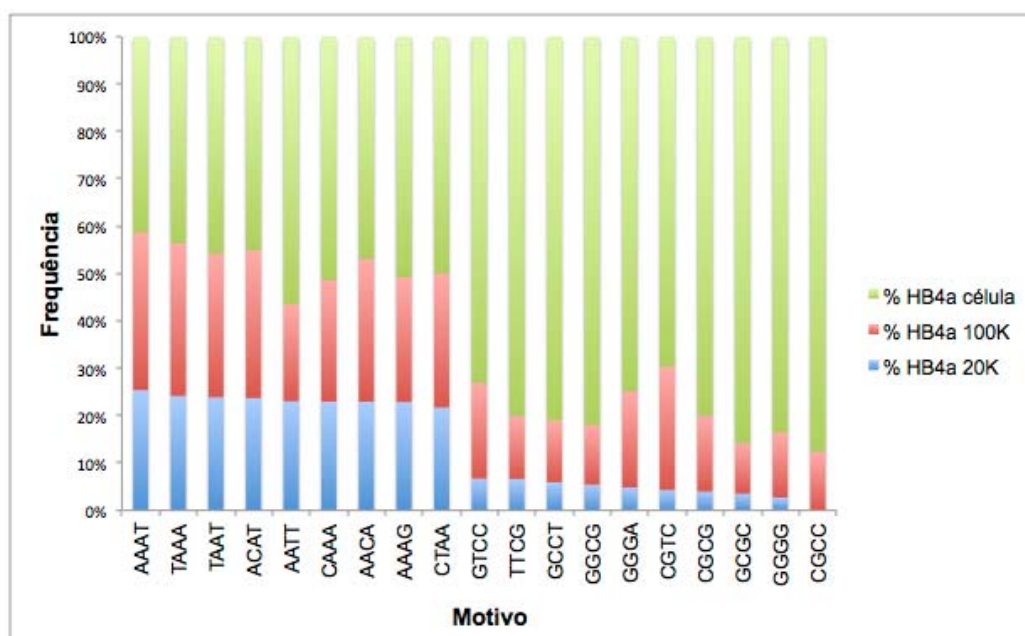
**Figura 22** - Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em C5.2 100K.



Conforme **Tabela 12** e **Figura 23**, na população HB4a 20K os motivos 5% mais frequentes (n=9) tem frequências acima de 21,7%, e os motivos 5% menos frequentes (n=10) tem frequências abaixo de 7,0%.

**Tabela 12** - Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 20K da linhagem HB4a.

| Motivo | % HB4a 20K | % HB4a 100K | % HB4a célula |
|--------|------------|-------------|---------------|
| AAAT   | 25,40      | 33,33       | 41,27         |
| TAAA   | 24,19      | 32,26       | 43,55         |
| TAAT   | 23,91      | 30,43       | 45,65         |
| ACAT   | 23,75      | 31,25       | 45,00         |
| AATT   | 23,08      | 20,51       | 56,41         |
| CAAA   | 22,97      | 25,68       | 51,35         |
| AACA   | 22,94      | 30,28       | 46,79         |
| AAAG   | 22,89      | 26,51       | 50,60         |
| CTAA   | 21,74      | 28,26       | 50,00         |
| GTCC   | 6,76       | 20,27       | 72,97         |
| TTCG   | 6,67       | 13,33       | 80,00         |
| GCCT   | 5,95       | 13,10       | 80,95         |
| GGCG   | 5,45       | 12,73       | 81,82         |
| GGGA   | 4,85       | 20,39       | 74,76         |
| CGTC   | 4,35       | 26,09       | 69,57         |
| CGCG   | 4,00       | 16,00       | 80,00         |
| GCGC   | 3,57       | 10,71       | 85,71         |
| GGGG   | 2,75       | 13,76       | 83,49         |
| CGCC   | 0,00       | 12,50       | 87,50         |

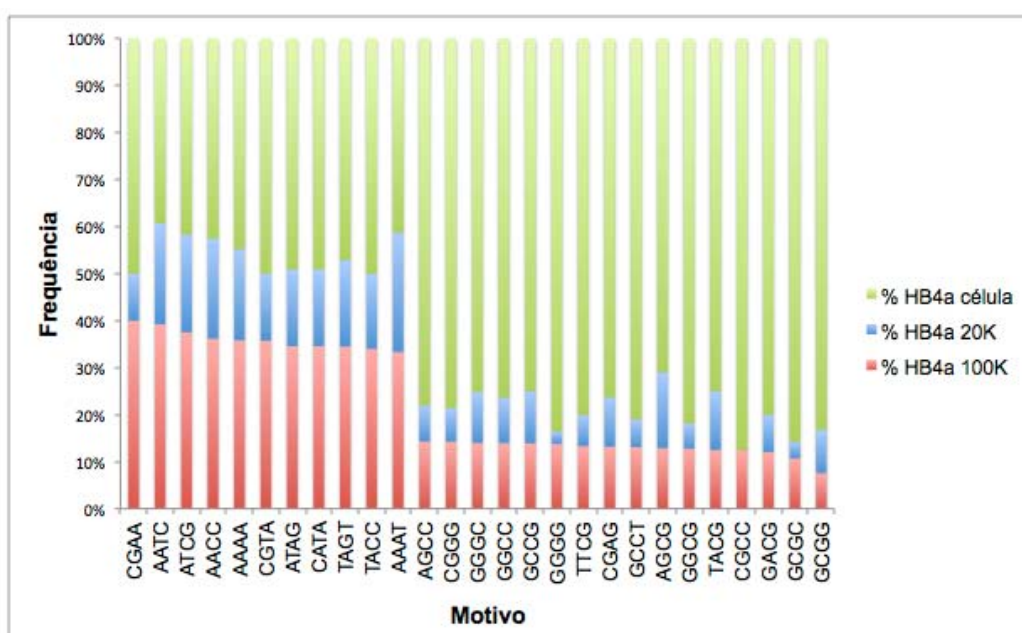


**Figura 23** - Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em HB4a 20K.

Conforme **Tabela 13** e **Figura 24**, na população HB4a 100K os motivos 5% mais frequentes (n=11) tem frequências acima de 33,1%, e os motivos 5% menos frequentes (n=16) tem frequências abaixo de 14,7%.

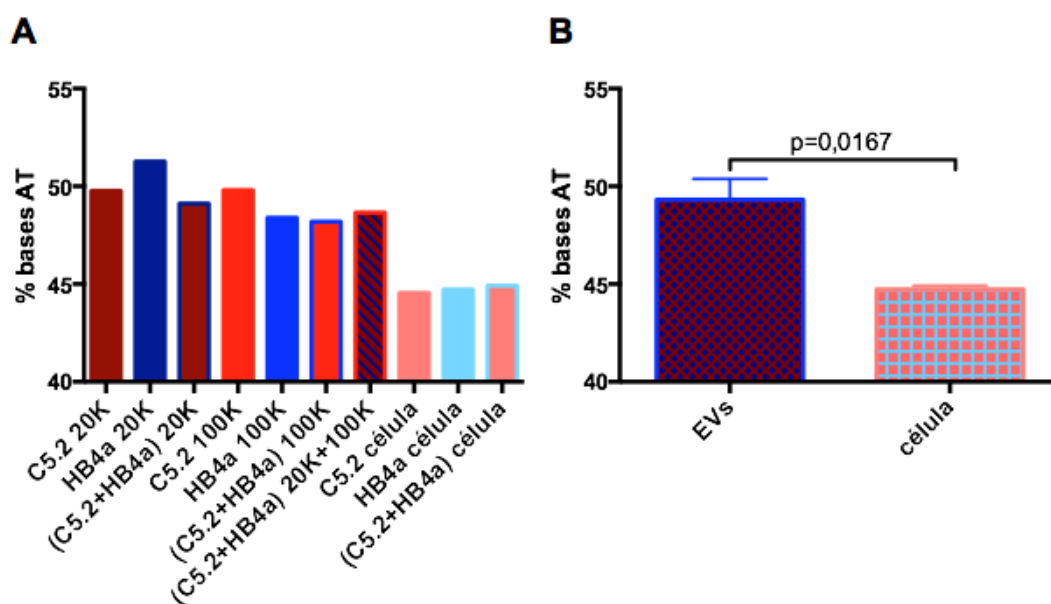
**Tabela 13** - Frequências de motivos enriquecidos e não-enriquecidos ordenados por sua frequência nas EVs 100K da linhagem HB4a.

| Motivo | % HB4a 20K | % HB4a 100K | % HB4a célula |
|--------|------------|-------------|---------------|
| CGAA   | 10,00      | 40,00       | 50,00         |
| AATC   | 21,62      | 39,19       | 39,19         |
| ATCG   | 20,83      | 37,50       | 41,67         |
| AACC   | 21,28      | 36,17       | 42,55         |
| AAAA   | 19,40      | 35,82       | 44,78         |
| CGTA   | 14,29      | 35,71       | 50,00         |
| ATAG   | 16,36      | 34,55       | 49,09         |
| CATA   | 16,36      | 34,55       | 49,09         |
| TAGT   | 18,39      | 34,48       | 47,13         |
| TACC   | 16,00      | 34,00       | 50,00         |
| AAAT   | 25,40      | 33,33       | 41,27         |
| AGCC   | 7,69       | 14,29       | 78,02         |
| CGGG   | 7,14       | 14,29       | 78,57         |
| GGGC   | 10,83      | 14,01       | 75,16         |
| GGCC   | 9,68       | 13,98       | 76,34         |
| GCCG   | 11,11      | 13,89       | 75,00         |
| GGGG   | 2,75       | 13,76       | 83,49         |
| TTCG   | 6,67       | 13,33       | 80,00         |
| CGAG   | 10,53      | 13,16       | 76,32         |
| GCCT   | 5,95       | 13,10       | 80,95         |
| AGCG   | 16,13      | 12,90       | 70,97         |
| GGCG   | 5,45       | 12,73       | 81,82         |
| TACG   | 12,50      | 12,50       | 75,00         |
| CGCC   | 0,00       | 12,50       | 87,50         |
| GACG   | 8,00       | 12,00       | 80,00         |
| GCGC   | 3,57       | 10,71       | 85,71         |
| GCGG   | 9,09       | 7,58        | 83,33         |



**Figura 24** - Representação gráfica de frequências de motivos enriquecidos e não-enriquecidos em HB4a 100K.

Ao compararmos qualitativamente os motivos mais frequentes nas EVs com os motivos mais frequentes nas células, notamos uma clara tendência para enriquecimento nas EVs de motivos ricos em AT e os motivos mais frequentes nas células ricos em GC. Realizamos então uma análise quantitativa da frequência de cada base (A, T, C, G) no conjunto de miRNAs anotados em cada biblioteca de EVs e células, separadamente. Conforme **Figura 25**, observamos uma diferença estatisticamente significativa ( $p=0,0167$ ) no % de bases AT/CG dos miRNAs encontrados nas células e nas EVs. Considerando as duas linhagens, temos em média 50,1% de bases AT nas EVs 20K e 48,8% de bases AT nas EVs 100K, comparado com 44,7% de bases AT nas células C5.2 e HB4a.



**Figura 25** - Comparação do percentual de AT dos miRNAs contidos em EVs e células. **A)** Representação gráfica da porcentagem de bases AT dos miRNAs contidos em cada população indicada. **B)** Em relação ao gráfico A) as 7 primeiras barras foram agrupadas como “EVs” e as três últimas como “célula” para teste estatístico *Mann Whitney* no programa *Prism 6*.

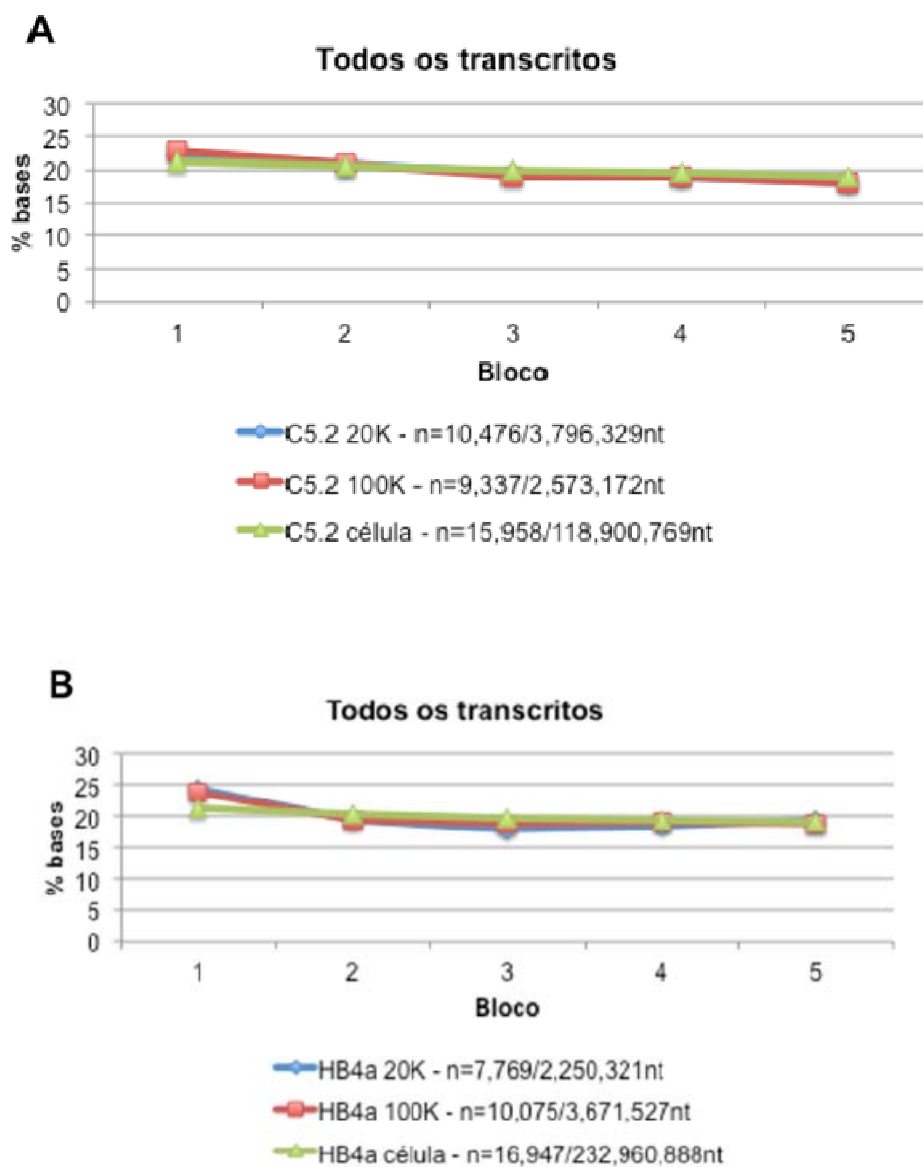


arranjo previamente publicados (plataforma *Agilent 44K whole genome*). Os autores analisaram a distribuição das sondas (de 60 oligonucleotídeos) ao longo dos transcritos que tiveram expressão aumentada nos exossomos em relação às células secretoras. Eles detectaram que os transcritos enriquecidos nos exossomos eram significativamente mais representados por sondas localizadas mais próximas da região 3' dos transcritos. Os achados foram validados por qRT-PCR para três transcritos. A região 3'-UTR de mRNAs é rica em sequências regulatórias, tais como sítios de ligação de miRNAs e de proteínas ligantes de RNA. Estas interações afetam a estabilidade e a eficiência de tradução dos mRNAs. Os autores concluíram que a transferência via exossomos de fragmentos de mRNAs contendo regiões 3'-UTR pode ser um importante mecanismo regulatório à distância, pois se estes fragmentos são internalizados na célula receptora eles podem competir para a ligação de miRNAs. A ligação de miRNAs a estes fragmentos levaria à diminuição dos níveis destas moléculas efetivamente disponíveis para ligação aos seus alvos endógenos, levando a uma desregulação da inibição de expressão gênica. Dado que a região 3'-UTR dos transcritos contêm sítios de ligação para vários miRNAs diferentes, pode-se imaginar que a transferência destas regiões tem o potencial de afetar a capacidade regulatória de certos alvos, de modo simultâneo para vários miRNAs. Experimentalmente já foi demonstrado, por exemplo, que a transfecção da 3'-UTR de CD44 em células da linhagem tumoral de mama MDA-MD-231 levou ao aumento sinérgico de três proteínas (CD44, Col1 $\alpha$ 1 e FN1) que são reguladas por miRNAs que têm sítios de ligação na mesma região do transcrito, resultando em uma maior motilidade, invasão e adesão celular nas células transfectadas, que quando injetadas em camundongos *nude* também possuíam maior capacidade de formação de metástases pulmonares em comparação com células não-transfectadas (YANG e RUTNAM 2012).

De maneira similar, mas utilizando o sequenciamento em larga escala, que é muito mais informativo que a metodologia de micro-arranjo, NOLTE-<sup>t</sup> HOEN et al. (2012) também observaram um enriquecimento em EVs de

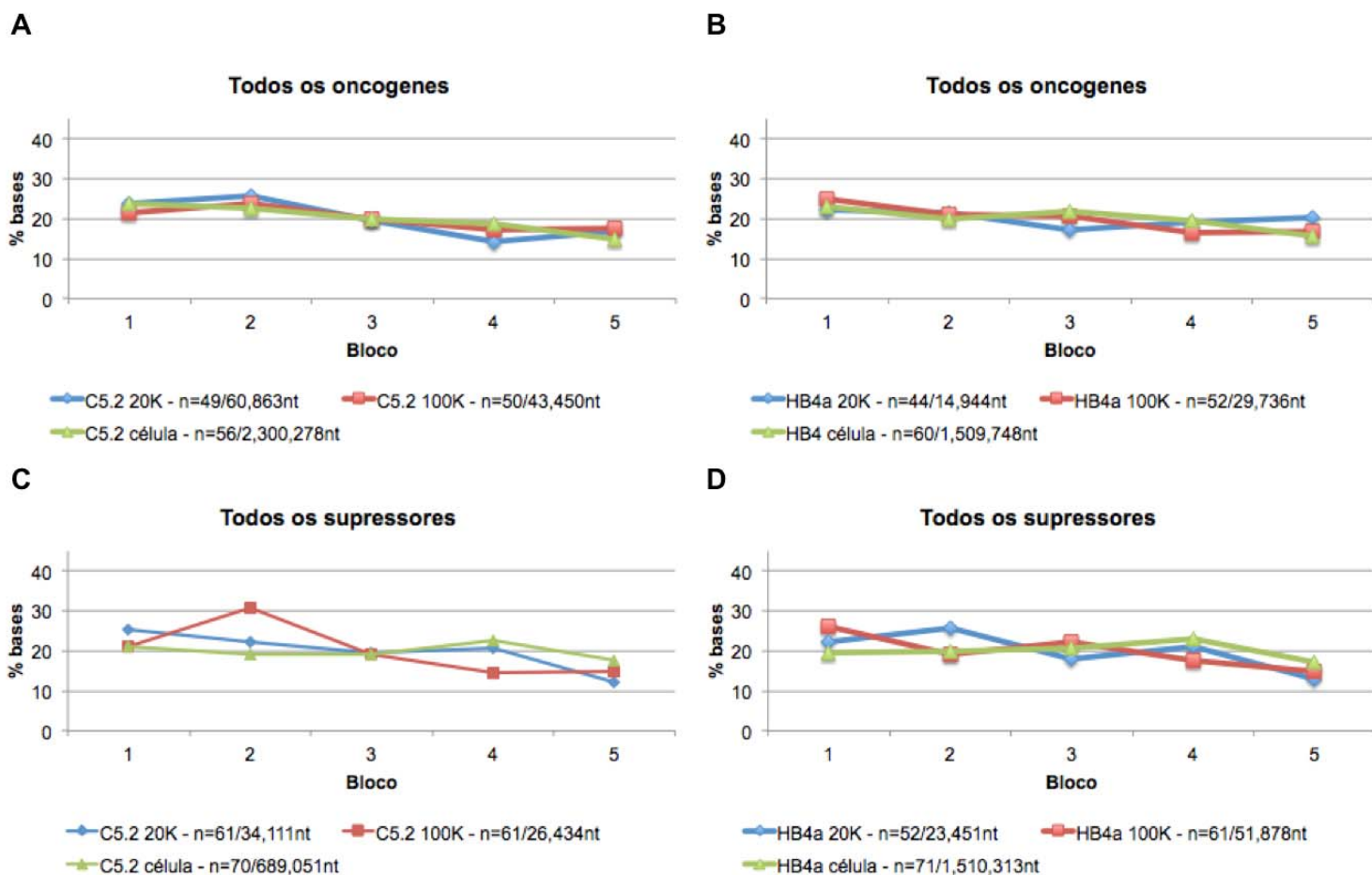
sequências que alinham nas regiões não traduzidas 5' e 3' de genes codificadores de proteína, em comparação com a célula secretora (no caso células dendríticas associadas a linfócitos T).

Nas nossas análises, avaliamos inicialmente a distribuição de bases ao longo de todos os transcritos anotados nas linhagens celulares HB4a e C5.2. Independentemente do tamanho, cada transcrito foi dividido em 5 blocos iguais, sendo o bloco 1 localizado na extremidade 5' do transcrito e o bloco 5 na extremidade 3'. A porcentagem de bases mapeadas em cada bloco foi calculada em relação ao total de bases mapeadas no transcrito todo (soma dos 5 blocos), possibilitando a visualização da distribuição posicional de bases ao longo de cada transcrito. Conforme **Figura 27**, quando avaliamos o conjunto total de transcritos codificadores de proteína anotados em cada biblioteca, vemos uma distribuição uniforme ao longo de todo o transcrito, sendo que cada bloco contém aproximadamente 20% das bases como esperado.



**Figura 27** - Distribuição posicional de todos os transcritos. n: número de transcritos avaliados/ nt: número de bases totais avaliadas.

Partimos então para uma avaliação de conjuntos específicos de transcritos, considerando separadamente oncogenes ou genes supressores tumorais, conforme lista previamente publicada (VOGELSTEIN et al. 2013). Novamente observamos uma distribuição relativamente homogênea das bases ao longo dos transcritos, com poucas variações da frequência de 20% esperada para cada bloco (**Figura 28**).

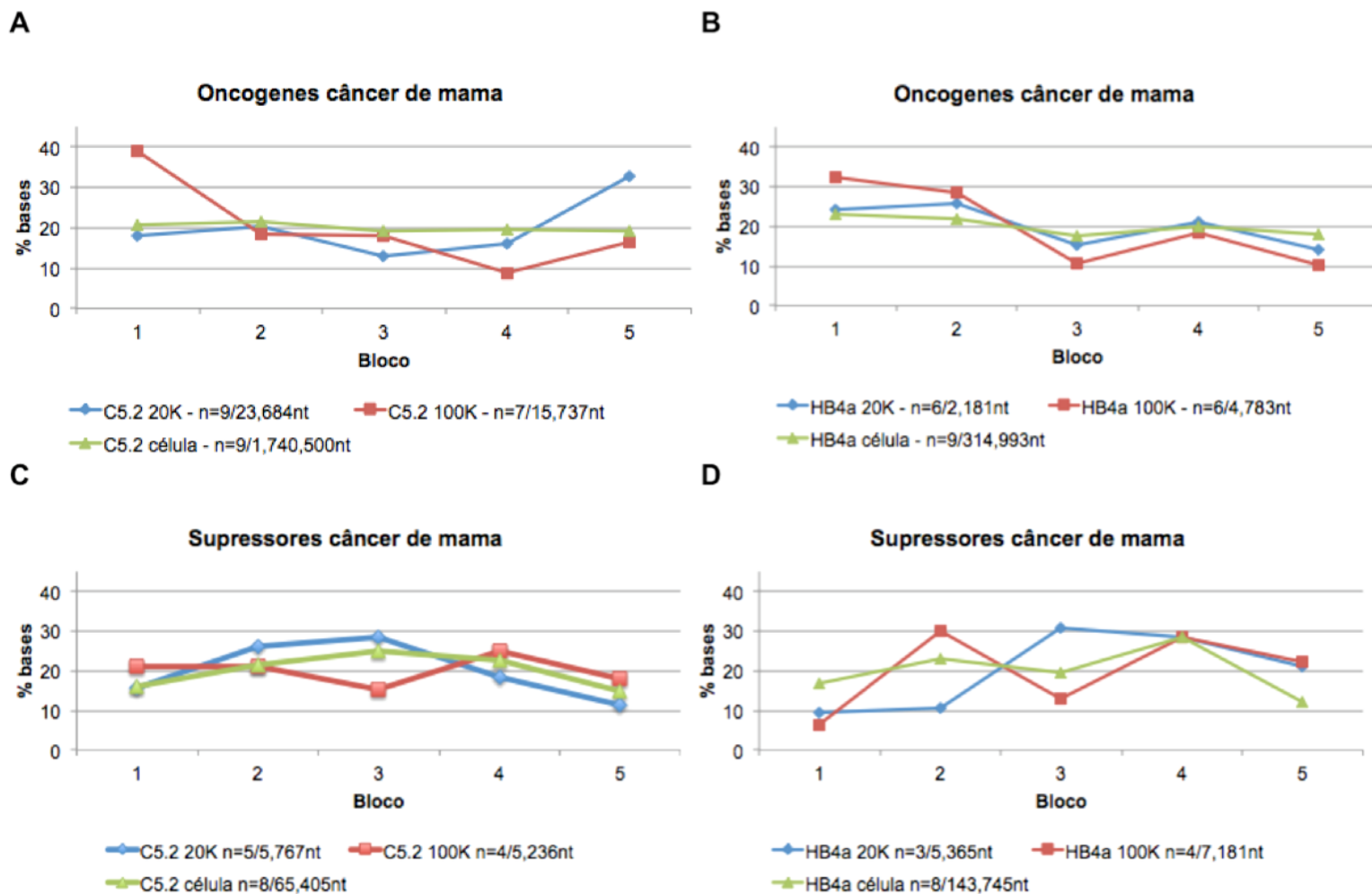


**Figura 28** - Distribuição posicional de oncogenes e genes supressores tumorais.

n: número de transcritos avaliados/ nt: número de bases totais avaliadas.

Avaliamos então, dentre o conjunto de todos os oncogenes e supressores tumorais, somente aqueles descritos como sendo de relevância clínica para o câncer de mama (OSBORNE et al. 2004). Conforme **Figura 29**, observamos um enriquecimento da porção 3' dos oncogenes importantes para a carcinogênese de mama nas EVs 20K derivadas da linhagem C5.2 em relação à célula secretora. Vemos também um enriquecimento da porção 5' dos supressores tumorais nas EVs 100K de C5.2. O possível significado funcional destes achados deve ser confirmado experimentalmente no futuro, mas podemos especular que o enriquecimento de regiões 3'-UTR de oncogenes em EVs de C5.2 possa contribuir para a transformação das células que internalizarem as EVs, uma vez que estas regiões ricas em sítios regulatórios competiriam para a ligação de miRNAs nas células receptoras. Tal ligação levaria a uma diminuição dos níveis de miRNAs livres para reprimir os seus alvos endógenos, efetivamente aumentando a expressão proteica destes oncogenes nas células receptoras.





**Figura 29** - Distribuição posicional de oncogenes e genes supressores tumorais de relevância clínica para o câncer de mama. n: número de transcritos avaliados/ nt: número de bases totais avaliadas.

## 4.2 CARACTERIZAÇÃO DE EVS DERIVADAS DO PLASMA DE PACIENTES

### 4.2.1 Pacientes e amostras

Recrutamos um total de 59 pacientes neste estudo prospectivo, sendo 49 pacientes do A.C. Camargo Cancer Center e 10 pacientes do Hospital de Câncer de Barretos. As pacientes foram divididas em três grupos de modo a refletir a evolução da doença de inicialmente restrita ao tumor primário (Grupo “N0”), avançando para comprometimento linfonodal (Grupo “N”), e finalmente evoluindo para metástases à distância (Grupo “M”). Dez pacientes de cada grupo foram incluídas no sequenciamento em larga escala, de acordo com a quantidade e a qualidade (ausência de hemólise) do plasma disponível. Os dados gerais destas pacientes são apresentados nas **Tabelas 14 e 15**. Em relação ao estadiamento das pacientes, 20,0% eram estágio I, 13,3% estágio IIA, 13,3% estágio IIB, 10,0% estágio IIIA, 10,0% estágio IIIB, e 33,3% estágio IV. Em relação ao subtipo molecular do tumor primário, 20,0% eram subtipo Her2 e 80,0% subtipo Luminal B. Em relação à idade, 60,0% das pacientes tinham menos de 50 anos. Usando uma análise de variância (ANOVA), não observamos diferença estatística na idade média das pacientes entre os três grupos (p-valor = 0,7635). Os dados clínicos de todas as pacientes recrutadas para este estudo encontram-se no **Anexo 6**, sendo indicadas em negrito as pacientes incluídas no sequenciamento.

**Tabela 14** - Resumo dos dados gerais das pacientes do estudo incluídas no sequenciamento em larga escala.

|                | N  | %    |
|----------------|----|------|
| <b>Estádio</b> |    |      |
| I              | 6  | 20,0 |
| IIA            | 4  | 13,3 |
| IIB            | 4  | 13,3 |
| IIIA           | 3  | 10,0 |
| IIIB           | 3  | 10,0 |
| IIIC           | 0  | 0,0  |
| IV             | 10 | 33,3 |
| <b>Subtipo</b> |    |      |
| Her2           | 6  | 20,0 |
| Luminal B      | 24 | 80,0 |
| <b>Idade</b>   |    |      |
| ≤ 50 anos      | 18 | 60,0 |
| > 50 anos      | 12 | 40,0 |

**Tabela 15** - Pacientes incluídas no sequenciamento em larga escala.

| Grupo | Amostra | Idade | Idade média | RE   | RP  | ki67 | HER2 | FISH (HER2/cr17) | Subtipo | Estadiamento |
|-------|---------|-------|-------------|------|-----|------|------|------------------|---------|--------------|
| N0    | Mama 19 | 54    |             | 95%  | 40% | 30%  | 3+   | -                | lumB    | I            |
|       | Mama 22 | 40    |             | 100% | 90% | 15%  | 2+   | 4                | lumB    | I            |
|       | Mama 30 | 48    |             | 5%   | 5%  | 20%  | 3+   | -                | lumB    | I            |
|       | Mama 31 | 48    |             | 90%  | 80% | 2%   | 3+   | -                | lumB    | I            |
|       | Mama 32 | 44    |             | 90%  | 80% | 60%  | 3+   | -                | lumB    | I            |
|       | Mama 44 | 54    |             | neg  | 30% | 30%  | 3+   | -                | lumB    | I            |
|       | Mama 33 | 36    |             | 90%  | 90% | 30%  | 3+   | -                | lumB    | IIA          |
|       | Mama 34 | 65    |             | neg  | neg | 50%  | 3+   | -                | Her2    | IIA          |
|       | Mama 36 | 28    |             | 95%  | 20% | 60%  | 3+   | -                | lumB    | IIA          |
|       | Mama 26 | 66    | 48.3        | neg  | 10% | 70%  | 3+   | -                | lumB    | IIA          |
| N     | Mama 38 | 40    |             | 90%  | 70% | 60%  | 3+   | -                | lumB    | IIB          |
|       | Mama 18 | 56    |             | neg  | 30% | 60%  | 3+   | -                | lumB    | IIB          |
|       | Mama 25 | 64    |             | 5%   | neg | 30%  | 3+   | -                | lumB    | IIB          |
|       | Mama 42 | 37    |             | 80%  | 10% | 70%  | 3+   | -                | lumB    | IIB          |
|       | Mama 24 | 66    |             | neg  | neg | 50%  | 3+   | -                | Her2    | IIIA         |
|       | Mama 28 | 46    |             | 70%  | 5%  | 40%  | 3+   | -                | lumB    | IIIA         |
|       | Mama 29 | 42    |             | 95%  | 95% | 40%  | 3+   | -                | lumB    | IIIA         |
|       | Mama 6  | 45    |             | 90%  | 5%  | 30%  | 3+   | -                | lumB    | IIIB         |
|       | Mama 41 | 42    |             | 50%  | neg | 20%  | 3+   | -                | lumB    | IIIB         |
|       | Mama 48 | 53    | 49.1        | 60%  | 60% | 80%  | 3+   | -                | lumB    | IIIB         |
| M     | Mama 21 | 42    |             | 70%  | 70% | 50%  | 3+   | -                | lumB    | IV           |
|       | Mama 37 | 53    |             | neg  | neg | 60%  | 3+   | -                | Her2    | IV           |
|       | Mama 46 | 49    |             | 2%   | 20% | 80%  | 3+   | -                | lumB    | IV           |
|       | Mama 50 | 34    |             | 60%  | 30% | 90%  | 3+   | -                | lumB    | IV           |
|       | Mama 51 | 42    |             | 100% | neg | pos  | 3+   | 3.5              | lumB    | IV           |
|       | Mama 52 | 34    |             | neg  | neg | 80%  | 3+   | -                | Her2    | IV           |
|       | Mama 53 | 51    |             | neg  | neg | pos  | 3+   | -                | Her2    | IV           |
|       | Mama 54 | 51    |             | neg  | 40% | 85%  | 3+   | -                | lumB    | IV           |
|       | Mama 55 | 55    |             | neg  | neg | 90%  | 3+   | -                | Her2    | IV           |
|       | Mama 56 | 48    | 45.9        | pos  | pos | 80%  | 3+   | -                | lumB    | IV           |

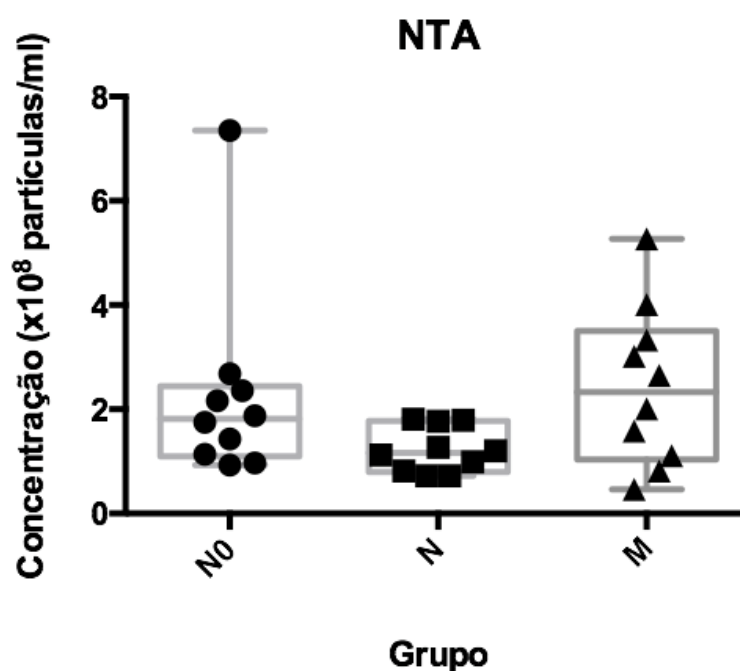
**Legenda** – Estadiamento de acordo com TNM.

**Fonte:** SOBIN et al. (2012)

Todas as pacientes são HER2+, subtipo Her2: ER/PR- e subtipo lumB: ER+ e/ou PR+.

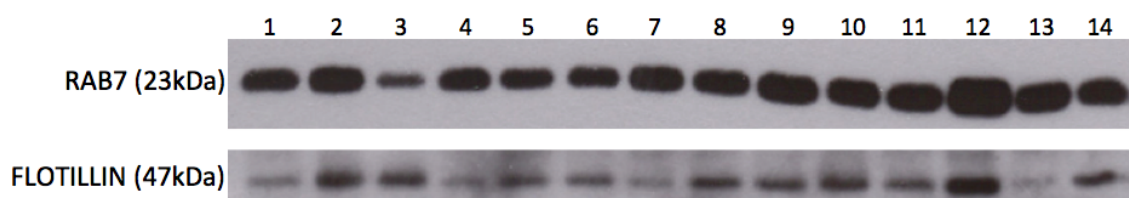
#### 4.2.2 Caracterização das EVs derivadas do plasma de pacientes por NTA e Western Blot

Utilizamos *Nanotracking Particle Analysis* (NTA) para investigar se havia diferença na concentração de EVs circulantes no plasma das pacientes dos três grupos. Isolamos as EVs por ultracentrifugação a partir de 100µl de plasma, ressuspensando o *pellet* em PBS filtrado, e realizamos leituras (de 60 segundos cada) em triplicata para cada amostra no equipamento NanoSight LM-10. Os parâmetros da câmera foram mantidos iguais para todas as amostras analisadas. Usando uma análise de variância (ANOVA), não observamos diferença estatística nas concentrações de EVs entres os grupos (p-valor = 0,1394), ressaltando a importância do conteúdo das EVs e não necessariamente da sua quantidade (**Figura 30**).



**Figura 30** - Avaliação por *Nanotracking Particle Analysis* (NTA) da concentração de EVs isoladas do plasma de pacientes dos três grupos de agressividade tumoral.

Devido ao grande número de pacientes que compõem os grupos e à quantidade limitante de plasma disponível para a realização de todos os experimentos deste trabalho, optamos por fazer as análises de *Western Blot* utilizando *pools* de amostras de plasma de três pacientes dentro de cada grupo de estadiamento tumoral. As EVs foram isoladas a partir de 4,5ml de plasma de três pacientes (1,5ml de cada) e dois marcadores clássicos de EVs, RAB7 e Flotilina, tiveram sua expressão avaliada. Apesar de algumas variações quantitativas, podemos observar a expressão de ambos marcadores em todos os conjuntos de amostras testadas, comprovando a eficácia do protocolo de isolamento utilizado para o enriquecimento de EVs (**Figura 31**).

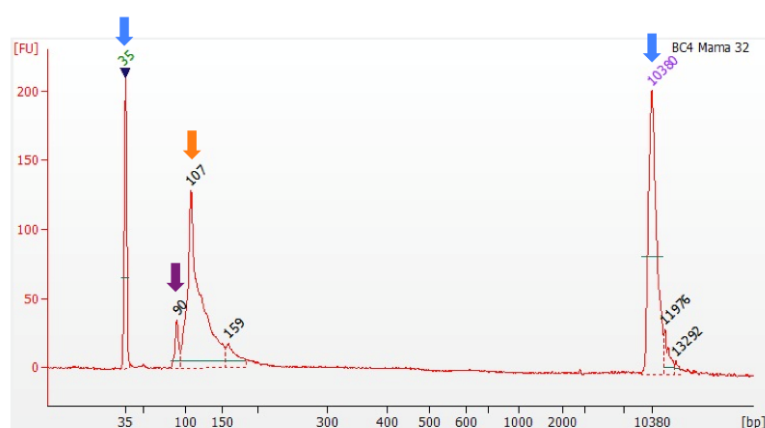


**Figura 31** - Imunodeteção por Western Blot dos marcadores de EVs RAB7 e Flotilina em EVs derivadas de *pools* de plasma de pacientes. 1: *Pool* de pacientes com tumores *in situ*, 2 a 7: *Pools* de pacientes do grupo N0, 8 a 13: *Pools* de pacientes do grupo N, 14: *Pool* de pacientes do grupo M. n=3 em cada *pool*.

#### 4.2.3 Caracterização do conjunto de RNAs de EVs derivadas do plasma de pacientes por NGS

Semelhante ao descrito acima para a análise das EVs derivadas do meio de cultura celular condicionado, isolamos as EVs por ultracentrifugação a partir de 2,5ml de plasma de cada paciente (n=10 para cada grupo), extraímos o RNA total, e construímos bibliotecas para a avaliação de todo o transcriptoma. A verificação do tamanho final das bibliotecas foi feita no equipamento *Bioanalyzer* utilizando o *High Sensitivity DNA Kit* (Agilent Technologies, California, EUA), que confirmou o tamanho de 107pb para o maior pico, que está dentro da faixa de tamanho esperado

de 94-114pb para fragmentos do tamanho de miRNAs ligados aos adaptadores (perfil representativo apresentado na **Figura 32** e os demais no **Anexo 7**, onde vemos que o tamanho do maior pico variou de 104 a 108pb nas 30 bibliotecas). Para as bibliotecas de pacientes observamos um pico de 90pb que corresponde a dímeros de adaptadores, possivelmente devido à quantidade ainda mais limitante de RNA que conseguimos isolar das EVs provenientes do plasma, em relação às EVs das linhagens.



**Figura 32** - Perfil representativo do *Bioanalyzer* de biblioteca para sequenciamento, obtido a partir de RNAs em EVs derivadas do plasma de paciente. As setas azuis correspondem a marcadores de tamanho (conforme eixo X), a seta laranja indica o pico obtido que corresponde ao tamanho esperado (107pb) após a ligação de adaptadores aos RNAs, seguido de PCR com a adição dos *barcodes*, e a seta roxa indica o pico obtido que corresponde ao tamanho esperado (90pb) de dímeros de adaptadores (que não contêm fragmentos de RNA de interesse). Paciente = Mama 32 (Grupo N0).

A seguir, construímos e sequenciamos na plataforma Ion Proton trinta bibliotecas, gerando no total cerca de 173 milhões de sequências (em média 5,7 milhões de sequências por paciente), com tamanho médio de 36,6nt, desvio padrão 22,6nt e variando entre 8-354nt. Conforme **Tabela 16**, em média 85% das sequências foram filtradas pelo FastQ Screen, sendo a maioria (51,9%) excluída por representarem rRNAs, de modo semelhante ao encontrado em EVs das linhagens celulares. Deste modo obtivemos uma média de 868,615 sequências para cada paciente, o que talvez seja

suficiente para representar pelo menos os transcritos mais abundantes desta fração.

Para a anotação dos transcriptoma, iniciamos avaliando os miRNAs maduros com o programa *miRDeep2*, utilizando para isto somente as sequências com tamanho entre 12 e 25 bases, o que corresponde a 232.590 sequências por paciente. Obtivemos uma média de 14.847 sequências por paciente anotando no miRBase e representando em média 167 miRNAs diferentes. Ao adotarmos um critério de pelo menos 5 *reads* comprovando um dado miRNA, temos então uma média de 93 miRNAs diferentes por paciente.

Esta quantidade de miRNAs identificados (93 a 167 miRNAs/paciente, conforme o critério de cobertura mínima adotado) está na mesma ordem de grandeza descrita na literatura. Os dois trabalhos principais de sequenciamento de miRNAs de EVs derivadas do plasma de indivíduos saudáveis identificaram aproximadamente 300 miRNAs, adotando um critério de no mínimo 5 reads por milhão (CHENG et al. 2014; HUANG et al. 2013). Variações na metodologia empregada podem explicar a diferença encontrada entre os nossos achados e os artigos mencionados. Por exemplo, um dos artigos utiliza uma resina comercial (ExoQuick) e não ultracentrifugação para isolar as EVs, possivelmente co-precipitando miRNAs livres, não-associados à EVs.

**Tabela 16** - Análise dos dados de sequenciamento em larga escala de miRNAs de EVs derivadas do plasma de pacientes.

| Grupo       | Amostra   | Total reads     | # reads após filtro FastQ Screen (%) | # reads usados miRDeep2 (12-25nt) | # reads mapeados miRBase | # miRNAs ≥ 1 read | # miRNAs ≥ 2 reads | # miRNAs ≥ 3 reads | # miRNAs ≥ 4 reads | # miRNAs ≥ 5 reads |
|-------------|-----------|-----------------|--------------------------------------|-----------------------------------|--------------------------|-------------------|--------------------|--------------------|--------------------|--------------------|
| NO          | Mama_19   | 5.429.152       | 756.132 (13,9%)                      | 212.785                           | 7.630                    | 145               | 115                | 98                 | 90                 | 82                 |
|             | Mama_22   | 7.297.514       | 898.097 (12,3%)                      | 249.614                           | 12.566                   | 153               | 122                | 104                | 93                 | 85                 |
|             | Mama_30   | 8.166.615       | 980.806 (12,0%)                      | 292.643                           | 15.900                   | 146               | 128                | 117                | 99                 | 93                 |
|             | Mama_32   | 7.483.101       | 843.234 (11,3%)                      | 232.822                           | 11.429                   | 153               | 122                | 105                | 95                 | 90                 |
|             | Mama_33   | 7.425.578       | 806.140 (10,9%)                      | 181.755                           | 7.844                    | 133               | 104                | 85                 | 76                 | 70                 |
|             | Mama_44   | 6.647.837       | 876.841 (13,2%)                      | 225.598                           | 13.508                   | 195               | 150                | 127                | 113                | 101                |
|             | Mama_26   | 6.618.094       | 902.057 (13,6%)                      | 260.494                           | 25.888                   | 179               | 141                | 121                | 109                | 100                |
|             | Mama_34   | 5.880.692       | 1.004.346 (17,1%)                    | 266.203                           | 9.206                    | 150               | 124                | 108                | 95                 | 85                 |
|             | Mama_36   | 7.417.349       | 894.304 (12,1%)                      | 244.235                           | 11.484                   | 148               | 111                | 98                 | 88                 | 81                 |
|             | Mama_31   | 4.926.499       | 731.592 (14,9%)                      | 229.166                           | 19.047                   | 200               | 162                | 141                | 123                | 111                |
|             | média     | 6.729.243       | 869.355 (12,9%)                      | 239.532                           | 13.450                   | 160               | 128                | 110                | 98                 | 90                 |
| N           | Mama_18   | 4.990.888       | 783.360 (15,7%)                      | 215.936                           | 11.414                   | 129               | 108                | 97                 | 86                 | 78                 |
|             | Mama_25   | 6.358.811       | 770.107 (12,1%)                      | 214.745                           | 16.375                   | 164               | 128                | 112                | 98                 | 92                 |
|             | Mama_38   | 8.355.382       | 917.173 (11,0%)                      | 225.553                           | 10.454                   | 162               | 129                | 106                | 95                 | 89                 |
|             | Mama_6    | 4.949.083       | 1.016.536 (20,5%)                    | 238.215                           | 26.248                   | 244               | 192                | 157                | 140                | 133                |
|             | Mama_28   | 6.214.640       | 945.230 (15,2%)                      | 244.015                           | 11.914                   | 157               | 125                | 107                | 99                 | 88                 |
|             | Mama_29   | 7.512.571       | 1.099.436 (14,6%)                    | 286.987                           | 18.198                   | 183               | 143                | 122                | 106                | 91                 |
|             | Mama_41   | 7.365.086       | 1.066.597 (14,5%)                    | 290.766                           | 18.350                   | 151               | 130                | 114                | 106                | 94                 |
|             | Mama_42   | 4.947.052       | 757.784 (15,3%)                      | 211.188                           | 10.186                   | 164               | 128                | 106                | 93                 | 88                 |
|             | Mama_48   | 4.112.776       | 711.702 (17,3%)                      | 209.469                           | 5.689                    | 134               | 104                | 85                 | 80                 | 73                 |
|             | Mama_24   | 6.162.086       | 926.697 (15,0%)                      | 249.329                           | 11.621                   | 151               | 120                | 102                | 90                 | 82                 |
|             | média     | 6.096.838       | 899.464 (14,8%)                      | 238.620                           | 14.045                   | 164               | 131                | 111                | 99                 | 91                 |
| M           | Mama_21   | 4.669.306       | 772.408 (16,5%)                      | 207.189                           | 6.957                    | 131               | 104                | 94                 | 85                 | 79                 |
|             | Mama_37   | 5.387.916       | 1.046.234 (19,4%)                    | 269.479                           | 11.125                   | 165               | 127                | 110                | 99                 | 91                 |
|             | Mama_46   | 3.981.004       | 680.919 (17,1%)                      | 172.802                           | 3.946                    | 126               | 98                 | 83                 | 72                 | 64                 |
|             | Mama_50   | 5.967.079       | 895.349 (15,0%)                      | 240.578                           | 21.897                   | 192               | 144                | 126                | 110                | 99                 |
|             | Mama_51   | 5.209.320       | 1.137.648 (21,8%)                    | 303.305                           | 58.955                   | 264               | 217                | 182                | 167                | 158                |
|             | Mama_52   | 4.141.561       | 819.962 (19,8%)                      | 213.082                           | 9.573                    | 183               | 143                | 119                | 102                | 96                 |
|             | Mama_53   | 4.383.627       | 805.556 (18,4%)                      | 221.387                           | 9.990                    | 170               | 131                | 114                | 101                | 98                 |
|             | Mama_54   | 3.679.994       | 750.821 (20,4%)                      | 176.406                           | 14.092                   | 191               | 143                | 117                | 107                | 99                 |
|             | Mama_55   | 2.333.196       | 473.705 (20,3%)                      | 120.876                           | 19.497                   | 163               | 132                | 114                | 95                 | 85                 |
|             | Mama_56   | 5.087.252       | 987.675 (19,4%)                      | 271.073                           | 14.425                   | 194               | 157                | 129                | 113                | 102                |
|             | média     | 4.484.026       | 837.028 (18,7%)                      | 219.618                           | 17.046                   | 178               | 140                | 119                | 105                | 97                 |
| média geral | 5.770.035 | 868.615 (15,1%) | 232.590                              | 14.847                            | 167                      | 133               | 113                | 101                | 93                 |                    |

Em relação à anotação dos demais transcritos (**Tabela 17**), de 853.768 sequências obtidas em média por paciente, 268.320 anotaram no banco de dados Ensembl, correspondendo em média a 6.787 transcritos codificadores de proteínas, 20 snoRNAs, 14 snRNAs, 151 lincRNAs e 213 outros transcritos classificados na categoria miscelânea (misc), considerando todos os transcritos com FPKM >0.



**Tabela 17** - Análise dos dados de sequenciamento em larga escala de outros RNAs de EVs derivadas do plasma de pacientes.

| Grupo       | Amostra | # reads não anotados no miRDeep2 | # reads usados Ensembl ( $\geq 20$ nt) | # reads mapeados Ensembl | # transcritos CodProt | # transcritos snoRNA | # transcritos snRNA | # transcritos lincRNA | # transcritos misc_RNA |
|-------------|---------|----------------------------------|--|--------------------------|-----------------------|----------------------|---------------------|-----------------------|------------------------|
| NO          | Mama_19 | 748.502                          | 600.506                                | 300.692                  | 6.501                 | 12                   | 10                  | 147                   | 198                    |
|             | Mama_22 | 885.531                          | 696.881                                | 300.007                  | 7.626                 | 38                   | 15                  | 148                   | 190                    |
|             | Mama_30 | 964.906                          | 758.671                                | 294.704                  | 7.170                 | 19                   | 14                  | 159                   | 190                    |
|             | Mama_32 | 831.805                          | 688.804                                | 250.930                  | 7.324                 | 25                   | 22                  | 145                   | 194                    |
|             | Mama_33 | 798.296                          | 703.003                                | 287.132                  | 7.838                 | 17                   | 16                  | 151                   | 185                    |
|             | Mama_44 | 863.333                          | 713.226                                | 293.553                  | 7.714                 | 26                   | 15                  | 169                   | 242                    |
|             | Mama_26 | 876.169                          | 630.722                                | 241.890                  | 7.726                 | 16                   | 12                  | 173                   | 212                    |
|             | Mama_34 | 995.140                          | 776.979                                | 303.444                  | 7.975                 | 11                   | 14                  | 185                   | 235                    |
|             | Mama_36 | 882.820                          | 684.155                                | 230.008                  | 6.884                 | 16                   | 18                  | 137                   | 203                    |
|             | Mama_31 | 712.545                          | 488.540                                | 193.031                  | 5.903                 | 22                   | 8                   | 128                   | 179                    |
|             | média   | 855.905                          | 674.149                                | 269.539                  | 7.266                 | 20                   | 14                  | 154                   | 203                    |
| N           | Mama_18 | 771.946                          | 554.082                                | 205.125                  | 6.698                 | 19                   | 17                  | 136                   | 175                    |
|             | Mama_25 | 753.732                          | 585.537                                | 224.111                  | 7.147                 | 25                   | 10                  | 148                   | 194                    |
|             | Mama_38 | 906.719                          | 779.403                                | 304.850                  | 7.784                 | 19                   | 22                  | 158                   | 210                    |
|             | Mama_6  | 990.288                          | 816.517                                | 528.195                  | 5.851                 | 19                   | 11                  | 126                   | 278                    |
|             | Mama_28 | 933.316                          | 613.461                                | 226.951                  | 6.847                 | 13                   | 16                  | 159                   | 181                    |
|             | Mama_29 | 1.081.238                        | 726.492                                | 283.287                  | 7.434                 | 22                   | 18                  | 176                   | 205                    |
|             | Mama_41 | 1.048.247                        | 642.146                                | 230.872                  | 5.720                 | 11                   | 11                  | 133                   | 201                    |
|             | Mama_42 | 747.598                          | 512.687                                | 214.888                  | 6.534                 | 18                   | 16                  | 153                   | 189                    |
|             | Mama_48 | 706.013                          | 462.359                                | 147.759                  | 4.965                 | 9                    | 7                   | 129                   | 150                    |
|             | Mama_24 | 915.076                          | 573.623                                | 201.708                  | 6.076                 | 20                   | 9                   | 135                   | 192                    |
|             | média   | 885.417                          | 626.631                                | 256.775                  | 6.506                 | 18                   | 14                  | 145                   | 198                    |
| M           | Mama_21 | 765.451                          | 468.166                                | 150.342                  | 5.434                 | 10                   | 6                   | 120                   | 162                    |
|             | Mama_37 | 1.035.109                        | 777.372                                | 371.891                  | 7.519                 | 29                   | 14                  | 166                   | 241                    |
|             | Mama_46 | 676.973                          | 515.888                                | 159.143                  | 6.020                 | 16                   | 9                   | 125                   | 180                    |
|             | Mama_50 | 873.452                          | 665.502                                | 235.618                  | 6.712                 | 20                   | 15                  | 164                   | 195                    |
|             | Mama_51 | 1.078.693                        | 968.963                                | 655.501                  | 7.480                 | 56                   | 28                  | 216                   | 475                    |
|             | Mama_52 | 810.389                          | 596.877                                | 259.411                  | 8.417                 | 21                   | 16                  | 205                   | 223                    |
|             | Mama_53 | 795.566                          | 500.370                                | 157.391                  | 5.881                 | 13                   | 18                  | 135                   | 221                    |
|             | Mama_54 | 736.729                          | 577.605                                | 354.885                  | 6.432                 | 16                   | 15                  | 136                   | 266                    |
|             | Mama_55 | 454.208                          | 360.387                                | 193.520                  | 4.926                 | 22                   | 8                   | 93                    | 182                    |
|             | Mama_56 | 973.250                          | 651.913                                | 248.767                  | 7.065                 | 25                   | 17                  | 184                   | 252                    |
|             | média   | 819.982                          | 608.304                                | 278.647                  | 6.589                 | 23                   | 15                  | 154                   | 240                    |
| média geral | 853.768 | 636.361                          | 268.320                                | 6.787                    | 20                    | 14                   | 151                 | 213                   |                        |

Após análise detalhada dos *reads* que não anotaram em nenhum banco e re-mapeamento contra rRNAs, vemos que em média 29% dos reads das bibliotecas de pacientes com qualidade ( $\geq Q20$ ) permanecem ao final do *pipeline* sem anotação (Tabela 18).

**Tabela 18** - Análise dos *reads* de EVs derivadas do plasma de pacientes que não anotaram após alinhamento contra miRBase e hg19/Ensembl.

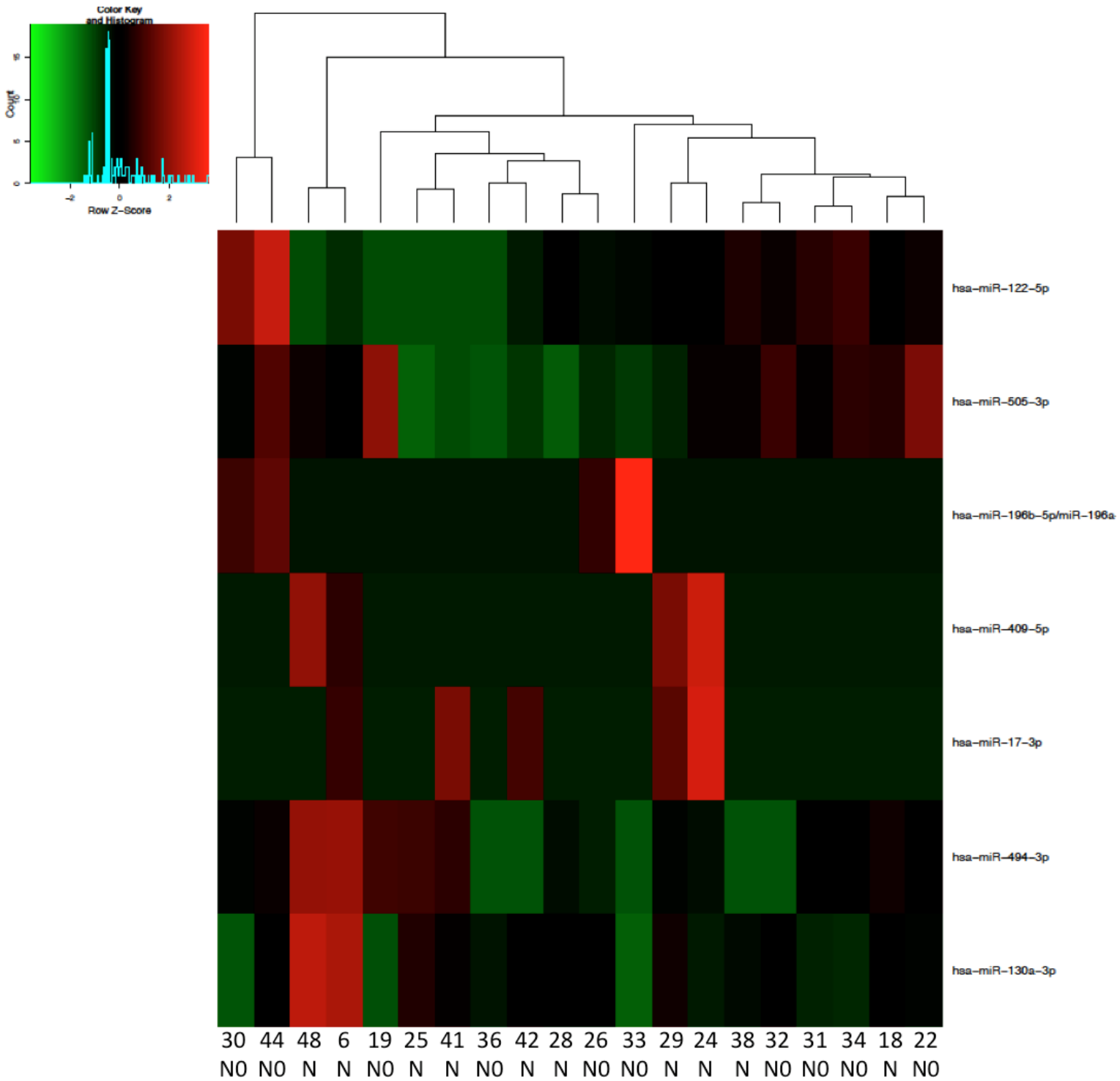
| Grupo       | Amostra | # reads não anotados no Ensembl | # reads mapeados rRNA | # reads que não anotam | % reads que não anotam | # reads que não anotam e tem qualidade (≥Q20) | % reads que não anotam e tem qualidade (≥Q20) |
|-------------|---------|---------------------------------|-----------------------|------------------------|------------------------|---|---|
| NO          | Mama_19 | 447.810                         | 148.249               | 299.561                | 66,9                   | 144.916                                       | 32,4  |
|             | Mama_22 | 585.524                         | 182.010               | 403.514                | 68,9                   | 192.039                                       | 32,8  |
|             | Mama_30 | 670.202                         | 176.197               | 494.005                | 73,7                   | 266.653                                       | 39,8  |
|             | Mama_32 | 580.875                         | 145.739               | 435.136                | 74,9                   | 223.104                                       | 38,4  |
|             | Mama_33 | 511.164                         | 116.076               | 395.088                | 77,3                   | 171.353                                       | 33,5  |
|             | Mama_44 | 569.780                         | 148.764               | 421.016                | 73,9                   | 175.801                                       | 30,9  |
|             | Mama_26 | 634.279                         | 218.557               | 415.722                | 65,5                   | 177.546                                       | 28,0  |
|             | Mama_34 | 691.696                         | 216.589               | 475.107                | 68,7                   | 139.322                                       | 20,1  |
|             | Mama_36 | 652.812                         | 176.612               | 476.200                | 72,9                   | 225.179                                       | 34,5  |
|             | Mama_31 | 519.514                         | 173.682               | 345.832                | 66,6                   | 160.774                                       | 30,9  |
|             | média   | 586.366                         | 170.248               | 416.118                | 70,9                   | 187.669                                       | 32,1  |
| N           | Mama_18 | 566.821                         | 189.643               | 377.178                | 66,5                   | 178.983                                       | 31,6  |
|             | Mama_25 | 529.621                         | 153.309               | 376.312                | 71,1                   | 178.516                                       | 33,7  |
|             | Mama_38 | 601.869                         | 146.410               | 455.459                | 75,7                   | 210.453                                       | 35,0  |
|             | Mama_6  | 462.093                         | 151.151               | 310.942                | 67,3                   | 125.326                                       | 27,1  |
|             | Mama_28 | 706.365                         | 268.636               | 437.729                | 62,0                   | 181.605                                       | 25,7  |
|             | Mama_29 | 797.951                         | 319.878               | 478.073                | 59,9                   | 217.671                                       | 27,3  |
|             | Mama_41 | 817.375                         | 312.092               | 505.283                | 61,8                   | 266.571                                       | 32,6  |
|             | Mama_42 | 532.710                         | 209.233               | 323.477                | 60,7                   | 138.284                                       | 26,0  |
|             | Mama_48 | 558.254                         | 186.701               | 371.553                | 66,6                   | 160.976                                       | 28,8  |
|             | Mama_24 | 713.368                         | 269.909               | 443.459                | 62,2                   | 198.942                                       | 27,9  |
|             | média   | 628.643                         | 220.696               | 407.947                | 65,4                   | 185.733                                       | 29,6  |
| M           | Mama_21 | 615.109                         | 218.941               | 396.168                | 64,4                   | 163.718                                       | 26,6  |
|             | Mama_37 | 663.218                         | 218.452               | 444.766                | 67,1                   | 136.248                                       | 20,5  |
|             | Mama_46 | 517.830                         | 128.996               | 388.834                | 75,1                   | 129.176                                       | 24,9  |
|             | Mama_50 | 637.834                         | 186.433               | 451.401                | 70,8                   | 202.615                                       | 31,8  |
|             | Mama_51 | 423.192                         | 103.712               | 319.480                | 75,5                   | 102.293                                       | 24,2  |
|             | Mama_52 | 550.978                         | 200.844               | 350.134                | 63,5                   | 102.205                                       | 18,5  |
|             | Mama_53 | 638.175                         | 218.481               | 419.694                | 65,8                   | 157.989                                       | 24,8  |
|             | Mama_54 | 381.844                         | 132.108               | 249.736                | 65,4                   | 101.720                                       | 26,6  |
|             | Mama_55 | 260.688                         | 79.584                | 181.104                | 69,5                   | 63.025  | 24,2  |
|             | Mama_56 | 724.483                         | 265.992               | 458.491                | 63,3                   | 143.845                                       | 19,9  |
|             | média   | 541.335                         | 175.354               | 365.981                | 68,0                   | 130.283                                       | 24,2  |
| média geral |         | 585.448                         | 188.766               | 396.682                | 68,1                   | 167.895                                       | 28,6  |

## **A Análise de expressão diferencial dos transcritos identificados**

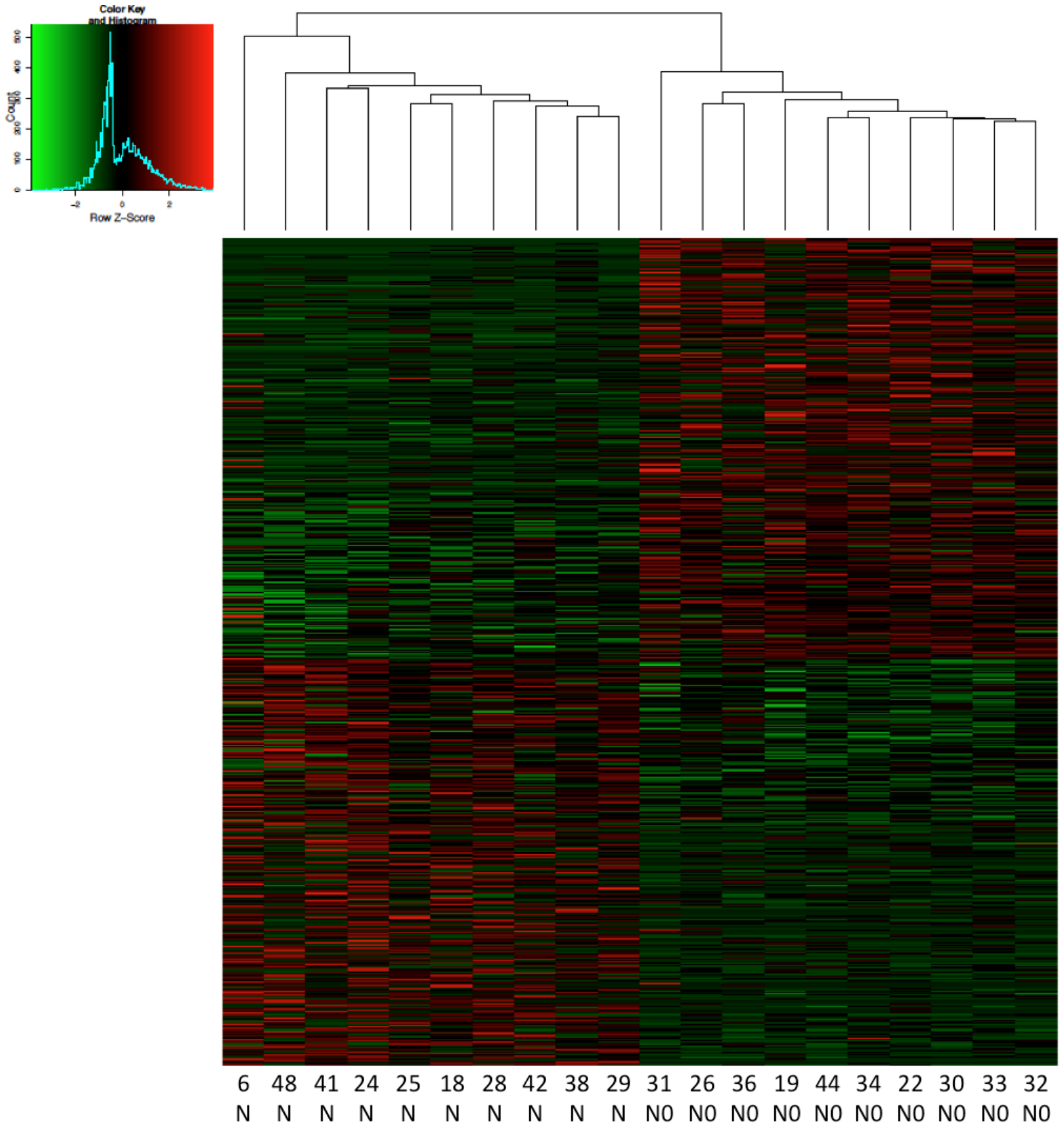
Para identificar um potencial biomarcador de estadiamento tumoral, realizamos a análise de expressão diferencial de todos os transcritos (miRNAs e outros RNAs) entre os três grupos de pacientes, comparando dois a dois, utilizando o pacote estatístico *DESeq2* do *Bioconductor* (LOVE et al. 2014).

### ***Transcritos diferencialmente representados entre N0 e N***

Na primeira comparação, entre os grupos N0 e N, encontramos 7 miRNAs diferencialmente representados com p-valor  $\leq 0,05$ , sendo 3 (42,9%) enriquecidos em pacientes do grupo N0 e 4 (57,1%) em pacientes do grupo N (**Figura 33**), porém os níveis de expressão destes 7 miRNAs não agrupam adequadamente os dois grupos no *heat map*. Dentre os 593 outros RNAs encontrados diferencialmente representados com p-valor  $\leq 0,05$ , 302 (50,9%) encontram-se enriquecidos em pacientes do grupo N0 e 291 (49,1%) em pacientes do grupo N (**Figura 34**). A distribuição de categorias destes 593 transcritos diferencialmente representados encontra-se na **Figura 35**: 86% são codificadores de proteína, 7% são *short noncoding*, 4% são *long noncoding* e 3% são *pseudogenes*.

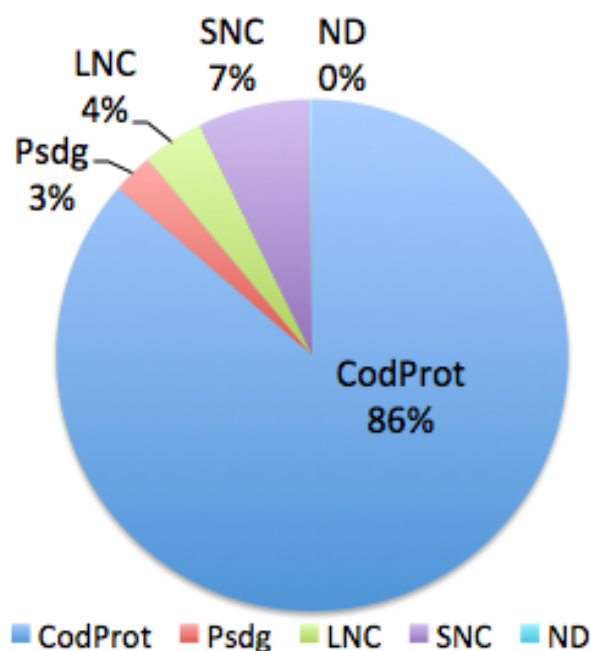


**Figura 33** - Heat map dos 7 miRNAs diferencialmente representados entre os grupos de pacientes N0 e N.

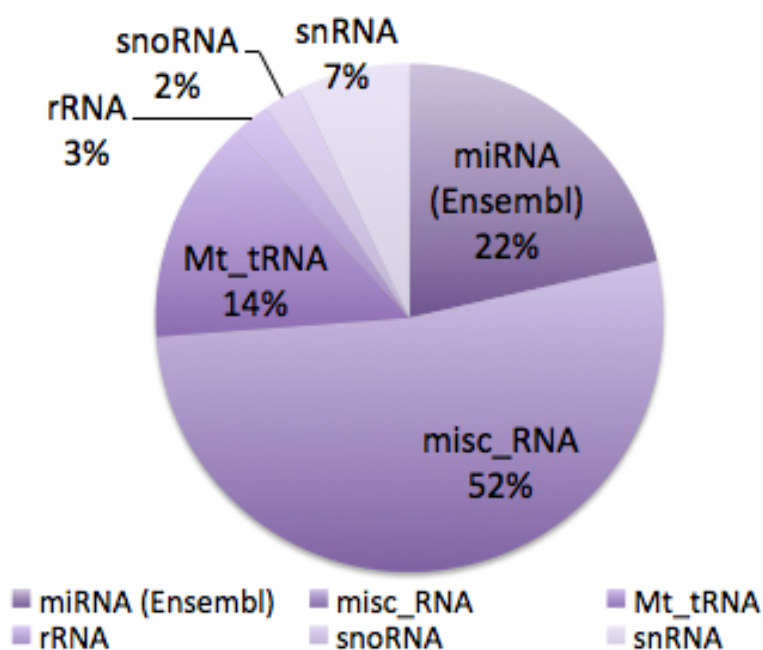


**Figura 34** - *Heat map* dos 593 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e N.

A



B



**Figura 35** - Categorias dos 593 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e N. **A)** Distribuição de todos os transcritos diferencialmente representados. CodProt = transcritos codificadores de proteína; Psdg = Pseudogenes; LNC = transcritos não-codificadores de proteínas longos (*long noncoding*); SNC = transcritos não-codificadores de proteínas curtos (*short noncoding*); ND = não definido. **B)** Distribuição dos transcritos na categoria *short noncoding*.

Dentre as classes que compõem a categoria *short noncoding*, a mais representada é miscRNA, com 52% dos transcritos. Interessantemente, dos 22 miscRNAs diferencialmente representados, 20 (90,9%) estão enriquecidos no grupo N e somente 2 (9,1%) estão enriquecidos no grupo N0. Dentre os miscRNAs diferencialmente representados, 77,3% são Y-RNA.

Os YRNAs foram primeiro descritos na década de 80, são altamente conservados e em humanos são transcritos pela RNA polimerase III a partir de quatro genes específicos e de quase mil pseudogenes (KÖHN et al. 2013; PERREAULT et al. 2005, 2007; WOLIN e STEITZ 1983).

As funções biológicas dos YRNAs ainda não foram inteiramente elucidadas, mas sabe-se que eles são essenciais para a replicação de DNA cromossomal (CHRISTOV et al. 2006) e tem significância clínica por fazerem parte do complexo ribonucleoproteico Ro, que é reconhecido por anticorpos em pacientes com doenças autoimunes como lúpus (LERNER et al. 1981).

Também sabe-se que os YRNAs são superexpressos em tumores sólidos em relação ao tecido normal correspondente (como em bexiga, colo do útero, cólon, rim, pulmão e próstata) e que eles são necessários para a proliferação celular, pois o silenciamento dos quatro YRNAs humanos por RNAi levou à inibição da proliferação das linhagens celulares HeLa e ME180 de carcinoma cervical, EJ30 de carcinoma de bexiga, DU145 de carcinoma de próstata, e WI38 de fibroblastos de pulmão (CHRISTOV et al. 2008).

DHAHBI et al. (2013) descreveram que fragmentos 5' de YRNA são abundantes no soro e plasma de indivíduos saudáveis. Usando NGS de bibliotecas de pequenos RNAs estes autores detectaram os YRNAs como a segunda classe mais abundante de transcritos identificados, correspondendo a 33% de todos os *reads* anotados, atrás somente de miRNAs com 44% dos *reads*. Por análise de *Northern Blot* a partir do *pellet* e sobrenadante de ultracentrifugação (110.000xg por 2 horas) do soro eles concluem que os YRNAs não estão associados a exossomos ou microvesículas, porém eles utilizaram sondas para um único YRNA específico (RNY4). Já no estudo de NOLTE-'T HOEN et al. (2012), foi identificado, também por NGS, um enriquecimento de YRNAs nos

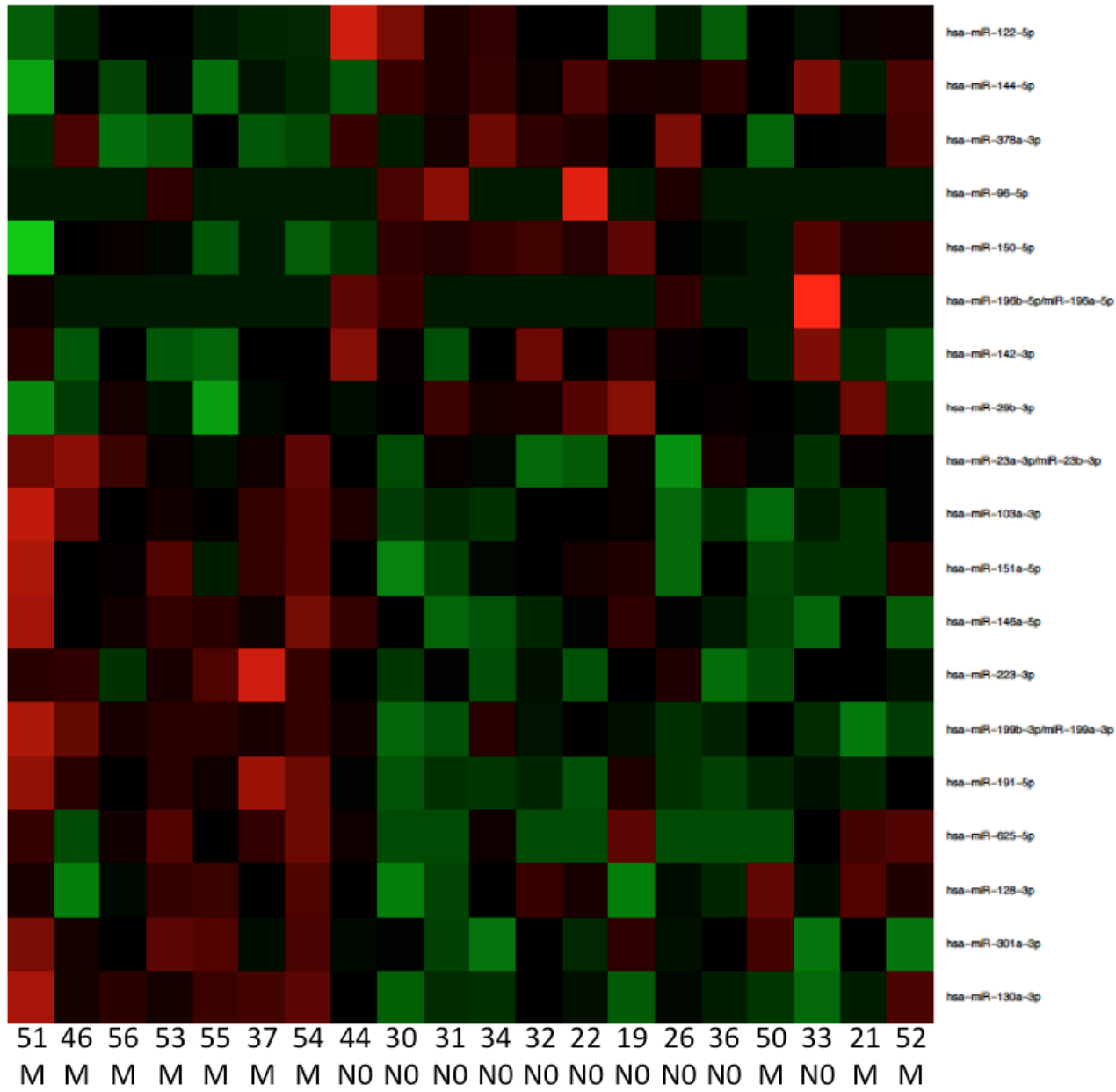
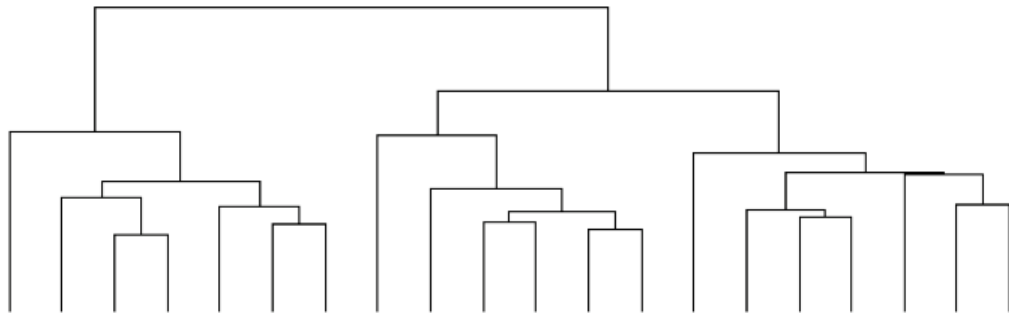
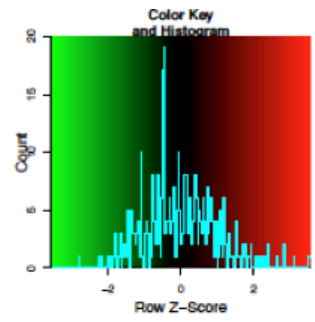
exossomos derivados do meio de cultura condicionado de células dendríticas em relação as células secretoras.

Em um outro estudo de DHAHBI et al. (2014) os autores descreveram a expressão diferencial de fragmentos de 25 YRNAs entre cinco pacientes com câncer de mama e cinco controles. Eles então analisaram dados públicos de pequenos RNAs contidos no soro de 42 pacientes diagnosticados com câncer de mama, e com acompanhamento clínico. Estas análises encontraram 36 YRNAs diferencialmente representados entre os pacientes de diferentes subtipos de câncer de mama (ER/PR/HER2). Os pacientes que tiveram relapso da doença, por exemplo, tinham expressão aumentada de 31 YRNAs em relação aos pacientes que não tiveram relapso. Ao compararmos a lista deste estudo mencionado com os YRNAs encontrados diferencialmente representados entre os três grupos de pacientes do nosso trabalho encontramos 9 transcritos em comum.

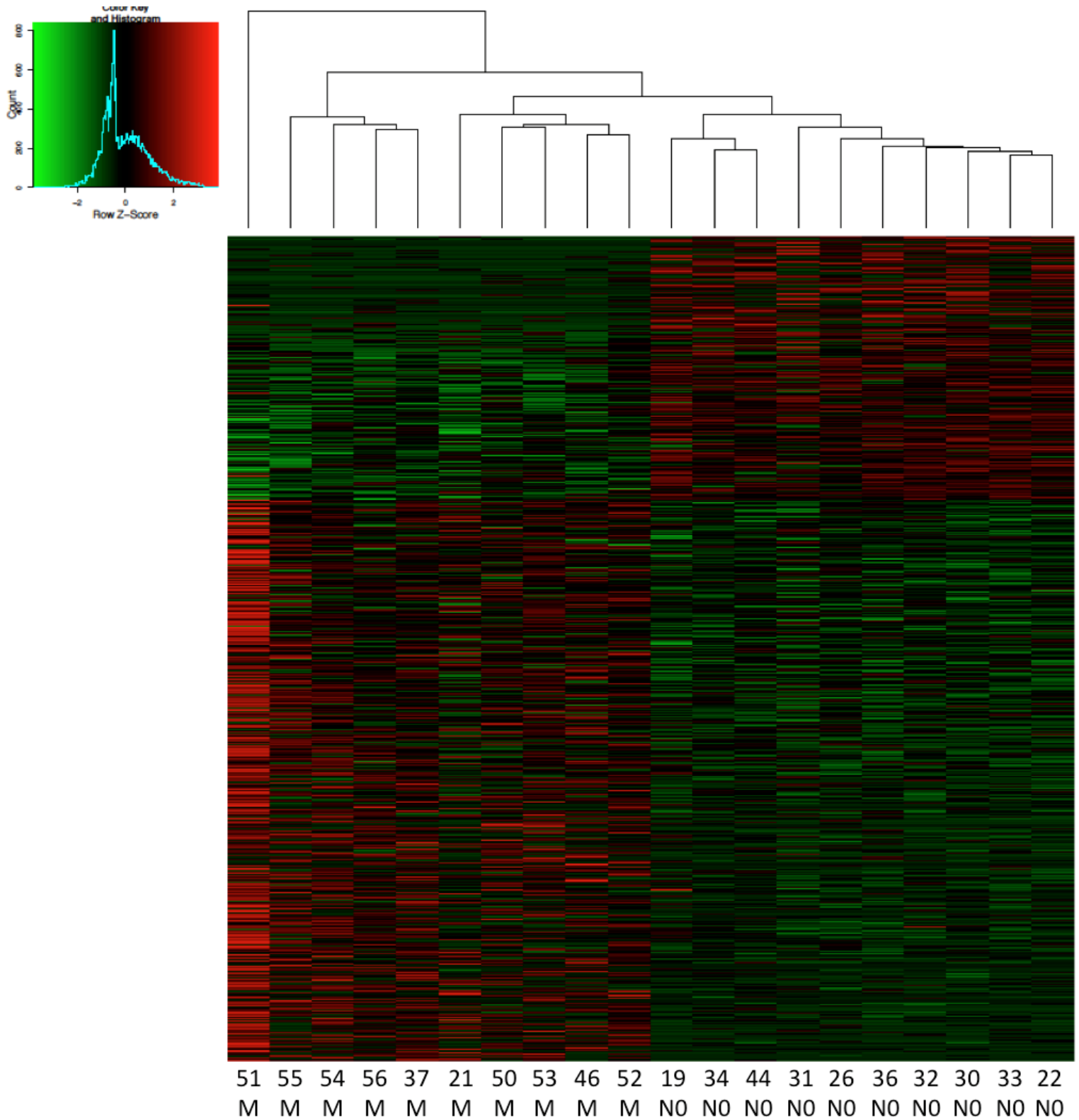
### ***Transcritos diferencialmente representados entre N0 e M***

Na segunda comparação, entre os grupos N0 e M, encontramos 19 miRNAs diferencialmente representados com p-valor  $\leq 0,05$ , sendo 8 (42,1%) enriquecidos em pacientes do grupo N0 e 11 (57,9%) em pacientes do grupo M (**Figura 36**). Dentre os 1.016 outros RNAs encontrados diferencialmente representados com p-valor  $\leq 0,05$ , 324 (31,9%) encontram-se enriquecidos em pacientes do grupo N0 e 692 (68,1%) em pacientes do grupo M (**Figura 37**). A distribuição de categorias destes 1.016 transcritos diferencialmente representados encontra-se na **Figura 38**: 79% são codificadores de proteína, 16% são *short noncoding*, 3% são *long noncoding* e 2% são *pseudogene*. Dentre as classes que compõem a categoria *short noncoding*, a mais representada é miscRNA, com 63% dos transcritos. Interessantemente, dos 105 miscRNAs diferencialmente representados, 103 (98,1%) estão enriquecidos no grupo M e somente 2 (1,9%) estão enriquecidos no grupo N0. Mais uma vez, dentre os miscRNAs diferencialmente representados, 87,6% são Y-RNAs.



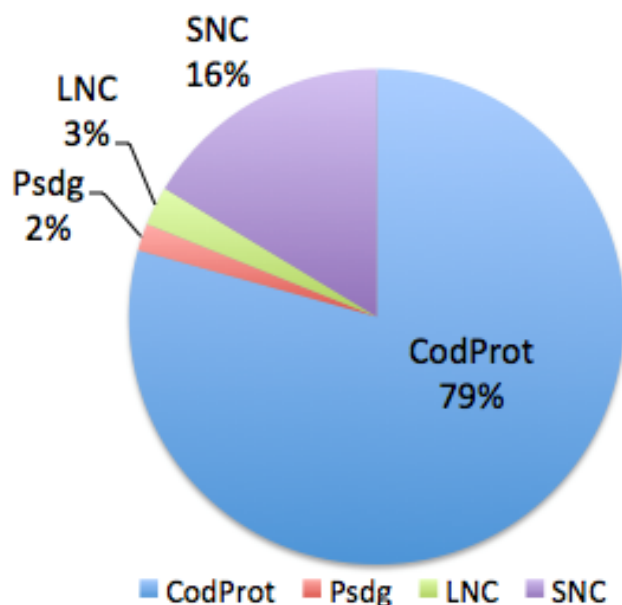


**Figura 36** - Heat map dos 19 miRNAs diferencialmente representados entre os grupos de pacientes N0 e M.

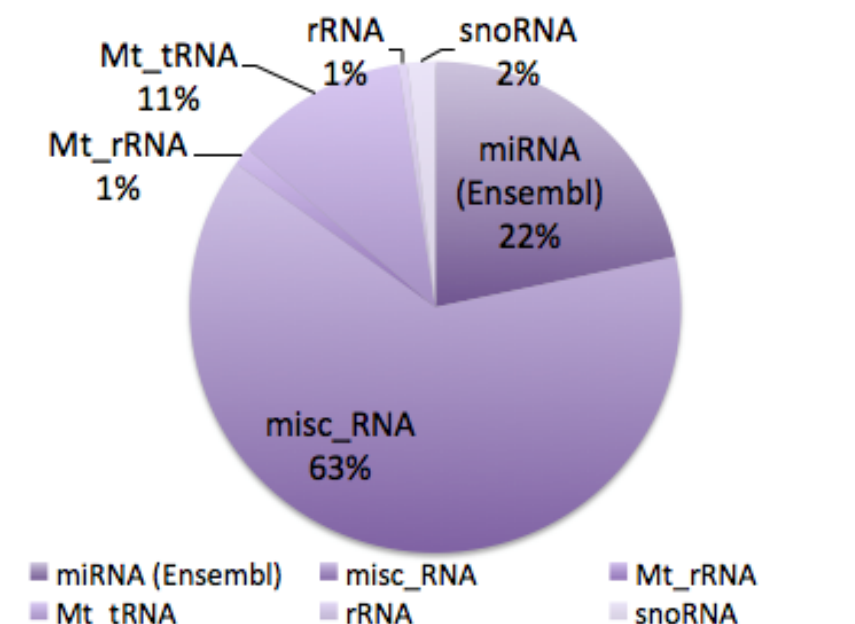


**Figura 37** - *Heat map* dos 1.016 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e M.

A



B

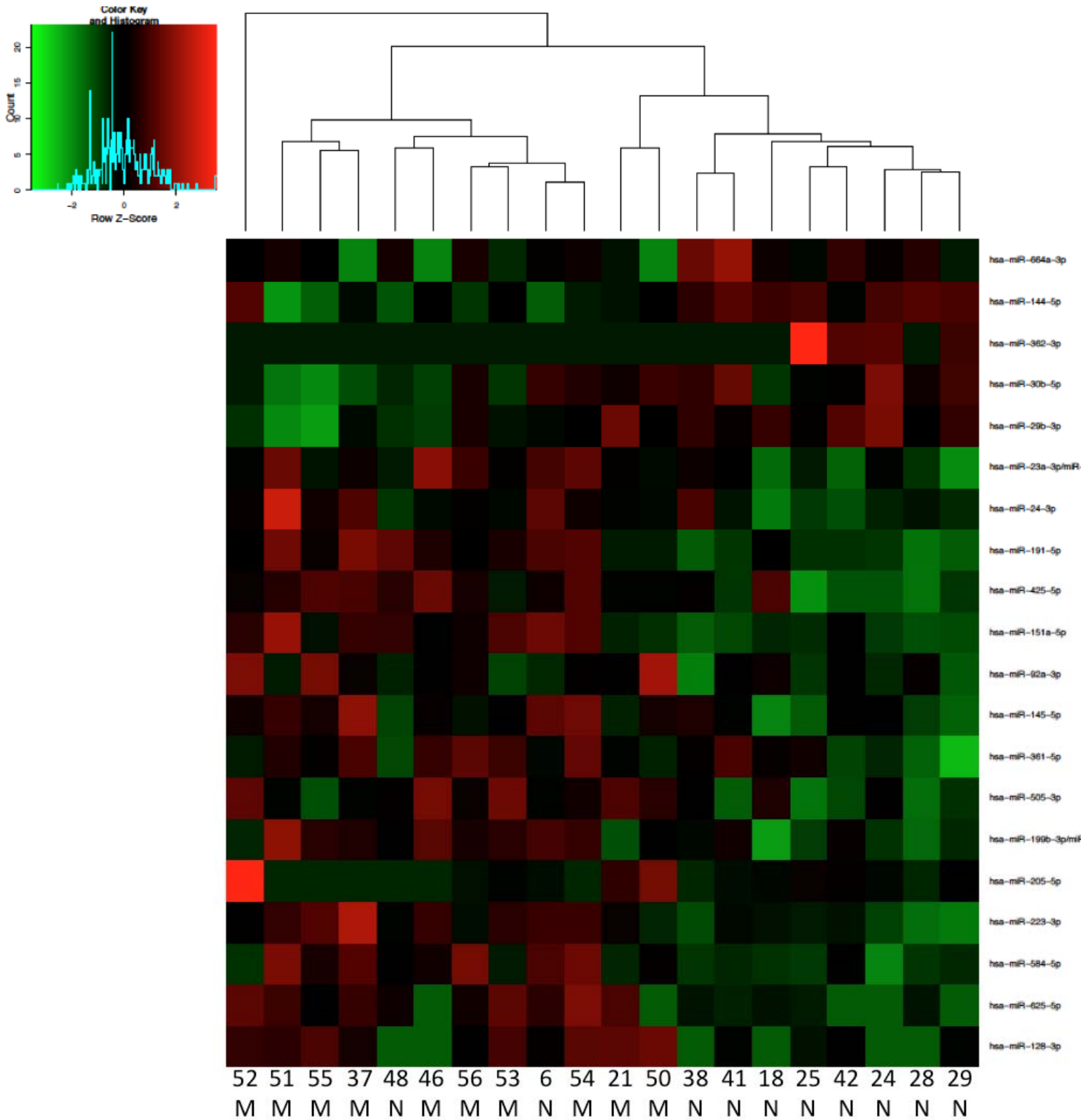


**Figura 38** - Categorias dos 1.016 outros RNAs diferencialmente representados entre os grupos de pacientes N0 e M. **A)** Distribuição de todos os transcritos diferencialmente representados. CodProt = transcritos codificadores de proteína; Psdg = Pseudogenes; LNC = transcritos não-codificadores de proteínas longas (*long noncoding*); SNC = transcritos não-codificadores de proteínas curtos (*short noncoding*). **B)** Distribuição dos transcritos na categoria *short noncoding*.

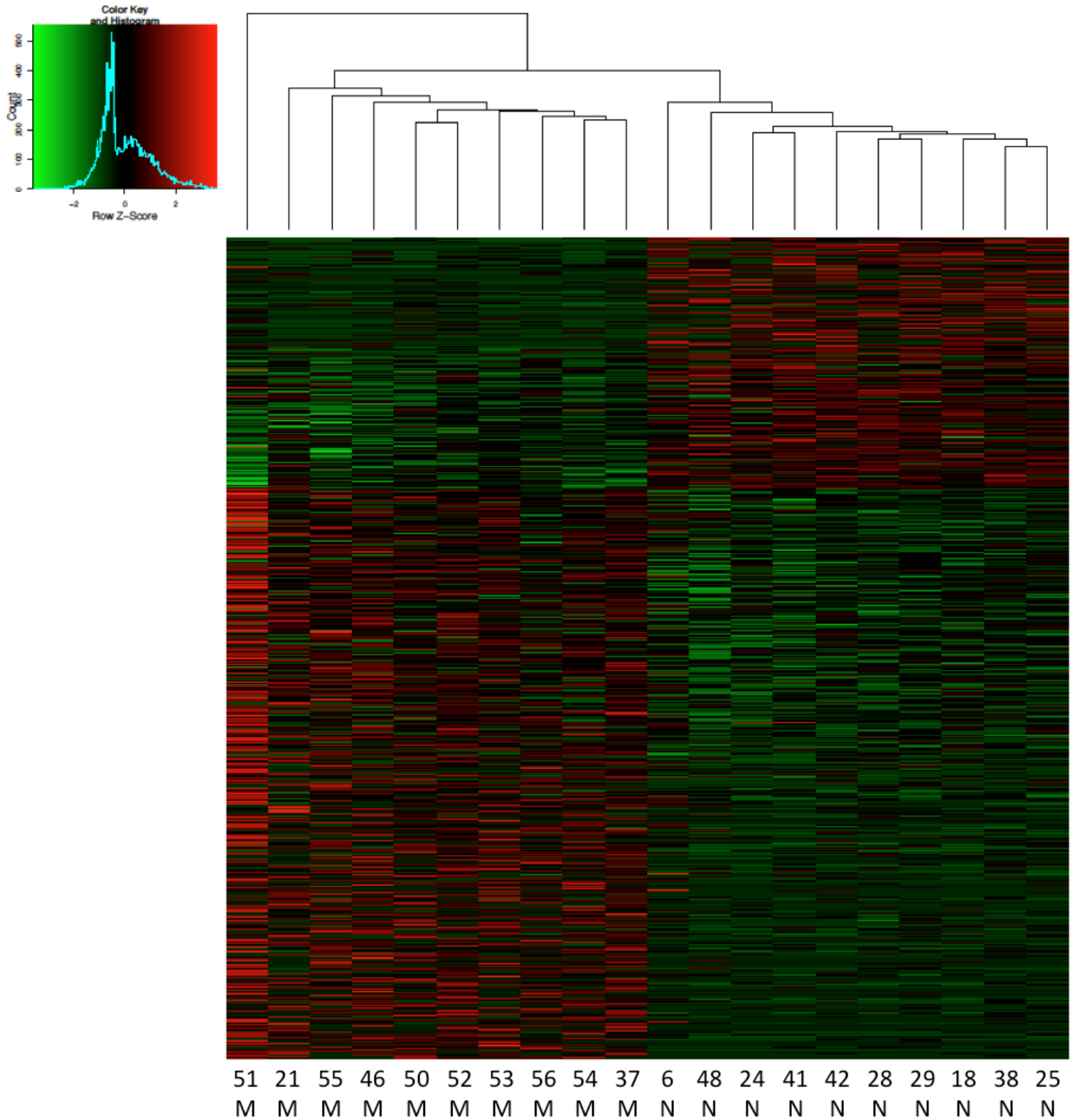
### ***Transcritos diferencialmente representados entre N e M***

Na terceira e última comparação, entre os grupos N e M, encontramos 20 miRNAs diferencialmente representados com p-valor  $\leq 0,05$ , sendo 5 (25,0%) enriquecidos em pacientes do grupo N e 15 (75,0%) em pacientes do grupo M (**Figura 39**). Dentre os 708 outros RNAs encontrados diferencialmente representados com p-valor  $\leq 0,05$ , 216 (30,5%) encontram-se enriquecidos em pacientes do grupo N e 492 (69,5%) em pacientes do grupo M (**Figura 40**). A distribuição de categorias destes 708 transcritos diferencialmente representados encontra-se na **Figura 41**: 88% são codificadores de proteína, 7% são *short noncoding*, 3% *long noncoding* e 2% são *pseudogene*. Dentre as classes que compõem a categoria *short noncoding*, a mais representada é miscRNA, com 80% dos transcritos. Interessantemente, dos 41 miscRNAs diferencialmente representados, 40 (97,6%) estão enriquecidos no grupo M e somente 1 (2,4%) está enriquecido no grupo N. Novamente, dentre os miscRNAs diferencialmente representados, 80,5% são Y-RNAs.

Apesar do entendimento sobre a significância fisiológica dos YRNAs ainda ser escasso, eles parecem ser promissores candidatos a biomarcadores tumorais, uma vez que foram encontrados superexpressos em diversos tumores epiteliais (CHRISTOV et al. 2008), e potencialmente a biomarcadores de estadiamento tumoral, uma vez que nossas análises indicaram que eles estão enriquecidos nos grupos N e M em relação ao grupo N0. Se confirmado o papel potencial de YRNAs como biomarcadores tumorais de diagnóstico e prognóstico, a alta abundância no plasma facilitaria o seu uso na rotina clínica, uma vez que não seria necessário o uso de material tumoral e a sua detecção poderia ser feita a partir de uma simples coleta de sangue.

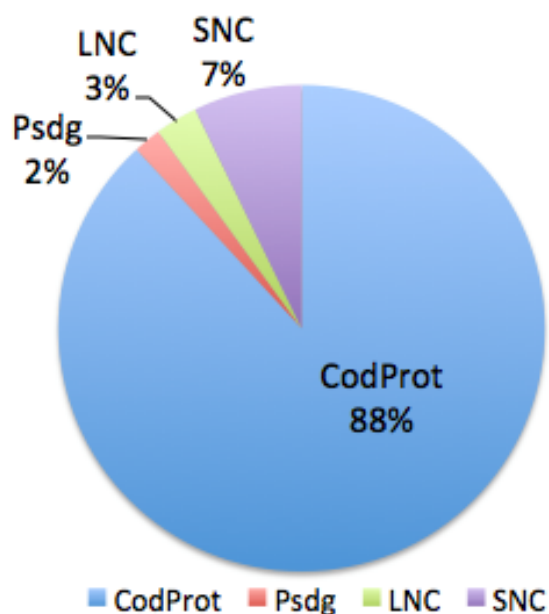


**Figura 39** - Heat map dos 20 miRNAs diferencialmente representados entre os grupos de pacientes N e M.

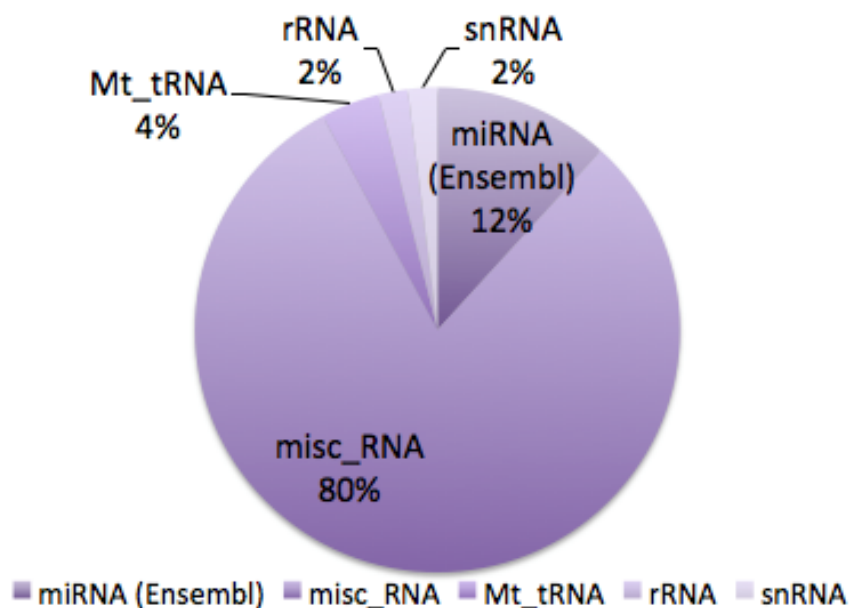


**Figura 40** - *Heat map* dos 708 outros RNAs diferencialmente representados entre os grupos de pacientes N e M.

A

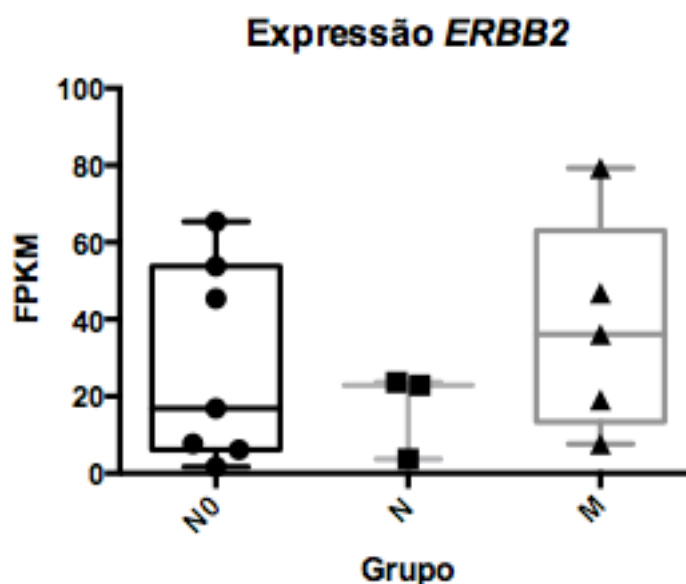


B



**Figura 41** - Categorias dos 708 outros RNAs diferencialmente representados entre os grupos de pacientes N e M. **A)** Distribuição de todos os transcritos diferencialmente representados. CodProt = transcritos codificadores de proteína; Psdg = Pseudogenes; LNC = transcritos não-codificadores de proteínas longos (*long noncoding*); SNC = transcritos não-codificadores de proteínas curtos (*short noncoding*). **B)** Distribuição dos transcritos na categoria *short noncoding*.

Analisamos a expressão de *ERBB2* nas amostras dos três grupos (**Figura 42**), com o intuito de avaliarmos a nossa sensibilidade de detecção de EVs derivadas das células tumorais. A expressão de *ERBB2* foi detectada em 7 pacientes do grupo N0, 3 pacientes do grupo N e 5 pacientes do grupo M. Apesar de haver uma tendência de maior expressão no grupo M, segundo uma análise de variância (ANOVA) não há diferença estatisticamente significativa entre os grupos ( $p$ -valor = 0,5229). Considerando que a presença ou não da molécula de mRNA nas EVs não se correlaciona necessariamente com a presença ou não da proteína deste oncogene, não podemos concluir sobre a sensibilidade de detecção de EVs tumorais, uma vez que a molécula efetora é a proteína e não temos como avaliá-la no momento.

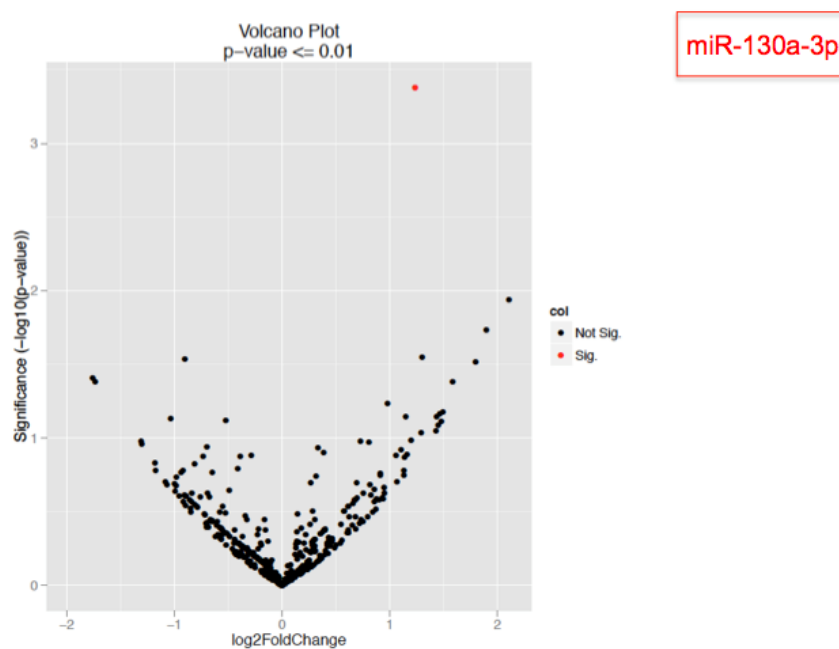


**Figura 42** - Expressão do mRNA de *ERBB2* entre os grupos de pacientes. FPKM = *Fragments Per Kilobase of transcript per Million mapped reads*.

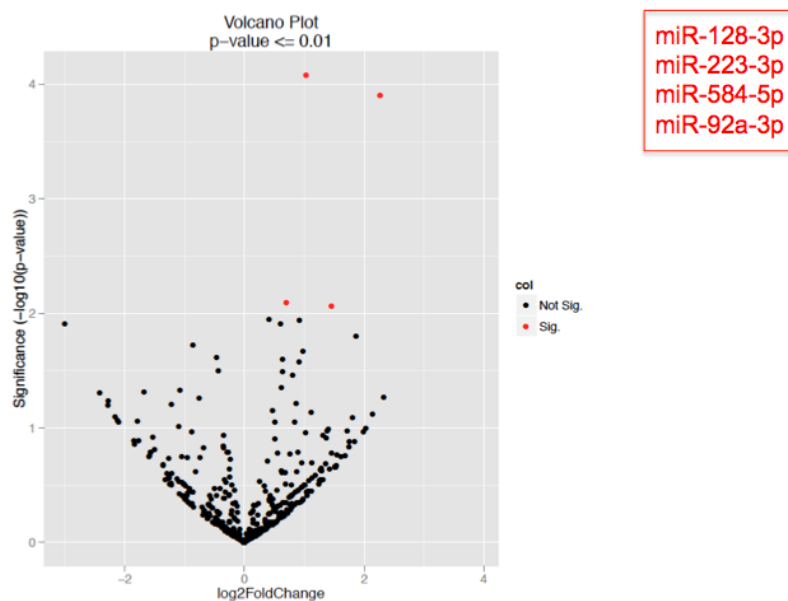


#### 4.2.4 Validação dos miRNAs diferencialmente representados em EVs derivadas do plasma de pacientes

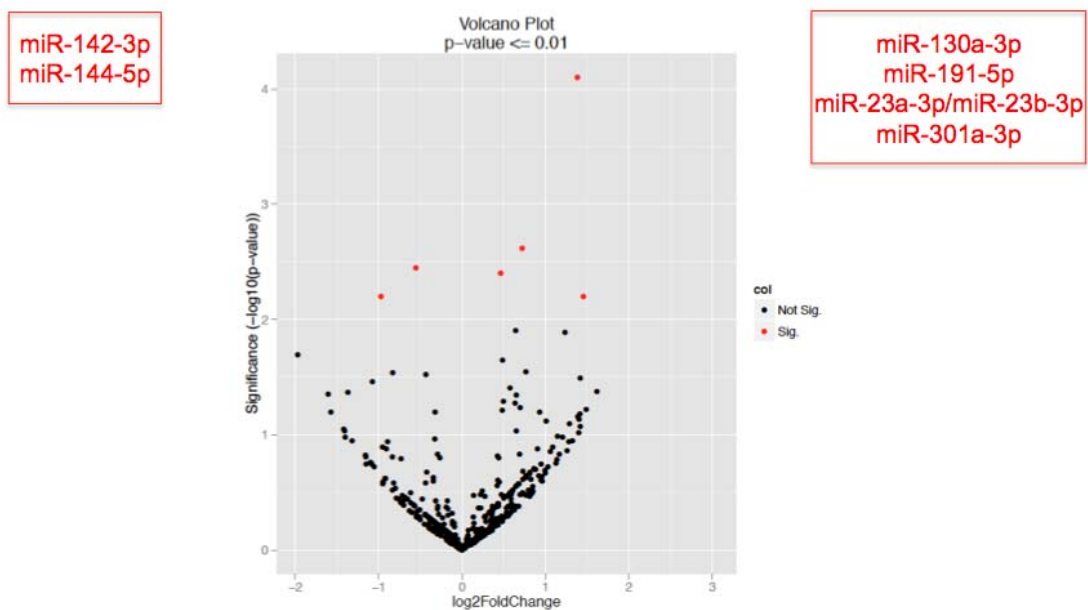
Na etapa de validação decidimos focar inicialmente nos miRNAs. Considerando um p-valor  $\leq 0,01$  para significância estatística, mais rigoroso que o adotado para gerar os *heat maps*, temos 1 miRNA diferencialmente representado entre os grupos N e N0 (**Figura 43**), 4 miRNAs diferencialmente representados entre os grupos M e N (**Figura 44**), e 6 miRNAs diferencialmente representados entre os grupos M e N0 (**Figura 45**). Nos *volcano plots* abaixo, os miRNAs com significado estatístico entre os grupos estão marcados em vermelho.



**Figura 43** - Análise de miRNAs diferencialmente representados entre os grupos de pacientes N e N0.

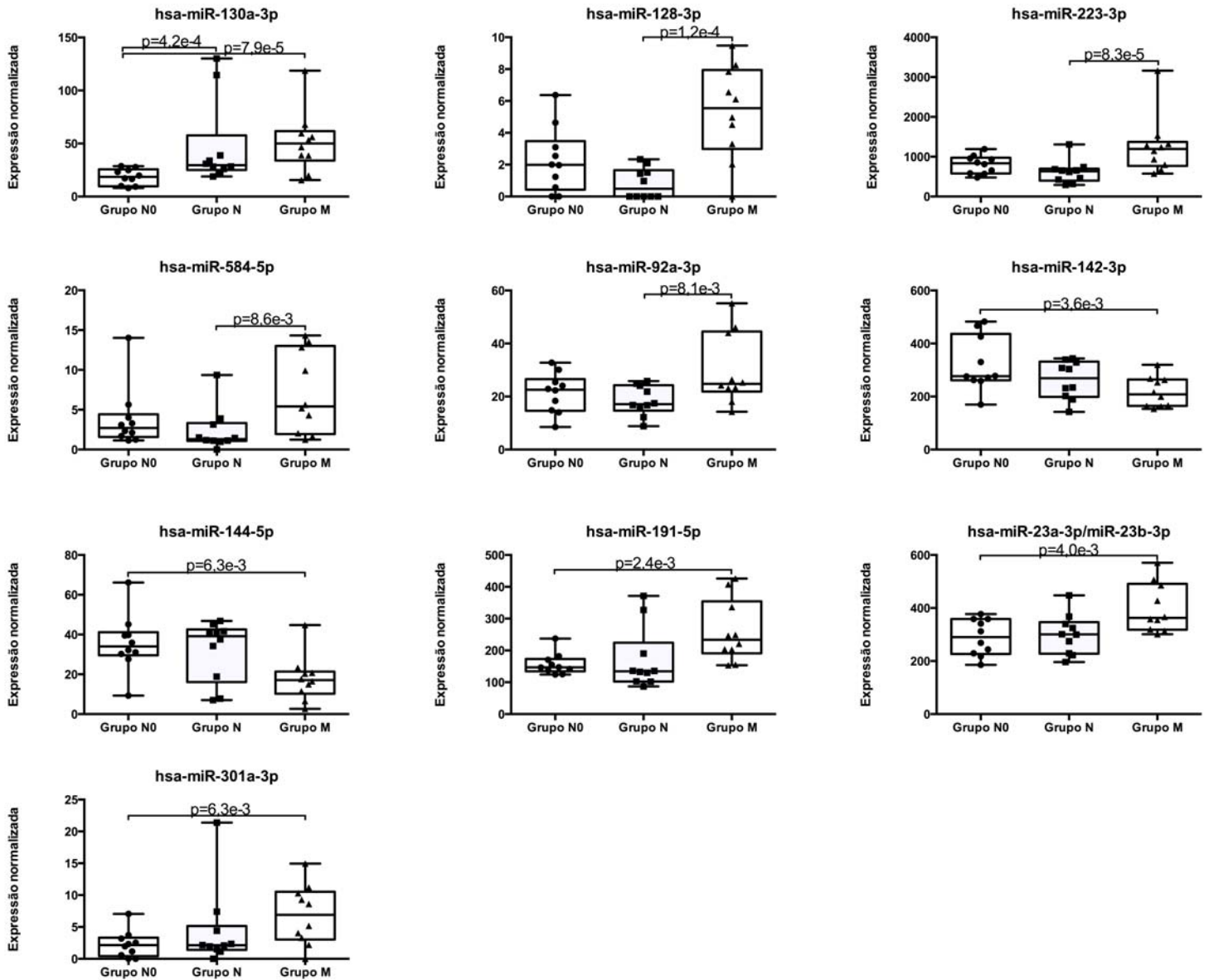


**Figura 44** - Análise de miRNAs diferencialmente representados entre os grupos de pacientes M e N.



**Figura 45** - Análise de miRNAs diferencialmente representados entre os grupos de pacientes M e N0.

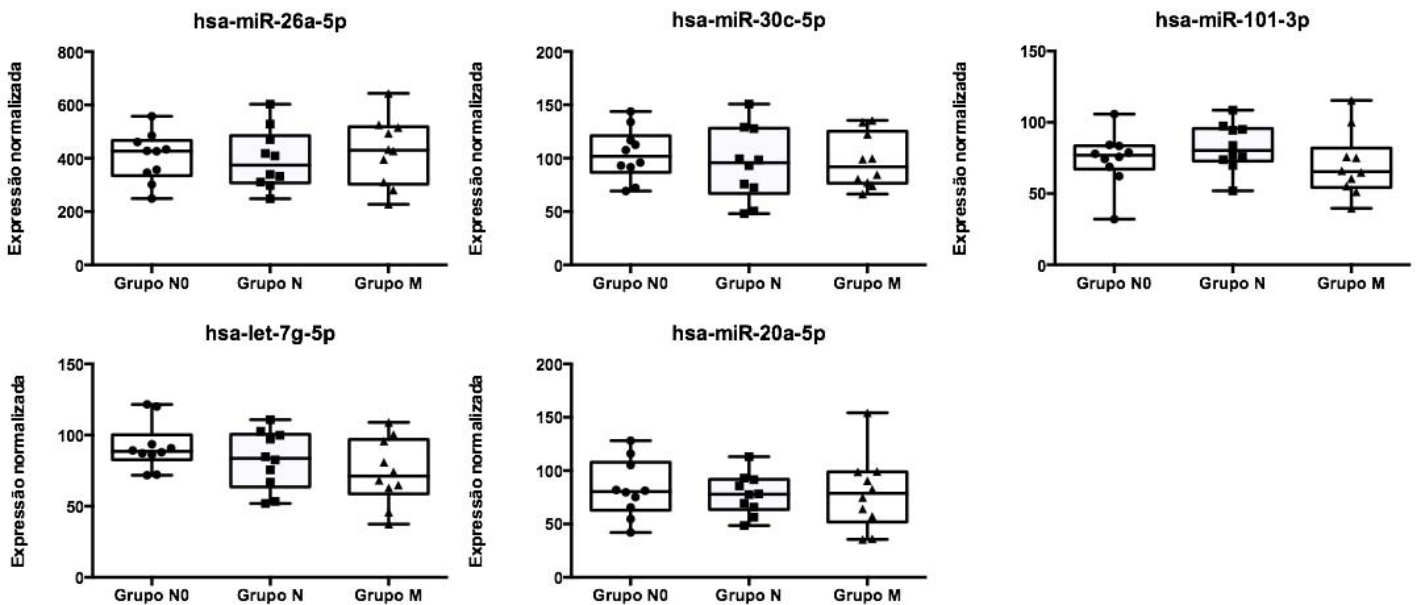
A expressão diferencial entre os grupos dos 10 miRNAs encontrados diferencialmente representados foi representada por *boxplots* (Figura 46).



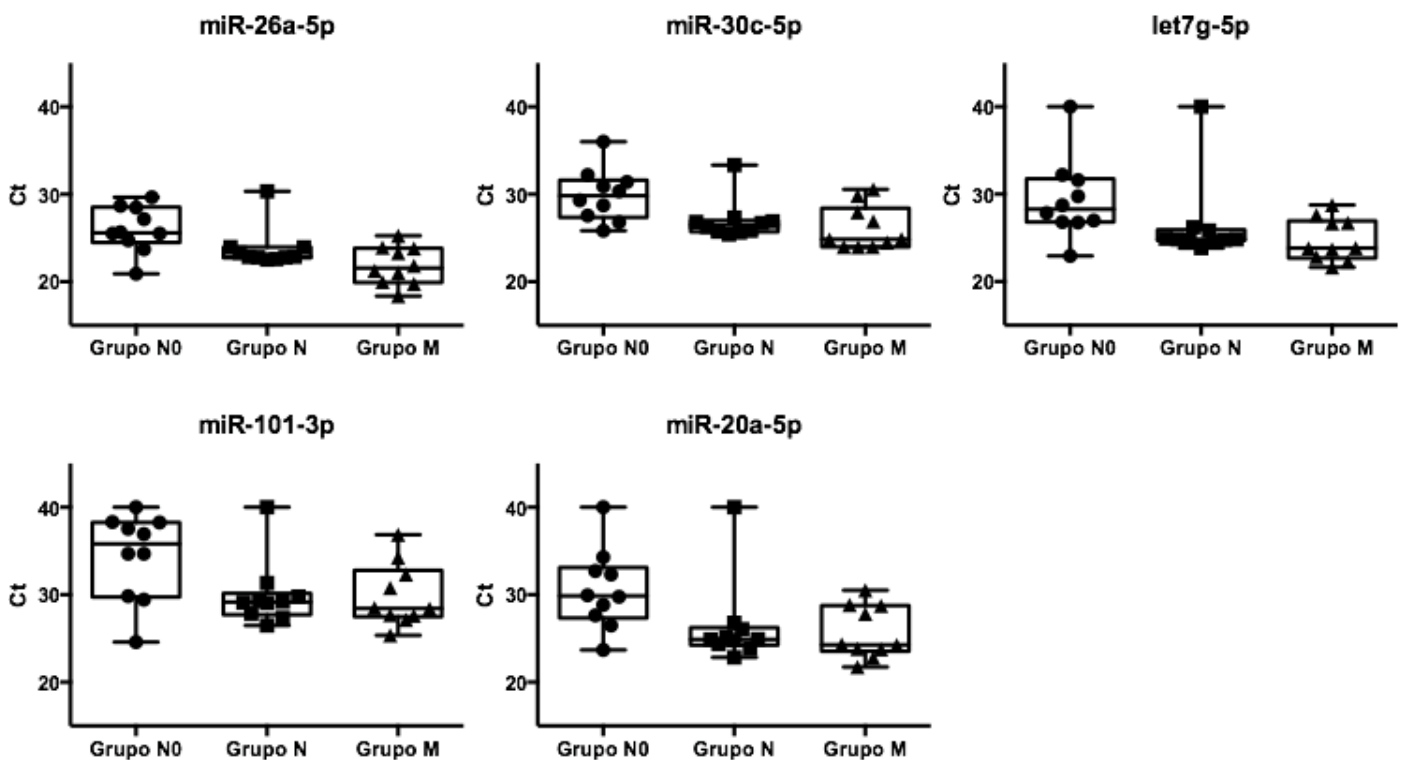
**Figura 46** - *Boxplots* dos miRNAs diferencialmente representados entre os grupos de pacientes. A expressão normalizada foi calculada a partir dos *reads* brutos pelo programa estatístico DESeq2, que leva em consideração a profundidade de sequenciamento de cada amostra para calcular fatores de normalização específicos.

Para validarmos os achados de NGS utilizamos uma abordagem de PCR quantitativo em tempo real (qRT-PCR), com ensaios específicos para cada miRNA indicado como diferencialmente expresso pelas análises de NGS e usando controles endógenos conforme descrito na sessão de Materiais e Métodos. A expressão dos 5 miRNAs (miR-26a-5p, miR-30c-5p, let-7g-5p, miR-101-3p, e miR-20a-5p) que tiveram

expressão pouco variável entre os três grupos, e que foram considerados mais estáveis segundo análise da expressão de NGS no programa *geNorm* (VANDESOMPELE et al. 2002) é mostrada na **Figura 47**. Todas as amostras foram sempre avaliadas em triplicata e fizemos novamente uma análise no *geNorm* com os dados de expressão obtidos no qRT-PCR (**Figura 48**) e obtivemos dois dos controles endógenos mais robustos (miR-26a-5p, miR-30c-5p) ( $M < 1,5$  - conforme definição do programa). Utilizamos então a média da expressão destes dois endógenos para a normalização pelo método de delta delta Ct, considerando o grupo N0 como referência.



**Figura 47** - *Boxplots* dos miRNAs selecionados como controles endógenos a partir dos dados de NGS. A expressão normalizada foi calculada a partir dos *reads* brutos pelo programa estatístico DESeq2, que leva em consideração a profundidade de sequenciamento de cada amostra para calcular fatores de normalização específicos.



**Figura 48** - *Boxplots* dos valores de Ct dos miRNAs inicialmente selecionados como controles endógenos.

O *fold-change* (definido como  $2^{-\Delta\Delta Ct}$ ) por grupo dos 11 miRNAs avaliados foi representado por *boxplots* (**Figura 49**). A **Tabela 19** compara as diferenças de expressão (*fold-change*) obtidas no sequenciamento (NGS) e na validação (qRT-PCR). Para a maioria dos ensaios avaliados (8/12) obtivemos resultados concordantes no NGS e qRT-PCR. Se considerarmos um  $p\text{-valor} \leq 0,01$ , temos um candidato (miR-223-3p) com *fold change* estatisticamente significativo pelo qRT-PCR, com expressão aumentada no grupo M em comparação com o grupo N, efetivamente validando o resultado de NGS.

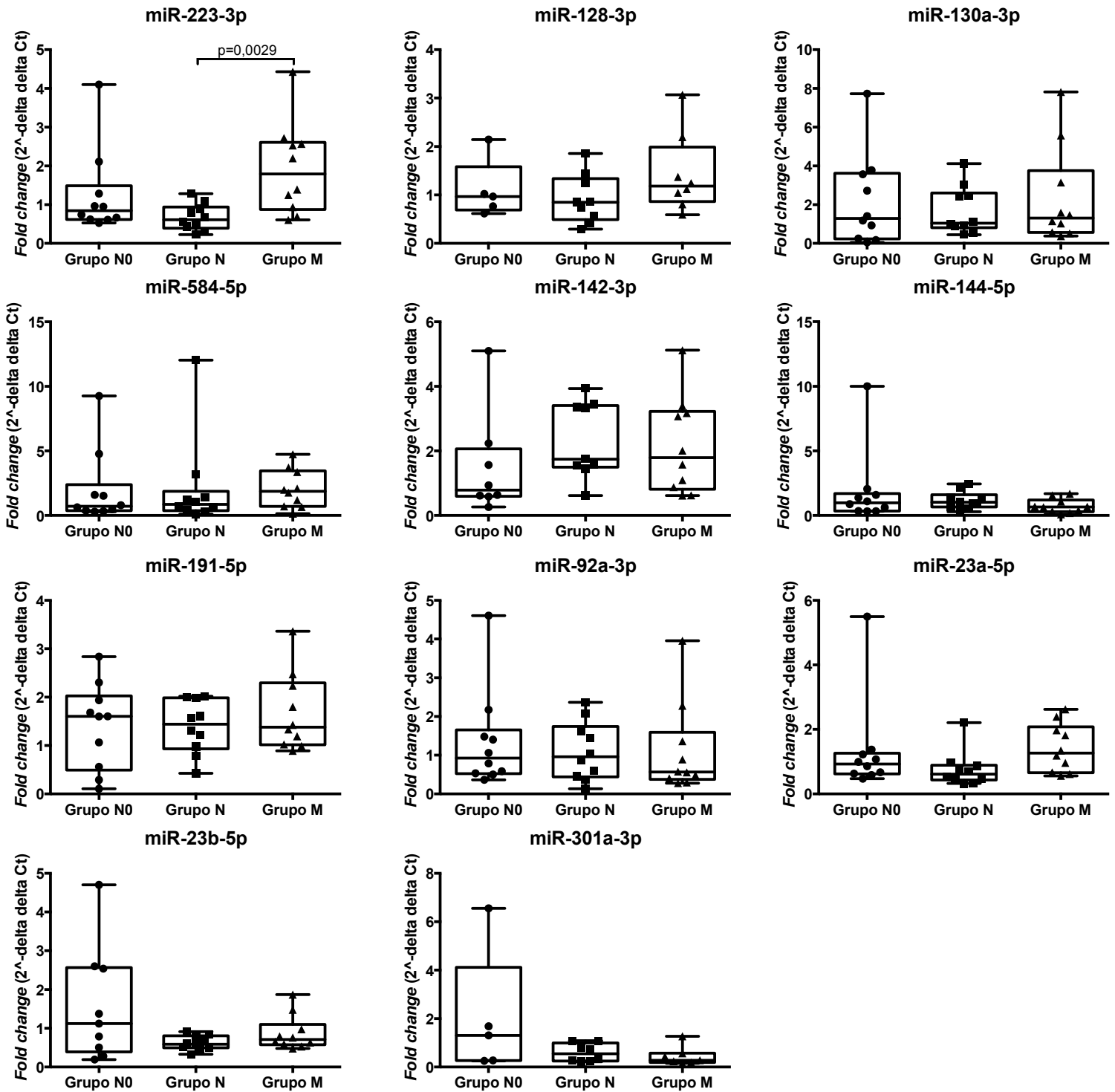


Figura 49 - Boxplots dos miRNAs avaliados por qRT-PCR.

**Tabela 19** - Comparação dos *fold-changes* obtidos por sequenciamento (NGS) e na validação por qRT-PCR.

| Grupos comparados | miRNA           | Média de expressão normalizada NGS (todas pacientes) | Fold change |         |
|-------------------|-----------------|--|-------------|---------|
|                   |                 |  | NGS         | qRT-PCR |
| N x NO            | hsa-miR-130a-3p | 39,09  | 2,36        | 1,32    |
| M x NO            | hsa-miR-130a-3p | 39,09  | 2,61        | 1,45    |
|                   | hsa-miR-142-3p  | 267,13   | 0,68        | 1,70    |
|                   | hsa-miR-144-5p  | 28,57  | 0,51        | 0,54    |
|                   | hsa-miR-191-5p  | 196,13   | 1,65        | 1,53    |
|                   | hsa-miR-23a-3p  | 330,12   | 1,38        | 1,23    |
|                   | hsa-miR-23b-3p  | 4,53   | 2,74        | 0,79    |
|                   | hsa-miR-301a-3p | 4,53   | 2,74        | 0,35    |
| M x N             | hsa-miR-128-3p  | 2,80   | 4,82        | 1,59    |
|                   | hsa-miR-223-3p  | 895,69   | 2,04        | 2,72    |
|                   | hsa-miR-584-5p  | 4,43   | 2,74        | 1,57    |
|                   | hsa-miR-92a-3p  | 23,25  | 1,62        | 0,91    |

Dentre todos os miRNAs que identificamos por NGS, miR-223-3p foi o segundo mais abundante, tendo sido identificado em todas as trinta pacientes por 1.180 *reads* em média para cada paciente. A alta abundância deste miRNA em EVs já foi relatada em vários artigos, independentemente da metodologia de análise de expressão empregada. Utilizando microarranjo de PCR em tempo real, CHEVILLET et al. (2014) identificaram miR-223 como sendo o miRNA mais abundante em exossomos derivados de plasma de indivíduos controles e de indivíduos com câncer de próstata metastático. Os autores também quantificaram a expressão deste miRNA por *Droplet Digital PCR* (ddPCR), uma técnica bastante sensível que permite a determinação do número absoluto de cópias sem o uso de curva padrão, obtendo valores muito semelhantes (variação menor que 1%) aos obtidos por PCR em tempo real (CHEVILLET et al. 2014). Outro grupo obteve resultados concordantes utilizando sequenciamento em larga escala na plataforma Ion. Eles identificaram miR-223-3p como sendo o miRNA mais abundante em exossomos derivados de plasma de indivíduos-controle isolados com um kit comercial da Norgen baseado em separação por coluna, e o segundo mais abundante em exossomos isolados por ultracentrifugação (CHENG et al. 2014). Além de abundante em EVs provenientes do plasma, miR-223-3p também foi identificado por RNA-Seq como sendo o miRNA mais abundante na saliva de indivíduos controles (BAHN et al. 2015).

O papel do miR-223-3p em câncer de mama já foi investigado em diversos trabalhos. PINATEL et al. (2014) utilizaram dados de quatro datasets públicos contendo dados de expressão gênica e informações clínicas de mais de 900 amostras. A partir de uma lista dos genes diferencialmente expressos entre os pacientes livres de doença em 5 anos e os pacientes onde houve relapso da doença, foi feita uma análise *in silico* para prever sítios de ligação de miRNAs enriquecidos no 3'-UTR dos genes downregulados nos pacientes com pior prognóstico (PINATEL et al. 2014). Eles então chegaram a uma lista com 6 miRNAs (miR-19ab-3p, miR-200bc, miR-203a, miR-223-3p, miR-21-5p, miR-340-5p), e decidiram focar no miR-223-3p para avaliação funcional. A superexpressão deste miRNA em células MDA-MB-231 resultou em uma diminuição na capacidade de migração e invasão das células, além de um aumento na morte celular induzida por perda de adesão (anoikis) e por tratamento com paclitaxel. A expressão proteica de 3 potenciais alvos de miR-223-3p, ITGA3, NRAS e STAT5A, foi encontrada diminuída nas células com superexpressão deste miRNA, e foi comprovada uma interação direta do miRNA com o 3'-UTR do gene STAT5A por ensaio de luciferase. Uma redução na invasão e migração e um aumento de morte celular também foram constatados em células com silenciamento de STAT5A, sugerindo que os efeitos biológicos do miR-223-3p são em grande parte devido à supressão do seu alvo STAT5A. Em termos do seu papel na progressão tumoral do câncer de mama, os autores concluem que o miR-223-3p atua como supressor tumoral. Este achado é apoiado pelos dados de GONG et al. (2013) que detectaram uma diminuição na expressão de miR-223 conforme a capacidade proliferativa e invasiva de linhagens celulares, sendo este miRNA altamente expresso na linhagem não-tumoral MCF-10A, moderadamente expresso nas linhagens tumorais MCF-7 e T-47D, e pouco expresso na linhagem tumoral MDA-MB-231, que é a mais agressiva. (GONG et al. 2013) Em leucemia linfóide crônica, foi observada uma relação entre a baixa expressão de miR-223 e uma maior agressividade da doença e pior sobrevida (STAMATOPOULOS et al. 2009).



A menor expressão de miR-223-3p nas células de um tumor pode ser ocasionada por diversos mecanismos, mas uma possibilidade é a eliminação deste miRNA das células tumorais pela sua secreção em EVs. Neste sentido, nós encontramos um aumento significativo deste miRNA em EVs circulantes em pacientes com doença metastática (Grupo M) em comparação com pacientes sem doença sistêmica (Grupo N), reforçando a hipótese que este miRNA possa ser exportado das células tumorais via EVs, gerando então uma diminuição da sua expressão intratumoral, evento biologicamente relevante para a progressão tumoral. Alternativamente, a quantidade elevada de miR-223-3p encontrada em EVs circulantes pode estar vindo de outras células não-tumorais, uma vez que isolamos do plasma uma população heterogênea de vesículas secretadas por diversos tipos celulares, incluindo plaquetas, células endoteliais e uma diversa gama de outras células e tecidos. Neste sentido, já foi demonstrado que exossomos secretados por macrófagos ativados por IL-4 contêm miR-223 (YANG et al. 2011). Em um sistema de co-cultura em câmara de Boyden de macrófagos (derivados de cultura primária de monócitos humanos) com as linhagens tumorais de mama SKBR3 e MDA-MD-231, que mimetiza o microambiente de macrófagos associados ao tumor, foi demonstrada a transferência de miR-223 dos macrófagos para as células tumorais. Curiosamente, diferentemente do encontrado nos dois trabalhos discutidos acima, a superexpressão de miR-223 aumentou a capacidade de invasão das células MDA-MB-231.

Acreditamos que os demais miRNAs encontrados diferencialmente representados no NGS não foram validados com significância estatística pelo qRT-PCR devido à sua baixa expressão e/ou pequeno *fold change*. Apesar de ser a metodologia padrão utilizada para a validação de dados de NGS, é sabido que o qRT-PCR não é indicado para analisar variações com *fold change* <2, já que um aumento/diminuição de um Ct corresponde a metade/dobro de quantidade, respectivamente. O miRNA diferencialmente expresso com maior *fold change* no NGS foi o miR-128-3p, com um aumento de 4,82 vezes no grupo M em relação ao grupo N, porém ele foi o

menos abundante dentre todos os candidatos, sendo que a amostra com maior expressão teve menos de 10 *reads* normalizados (DESeq2). Sabemos que a sensibilidade e o intervalo dinâmico do qRT-PCR são muito menores que do NGS, por isso tivemos que realizar uma pré-amplificação do cDNA utilizando um *pool* de todos os 16 ensaios antes do qRT-PCR. Selecionamos candidatos diferencialmente representados mas com baixa expressão pois foi descrito recentemente que a maioria dos exossomos contém menos de uma molécula de miRNA. Em uma análise estequiométrica entre o número de moléculas de miRNAs e a concentração de exossomos isolados de cinco diferentes fontes (plasma de indivíduos saudáveis e de pacientes com câncer de próstata, fluido seminal, e meio de cultura condicionado de células dendríticas, mastócitos e carcinoma de ovário), CHEVILLET et al. (2014) determinaram que em média há 0,00825 moléculas de miRNAs/exossomo. A baixa expressão de uma molécula não necessariamente se correlaciona com ausência de função biológica, pois devemos sempre pensar no contexto da transferência de moléculas da vesícula internalizada para a célula receptora, e não no contexto de uma vesícula em meio a milhões de outras vesículas, que é o caso quando analisamos a população total de EVs circulantes no plasma das pacientes.

Neste sentido, quando consideramos que a metodologia de descoberta (NGS) é mais sensível que a metodologia padrão de validação (qRT-PCR), percebemos que este não é o cenário ideal para identificar e comprovar potenciais biomarcadores a partir de moléculas contidas em EVs. Uma possibilidade seria validar os dados de NGS a partir do sequenciamento NGS de outra biblioteca construída da mesma amostra inicial, porém isso seria economicamente inviável. Talvez um possível método de validação alternativo seja a PCR digital, que vem sendo cada vez mais utilizada para este propósito, pois teria, teoricamente, capacidade de detectar de forma absoluta alvos expressos desde uma única molécula.

### 4.3 PREDIÇÃO DE NOVOS miRNAs

O sequenciamento de nova geração é vantajoso em relação às plataformas fechadas (microarranjos) pois possibilita, além da identificação de alterações quantitativas, a observação de mutações e edições nos diversos transcritos, permitindo ainda a predição de novos miRNAs. Para este fim, o programa mais utilizado é o miRDeep2 (MACKOWIAK 2011), que contém um algoritmo que avalia o enquadramento dos RNAs sequenciados baseado na biogênese dos miRNAs. O programa descreve a localização genômica do potencial novo precursor, apresenta a estrutura predita e demonstra o alinhamento das nossas sequências nas duas fitas (5p e 3p) do maduro (**Figura 50**). Neste projeto, avaliamos todas as sequências geradas usando o miRDeep2, realizando um estudo de potenciais miRNAs novos dentre os transcritos aqui sequenciados. Quando consideramos uma *false-discovery rate (FDR)* igual a zero, foi possível prevermos 3 potenciais novos miRNAs a partir dos RNAs contidos nas EVs isoladas do meio de cultura celular condicionado das linhagens C5.2 e HB4a (**Tabela 20**) e 11 potenciais novos miRNAs a partir dos RNAs contidos nas EVs isoladas do sangue periférico das pacientes (**Tabela 21**). A maioria dos candidatos foram identificados em amostras únicas, o que sugere sua baixa expressão e de certo modo ajuda a justificar o ineditismo dos mesmos. Todavia, um dos candidatos (mapeado no cromossomo 12 de *Homo sapiens*) foi identificado de forma recorrente em 5 pacientes distintas, sendo 3 do grupo N0 e 2 do grupo N, sugerindo a sua expressão ubíqua no sangue periférico destas mulheres. As sequências aqui geradas, incluindo-se os potenciais novos miRNAs, serão depositadas em bancos de dados públicos, permitindo acesso da comunidade científica internacional.

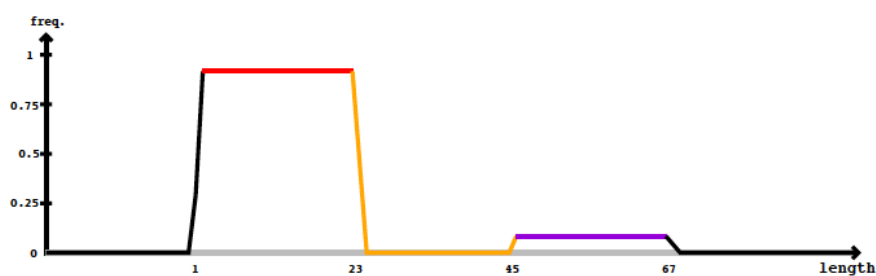
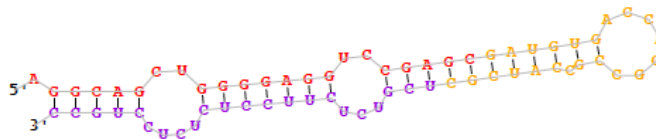
**Tabela 20** - Potenciais novos miRNAs preditos pelo programa miRDeep2 em EVs derivadas das linhagens C5.2 e HB4a.

| Amostra   | Nome provisório | miRDeep2 score | Probabilidade do candidato ser verdadeiro positivo | Total reads | Nº reads maduro | Nº reads loop | Nº reads star | Localização do precursor     |
|-----------|-----------------|----------------|--|-------------|-----------------|---------------|---------------|------------------------------|
| C5.2 100K | chr13_27225     | 6,6            | 0,85 ± 0,26  | 3           | 2               | 0             | 1             | chr13:113977112..113977182:+ |
| HB4a 100K | chr19_33656     | 2,10E+01       | 0,69 ± 0,46  | 47          | 41              | 2             | 4             | chr19:24183657..24183723:-   |
| HB4a 100K | chr11_20546     | 6,3            | 0,76 ± 0,32  | 4           | 2               | 1             | 1             | chr11:65269487..65269555:+   |

**Tabela 21** - Potenciais novos miRNAs preditos pelo programa miRDeep2 em EVs derivadas do plasma de pacientes.

| Grupo | Amostra | Nome provisório | miRDeep2 score | Probabilidade do candidato ser verdadeiro positivo | Total reads | Nº reads maduro | Nº reads loop | Nº reads star | Localização do precursor    |
|-------|---------|-----------------|----------------|--|-------------|-----------------|---------------|---------------|-----------------------------|
| N0    | Mama 30 | chr12_7499      | 1,70E+01       | 0,74 ± 0,44  | 24          | 22              | 0             | 2             |                             |
| N0    | Mama 31 | chr12_5074      | 6,9            | 0,73 ± 0,45  | 4           | 2               | 0             | 2             |                             |
| N0    | Mama 34 | chr12_10025     | 9,4            | 0,74 ± 0,44  | 9           | 5               | 0             | 4             | chr12:92539495..92539561:-  |
| N     | Mama 41 | chr12_5854      | 5,3            | 0,59 ± 0,49  | 8           | 7               | 0             | 1             |                             |
| N     | Mama 6  | chr12_6908      | 6,4            | 0,67 ± 0,47  | 3           | 1               | 1             | 1             |                             |
| N0    | Mama 34 | chr19_13685     | 6,8            | 0,87 ± 0,20  | 4           | 3               | 0             | 1             | chr19:24185390..24185474:-  |
| N0    | Mama 34 | chr19_13674     | 6,2            | 0,87 ± 0,20  | 18          | 12              | 0             | 6             | chr19:24183657..24183723:-  |
| N0    | Mama 34 | chr1_744        | 5,1            | 0,89 ± 0,17  | 50          | 47              | 3             | 0             | chr1:153549489..153549581:+ |
| N0    | Mama 36 | chr17_8780      | 5,3            | 0,77 ± 0,42  | 12          | 5               | 0             | 7             | chr17:67352297..67352355:+  |
| N     | Mama 28 | chr3_2296       | 5,7            | 0,67 ± 0,47  | 3           | 2               | 0             | 1             | chr3:183408907..183408957:+ |
| N     | Mama 18 | chr17_9473      | 5,5            | 0,72 ± 0,45  | 14          | 14              | 0             | 0             | chr17:73836822..73836896:-  |
| N     | Mama 42 | chr19_9781      | 5,1            | 0,75 ± 0,36  | 16          | 16              | 0             | 0             | chr19:6589051..6589124:+    |
| M     | Mama 51 | chr3_4060       | 7,2            | 0,61 ± 0,49  | 12          | 11              | 0             | 1             | chr3:49058066..49058131:+   |
| M     | Mama 21 | chr20_7268      | 7              | 0,70 ± 0,46  | 6           | 2               | 0             | 4             | chr20:17497699..17497765:+  |
| M     | Mama 52 | chr15_12483     | 5              | 0,68 ± 0,47  | 4           | 2               | 0             | 2             | chr15:83424761..83424821:+  |

Provisional ID : chr12\_7499  
 Score total : 17.1  
 Score for star read(s) : 3.9  
 Score for read counts : 6.2  
 Score for mfe : 2.3  
 Score for randfold : 1.6  
 Score for cons. seed : 3  
 Total read count : 24  
 Mature read count : 22  
 Loop read count : 0  
 Star read count : 2



**Mature** **Star**

```

5'- ugcugcgagcuggagcggaggcagcuggggagguccgagcgaugugaccaggccgccaucgucgucucucuccucucugcgcgccuccugucgaaaaaacuuuuu -3'  obs
   ugcugcgagcuggagcggaggcagcuggggagguccgagcgaugugaccaggccgccaucgucgucucucuccucucugcgcgccuccugucgaaaaaacuuuuu  exp
   .....(((.....(((.....(((.....(((.....(((.....(((.....(((.....(((.....(((.....(((.....(((.....(((.....(((.....))).....))).....))).....))).....))).....))).....))).....
   .....gaggcagcuggggagguccgagc.....
   .....aggcagcuggggagguccgagc.....
   .....ucgucucucuccucucugc.....

```

| reads | mm | sample |
|-------|----|--------|
| 7     | 0  | seq    |
| 15    | 0  | seq    |
| 2     | 0  | seq    |

**Figura 50** – Estrutura predita e número de *reads* de um dos potenciais precursores de novos miRNAs, mostrando alinhamento das sequências aqui obtidas para os miRNAs 5p e 3p a partir do precursor predito.

#### 4.4 IDENTIFICAÇÃO DE RNAs CIRCULARES EM EVs

Milhares de RNAs circulares (circRNAs) foram recentemente descritos como sendo expressos em células eucarióticas, e em centenas de transcritos os circRNAs parecem ser a isoforma predominante (JECK et al. 2013; MEMCZAK et al. 2013; SALZMAN et al. 2012). Por requerem algoritmos de análise de dados que permitam o alinhamento dos *reads* de forma invertida no genoma humano (exons em sequência não-canônica), os circRNAs passaram despercebidos por muito tempo, sendo que a primeira descrição deles em um biofluido humano (saliva) ocorreu somente no final de 2014 (BAHN et al. 2015).

A identificação de circRNAs nos nossos dados de NGS foi feita utilizando *scripts* desenvolvidos por MEMCZAK et al. (2013) (<http://circbase.org/cgi-bin/downloads.cgi>). Obtivemos 1.030 reads que

anotaram como RNAs circulares nas EVs derivadas do plasma de pacientes, com tamanho médio de 68,6nt, desvio padrão 20,2nt, e variando entre 36-160nt. Encontramos 513 circRNAs diferentes, dos quais 6 (1,2%) estavam representados por mais de 10 *reads* (**Tabela 22**). Destes 6 circRNAs mais abundantes, um (gene *UBXN7*) foi identificado em 8 amostras, um (gene *CORO1C*) em 6 amostras, um (gene *VAMP3*) em 4 amostras e três (genes *MTRF1*, *MTR*, *ATP9B*) em somente uma amostra cada (**Quadro 6**).

**Tabela 22** - Distribuição do número de *reads* dos 513 RNAs circulares (circRNAs) identificados nas EVs derivadas do plasma de pacientes.

| # circRNAs | %    | # reads |
|------------|------|---------|
| 1          | 0,2  | 24      |
| 1          | 0,2  | 13      |
| 2          | 0,4  | 12      |
| 1          | 0,2  | 11      |
| 1          | 0,2  | 10      |
| 1          | 0,2  | 9       |
| 5          | 1,0  | 8       |
| 4          | 0,8  | 7       |
| 10         | 1,9  | 6       |
| 11         | 2,1  | 5       |
| 31         | 6,0  | 4       |
| 37         | 7,2  | 3       |
| 113        | 22,0 | 2       |
| 295        | 57,5 | 1       |
| 513        | 100  | 1.030   |

**Quadro 6** - Distribuição dos seis RNAs circulares (circRNAs) mais abundantes no sequenciamento das EVs derivadas do plasma de pacientes.

| Gene          | Amostra | # Reads | Grupo |
|---------------|---------|---------|-------|
| <i>CORO1C</i> | Mama 19 | 1       | NO    |
|               | Mama 32 | 3       | NO    |
|               | Mama 44 | 3       | NO    |
|               | Mama 25 | 3       | N     |
|               | Mama 38 | 1       | N     |
|               | Mama 51 | 13      | M     |
|               | total   |         | 24    |
| <i>MTRF1</i>  | Mama 56 | 13      | M     |
| <i>VAMP3</i>  | Mama 37 | 1       | M     |
|               | Mama 51 | 8       | M     |
|               | Mama 52 | 1       | M     |
|               | Mama 53 | 2       | M     |
|               | total   |         | 12    |
| <i>UBXN7</i>  | Mama 33 | 2       | NO    |
|               | Mama 44 | 3       | NO    |
|               | Mama 25 | 1       | N     |
|               | Mama 29 | 1       | N     |
|               | Mama 37 | 1       | M     |
|               | Mama 46 | 1       | M     |
|               | Mama 50 | 2       | M     |
|               | Mama 54 | 1       | M     |
| total         |         | 12      |       |
| <i>MTR</i>    | Mama 28 | 11      | N     |
| <i>ATP9B</i>  | Mama 30 | 10      | NO    |

Também investigamos a presença de circRNAs no dados de sequenciamento de EVs e células das linhagens C5.2 e HB4a. Obtivemos 444 *reads* anotados como circRNAs, com tamanho médio 58,6nt, desvio padrão 16,1nt e variando entre 36-126nt, identificando um total de 365 circRNAs diferentes nestas bibliotecas. Destes, 15 (4,1%) estavam representados por mais de 3 *reads* (**Tabela 23**), sendo a grande maioria expressos somente nas células e não nas EVs (**Quadro 7**).

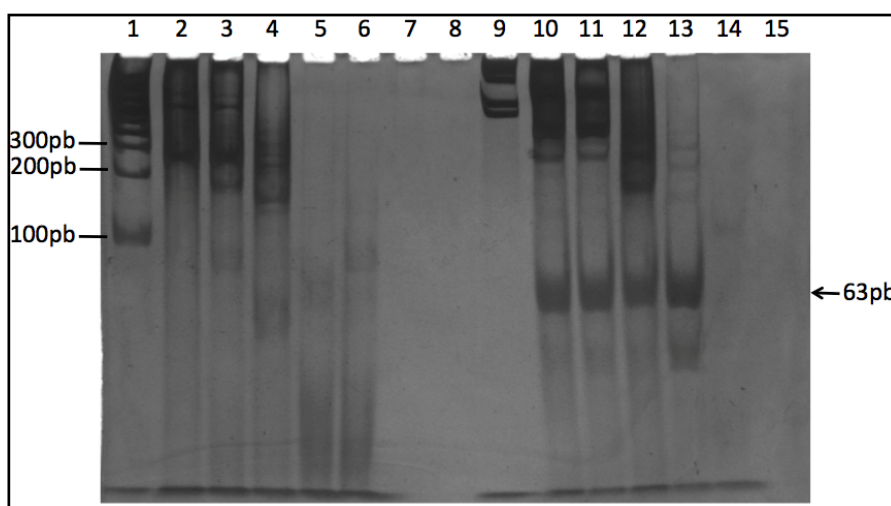
**Tabela 23** - Distribuição dos *reads* dos 365 RNAs circulares (circRNAs) identificados em C5.2 e HB4a e EVs derivadas do meio condicionado.

| # circRNAs | %    | # reads |
|------------|------|---------|
| 1          | 0,3  | 8       |
| 2          | 0,5  | 7       |
| 1          | 0,3  | 5       |
| 4          | 1,1  | 4       |
| 7          | 1,9  | 3       |
| 30         | 8,2  | 2       |
| 320        | 87,7 | 1       |
| 365        | 100  | 444     |

**Quadro 7** - Distribuição dos quinze RNAs circulares (circRNAs) mais abundantes em C5.2 e HB4a e nas EVs do meio condicionado.

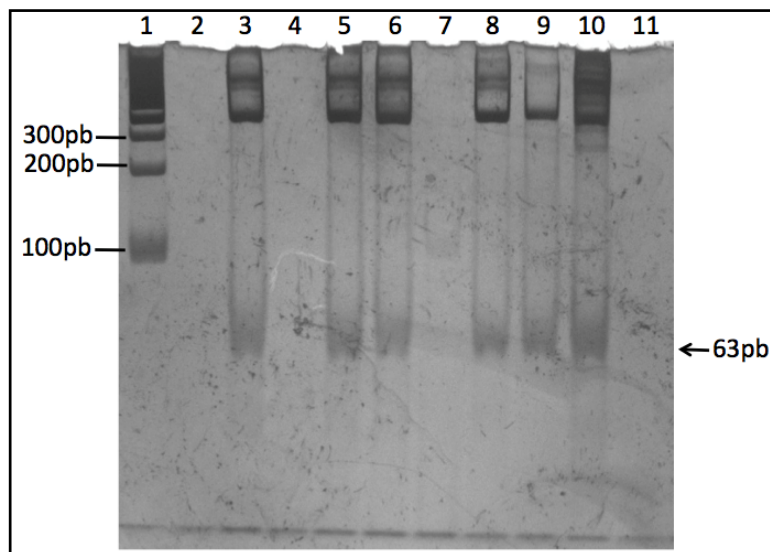
| Gene        | Amostra     | # Reads |
|-------------|-------------|---------|
| HIPK3       | C5.2 célula | 2       |
|             | HB4a célula | 6       |
|             | total       | 8       |
| ARHGAP12    | C5.2 célula | 3       |
|             | HB4a célula | 4       |
|             | total       | 7       |
| ITGAV       | C5.2 célula | 6       |
|             | HB4a célula | 1       |
|             | total       | 7       |
| UBXN7       | C5.2 célula | 1       |
|             | HB4a célula | 4       |
|             | total       | 5       |
| SEPSECS     | C5.2 célula | 2       |
|             | HB4a célula | 2       |
|             | total       | 4       |
| ZZEF1       | HB4a 100K   | 4       |
| NAB1        | C5.2 célula | 2       |
|             | HB4a célula | 2       |
|             | total       | 4       |
| PMS1        | C5.2 célula | 1       |
|             | HB4a célula | 2       |
|             | HB4a 100K   | 1       |
|             | total       | 4       |
| UBR2/TRERF1 | C5.2 célula | 2       |
|             | HB4a célula | 1       |
|             | total       | 3       |
| EHMT1       | C5.2 célula | 3       |
| GSK3B       | HB4a 20K    | 3       |
| LARP4B      | C5.2 célula | 3       |
| ARFGEF1     | C5.2 célula | 3       |
| MPP6        | HB4a célula | 1       |
|             | HB4a 100K   | 2       |
|             | total       | 3       |
| EGFR        | HB4a célula | 3       |

Selecionamos o circRNA *UBXN7*, que foi identificado em 8 amostras de pacientes e também nas células C5.2 e HB4a, para validação experimental. Desenhamos dois pares de iniciadores, um que amplifica somente a isoforma circular e o outro específico para a isoforma linear. A análise dos *amplicons* em gel de poliacrilamida 8% corado com prata confirmou a banda do tamanho esperado de 63 pb para a isoforma circular em cDNA proveniente da linhagem HB4a (**Figura 51**) e em 5/8 (62%) das pacientes (**Figura 52**), e a banda do tamanho esperado de 228pb para a isoforma linear (**Figura 53**).

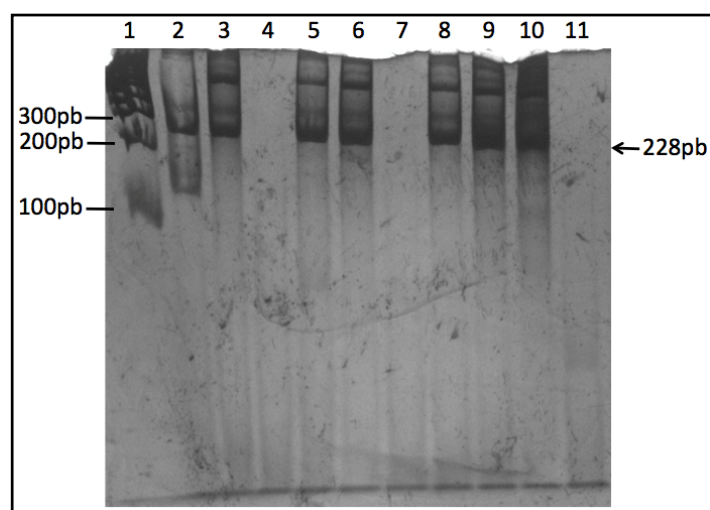


**Figura 51** - Visualização em gel de poliacrilamida dos produtos de amplificação do transcrito *UBXN7* em cDNA de célula HB4a. **1**: Ladder 100pb; **9**: Ladder 100kb; **2 a 6**: iniciadores que amplificam a isoforma linear; **10 a 14**: iniciadores que amplificam a isoforma circular; **2 e 10**: cDNA Superscript sem diluir; **3 e 11**: cDNA Superscript 1:10; **4 e 12**: cDNA biblioteca sem diluir; **5 e 13**: cDNA biblioteca 1:10; **6 e 14**: controles negativos. A banda esperada para o circRNA é de 63pb, e está indicada por seta no gel.





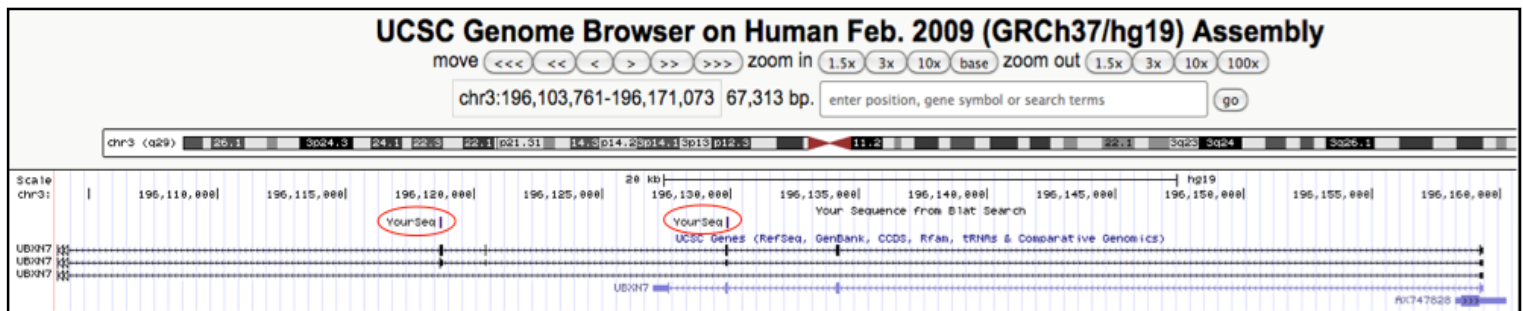
**Figura 52** - Visualização em gel de poliacrilamida dos produtos de amplificação de *UBXN7* (isoforma circular) em cDNA de EVs derivadas do plasma de pacientes. 1: Ladder 100pb; 2: Mama 33; 3: Mama 44; 4: Mama 25; 5: Mama 29; 6: Mama 37; 7: Mama 46; 8: Mama 50; 9: Mama 54; 10: controle positivo (cDNA célula HB4a); 11: controle negativo. A banda esperada para o circRNA é de 63pb, e está indicada por seta no gel.



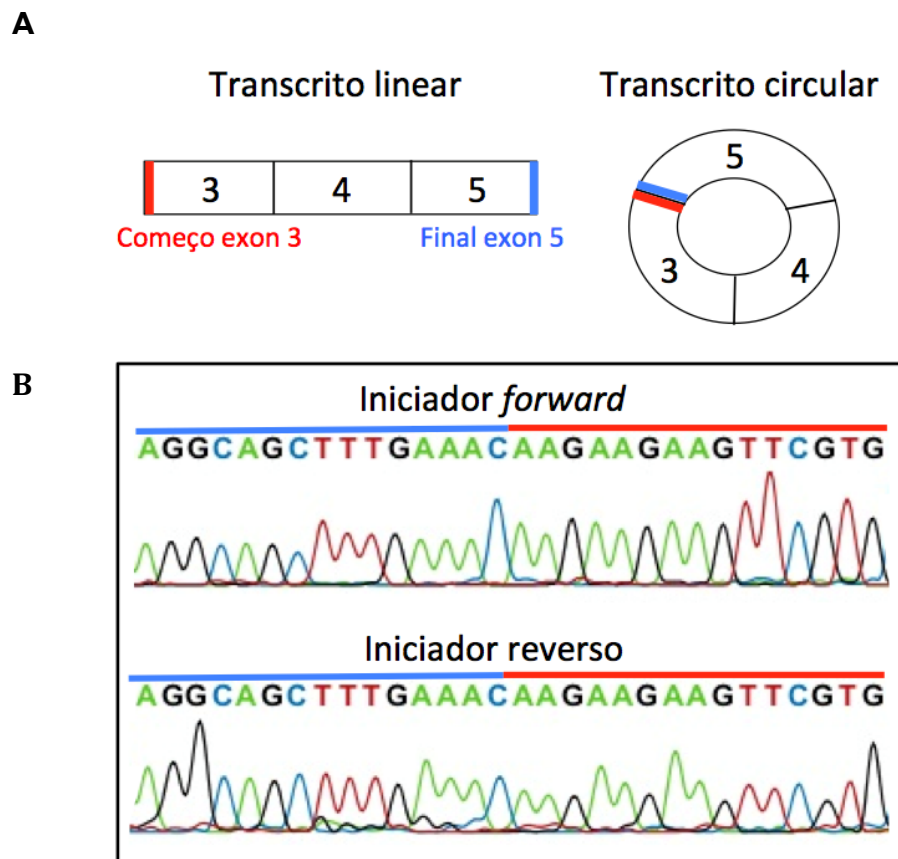
**Figura 53** - Visualização em gel de poliacrilamida dos produtos de amplificação de *UBXN7* (isoforma linear) em cDNA de EVs derivadas do plasma de pacientes. 1: Ladder 100pb; 2: Mama 33; 3: Mama 44; 4: Mama 25; 5: Mama 29; 6: Mama 37; 7: Mama 46; 8: Mama 50; 9: Mama 54; 10: controle positivo (cDNA célula HB4a); 11: controle negativo. A banda esperada para a isoforma linear é de 228pb, e está indicada por seta no gel.

Sequenciamos o produto de PCR pelo método de Sanger e a análise dos cromatogramas validou a expressão do circRNA *UBXN7*, confirmando a junção observada no NGS (**Figura 54**) do começo do exon 3 se unindo ao final do exon 5, em orientação que só é possível na isoforma circular (**Figura 55**).

A identificação de circRNAs em EVs é inédita, e a confirmação experimental da expressão de um circRNA em EVs foi feita tanto em vesículas provenientes do meio de cultura celular condicionado, quanto derivadas do plasma de pacientes. A definição das funções biológicas destas moléculas ainda requer investigação, mas existem evidências que os circRNAs possam atuar como esponjas de miRNAs, por competirem com os alvos endógenos para a ligação de miRNAs (HANSEN et al. 2013). A transferência de circRNAs entre células via EVs pode ser um novo mecanismo importante para a regulação de expressão gênica. Realizamos uma análise de predição de sítios de ligação de miRNAs no circRNA *UBXN7* e identificamos 67 potenciais sítios de ligação para diferentes miRNAs.



**Figura 54** - Análise de um dos *reads* do circRNA *UBXN7* no UCSC Genome Browser. Em vermelho está indicado o mapeamento de um dos *reads* obtidos por NGS (GGAATTGGGGCACGAACCTTCTTCTTGTTCCTTTCAAAGCTGCCTTTATGCATCAAATCAATGGGTGGCCGGA ATAGATCTGCAAGGGTAGTTAATTTCTTATCGATAGCT) em dois pedaços: nos exons 3 e 5



**Figura 55** - Resultado da análise por sequenciamento Sanger do RNA circular do transcrito *UBXN7* em cDNA de célula HB4a. **A:** Diagramas ilustrando o posicionamento esperado dos exons nos transcritos linear e circular. **B:** Cromatogramas demonstrando que amplificamos o transcrito circular, pois a junção do começo do exon 3 com o final do exon 5 (detectada no NGS e confirmada aqui pelo Sanger) só é possível no RNA circular.

## 5 CONCLUSÕES

Diante dos resultados apresentados nesta tese concluímos que:

- A superexpressão de ERBB2/HER2 na linhagem C5.2 altera o perfil proteômico de duas populações de EVs (20K – enriquecida para microvesículas e 100K – enriquecida para exossomos). Proteínas com capacidade demonstrada de induzir transformação maligna foram encontradas superexpressas nas duas populações de EVs da linhagem C5.2, quando comparadas com as EVs da linhagem HB4a. Para estas linhagens o oncogene HER2 foi predominantemente encontrado nas EVs, quando comparado às células secretoras.
- Padronizamos uma metodologia de construção de bibliotecas para NGS que possibilita a avaliação simultânea de todo o transcriptoma (pequenos e longos RNAs) derivado de EVs.
- Análises do conteúdo de ácidos nucleicos determinados por NGS nas EVs de C5.2 em relação as EVs de HB4a, permitiram encontrar 16 miRNAs e 644 outros RNAs diferencialmente representados, que podem estar direta ou indiretamente sendo afetados pela superexpressão de ERBB2/HER2.
- Demonstramos que existe uma incorporação preferencial de determinadas proteínas e ácidos nucleicos nas EVs em relação às células secretoras. Na análise proteômica, vimos que 74% das proteínas identificadas nas células de ambas linhagens não puderam ser encontradas nas EVs, e que 43% das proteínas identificadas nas duas populações de EVs de ambas linhagens não foram encontradas nas células.
- Ao compararmos o conteúdo de ácidos nucleicos encontrados por NGS nas duas populações de EVs de ambas linhagens, com o encontrado nas duas linhagens celulares, vimos 100 miRNAs e 3.190

outros RNAs diferencialmente representados, demonstrando que o conteúdo vesicular é diferente do celular.

- Identificamos diversos motivos de sequências enriquecidos nos miRNAs contidos nas EVs em relação a aqueles presentes nas células de origem, que podem estar potencialmente envolvidos com o mecanismo de endereçamento de miRNAs para as EVs.
- Observamos uma distribuição posicional preferencial de transcritos contendo preferencialmente a porção 3' de oncogenes relacionados ao câncer de mama em EVs da linhagem C5.2, sendo este um mecanismo potencial para a regulação destes oncogenes nas células-alvo.
- Análises por *Nanoparticle Tracking Analysis* não revelaram diferenças quantitativas significantes nas EVs derivadas de grupos de pacientes, que pudessem refletir um estadiamento progressivo da doença.
- Diversos transcritos obtidos a partir de EVs derivadas do plasma das pacientes sugeriram possível correlação com os grupos de progressão tumoral.
- Validamos experimentalmente um miRNA (miR-223-3p) que pode estar envolvido com o desenvolvimento de metástase. Encontramos 1.808 outros RNAs únicos diferencialmente representados entre os três grupos de pacientes, e que podem ser potenciais marcadores de estadiamento tumoral.
- Identificamos, pela primeira vez em EVs, a presença de centenas de RNAs circulares (circRNAs).
- Descobrimos 14 possíveis novos miRNAs humanos derivados das amostras estudadas nesta tese.

## 6 PERSPECTIVAS FUTURAS

Este estudo revelou alguns achados de interesse e abriu diversas novas perspectivas de estudo. A seguir relacionamos algumas das análises que faremos para complementar o estudo atual e para prosseguirmos com experimentos que talvez demonstrem a importância de alguns dos achados desta tese:

- Realizar análise proteômica de EVs derivadas do plasma das pacientes em colaboração com o Dr. Daniel Martins-de-Souza, atualmente docente na UNICAMP, a partir de frações específicas obtidas pela ultracentrifugação das EVs em gradiente de densidade de iodixanol, processo que elimina proteínas plasmáticas abundantes e já foi padronizado no laboratório.
- Coletar sangue periférico de um grupo de pelo menos 10 mulheres saudáveis, avaliadas pelo serviço de prevenção do AC Camargo Cancer Center (adendo ao projeto já foi aprovado pelo CEP) pareadas em idade com as pacientes do estudo, para a avaliação por NGS das EVs derivadas do plasma de indivíduos sem câncer.
- Validação técnica por qRT-PCR do miRNA miR-223-3p – encontrado aumentado por NGS nas EVs do grupo M em relação ao grupo N e subsequentemente validado por qRT-PCR nas mesmas amostras – em um grupo independente de amostras. Tais amostras incluirão EVs derivadas do plasma de mulheres saudáveis, EVs derivadas do plasma de pacientes com câncer de mama com as mesmas características clínico-patológicas das pacientes que foram sequenciadas, assim como amostras provenientes do tumor primário das mesmas pacientes.
- Buscar genes-alvos regulados pelo miR-223-3p através de análises *in silico* e *in vitro*.

- Validação técnica por PCR digital dos outros miRNAs encontrados diferencialmente representados pelo NGS, mas que não foram validados com significância estatística pelo qRT-PCR, provavelmente por serem pouco expressos ou terem *fold changes* baixos.
- Análise por *Ingenuity Pathway Analysis (IPA)* de enriquecimento de vias metabólicas representadas pelos transcritos encontrados diferencialmente representados entre os três grupos de pacientes pelo NGS, com o intuito de selecionar os melhores candidatos a biomarcadores.
- Validação técnica por qRT-PCR dos transcritos selecionados no item acima, nas mesmas amostras já sequenciadas e em um grupo independente de amostras.
- Validação técnica por sequenciamento Sanger de outros circRNAs identificados por NGS, assim como análise de sítios de ligação de miRNAs nos circRNAs validados. Caso algum circRNA apresente um enriquecimento de sítios de ligação, avaliar o seu potencial para atuar como esponja de miRNAs por ensaios de *gene reporter*.
- Validação por *Northern Blot* ou qRT-PCR da distribuição posicional preferencial em EVs do 3'-UTR de oncogenes relacionados ao câncer de mama.
- Buscar possíveis proteínas ligantes de RNA que reconheçam os motivos de sequência encontrados enriquecidos em miRNAs contidos em EVs, em relação a aqueles contidos nas células secretoras.
- Realizar análises com as sequências geradas de boa qualidade que não foram anotadas em nenhum banco de referência, visando a descoberta e anotação de novas moléculas.
- Deposição das sequências dos potenciais novos miRNAs identificados neste estudo em bancos de dados públicos.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

Abusamra AJ, Zhong Z, Zheng X, et al. Tumor exosomes expressing Fas ligand mediate CD8+ T-cell apoptosis. **Blood Cells Mol Dis** 2005; 35:169–73.

Adachi R, Horiuchi S, Sakurazawa Y, Hasegawa T, Sato K, Sakamaki T. ErbB2 down-regulates microRNA-205 in breast cancer. **Biochem Biophys Res Commun** 2011; 411:804–8.

Admyre C, Johansson SM, Qazi KR, et al. Exosomes with immune modulatory features are present in human breast milk. **J Immunol** 2007; 179:1969–78.

Al-Nedawi K, Meehan B, Kerbel RS, Allison AC, Rak J. Endothelial expression of autocrine VEGF upon the uptake of tumor-derived microvesicles containing oncogenic EGFR. **Proc Natl Acad Sci U S A** 2009; 106:3794–9.

Al-Nedawi K, Meehan B, Micallef J, et al. Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. **Nat Cell Biol** 2008; 10:619–24.

Allison KH, Dintzis SM, Schmidt R A. Frequency of HER2 heterogeneity by fluorescence in situ hybridization according to CAP expert panel recommendations: time for a new look at how to report heterogeneity. **Am J Clin Pathol** 2011; 136:864–71.

Amorim M, Fernandes G, Oliveira P, Martins-de-Souza D, Dias-Neto E, Nunes D. The overexpression of a single oncogene (ERBB2/HER2) alters the proteomic landscape of extracellular vesicles. **Proteomics** 2014; 14:1472–9.

Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. **Bioinformatics** 2014; 31:166–169.

Andre F, Scharz NEC, Movassagh M, et al. Malignant effusions and immunogenic tumour-derived exosomes. **Lancet** 2002; 360:295–305.

Antonyak MA, Li B, Boroughs LK, et al. Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells. **Proc Natl Acad Sci U S A** 2011; 108:4852–7.



Appaiah HN, Goswami CP, Mina L A, et al. Persistent upregulation of U6:SNORD44 small RNA ratio in the serum of breast cancer patients. **Breast Cancer Res** 2011; 13:R86.

Arroyo JD, Chevillet JR, Kroh EM, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. **Proc Natl Acad Sci U S A** 2011; 108:5003–5008.

Bahn JH, Zhang Q, Li F, et al. The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. **Clin Chem** 2015; 61:221–30.

Balaj L, Lessard R, Dai L, Cho Y. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. **Nature** 2011; 2:180.

Barteneva NS, Fasler-Kan E, Bernimoulin M, et al. Circulating microparticles: square the circle. **BMC Cell Biol** 2013; 14:23.

Bartsch R, Wenzel C, Steger GG. Trastuzumab in the management of early and advanced stage breast cancer. **Biologics** 2007; 1:19–31.

Batagov AO, Kurochkin I V. Exosomes secreted by human cells transport largely mRNA fragments that are enriched in the 3'-untranslated regions. **Biol Direct** 2013; 8:12.

Batagov AO, Kuznetsov V A, Kurochkin I V. Identification of nucleotide patterns enriched in secreted RNAs as putative cis-acting elements targeting them to exosome nano-vesicles. **BMC Genomics** 2011; 12 Suppl 3:S18.

Battke C, Ruiss R, Welsch U, et al. Tumour exosomes inhibit binding of tumour-reactive antibodies to tumour cells and reduce ADCC. **Cancer Immunol Immunother** 2011; 60:639–48.

Bazzini AA, Lee MT, Giraldez AJ. Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. **Science** 2012; 336:233–237.

Bernal-Mizrachi L, Jy W, Jimenez JJ, et al. High levels of circulating endothelial microparticles in patients with acute coronary syndromes. **Am Heart J** 2003; 145:962–70.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. **Nature** 2001; 409:363–6.

Bianco F, Perrotta C, Novellino L, et al. Acid sphingomyelinase activity triggers microparticle release from glial cells. **EMBO J** 2009; 28:1043–1054.

Birkenfeld J, Betz H, Roth D. Identification of cofilin and LIM-domain-containing protein kinase 1 as novel interaction partners of 14-3-3 zeta. **Biochem J** 2003; 369:45–54.

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

Boelens MC, Wu TJ, Nabet BY, et al. Exosome Transfer from Stromal to Breast Cancer Cells Regulates Therapy Resistance Pathways. **Cell** 2014; 159:499–513.

Bolukbasi MF, Mizrak A, Ozdener GB, et al. miR-1289 and “Zipcode”-like Sequence Enrich mRNAs in Microvesicles. **Mol Ther - Nucleic acids** 2012; 1:e10.

Bourguignon LYW. Interaction between the Adhesion Receptor, CD44, and the Oncogene Product, p185HER2, Promotes Human Ovarian Tumor Cell Activation. **J Biol Chem** 1997; 272:27913–27918.

Braga S, Dal Lago L, Bernard C, Cardoso F, Piccart M. Use of trastuzumab for the treatment of early stage breast cancer. **Expert Rev Anticancer Ther** 2006; 6:1153–64.

Brouwers JF, Aalberts M, Jansen JW A, et al. Distinct lipid compositions of two types of human prostasomes. **Proteomics** 2013; 13:1660–6.

Brown CJ, Hendrich BD, Rupert JL, et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. **Cell** 1992; 71:527–42.

Bunkenborg J, García GE, Paz MIP, Andersen JS, Molina H. The minotaur proteome: avoiding cross-species identifications deriving from bovine serum in cell culture models. **Proteomics** 2010; 10:3040–4.

Burstein H, Harris J, Morrow M. Malignant Tumors of the Breast. In: DeVita Jr VT, Lawrence TS, Rosenberg SA. editors. **DeVita, Hellman, Rosenberg’s Cancer principles & practice of oncology**. 9<sup>th</sup> ed. Philadelphia: Lippincott Williams & Wilkins, 2011. p.1401–46.

Caby M-P, Lankar D, Vincendeau-Scherrer C, Raposo G, Bonnerot C. Exosomal-like vesicles are present in human blood plasma. **Int Immunol** 2005; 17:879–87.

Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. **Proc Natl Acad Sci U S A** 2002; 99:15524–9.

Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. **Proc Natl Acad Sci U S A** 2004; 101:2999–3004.

Camps C, Buffa FM, Colella S, et al. hsa-miR-210 is induced by hypoxia and is an independent prognostic factor in breast cancer. **Clin Cancer Res** 2008; 14:1340–8.

Chaput N, Théry C. Exosomes: immune properties and potential clinical implementations. **Semin Immunopathol** 2011; 33:419–440.

Cheng L, Sharples RA, Scicluna BJ, Hill AF. Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood. **J Extracell vesicles** 2014; 3:1–14.

Chevillet JR, Kang Q, Ruf IK, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. **Proc Natl Acad Sci U S A** 2014; 111:14888–93.

Christov CP, Gardiner TJ, Szüts D, Krude T. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. **Mol Cell Biol** 2006; 26:6993–7004.

Christov CP, Trivier E, Krude T. Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. **Br J Cancer** 2008; 98:981–8.

Ciravolo V, Huber V, Ghedini GC, et al. Potential role of HER2-overexpressing exosomes in countering trastuzumab-based therapy. **J Cell Physiol** 2012; 227:658–67.

Colombo M, Raposo G, Théry C. Biogenesis, Secretion, and Intercellular Interactions of Exosomes and Other Extracellular Vesicles. **Annu Rev Cell Dev Biol** 2014; 30:255–289.

Deregibus MC, Cantaluppi V, Calogero R, et al. Endothelial progenitor cell derived microvesicles activate an angiogenic program in endothelial cells by a horizontal transfer of mRNA. **Blood** 2007; 110:2440–8.

Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and YRNA Fragment Expression Associated with Breast Cancer. **Biomark Cancer** 2014; 6:37–47.

Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Mote P, Martin DIK. 5'-YRNA fragments derived by processing of transcripts from specific YRNA genes and pseudogenes are abundant in human serum and plasma. **Physiol Genomics** 2013; 45:990–8.

Djuranovic S, Nahvi A, Green R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. **Science** 2012; 336:237–40.

Dolo V, Ginestra A, Cassarà D, et al. Selective localization of matrix metalloproteinase 9, beta1 integrins, and human lymphocyte antigen class I molecules on membrane vesicles shed by 8701-BC breast carcinoma cells. **Cancer Res** 1998; 58:4468–4474.

Dong X-Y, Guo P, Boyd J, et al. Implication of snoRNA U50 in human breast cancer. **J Genet genomics** 2009; 36:447–54.

Dvorak HF, Quay SC, Orenstein NS, et al. Tumor shedding and coagulation. **Science** 1981; 212:923–4.

Eichhorn SW, Guo H, McGeary SE, et al. mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues. **Mol Cell** 2014;

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

Emmanouilidou E, Melachroinou K, Roumeliotis T, et al. Cell-produced alpha-synuclein is secreted in a calcium-dependent manner by exosomes and impacts neuronal survival. **J Neurosci** 2010; 30:6838–51.

Esteller M. Non-coding RNAs in human disease. **Nat Rev Genet** 2011; 12:861–74.

Ferlay J, Soerjomataram I, Ervik M, et al. **GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality WorldWide: IARC CancerBase No. 11 [Internet]**. Disponível em: <<http://globocan.iarc.fr>>. Acesso em: 19 fev. 2015.

Fevrier B, Vilette D, Archer F, et al. Cells release prions in association with exosomes. **Proc Natl Acad Sci U S A** 2004; 101:9683–8.

Fidler IJ, Poste G. The “seed and soil” hypothesis revisited. **Lancet Oncol** 2008; 9:808.

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

Folkman J, Merler E, Abernathy C, Williams G. Isolation of a tumor factor responsible for angiogenesis. **J Exp Med** 1971; 133:275–88.

Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. **Genome Res** 2009; 19:92–105.

Gao L, Ma J, Mannoor K, et al. Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. **Int J cancer** 2015; 136:E623–9.

Gee HE, Buffa FM, Camps C, et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. **Br J Cancer** 2011; 104:1168–77.

Ginestra A, La Placa MD, Saladino F, Cassarà D, Nagase H, Vittorelli ML. The amount and proteolytic content of vesicles shed by human cancer cell lines correlates with their in vitro invasiveness. **Anticancer Res** 1998; 18:3433–7.

Gong B, Hu H, Chen J, et al. Caprin-1 is a novel microRNA-223 target for regulating the proliferation and invasion of human breast cancer cells. **Biomed Pharmacother** 2013; 67:629–36.

Gregory P A, Bert AG, Paterson EL, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. **Nat Cell Biol** 2008; 10:593–601.

Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. **Nucleic Acids Res** 2006; 34:D140–4.

Günthert U, Hofmann M, Rudy W, et al. A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. **Cell** 1991; 65:13–24.

Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. **Nature** 2010; 466:835–40.

Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. **Nature** 2010; 464:1071–6.

György B, Pálóczi K, Kovács A, et al. Improved circulating microparticle analysis in acid-citrate dextrose (ACD) anticoagulant tube. **Thromb Res** 2014; 133:285–92.

Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. **Nature** 2013; 495:384–8.

Harding C, Heuser J, Stahl P. Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes. **J Cell Biol** 1983; 97:329–39.

Harding C, Heuser J, Stahl P. Endocytosis and intracellular processing of transferrin and colloidal gold-transferrin in rat reticulocytes: demonstration of a pathway for receptor shedding. **Eur J Cell Biol** 1984; 35:256–63.

Harris R A, Eichholtz TJ, Hiles ID, Page MJ, O'Hare MJ. New model of ErbB-2 over-expression in human mammary luminal epithelial cells. **Int J cancer** 1999; 80:477–84.

Heijnen HF, Schiel AE, Fijnheer R, Geuze HJ, Sixma JJ. Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules. **Blood** 1999; 94:3791–3799.

Hla T, Lee MJ, Ancellin N, et al. Sphingosine-1-phosphate signaling via the EDG-1 family of G-protein-coupled receptors. **Ann N Y Acad Sci** 2000; 905:16–24.

Hofstatter E, Chung G, Harris L. Molecular Biology of Breast Cancer. In: DeVita Jr VT, Lawrence TS, Rosenberg SA. editors. **DeVita, Hellman, Rosenberg's Cancer principles & practice of oncology**. 9<sup>th</sup> ed. Philadelphia: Lippincott Williams & Wilkins, 2011. p.1392–400.

Hsu MT, Coca-Prados M. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. **Nature** 1979; 280:339–40.

Huang X, Yuan T, Tschannen M, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. **BMC Genomics** 2013; 14:319.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. **Nature** 2004; 431:931–45.

Ivanov A, Memczak S, Wylter E, et al. Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals. **Cell Rep** 2014; 10:170–177.

Janas MM, Novina CD. Not lost in translation: stepwise regulation of microRNA targets. **EMBO J** 2012; 31:2446–7.

Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. **Nat Biotechnol** 2014; 32:453–61.

Jeck WR, Sorrentino JA, Wang K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. **RNA** 2013; 19:141–57.

Jenjaroenpun P, Kremenska Y, Nair VM, Kremenskoy M, Joseph B, Kurochkin I V. Characterization of RNA in exosomes secreted by human breast cancer cell lines using next-generation sequencing. **PeerJ** 2013; 1:e201.

Ji H, Chen M, Greening DW, et al. Deep Sequencing of RNA from Three Different Extracellular Vesicle (EV) Subtypes Released from the Human LIM1863 Colon Cancer Cell Line Uncovers Distinct Mirna-Enrichment Signatures. **PLoS One** 2014; 9:e110314.

Johnstone RM, Adam M, Hammond JR, Orr L, Turbide C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). **J Biol Chem** 1987; 262:9412–20.

Jung E-J, Santarpia L, Kim J, et al. Plasma microRNA 210 levels correlate with sensitivity to trastuzumab and tumor presence in breast cancer patients. **Cancer** 2012; 118:2603–14.

Kahlert C, Melo S A, Protopopov A, et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. **J Biol Chem** 2014; 289:3869–75.

Kalra H, Simpson RJ, Ji H, et al. Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. **PLoS Biol** 2012; 10:e1001450.

Kim CW, Lee HM, Lee TH, Kang C, Kleinman HK, Gho YS. Extracellular membrane vesicles from tumor cells promote angiogenesis via sphingomyelin. **Cancer Res** 2002; 62:6312–7.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome Biol** 2013a; 14:R36.

Kim D-K, Kang B, Kim OY, et al. EVpedia: an integrated database of high-throughput data for systemic analyses of extracellular vesicles. **J Extracell vesicles** 2013b; 2:20384.

Kim JW, Wieckowski E, Taylor DD, Reichert TE, Watkins S, Whiteside TL. Fas ligand-positive membranous vesicles isolated from sera of patients with oral cancer induce apoptosis of activated T lymphocytes. **Clin Cancer Res** 2005; 11:1010–20.

King HW, Michael MZ, Gleagle JM. Hypoxic enhancement of exosome release by breast cancer cells. **BMC Cancer** 2012; 12:421.

Koga K, Matsumoto K, Akiyoshi T, et al. Purification, characterization and biological significance of tumor-derived exosomes. **Anticancer Res** 2005; 25:3703–7.

Köhn M, Pazaitis N, Hüttelmaier S. Why YRNAs? About Versatile RNAs and Their Functions. **Biomolecules** 2013; 3:143–56.

Koppers-Lalic D, Hackenberg M, Bijnsdorp I V, et al. Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. **Cell Rep** 2014; 8:1649–58.

Kosaka N, Iguchi H, Yoshioka Y, Takeshita F, Matsuki Y, Ochiya T. Secretory mechanisms and intercellular transfer of microRNAs in living cells. **J Biol Chem** 2010; 285:17442–52.

Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. **Nucleic Acids Res** 2014; 42:D68–73.

Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. **Nat Genet** 2005; 37:495–500.

Lacroix R, Judicone C, Poncelet P, et al. Impact of pre-analytical parameters on the measurement of circulating microparticles: towards standardization of protocol. **J Thromb Haemost** 2012; 10:437–46.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. **Science** 2001; 294:853–8.

Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. **Nature** 2001; 409:860–921.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. **Nat Methods** 2012; 9:357–9.

Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. **Science** 2001; 294:858–62.

Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. **Science** 2001; 294:862–4.

Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. **Cell** 1993; 75:843–54.

Lee Y, Ahn C, Han J, et al. The nuclear RNase III Drosha initiates microRNA processing. **Nature** 2003; 425:415–9.



Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. **EMBO J** 2004; 23:4051–60.

Lerner MR, Boyle JA, Hardin JA, Steitz JA. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. **Science** 1981; 211:400–2.

Levine AJ, Oren M. The first 30 years of p53: growing ever more complex. **Nat Rev Cancer** 2009; 9:749–758.

Li B, Antonyak MA, Zhang J, Cerione RA. RhoA triggers a specific signaling pathway that generates transforming microvesicles in cancer cells. **Oncogene** 2012; 31:4740–9.

Li CCY, Eaton S A, Young PE, et al. Glioma microvesicles carry selectively packaged coding and non-coding RNAs which alter gene expression in recipient cells. **RNA Biol** 2013; 10:1333–44.

Li YM, Zhou BP, Deng J, Pan Y, Hay N, Hung M-C. A hypoxia-independent hypoxia-inducible factor-1 activation pathway induced by phosphatidylinositol-3 kinase/Akt in HER2 overexpressing cells. **Cancer Res** 2005; 65:3257–63.

Liebhardt S, Ditsch N, Nieuwland R, et al. CEA-, Her2/neu-, BCRP- and Hsp27-positive microparticles in breast cancer patients. **Anticancer Res** 2010; 30:1707–12.

Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. **Nature** 2005; 433:769–73.

Lipinski B, Egyud LG. Resistance of cancer cells to immune recognition and killing. **Med Hypotheses** 2000; 54:456–60.

Long JS, Fujiwara Y, Edwards J, et al. Sphingosine 1-phosphate receptor 4 uses HER2 (ERBB2) to regulate extracellular signal regulated kinase-1/2 in MDA-MB-453 breast cancer cells. **J Biol Chem** 2010; 285:35957–66.

Louro R, El-Jundi T, Nakaya HI, Reis EM, Verjovski-Almeida S. Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. **Genomics** 2008; 92:18–25.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biol** 2014; 15:550.

Lu J, Getz G, Miska E A, et al. MicroRNA expression profiles classify human cancers. **Nature** 2005; 435:834–8.

Luga V, Zhang L, Vitoria-Petit AM, et al. Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. **Cell** 2012; 151:1542–56.

Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. **Science** 2004; 303:95–8.

Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. **Nature** 2007; 449:682–8.

Maccarrone G, Rewerts C, Lebar M, Turck CW, Martins-de-Souza D. Proteome profiling of peripheral mononuclear cells from human blood. **Proteomics** 2013; 13:893–7.

Maccarrone G, Turck CW, Martins-de-Souza D. Shotgun mass spectrometry workflow combining IEF and LC-MALDI-TOF/TOF. **Protein J** 2010; 29:99–102.

Mackowiak SD. Identification of novel and known miRNAs in deep-sequencing data with miRDeep2. **Curr Protoc Bioinforma** 2011; Chapter 12:Unit 12.10.

Mannoor K, Liao J, Jiang F. Small nucleolar RNAs in cancer. **Biochim Biophys Acta** 2012; 1826:121–8.

Marleau AM, Chen C-S, Joyce JA, Tullis RH. Exosome removal as a therapeutic adjuvant in cancer. **J Transl Med** 2012; 10:134.

Matera AG, Wang Z. A day in the life of the spliceosome. **Nat Rev Mol Cell Biol** 2014; 15:108–21.

Mathivanan S, Ji H, Simpson RJ. Exosomes: extracellular organelles important in intercellular communication. **J Proteomics** 2010; 73:1907–20.

Mathivanan S, Simpson RJ. ExoCarta: A compendium of exosomal proteins and RNA. **Proteomics** 2009; 9:4997–5000.

McCready J, Sims JD, Chan D, Jay DG. Secretion of extracellular hsp90alpha via exosomes increases cancer cell motility: a role for plasminogen activation. **BMC Cancer** 2010; 10:294.

McDougall SR, Anderson AR A, Chaplain M A J. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. **J Theor Biol** 2006; 241:564–89.

Melo SA, Sugimoto H, O'Connell JT, et al. Cancer Exosomes Perform Cell-Independent MicroRNA Biogenesis and Promote Tumorigenesis. **Cancer Cell** 2014; 26:707–721.

Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. **Nature** 2013; 495:333–8.

Miller TE, Ghoshal K, Ramaswamy B, et al. MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. **J Biol Chem** 2008; 283:29897–903.

Ministério da Saúde. **Estimativa / 2014 Incidência de Câncer no Brasil**. Rio de Janeiro: INCA, 2013.

Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. **Oncogene** 2007; 26:6469–87.

Montecalvo A, Larregina AT, Shufesky WJ, et al. Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. **Blood** 2012; 119:756–66.

Morello M, Minciacchi VR, De Candia P, et al. Large oncosomes mediate intercellular transfer of functional microRNA. **Cell Cycle** 2013; 12:3526–36.

Morris K V, Mattick JS. The rise of regulatory RNA. **Nat Rev Genet** 2014; 15:423–37.

Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. **Oncogene** 2009; 28:195–208.

Nguyen DG, Booth A, Gould SJ, Hildreth JEK. Evidence that HIV budding in primary macrophages occurs through the exosome release pathway. **J Biol Chem** 2003; 278:52347–54.

Nolte-'t Hoen ENM, Buermans HPJ, Waasdorp M, Stoorvogel W, Wauben MHM, 't Hoen P A C. Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. **Nucleic Acids Res** 2012; 40:9272–85.

Normoyle KPM, Briehner WM. Cyclase-associated protein (CAP) acts directly on F-actin to accelerate cofilin-mediated actin severing across the range of physiological pH. **J Biol Chem** 2012; 287:35722–32.

Obermajer N, Doljak B, Kos J. Cytokeratin 8 ectoplasmic domain binds urokinase-type plasminogen activator to breast tumor cells and modulates their adhesion, growth and invasiveness. **Mol Cancer** 2009; 8:88.

Ogawa Y, Kanai-Azuma M, Akimoto Y, Kawakami H, Yanoshita R. Exosome-like vesicles with dipeptidyl peptidase IV in human saliva. **Biol Pharm Bull** 2008; 31:1059–62.

Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. **Oncogene** 2014; 33:5311–8.

Osborne C, Wilson P, Tripathy D. Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. **Oncologist** 2004; 9:361–77.

Ostrowski M, Carmo NB, Krumeich S, et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. **Nat Cell Biol** 2010; 12:19–30; sup pp 1–13.

Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? **Front Genet** 2015; 6:1–11.

Pan BT, Johnstone RM. Fate of the transferrin receptor during maturation of sheep reticulocytes in vitro: selective externalization of the receptor. **Cell** 1983; 33:967–78.

Pan BT, Teng K, Wu C, Adam M, Johnstone RM. Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. **J Cell Biol** 1985; 101:942–8.

Parolini I, Federici C, Raggi C, et al. Microenvironmental pH is a key factor for exosome traffic in tumor cells. **J Biol Chem** 2009; 284:34211–22.

Pasquinelli AE, Reinhart BJ, Slack F, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. **Nature** 2000; 408:86–9.

Peinado H, Alečković M, Lavotshkin S, et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. **Nat Med** 2012; 18:883–91.

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

Perreault J, Noël J-F, Brière F, et al. Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs. **Nucleic Acids Res** 2005; 33:2032–41.

Perreault J, Perreault J-P, Boire G. Ro-associated Y RNAs in metazoans: evolution and diversification. **Mol Biol Evol** 2007; 24:1678–89.

Pinatel EM, Orso F, Penna E, et al. miR-223 is a coordinator of breast cancer progression as revealed by bioinformatics predictions. **PLoS One** 2014; 9:e84859.

Pisitkun T, Shen R, Knepper MA. Identification and proteomic profiling of exosomes in human urine. **Proc Natl Acad Sci U S A** 2004; 101:13368–73.

Ponta H, Sherman L, Herrlich P A. CD44: from adhesion molecules to signalling regulators. **Nat Rev Mol Cell Biol** 2003; 4:33–45.

Rajendran L, Hoshino M, Zahn TR, et al. Alzheimer's disease beta-amyloid peptides are released in association with exosomes. **Proc Natl Acad Sci U S A** 2006; 103:11172–7.

Raposo G, Nijman HW, Stoorvogel W, et al. B lymphocytes secrete antigen-presenting vesicles. **J Exp Med** 1996; 183:1161–72.

Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. **J Cell Biol** 2013; 200:373–83.

Reinhart BJ, Slack FJ, Basson M, et al. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. **Nature** 2000; 403:901–6.

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

Reis EM, Verjovski-Almeida S. Perspectives of Long Non-Coding RNAs in Cancer Diagnostics. **Front Genet** 2012; 3:32.

Ronchetti D, Todoerti K, Tuana G, et al. The expression pattern of small nucleolar and small Cajal body-specific RNAs characterizes distinct molecular subtypes of multiple myeloma. **Blood Cancer J** 2012; 2:e96.

Rottiers V, Näär AM. MicroRNAs in metabolism and metabolic disorders. **Nat Rev Mol Cell Biol** 2012; 13:239–50.

Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. **Nature** 2007; 448:83–6.

Ruoslahti E. Fibronectin in cell adhesion and invasion. **Cancer Metastasis Rev** 1984; 3:43–51.

Rusten TE, Vaccari T, Stenmark H. Shaping development with ESCRTs. **Nat Cell Biol** 2012; 14:38–45.

Sabatier F, Darmon P, Hugel B, et al. Type 1 and type 2 diabetic patients display different patterns of cellular microparticles. **Diabetes** 2002; 51:2840–5.

Safaei R, Larson BJ, Cheng TC, et al. Abnormal lysosomal trafficking and enhanced exosomal export of cisplatin in drug-resistant human ovarian carcinoma cells. **Mol Cancer Ther** 2005; 4:1595–604.

Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. **PLoS One** 2012; 7:e30733.

Sanguinetti CJ, Dias Neto E, Simpson AJ. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. **Biotechniques** 1994; 17:914–21.

Savina A, Furlán M, Vidal M, Colombo MI. Exosome release is regulated by a calcium-dependent mechanism in K562 cells. **J Biol Chem** 2003; 278:20083–90.

Shedden K, Xie XT, Chandaroy P, Chang YT, Rosania GR. Expulsion of small molecules in vesicles shed by cancer cells: association with gene expression and chemosensitivity profiles. **Cancer Res** 2003; 63:4331–7.

Sims JD, McCready J, Jay DG. Extracellular heat shock protein (Hsp)70 and Hsp90 $\alpha$  assist in matrix metalloproteinase-2 activation and breast cancer cell migration and invasion. **PLoS One** 2011; 6:e18848.

Skog J, Würdinger T, Van Rijn S, et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. **Nat Cell Biol** 2008; 10:1470–6.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. **Science** 1987; 235:177–82.

Smith CM, Steitz JA. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. **Mol Cell Biol** 1998; 18:6897–6909.

Sobin LH, Gospodarowicz M, Wittekind Ch. **TNM Classificação dos tumores malignos**. [tradução de A L M Eisenberg]. 7<sup>a</sup> ed. Rio de Janeiro: INCA, 2012. Tumores de mama; p.187-99.

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

Stamatopoulos B, Meuleman N, Haibe-Kains B, et al. microRNA-29c and microRNA-223 down-regulation has in vivo significance in chronic lymphocytic leukemia and improves disease risk stratification. **Blood** 2009; 113:5237–45.

Stamps AC, Davies SC, Burman J, O'Hare MJ. Analysis of proviral integration in human mammary epithelial cell lines immortalized by retroviral infection with a temperature-sensitive SV40 T-antigen construct. **Int J cancer** 1994; 57:865–74.

Starke S, Jost I, Rossbach O, et al. Exon Circularization Requires Canonical Splice Signals. **Cell Rep** 2014; 10:103–111.

Stoek A, Keller S, Riedle S, et al. A role for exosomes in the constitutive and stimulus-induced ectodomain cleavage of L1 and CD44. **Biochem J** 2006; 393:609–18.

Szajnik M, Czystowska M, Szczepanski MJ, Mandapathil M, Whiteside TL. Tumor-derived microvesicles induce, expand and up-regulate biological activities of human regulatory T cells (Treg). **PLoS One** 2010; 5:e11469.

Szostak N, Royo F, Rybarczyk A, et al. Sorting signal targeting mRNA into hepatic extracellular vesicles. **RNA Biol** 2014; 11:836–44.

Tavazoie SF, Alarcón C, Oskarsson T, et al. Endogenous human microRNAs that suppress breast cancer metastasis. **Nature** 2008; 451:147–52.

Taylor DD, Homesley HD, Doellgast GJ. Binding of specific peroxidase-labeled antibody to placental-type phosphatase on tumor-derived membrane fragments. **Cancer Res** 1980; 40:4064–4069.

Tesselaar MET, Romijn FPHTM, Van Der Linden IK, Prins FA, Bertina RM, Osanto S. Microparticle-associated tissue factor activity: A link between cancer and thrombosis? **J Thromb Haemost** 2007; 5:520–527.

Thakur BK, Zhang H, Becker A, et al. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. **Cell Res** 2014; 1–4.

The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. **Nature** 2012; 489:57–74.

Théry C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. **Curr Protoc cell Biol** 2006; Chapter 3:3.22.1–3.22.29.

Timms JF, White SL, O'Hare MJ, Waterfield MD. Effects of ErbB-2 overexpression on mitogenic signalling and cell cycle progression in human breast luminal epithelial cells. **Oncogene** 2002; 21:6573–86.

- Toth B, Liebhardt S, Steinig K, et al. Platelet-derived microparticles and coagulation activation in breast cancer patients. **Thromb Haemost** 2008; 100:663–9.
- Trajkovic K, Hsu C, Chiantia S, et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. **Science** 2008; 319:1244–7.
- Trapnell C, Williams B A, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nat Biotechnol** 2010; 28:511–5.
- Untch M, Gerber B, Harbeck N, et al. 13th st. Gallen international breast cancer conference 2013: primary therapy of early breast cancer evidence, controversies, consensus - opinion of a german team of experts (zurich 2013). **Breast Care** 2013; 8:221–9.
- Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. **Nat Cell Biol** 2007; 9:654–9.
- Valastyan S, Benaich N, Chang A, Reinhardt F, Weinberg R A. Concomitant suppression of three target genes can explain the impact of a microRNA on metastasis. **Genes Dev** 2009a; 23:2592–7.
- Valastyan S, Reinhardt F, Benaich N, et al. A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. **Cell** 2009b; 137:1032–46.
- Valenti R, Huber V, Filipazzi P, et al. Human tumor-released microvesicles promote the differentiation of myeloid cells with transforming growth factor-beta-mediated suppressive activity on T lymphocytes. **Cancer Res** 2006; 66:9290–8.
- Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome Biol** 2002; 3:RESEARCH0034.
- Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. **Science** 2001; 291:1304–51.
- Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. **Nat Cell Biol** 2011; 13:423–33.
- Villarroya-Beltri C, Gutiérrez-Vázquez C, Sánchez-Cabo F, et al. Sumoylated hnRNP A2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. **Nat Commun** 2013; 4:2980.



Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. **Science** 2013; 339:1546–58.

Voinnet O. Origin, biogenesis, and activity of plant microRNAs. **Cell** 2009; 136:669–87.

Wang PL, Bao Y, Yee M-C, et al. Circular RNA is expressed across the eukaryotic tree of life. **PLoS One** 2014a; 9:e90859.

Wang W, Eddy R, Condeelis J. The cofilin pathway in breast cancer invasion and metastasis. **Nat Rev Cancer** 2007; 7:429–440.

Wang W, Li J, Zhu W, et al. MicroRNA-21 and the clinical outcomes of various carcinomas: a systematic review and meta-analysis. **BMC Cancer** 2014b; 14:819.

Williams GT, Farzaneh F. Are snoRNAs and snoRNA host genes new players in cancer? **Nat Rev Cancer** 2012; 12:84–8.

Witwer KW, Buzás EI, Bemis LT, et al. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. **J Extracell vesicles** 2013; 2:

Wolf P. The nature and significance of platelet products in human plasma. **Br J Haematol** 1967; 13:269–88.

Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. **Arch Pathol Lab Med** 2014; 138:241–56.

Wolin SL, Steitz JA. Genes for two small cytoplasmic Ro RNAs are adjacent and appear to be single-copy in the human genome. **Cell** 1983; 32:735–44.

Wright MW, Bruford EA. Naming “junk”: human non-protein coding RNA (ncRNA) gene nomenclature. **Hum Genomics** 2011; 5:90–8.

Yan L-X, Huang X-F, Shao Q, et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. **RNA** 2008; 14:2348–60.

Yang BB, Rutnam ZJ. The non-coding 3' UTR of CD44 induces metastasis by regulating extracellular matrix functions. **J Cell Sci** 2012; 125:2075–2085.

Yang M, Chen J, Su F, et al. Microvesicles secreted by macrophages shuttle invasion-potentiating microRNAs into breast cancer cells. **Mol Cancer** 2011; 10:117.

Yu X, Harris SL, Levine AJ. The regulation of exosome secretion: a novel function of the p53 protein. **Cancer Res** 2006; 66:4795–801.

Zhao J, Sun BK, Erwin JA, Song J, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. **Science** 2008; 322:750–6.

Zitvogel L, Regnault A, Lozier A, et al. Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes. **Nat Med** 1998; 4:594–600.

**Anexo 1 – Carta de aprovação do Comitê de Ética e Pesquisa do Hospital  
A.C. Camargo**



**Comitê de Ética em  
Pesquisa - CEP**

São Paulo, 20 de agosto de 2012.

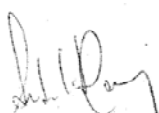
Ao  
**Dr. Emmanuel Dias Neto.**

**Ref.: Projeto de Pesquisa nº. 1554/11**  
**“Avaliação de Micropartículas Circulantes em Pacientes com Câncer de Mama Ductal Invasivo Her2+ e sua Possível Relação com a Agressividade Tumoral”.**

Os membros do Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antonio Prudente – Hospital do Câncer - A.C. Camargo/SP, em sua última reunião de 14/08/2012, após analisarem as respostas aos questionamentos realizados em reunião de 29/05/2012, **tomaram conhecimento e aprovaram** a seguinte documentação:


- Emenda ao projeto – versão 4, datada de 30 de julho de 2012;
- Termo de Consentimento Livre e Esclarecido – versão 6.0.

Atenciosamente,



**Dr. Antônio Hugo José Fróes Marques Campos**  
**2º Vice-Coordenador do Comitê de Ética em Pesquisa**

## Anexo 2 – Carta de aprovação do Comitê de Ética e Pesquisa do Hospital de Câncer de Barretos


|   |  |
|---|--|
|    | <b>Comitê de Ética em Pesquisa<br/>CEP</b> |
| Para: José Humberto Tavares Guerreiro Fregnani                                      |  |
| De: Ednise Woyciechowski<br><i>Vice Coordenadora do Comitê de Ética em Pesquisa</i> |  |
| Data: 14/02/2013  |  |
| Projeto de Pesquisa: 689/2013   |  |

Prezado (a) Senhor (a),

Vimos, por meio desta, informar que o Comitê de Ética em Pesquisa da Fundação Pio XII – Hospital de Câncer de Barretos analisou em reunião ordinária de 14/02/2013 o projeto de pesquisa 689/2013 "Avaliação de Micropartículas Circulantes em Pacientes com Câncer de Mama Ductal Her2+ e sua Possível Relação com a Agressividade Tumoral.", decidindo que o mesmo encontra-se: **"Aprovado"**.

*Solicitamos que sejam encaminhadas ao CEP relatórios semestral e final, bem como possíveis emendas e novos termos de consentimento livre e esclarecido. Notifique qualquer evento adverso sério ocorrido no centro e novas informações sobre a segurança do estudo a fim de se fazer o devido acompanhamento. Os dados individuais de todas as etapas da pesquisa devem ser mantidos em local seguro por 5 anos, após a conclusão da pesquisa, para possível auditoria dos órgãos competentes.*

Atenciosamente,



**Ednise Woyciechowski**  
*Vice Coordenadora do Comitê de Ética em Pesquisa*  
**Fundação Pio XII - Hospital de Câncer de Barretos**

---

Fundação Pio XII – Hospital de Câncer de Barretos  
Rua Antenor Duarte Villela, 1331 – B. Dr. Paulo Prata – Barretos – SP CEP 14784-400 – CNPJ 49.150.352/0001-12  
Tel. 17 33216600 R. 6894 – [cep@hcancerbarretos.com.br](mailto:cep@hcancerbarretos.com.br)



## Centro de Tratamento e Pesquisa Hospital do Câncer A. C. Camargo

Fundação Antonio Prudente - *Departamento de Mastologia*

Rua Professor Antônio Prudente, 211 - 01509-900 - São Paulo – SP - Fone: 2189-5150

### Termo de Consentimento Livre e Esclarecido (TCLE)

---

#### DADOS DE IDENTIFICAÇÃO DO PACIENTE OU RESPONSÁVEL LEGAL

Nome do Paciente: \_\_\_\_\_

Documento de identidade nº: \_\_\_\_\_ Sexo: M/F Nascimento: \_\_/\_\_/\_\_

Endereço: \_\_\_\_\_ nº \_\_\_\_ Compl. \_\_\_\_ CEP \_\_\_\_\_

Cidade: \_\_\_\_\_ Estado: \_\_\_\_\_ Telefone: \_\_\_\_\_

---

#### DADOS SOBRE O ESTUDO

O presente estudo, intitulado “Avaliação de Micropartículas Circulantes em Pacientes com Câncer de Mama Ductal (*In situ* e Invasivo) Her2+ e sua Possível Relação com a Agressividade Tumoral”, é um projeto de pesquisas do Centro Internacional de Pesquisas e Ensino do Hospital AC Camargo e do Departamento de Mastologia deste mesmo Hospital.

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antônio Prudente – Hospital do Câncer – AC Camargo – SP.

---

#### EXPLICAÇÕES DO PESQUISADOR AO PACIENTE OU SEU REPRESENTANTE LEGAL

##### Justificativas da pesquisa

Micropartículas são pequenas estruturas produzidas pelas células das pessoas, sejam elas saudáveis ou acometidas por alguma doença. Estudos iniciais sugerem que estas micropartículas possam ter algum papel no desenvolvimento do câncer. No entanto, mais estudos devem ser feitos para que o papel destas micropartículas seja melhor compreendido. Neste projeto, investigaremos as micropartículas em amostras de sangue de pacientes com tumores de mama, com o objetivo de conhecer melhor a sua função nesta doença. Os resultados deste estudo não trarão benefícios diretos a você, mas talvez sejam úteis no futuro, permitindo um diagnóstico mais informativo e um maior conhecimento a respeito desta doença.

##### Objetivos da pesquisa

Avaliar se micropartículas encontradas no sangue de pacientes com câncer de mama tem algum papel durante a evolução da doença.

## **Procedimentos que serão utilizados e propósitos**

Para participar do estudo você (a paciente) deverá ter:

- Diagnóstico de câncer de mama do tipo ductal (*in situ* ou invasivo) Her2+.
- Não ter recebido qualquer tratamento prévio (cirurgia, radioterapia ou quimioterapia).

Caso você concorde em participar deste estudo, será feita uma coleta de quinze mililitros (15mL) de sangue por um profissional treinado. Esta coleta deverá ocorrer quando você for fazer coleta de sangue para outros exames solicitados pelo seu médico, portanto você não terá despesas adicionais decorrentes da participação nesta pesquisa, e deste modo não será ressarcida de gastos com transporte ou alimentação no dia da coleta de sangue. Este será um procedimento absolutamente normal de coleta de sangue, geralmente tem uma duração de menos de 3 minutos e não há necessidade de uso de qualquer tipo de anestesia. O sangue coletado será processado para a separação de plasma e de células brancas. O plasma ficará armazenado até a sua utilização, em freezers apropriados, localizados no Centro de Pesquisas (sob responsabilidade do Dr. Emmanuel Dias-Neto, PhD). As células brancas vão ficar armazenadas no banco de tumores do Hospital AC Camargo (sob a responsabilidade do Dr. Antônio Hugo JFM Campos, MD, PhD). Havendo sobra de material, este poderá ser utilizado para outros estudos futuros visando a maior compreensão do câncer humano. Você pode optar, neste momento, pelo direito de ser contactado para fornecer autorização para cada nova pesquisa que venha a fazer uso do seu material biológico ou autorizar o uso futuro do mesmo sem a necessidade de um novo contato.

Você tem a absoluta liberdade de participar ou não deste estudo sem que isto acarrete qualquer comprometimento ou alteração em seu tratamento. Caso você concorde em participar, não haverá nenhum gasto extra para você ou para a sua seguradora de saúde. Do mesmo modo, a sua participação é uma colaboração voluntária, e não haverá nenhuma forma de remuneração. Você pode desistir de participar deste estudo em qualquer momento, mesmo que você já tenha assinado este termo. Sua desistência em participar do estudo, assim como seu pedido de retirada de qualquer material armazenado, podem ser feitos a qualquer momento, e não irão trazer qualquer prejuízo ao seu tratamento. Todas as informações obtidas são sigilosas e não serão usadas para nenhum outro fim que não os objetivos desta pesquisa. Sua identidade será sempre preservada.

Se você concordar em participar, este Termo de Consentimento deverá ser assinado em 2 vias, sendo que uma via permanecerá com você e a outra com o pesquisador responsável.

## **Complicações e riscos esperados**

Os riscos e complicações físicas associadas com a participação neste estudo são decorrentes da coleta de sangue, que será feita por profissional treinado seguindo procedimento padrão. É pouco provável que a coleta de sangue cause desconforto prolongado. Porém, se isto ocorrer, este deve ser bastante leve conseqüente apenas de uma picada de uma agulha. Todas as perguntas aqui feitas, e também o método de coleta que será utilizado neste estudo foram previamente empregados em outros projetos feitos neste e em outros hospitais do Brasil e do exterior. Todos os dados aqui levantados serão usados apenas para pesquisa científica, visando a melhor compreensão do câncer de mama. Todos os dados serão mantidos confidenciais e sua identidade será preservada em qualquer forma de divulgação científica dos mesmos.

### **Benefícios que poderão ser obtidos**

A sua participação neste estudo não irá te trazer nenhum benefício pessoal direto. Contudo, os resultados podem contribuir para o maior entendimento dos fatores que determinam o avanço e a agressividade do câncer de mama, e desta forma podem ajudar outras pessoas no futuro.

#### **Por favor, responda sim ou não às questões a seguir:**

Foi garantido a você o direito de receber resposta a qualquer pergunta ou esclarecimento acerca dos procedimentos, riscos, benefícios e outros assuntos relacionados com a pesquisa e o tratamento?

**SIM NÃO**

Ficou esclarecida a você a liberdade de retirar o seu consentimento a qualquer momento, sem que isso traga prejuízo quanto à continuidade do seu tratamento neste hospital?

**SIM NÃO**

Ficou esclarecido que não haverá ressarcimento ou qualquer tipo de pagamento referente à realização deste estudo?

**SIM NÃO**

Ficou esclarecido que não haverá indenização além das previstas pela lei, em reparação a dano imediato ou tardio, causado pela pesquisa em questão?

**SIM NÃO**

Ficou esclarecido que a sua identidade será preservada, mantendo-se todas as informações em caráter confidencial, sob guarda do pesquisador principal, e que os dados serão utilizados somente para análises estatísticas e posteriormente para publicações científicas, sem qualquer informação que possa identificar o participante?

**SIM NÃO**

Ficou esclarecido a você os possíveis riscos e complicações associadas com a participação neste estudo, decorrentes da coleta de sangue?

**SIM NÃO**

## CONSENTIMENTO PÓS-INFORMADO

De acordo com a Resolução nº 441/2011 do Conselho Nacional de Saúde, item 5: “o consentimento livre e esclarecido referente à coleta, depósito, armazenamento e utilização de material biológico humano em Biobanco é formalizado através de TCLE, por meio do qual o sujeito da pesquisa deve se manifestar expressamente quanto às seguintes alternativas, excludentes entre si: I - necessidade de novo consentimento a cada pesquisa; e II - dispensa de novo consentimento a cada pesquisa”.

Declaro que, após ter sido convenientemente esclarecida dos riscos e benefícios deste estudo, consinto em participar, na qualidade de paciente, deste projeto de pesquisa, e opto pela alternativa marcada abaixo:

( ) Quero ser contactada para fornecer novo consentimento a cada pesquisa futura que venha a fazer uso do meu material biológico.

( ) Dispensio ser contactada para fornecer novo consentimento a cada pesquisa futura que venha a fazer uso do meu material biológico .

Ainda de acordo com esta mesma resolução, item 5b: “o TCLE deve conter a garantia expressa da possibilidade de acesso pelo sujeito da pesquisa, inclusive a(s) forma(s) de contato para tal, ao conhecimento dos resultados obtidos com a utilização do seu material biológico e às orientações quanto as suas implicações, incluindo aconselhamento genético quando aplicável, a qualquer tempo”.

Deste modo, garantimos que voce terá acesso, a qualquer momento, a todos os resultados gerados a partir das análises de sua amostra de sangue. Do mesmo modo é seu direito obter orientação, quando cabível, incluindo aconselhamento genético, se achados aplicáveis forem obtidos a partir da sua amostra.

Em casos de dúvidas e para ter acesso aos resultados da pesquisa, assim como às orientações quanto as implicações dos resultados obtidos, contatar o pesquisador principal, Dr. Emmanuel Dias-Neto no Centro de Pesquisas do Hospital do Câncer A.C. Camargo pelo telefone 2189-5000 (ramal 5023). Se o pesquisador principal não fornecer as informações e/ou esclarecimentos suficientes, por favor, entre em contato com o Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antônio Prudente – Hospital do Câncer - AC Camargo/SP pelo telefone 2189-5000 (ramais 2069 ou 5020), de segunda-feira à quinta-feira das 7 horas às 18 horas e sexta-feira das 7 horas às 16 horas.

São Paulo, de \_\_\_\_\_ de 201\_\_\_\_

\_\_\_\_\_  
Assinatura do paciente ou responsável legal

\_\_\_\_\_  
Assinatura do pesquisador



## RESEARCH ARTICLE

# The overexpression of a single oncogene (ERBB2/HER2) alters the proteomic landscape of extracellular vesicles

Maria Amorim<sup>1,2</sup>, Gustavo Fernandes<sup>1</sup>, Paulo Oliveira<sup>3</sup>, Daniel Martins-de-Souza<sup>4,5\*</sup>, Emmanuel Dias-Neto<sup>1,4\*</sup> and Diana Nunes<sup>1</sup>

<sup>1</sup> Laboratory of Medical Genomics, AC Camargo Cancer Center, São Paulo, SP, Brazil

<sup>2</sup> Pós-graduação em Oncologia, Fundação Antônio Prudente, São Paulo, SP, Brazil

<sup>3</sup> Laboratório Nacional de Biotecnologia (LNBio), Campinas, SP, Brazil

<sup>4</sup> Laboratório de Neurociências Alzira Denise Hertzog Silva (LIM27), Instituto de Psiquiatria, FMUSP, São Paulo, SP, Brazil

<sup>5</sup> Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-University Munich, Munich, Germany

*ERBB2/HER2* amplification activates signaling cascades that lead to a tumor cell phenotype. However, despite its remarkable importance in oncology, the consequences of HER2 amplification over the extracellular vesicles (EVs) content have not yet been investigated. Here, we isolated EVs secreted by HB4a, a mammary luminal epithelial cell line and C5.2, its HER2-overexpressing clone. We isolated two EV sets (20 and 100 K) by ultracentrifugation and used electron microscopy and nanoparticle tracking analysis for their morphological characterization. We employed GeLC-MS/MS combined with isotope-coded protein labeling to evaluate cell-derived proteins and LC-MS/MS label free spectral counting to quantify the EVs proteome. We found higher HER2 levels in both C5.2-derived EVs when compared with C5.2 cells, suggesting its preferential shuttling. Proteins capable of inducing malignant transformation are enriched in both C5.2 EV subsets, including two HER2-related proteins involved in cell motility and invasion, cofilin and CD44. MetaCore™ analysis indicated an enrichment of cell adhesion and cytoskeleton-remodeling pathways in C5.2 EVs, as well as proteins related to HER2 signaling, such as sphingosine-1-phosphate pathway. Together, our data indicate that in terms of protein content, distinct vesicle sets reinforce and complement each other. Our results also suggest that HER2-upregulated proteins from EVs may be relevant for cellular malignancy and can be potential biomarkers for HER2<sup>+</sup> cancer patients.

Received: November 2, 2013

Revised: March 7, 2014

Accepted: March 27, 2014

## Keywords:

Cell biology / Exosomes / Extracellular vesicles / HER2 / Microvesicles



Additional supporting information may be found in the online version of this article at the publisher's web-site

## 1 Introduction

Extracellular vesicles (EVs) secreted by cells are important mediators of cell-cell communication [1]. EVs are diverse in terms of size and intracellular origin and their two most

studied populations are generated through fundamentally different mechanisms of cellular vesiculation [2]: microvesicles (MVs) are formed through the outward blebbing of the plasma membrane, and exosomes (EXs) are derived from the inward budding of the plasma membrane in a process coordinated by the endosomal-sorting complex required for transport machinery [3]. MVs are between 0.1 and 1 μm in size and EXs are smaller (50–100 nm), but both have been shown to

**Correspondence:** Dr. Diana Nunes, Laboratory of Medical Genomics, AC Camargo Cancer Center, Rua Taguá 440, 1st floor, São Paulo, SP 01508-010, Brazil

**E-mail:** dianann@gmail.com

**Abbreviations:** EV, extracellular vesicle; EX, exosome; ICPL, isotope-coded protein label; MV, microvesicle; S1P, sphingosine-1-phosphate; RT, room temperature

\*Additional corresponding authors: Dr. Daniel Martins-de-Souza and Dr. Emmanuel Dias-Neto, E-mail: danms90@gmail.com; emmanuel@cipe.accamargo.org.br

**Colour Online:** See the article online to view Fig. 1 in colour.

transport RNAs and proteins that are functionally exchanged between cells [4, 5]. Importantly, comparisons between EVs and their originating cells have shown that the incorporation of mRNAs, microRNAs, and proteins into these organelles appears to happen in a selective manner [6–8]. Tumor-derived EVs carry molecules that are relevant to cancer biology aspects including angiogenesis [9, 10], extracellular matrix remodeling [11], and oncogene transfer [4].

Here, we studied the EVs derived from the immortalized human mammary luminal epithelial cell-line HB4a and its derived clone, C5.2, which overexpresses *HER2/ERBB2* [12, 13], with the aim of evaluating if the overexpression of a single oncogene affects the EVs proteome. Given that these two cell lines primarily differ only by the expression level of this oncogene, this is a suitable model to determine a set of EVs-contained, HER2-mediated proteins, which may be related to the biology of HER2<sup>+</sup> tumors.

## 2 Materials and methods

### 2.1 Cell lines

HB4a and C5.2 cell lines were cultivated at 37°C and 5% CO<sub>2</sub> in RPMI 1640 medium (Invitrogen, Carlsbad, CA, EUA) supplemented with 10% FBS (Invitrogen), 1% antibiotic-antimycotic (penicillin/streptomycin/amphotericin B; Invitrogen) and 5 µg/mL hydrocortisone (Sigma-Aldrich, St. Louis, MO, EUA) as described before [14]. Complete medium was depleted of serum-derived EVs by overnight centrifugation at 100 000 × g.

### 2.2 Isolation of EVs

EVs were isolated from cell-conditioned medium as described [15]. Briefly, the medium was cleared of cells and debris by the following centrifugations steps: 300 × g for 10 min, 2000 × g for 20 min, 10 000 × g for 30 min; followed by final centrifugations at 20 000 × g (20 K EVs) and 100 000 × g (100 K EVs) for 60 min each.

### 2.3 Protein extraction

Proteins were extracted by resuspending the 20 and 100 K pellets in 100 µL lysis buffer containing 300 mM NaCl, 50 mM Tris pH 7.4, 0.5% NP-40 and antiproteases cocktail (Roche, Indianapolis, IN, USA). Cellular extracts were prepared by incubating cell pellets with 200 µL lysis buffer for 20 min on ice, and then collecting the supernatant of 20 000 × g centrifugation for 20 min.

### 2.4 Western blot

Protein concentrations were determined by Bradford assay (Bio-Rad, Hercules, CA, USA), and 1 µg total protein was submitted to SDS-PAGE. Primary antibodies used

were: HER2 1:10,000 (2165, Cell Signaling, Danvers, MA, USA), HSP70 1:500 (EXOAB-Hsp70A-1, System Biosciences, Mountain View, CA, USA), Flotillin 1:1,000 (ab41927, Abcam, Cambridge, MA, USA), and RAB27B 1:250 (HPA019849, Sigma-Aldrich).

### 2.5 Nanosight analysis

EV pellets were resuspended in 100 µL of PBS, diluted 1:10, and introduced into the NanoSight LM10 (NanoSight Ltd., Amesbury, UK) instrument's viewing unit with a disposable syringe. Images were acquired for 60 s (in triplicates for each sample) with the following parameters: camera shutter, 414; camera gain, 196; detection threshold, 10; min expected particle size, 30 nm.

### 2.6 Transmission electron microscopy

The ultracentrifugation pellets (20 and 100 K) were fixed overnight with 2% glutaraldehyde/paraformaldehyde at 4°C, washed with PBS, postfixed with 1% osmium tetroxide and 1% uranyl acetate for 10 min each, and dehydrated successively with 50, 70, 90, and 100% ethanol for 1 min each. Samples were then embedded in Hard Plus Resin 812 (EMS, Halfield, PA, USA), which polymerized overnight at 60°C. Ultrathin sections were cut and deposited on 200-mesh copper grids, stained with 0.5% uranyl acetate for 5 min, followed by lead for 4 min and analyzed with a JEM-1010 electron microscope (JEOL, Tokyo, Japan) at 80 kV.

### 2.7 ICPL labeling and prefractionation of cellular proteomes

Proteomes of the cells were labeled with isotope-coded protein label (ICPL) 4-plex (SERVA Electrophoresis, Heidelberg, Germany), as previously described [16]. Briefly, pools of 100 µg of total protein from each of the four sample groups (two cell lines in duplicate) were reduced for 30 min at 60°C and cooled to room temperature (RT). Next, free thiol groups were alkylated in the dark for 30 min at RT with 1 mL of 0.4 M iodoacetamide, quenched with 1 mL of 0.5 M *N*-acetylcysteine. For protein labeling, stable isotope tags—ICPL 0, ICPL 4, ICPL 6, and ICPL 10—were added with a tenfold molar excess to each of the cell samples and left to react for 2 h at RT. Four microliters of 1.5 M hydroxylamine was added to each sample for labeling inactivation. Remaining esters were hydrolyzed by raising the pH to 11–12 for 20 min. All four samples (two replicates of each cell line) were combined and prefractionated on a 12% SDS-PAGE mini-gel. Proteins were visualized using standard Coomassie brilliant blue staining. Each stained gel lane was sliced in 20 pieces, and subjected to trypsin in-gel digestion and lyophilized for further use.

## 2.8 In-solution digestion of EVs proteomes

EVs proteomes were digested in solution using 1:50 of a trypsin to sample ratio. Digested proteins were lyophilized for further use.

## 2.9 LC-MS/MS

Each digested gel slice from the cells as well as each EVs digested proteome were dissolved in aqueous 0.1% formic acid solution and injected in a nano-LC system (Eksigent, Dublin, CA, USA) onto a Zorbax-300 SB-C18 trapping column (5  $\mu$ m, 5  $\times$  0.3 mm id, Agilent Technologies, Santa Clara, CA, USA) and separated in a nano-column, 75  $\mu$ m id  $\times$  15 cm packed in house with 3  $\mu$ m C18 coated particles. The eluted peptides were analyzed online in an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with a nano-ESI source in positive mode [17]. The chromatographic separation and spectra acquisition were automatically performed through Xcalibur software (version 2.0.7, Thermo Scientific, San José, CA, USA).

## 2.10 Data analysis

MS raw data were processed directly in MASCOT Distiller (Matrix Sciences, London, UK) and searched against a decoy Uniprot human protein database (release 2013\_09). Search parameters were set as: peptide and fragment ion mass accuracy: 10 ppm and 0.6 Da, respectively; trypsin as enzyme allowing two missed cleavages; cysteine carbamidomethylation (fixed) and oxidized methionine (variable) as modifications. Experiments were performed in duplicates and only proteins identified by at least two distinct peptides with >95% confidence determined by MASCOT's algorithm were considered. Quantifications were also performed in MASCOT Distiller. By employing ICPL labeling for the quantitation of cells proteomes, the relative intensity of each precursor peptide in each labeled form was obtained from the extracted ion chromatograms and ratios were calculated by pairwise comparisons. For EVs, a spectral counting label-free analysis was performed using MASCOT Distiller's algorithm. Pathway analysis were performed using MetaCore<sup>TM</sup>.

# 3 Results and discussion

## 3.1 Characterization of EVs from HB4a and C5.2 cell lines

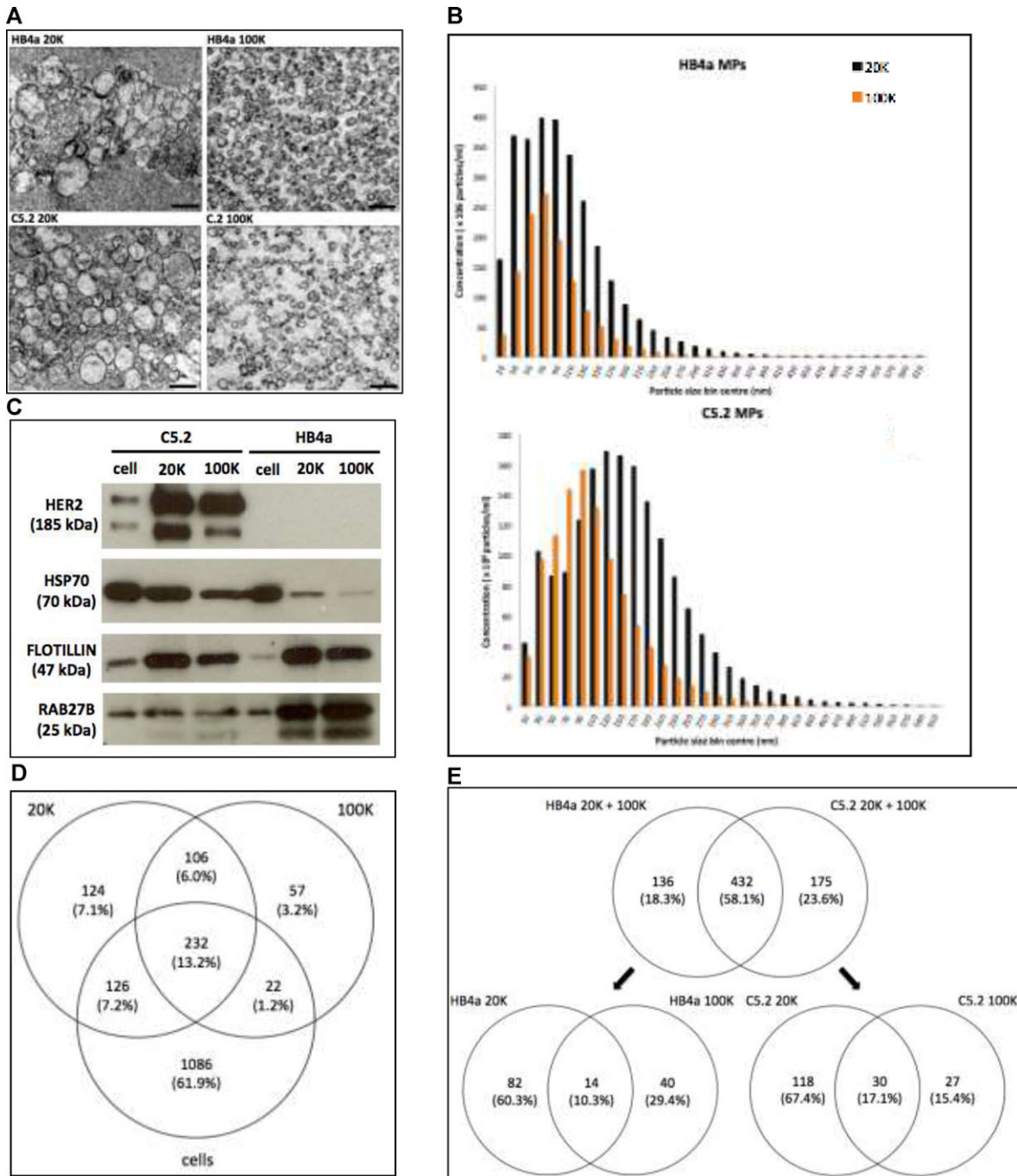
Electron microscopy revealed that both EV populations are morphologically distinct, with the 20 K fraction enriched for larger and more heterogeneous vesicles when compared to 100 K derived EVs, which displayed the EXs characteristic cup-shaped morphology [15] (Fig. 1A). These findings

were confirmed by nanoparticle tracking analysis that showed 20 K vesicles to be larger than 100 K vesicles for both cell lines (Wilcoxon rank test: W: 627, *p*-value: 0.04 for HB4a; W: 618, *p*-value: 0.05 for C5.2; Fig. 1B). The Western blot evaluation of vesicle-associated markers showed Flotillin to be present in all EV subpopulations from both cell lines, whereas HSP70 appeared to be more abundant in C5.2 (cells and both vesicle sets), when contrasted to HB4a. On the other hand, RAB27B was overexpressed in both 20 and 100 K subpopulations from HB4a when compared with C5.2 (Fig. 1C). HER2 was highly expressed in C5.2 cells, but appeared to be absent or very low in HB4a and its derived EVs (Fig. 1C), a finding also confirmed by real-time PCR (data not shown) and proteomics. Western blot analysis using normalized amounts of protein/well, showed HER2 to be much more abundant in both C5.2-derived EVs subpopulations, when compared with C5.2 cells (Fig. 1C).

## 3.2 Proteome analysis

Our analyses allowed the reliable identification of 1753 unique proteins; 83.6% of these (1466) were found in the cells (Supporting Information Tables 1 and 5), whereas 33.5% (588) could be identified in the 20 K (Supporting Information Table 2) and 23.8% (417) were found in the 100 K EVs (Supporting Information Table 3). The cell-exclusive component was comprised of 1086 proteins (61.9%), whereas 287 proteins (16.4%) were found only in EVs and 232 (13.2%) were present in both EVs subpopulations, as well as in the cells (Fig. 1D). The predicted biological processes and subcellular localizations of proteins exclusive of each EVs population from both cell lines are depicted in Supporting Information Fig. 1. Out of the 667 unique proteins identified by at least two peptides in both EVs subsets from both cell lines, 650 (97%) were already described as EVs components in EVpedia [18], suggesting that our EVs isolation protocol was well succeeded. Most of the non-EVpedia proteins (12/17) were exclusively found in 20 K derived vesicles, a finding likely to reflect the broader diversity of particles present in this fraction, as well as the exclusion of larger vesicles by filtering-based EXs isolation protocols [19].

In order to minimize the previously described serum-starvation stress [20] and focus our study on HER2-induced effects, HB4a and C5.2 cells were grown in FBS-containing medium, previously depleted of EVs. However, as proteins derived from FBS could potentially have been copurified together with the EVs we have flagged the most abundant proteins from FBS [21] that were also identified here (Table 1, Supporting Information Tables 1–6). Keratins, which were abundantly found in FBS, were also present in both EVs subsets. Interestingly, some studies have pointed to the functional role of these proteins in cytoskeleton remodeling, cell invasion, and metastasis [22]. In this sense, it is possible that the presence of keratins in EVs could indicate more than sample contamination.



**Figure 1.** Characterization of 20 and 100 K EVs. (A) Transmission electron microscopy showing the morphological heterogeneity of the EV populations; 20 K EVs are larger and more heterogeneous than 100 K EVs, which display the typical EX cup-shaped morphology. (B) Nanosight evaluation showing the size-shift between 20 K/100 K fractions for both cell lines. (C) Western blots showing the overexpression of HER2 in C5.2 cell line and EV subsets as well as the vesicle-associated markers Flotillin, RAB27B, and HSP-70. (D) Venn-diagram showing the overlap of proteins derived from the cell lines and their distinct EV fractions. (E) Venn diagrams showing the distribution of proteins exclusive to the EVs of each cell-line.

The proteome analysis of HB4a and C5.2 indicated 1466 proteins shared between both cell lines (Supporting Information Table 1). Interestingly, 74% of these proteins were not found in the EVs. On the other hand, of the 667 proteins

consistently found in all EVs sets (Supporting Information Tables 2 and 3), 43% were not identified in the original cells. Both findings reinforce the concept that EVs serve as shuttles of specific protein sets [6–8].

**Table 1.** Top 10 upregulated proteins of the 20 and 100 K EVs subsets in C5.2 (light gray) and HB4a (dark gray) cells

| Accession                 | Ratio C5.2/Hb4a | Number of Peptides | Description  | EVs subset |
|---------------------------|-----------------|--------------------|--|------------|
| COF1_HUMAN <sup>a)</sup>  | 95.96           | 3                  | Cofilin-1  | 20 K       |
| EF1A1_HUMAN               | 33.63           | 6                  | Elongation factor 1 alpha 1                          |            |
| FINC_HUMAN <sup>a)</sup>  | 26.83           | 71                 | Fibronectin  |            |
| EPS8_HUMAN                | 21.98           | 2                  | Epidermal growth factor receptor kinase substrate 8  |            |
| 1433Z_HUMAN <sup>a)</sup> | 20.38           | 6                  | 14–3–3 protein zeta/delta                            |            |
| RL31_HUMAN                | 19.62           | 2                  | 60S ribosomal protein L31                            |            |
| NUCL_HUMAN                | 18.83           | 4                  | Nucleolin  |            |
| THIL_HUMAN                | 18.30           | 7                  | Acetyl-CoA acetyltransferase, mitochondrial          |            |
| EF1D_HUMAN                | 16.11           | 5                  | Elongation factor 1 delta                            |            |
| CAP1_HUMAN <sup>a)</sup>  | 15.89           | 4                  | Adenylyl cyclase associated protein 1                |            |
| CD44_HUMAN                | 90.09           | 2                  | CD44 antigen   | 100 K      |
| RRAS_HUMAN                | 31.51           | 2                  | Ras-related protein R-Ras                            |            |
| IF4A3_HUMAN               | 25.81           | 2                  | Eukaryotic initiation factor 4A-III                  |            |
| AKA12_HUMAN               | 22.26           | 2                  | A-kinase anchor protein 12                           |            |
| CYBP_HUMAN                | 21.48           | 2                  | Calcyclin-binding protein                            |            |
| HIBCH_HUMAN               | 19.76           | 2                  | 3-hydroxyisobutyryl-CoA hydrolase, mitochondrial     |            |
| ERBB2_HUMAN               | 18.33           | 18                 | Receptor tyrosine-protein kinase erbB-2              |            |
| H3L_HUMAN                 | 18.14           | 2                  | Histone H3-like                                      |            |
| RL5_HUMAN                 | 16.60           | 2                  | 60S ribosomal protein L5                             |            |
| YBOX1_HUMAN               | 16.51           | 3                  | Nuclease-sensitive element-binding protein 1         |            |
| IF2A_HUMAN                | 0.50            | 2                  | Eukaryotic translation initiation factor 2 subunit 1 | 20 K       |
| NPM_HUMAN                 | 0.46            | 6                  | Nucleophosmin  |            |
| PARP1_HUMAN               | 0.45            | 2                  | Poly (ADP-ribose) polymerase 1                       |            |
| TCPG_HUMAN                | 0.44            | 11                 | T-complex protein 1 subunit gamma                    |            |
| TFR1_HUMAN                | 0.44            | 10                 | Transferrin receptor protein 1                       |            |
| K1109_HUMAN               | 0.41            | 2                  | Uncharacterized protein KIAA1109                     |            |
| RL12_HUMAN                | 0.41            | 2                  | 60S ribosomal protein L12                            |            |
| RHOA_HUMAN                | 0.40            | 2                  | Transforming protein RhoA                            |            |
| G6PD_HUMAN                | 0.40            | 7                  | Glucose-6-phosphate 1-dehydrogenase                  |            |
| SFPO_HUMAN                | 0.38            | 8                  | Splicing factor, proline- and glutamine rich         |            |
| FUMH_HUMAN                | 0.50            | 5                  | Fumarate hydratase, mitochondrial                    | 100 K      |
| AHNK_HUMAN                | 0.48            | 5                  | Neuroblast differentiation associated protein AHNK   |            |
| RAB35_HUMAN               | 0.46            | 4                  | Ras-related protein Rab-35                           |            |
| MOT4_HUMAN                | 0.45            | 4                  | Monocarboxylate transporter 4                        |            |
| TGM2_HUMAN                | 0.42            | 3                  | Protein-glutamine gamma-glutamyltransferase 2        |            |
| TFR1_HUMAN                | 0.42            | 5                  | Transferrin receptor protein 1                       |            |
| RL19_HUMAN                | 0.40            | 2                  | 60S ribosomal protein L19                            |            |
| RS14_HUMAN                | 0.39            | 2                  | 40S ribosomal protein S14                            |            |
| RAB3A_HUMAN               | 0.32            | 2                  | Ras-related protein Rab-3A                           |            |
| TCPB_HUMAN                | 0.31            | 7                  | T-complex protein 1 subunit beta OS = Homo sapiens   |            |

a) This protein was also found in FBS [21].

We compared the proteomes of EVs derived from the two cell lines to uncover proteins directly or indirectly affected by HER2 overexpression. A total of 136 (18.3%) proteins were exclusively found in HB4a-derived EVs, while 175 (23.6%) proteins were only found in C5.2-derived EVs, considering both 20 and 100 K subpopulations (Fig. 1E and Supporting Information Table 4). Proteomic alterations directly or indirectly correlated with HER2 overexpression were present in both 20 and 100 K vesicle populations, and their biological activities are mainly related to processes such as cell migration, malignant transformation, angiogenesis, metastasis, and cytoskeleton remodeling. Remarkably, despite the grouping of the top ten proteins from 20 and 100 K subpopulations in the same biological processes, they do not overlap (Table 1

and Supporting Information Table 6). This reinforces that the shuttling of proteins is not arbitrarily defined and suggests that the proteins carried by distinct vesicle-sets may have complementary roles at least for some biological pathways.

Proteins capable of inducing malignant transformation were found to be overrepresented in EVs from the HER2<sup>+</sup> cell line (C5.2), including the top proteins from 20 and 100 K subpopulations, both increased >90×: cofilin-1 and CD44, both involved in cell adhesion, motility and invasion, and cytoskeleton remodeling. HER2 regulates the cofilin pathway by activating PLC $\gamma$  to hydrolyze the PIP2-tethering cofilin to the cell membrane. Free cofilin leads to actin remodeling, which is essential for cell motility and invasion [23]. Cofilin inhibition was described to be involved in MVs

**Table 2.** Most represented pathways from HB4a and C5.2-derived EVs determined by MetaCore™ analysis

| Signaling and metabolic pathways  | 20 K            |        |                 |        | 100 K           |        |                 |        |
|---|-----------------|--------|-----------------|--------|-----------------|--------|-----------------|--------|
|   | HB4a            |        | C5.2            |        | HB4a            |        | C5.2            |        |
|   | <i>p</i> -value | PP     | <i>p</i> -value | PP     | <i>p</i> -value | PP     | <i>p</i> -value | PP     |
| Cell adhesion_Cadherin-mediated cell adhesion   | 1.61e-04        | 6/26   | 3.67e-07        | 9/26   | –               | 0/26   | –               | 0/26   |
| Cell adhesion_Chemokines and adhesion   | 2.25e-05        | 13/100 | 6.36e-15        | 26/100 | –               | 0/100  | –               | 0/100  |
| Cell adhesion_ECM remodeling  | <i>p</i> > 0.01 | 5/52   | 2.95e-05        | 10/52  | <i>p</i> > 0.01 | 5/52   | 5.71e-04        | 7/52   |
| Cell adhesion_Endothelial cell contacts by non-junctional mechanisms                        | 3.44e-08        | 9/24   | 7.79e-13        | 13/24  | 4.01e-13        | 12/24  | 1.22e-11        | 11/24  |
| Cell adhesion_Ephrin signaling  | <i>p</i> > 0.01 | 2/45   | 8.11e-03        | 6/45   | <i>p</i> > 0.01 | 3/45   | 8.16e-03        | 5/45   |
| Cell adhesion_Histamine H1 receptor signaling in the interruption of cell barrier integrity | 6.10e-04        | 7/45   | 7.61e-06        | 10/45  | –               | 0/45   | –               | 0/45   |
| Cell adhesion_Integrin-mediated cell adhesion and migration                                 | 1.52e-04        | 8/48   | 2.23e-10        | 15/48  | 6.40e-06        | 9/48   | 6.09e-08        | 11/48  |
| Cell adhesion_Role of CDK5 in cell adhesion   | <i>p</i> > 0.01 | 2/9    | 1.04e-05        | 5/9    | 1.66e-03        | 3/9    | 6.81e-05        | 4/9    |
| Cell adhesion_Role of tetraspanins in the integrin-mediated cell adhesion                   | 2.25e-06        | 9/37   | 2.84e-12        | 15/37  | 3.30e-09        | 11/37  | 1.80e-10        | 12/37  |
| Cytoskeleton remodeling_Cytoskeleton remodeling   | 3.07e-08        | 17/102 | 7.90e-13        | 24/102 | 1.50e-07        | 15/102 | 3.40e-10        | 18/102 |
| Cytoskeleton remodeling_FAK signaling   | <i>p</i> > 0.01 | 6/57   | 6.75e-05        | 10/57  | –               | 0/57   | –               | 0/57   |
| Cytoskeleton remodeling_Fibronectin-binding integrins in cell motility                      | 4.33e-07        | 9/31   | 2.54e-12        | 14/31  | 1.64e-06        | 8/31   | 3.48e-10        | 11/31  |
| Cytoskeleton remodeling_Integrin outside-in signaling                                       | 1.76e-04        | 8/49   | 3.10e-10        | 15/49  | –               | 0/49   | –               | 0/49   |
| Cytoskeleton remodeling_Keratin filaments   | –               | 0/36   | –               | 0/36   | 5.50e-05        | 7/36   | 4.50e-07        | 9/36   |
| Cytoskeleton remodeling_Neurofilaments  | 1.16e-03        | 5/25   | 2.49e-07        | 9/25   | –               | 0/25   | –               | 0/25   |
| Cytoskeleton remodeling_RaA regulation pathway  | <i>p</i> > 0.01 | 4/30   | 1.32e-04        | 7/30   | 9.60e-03        | 4/30   | 1.54e-04        | 6/30   |
| Cytoskeleton remodeling_Regulation of actin cytoskeleton by Rho GTPases                     | 4.04e-07        | 8/23   | 5.94e-09        | 10/23  | 2.21e-06        | 7/23   | 1.16e-07        | 8/23   |
| Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling                                | 1.55e-05        | 14/111 | 2.97e-07        | 18/111 | 1.15e-03        | 10/111 | 1.26e-05        | 13/111 |
| Development_Beta-adrenergic receptors regulation of ERK                                     | <i>p</i> > 0.01 | 4/47   | 4.49e-04        | 8/47   | –               | 0/47   | –               | 0/47   |
| Development_Ligand-independent activation of ESR1 and ESR2                                  | –               | 0/45   | –               | 0/45   | <i>p</i> > 0.01 | 3/45   | 8.16e-03        | 5/45   |
| Development_Lipoxin inhibitory action on PDGF, EGF and LTD4 signaling                       | <i>p</i> > 0.01 | 3/36   | 2.61e-03        | 6/36   | –               | 0/36   | –               | 0/36   |
| Development_Osteopontin signaling in osteoclasts  | 2.73e-03        | 5/30   | 1.46e-06        | 9/30   | 1.39e-03        | 5/30   | 1.17e-06        | 8/30   |
| Development_Regulation of telomere length and cellular immortalization                      | –               | 0/35   | –               | 0/35   | <i>p</i> > 0.01 | 3/35   | 2.72e-03        | 5/35   |
| Development_S1P1 receptor signaling via beta-arrestin                                       | <i>p</i> > 0.01 | 3/34   | 1.93e-03        | 6/34   | –               | 0/34   | –               | 0/34   |
| Development_S1P1 signaling pathway  | <i>p</i> > 0.01 | 4/44   | 4.43e-05        | 9/44   | <i>p</i> > 0.01 | 3/44   | 1.32e-03        | 6/44   |
| Development_S1P3 receptor signaling pathway   | <i>p</i> > 0.01 | 2/43   | 6.49e-03        | 6/43   | –               | 0/43   | –               | 0/43   |
| Development_S1P4 receptor signaling pathway   | –               | 0/22   | –               | 0/22   | <i>p</i> > 0.01 | 2/22   | 2.97e-03        | 4/22   |
| Development_Slit-Robo signaling   | 4.18e-05        | 7/30   | 1.22e-07        | 10/30  | 1.62e-04        | 6/30   | 7.93e-08        | 9/30   |
| Development_TGF-beta-dependent induction of EMT via MAPK                                    | <i>p</i> > 0.01 | 4/47   | 4.49e-04        | 8/47   | <i>p</i> > 0.01 | 4/47   | 1.88e-03        | 6/47   |
| Development_TGF-beta-dependent induction of EMT via RhoA, PI3K & ILK.                       | <i>p</i> > 0.01 | 4/46   | 2.02e-03        | 7/46   | 9.26e-03        | 5/46   | 3.54e-05        | 8/46   |
| Development_VEGF signaling via VEGFR2 – generic cascades                                    | <i>p</i> > 0.01 | 5/84   | 1.68e-03        | 10/84  | <i>p</i> > 0.01 | 6/84   | 5.46e-04        | 9/84   |
| Development_WNT signaling pathway. Part 2   | <i>p</i> > 0.01 | 3/53   | 4.58e-03        | 7/53   | –               | 0/53   | –               | 0/53   |
| Glycolysis and gluconeogenesis (short map)  | –               | 0/66   | –               | 0/66   | 2.86e-10        | 15/66  | 2.32e-07        | 12/66  |
| Glycolysis and gluconeogenesis p.3 / Human version  | 9.56e-04        | 5/24   | 2.75e-05        | 7/24   | –               | 0/24   | –               | 0/24   |
| LRRK2 in neurons in Parkinson's disease   | 6.11e-13        | 14/33  | 6.27e-16        | 17/33  | 7.80e-17        | 16/33  | 1.68e-12        | 13/33  |

Note: PP indicates the proteins identified over the total proteins in the pathway. Rows of the same colors were used to group the main MetaCore™ pathways.

formation, when activated Rho-associated protein kinase (ROCK) leads to LIM kinase (LIMK) phosphorylation, which ultimately phosphorylates cofilin that is prevented from severing actin, resulting in extended actin fibers [24]. The binding of hyaluronic acid to CD44 modulates CD44–HER2 interaction [25], which affects cell growth, survival, and differentiation [26, 27]. CD44 cleaved by ADAM10 was also reported to be shuttled in EVs, as a pathway to release a soluble form of this receptor [28]. Also noteworthy was the overexpression of fibronectin-1 in the 20 K subset of C5.2 cells, as it has been shown that when this protein is carried in tumor-derived MVs it is capable to transform nontumor cells [29]. We should note that fibronectin-1 is abundant in FBS [21]. However, its previous description in EVs from serum-starved cells [29] and its roles in cytoskeleton remodeling, cell migration, and metastasis [30] may be indicative of its biological role in this set of EVs and not a simple FBS contamination. From the top ten overexpressed proteins in C5.2 EVs (Table 1 and Supporting Information Table 6), some were described to interact with each other, such as 14–3–3 protein zeta/delta and cofilin-1 [31], or CAP-1 and cofilin [32], both associations being related to actin recycling and reorganization.

When HB4a- and C5.2-derived proteins were evaluated by MetaCore™ (Thomson-Reuters) the enrichment of cell adhesion and cytoskeleton remodeling proteins, as well as proteins grouped as distinct development-related pathways, was confirmed. These pathways corresponded to 32 of the 80 MetaCore-enriched signaling and metabolic pathways ( $p < 0.01$ ) (Table 2 and Supporting Information Table 7). The most enriched pathway, LRRK2 in neurons, was consistent in all EV fractions from both cells ( $p < 1.7 \times 10^{-12}$ ) and may be suggestive of its involvement in the EV synthesis pathway. Table 2 also shows pathways exclusively enriched in 20 or 100 K vesicles, as well as pathways that appear to be induced by HER2 overexpression, such as processes related to tumor invasion and cell migration, which may be relevant for the biology of HER2<sup>+</sup> tumors. An example is the enrichment of sphingosine-1-phosphate (S1P) signaling for both C5.2-derived EVs sets. Interestingly, among other cancer-related pathways affected by S1P, it appears to regulate angiogenesis and vascular endothelial cell function through its S1P3 receptor [33]. Together with HER2 and its S1P4 receptor, S1P is capable to activate the ERK-1/2 pathway in the MDA-MB-453 breast cancer cell, an event that is possibly related to tumor progression [34]. Curiously S1P3-pathway related proteins were only found in 20 K EVs, whereas S1P4-proteins were exclusively found in the 100 K set, reinforcing again the complementary nature of the 20 K and the 100 K EVs protein sets.

#### 4 Concluding remarks

Our results indicate that HER2 overexpression modifies the protein content of C5.2 cell line and of its EVs. The altered protein content of the EVs includes complementary protein sets,

which is reaffirmed by the presence of distinct proteins of the same pathway, linked to tumor-relevant phenotypes especially associated to tumor invasion. The investigation of these proteins in plasma-derived EVs from HER2 overexpressing tumors may reveal potential biomarkers that may be useful for diagnostic or prognostication purposes in a noninvasive fashion. It should also be noted that the higher levels of HER2 seen in C5.2 not only are reflected in its derived EVs but the C5.2-derived EVs shuttle even higher amounts of this oncogene. If this holds true for tumors, EVs may potentially act as HER2-shuttling agents, delivering it to distant sites, or even to other non-HER2 amplified tumor cells—a relevant issue as >20% of HER2<sup>+</sup> breast cancer patients show HER2 heterogeneity—i.e. only between 5 and 50% of the tumor cells carry the amplification of this oncogene [35].

*This work received financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and from Associação Beneficente Alzira Denise Hertzog Silva (ABADHS). EDN is a fellow from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil). We kindly thank Dr. Vilma Martins for the antibodies used in Western blotting.*

*The authors have declared no conflict of interest.*

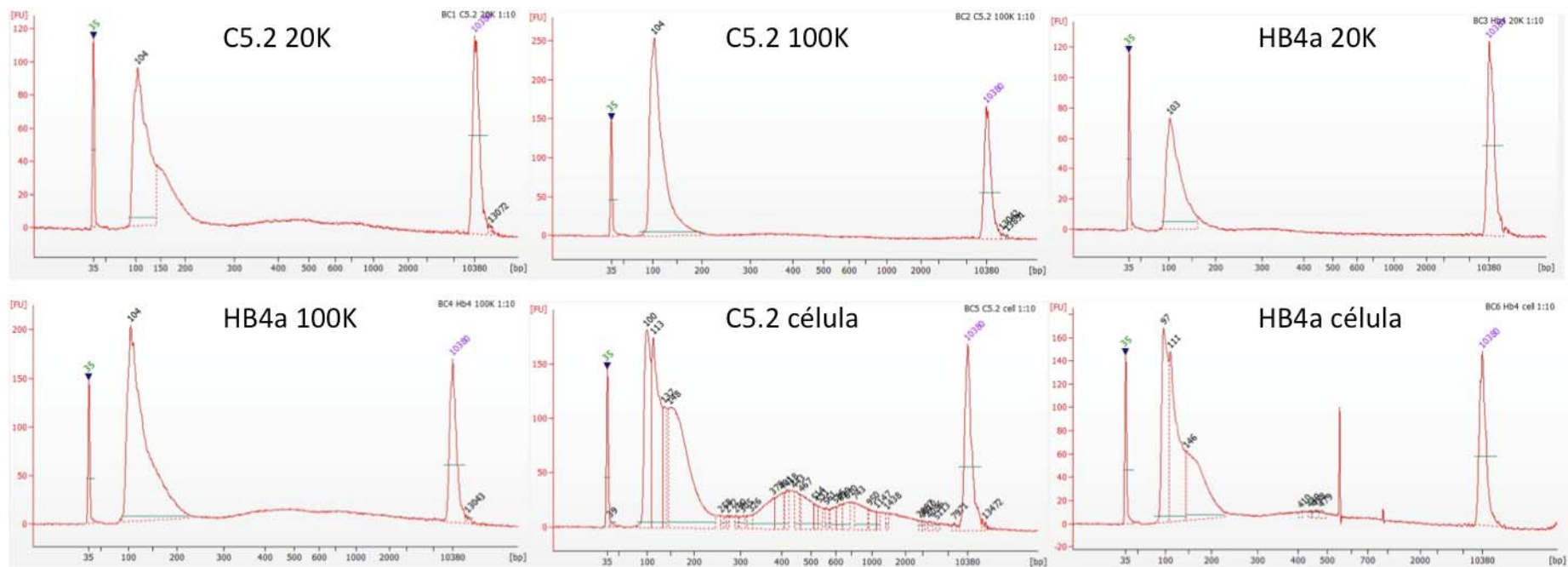
#### 5 References

- [1] Ratajczak, J., Wysoczynski, M., Hayek, F., Janowska-Wieczorek, A., Ratajczak, M. Z., Membrane-derived microvesicles: important and underappreciated mediators of cell-to-cell communication. *Leukemia* 2006, 20, 1487–1495.
- [2] Al-Nedawi, K., Meehan, B., Rak, J., Microvesicles: messengers and mediators of tumor progression. *Cell Cycle* 2009, 8, 2014–2018.
- [3] Williams, R. L., Urbé, S., The emerging shape of the ESCRT machinery. *Nat. Rev. Mol. Cell Biol.* 2007, 8, 355–368.
- [4] Al-Nedawi, K., Meehan, B., Micallef, J., Lhotak, V. et al., Inter-cellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat. Cell Biol.* 2008, 10, 619–624.
- [5] Peinado, H., Alečković, M., Lavotshkin, S., Matei, I. et al., Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat. Med.* 2012, 18, 883–891.
- [6] Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M. et al., Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* 2007, 9, 654–659.
- [7] Skog, J., Würdinger, T., Rijn, S., Van Meijer, D. H. et al., Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* 2008, 10, 1470–1476.
- [8] Nolte-’t Hoen, E. N. M., Buermans, H. P. J., Waasdorp, M., Stoorvogel, W. et al., Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. *Nucleic Acids Res.* 2012, 40, 9272–9285.

- [9] Deregibus, M. C., Cantaluppi, V., Calogero, R., Lo Iacono, M. et al., Endothelial progenitor cell derived microvesicles activate an angiogenic program in endothelial cells by a horizontal transfer of mRNA. *Blood* 2007, *110*, 2440–2448.
- [10] Choi, D., Lee, J., Park, G. W., Lim, H. et al., Proteomic analysis of microvesicles derived from human colorectal cancer cells. *J. Proteome Res.* 2007, *6*, 4646–4655.
- [11] Luga, V., Zhang, L., Vitoria-Petit, A. M., Ogunjimi, A. A. et al., Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. *Cell* 2012, *151*, 1542–1556.
- [12] Stamps, A. C., Davies, S. C., Burman, J., O'Hare, M. J., Analysis of proviral integration in human mammary epithelial cell lines immortalized by retroviral infection with a temperature-sensitive SV40 T-antigen construct. *Int. J. Cancer* 1994, *57*, 865–874.
- [13] Harris, R. A., Eichholtz, T. J., Hiles, I. D., Page, M. J., O'Hare, M. J., New model of ErbB-2 over-expression in human mammary luminal epithelial cells. *Int. J. Cancer* 1999, *80*, 477–484.
- [14] Carraro, D. M., Ferreira, E. N., de Campos Molina, G., Puga, R. D. et al., Poly (A) +transcriptome assessment of ERBB2-induced alterations in breast cell lines. *PLoS ONE* 2011, *6*, e21022.
- [15] Théry, C., Amigorena, S., Raposo, G., Clayton, A., Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* 2006, Chapter 3. Unit 3.22.1–3.22.29.
- [16] Maccarrone, G., Turck, C. W., Martins-de-Souza, D., Shotgun mass spectrometry workflow combining IEF and LC-MALDI-TOF/TOF. *Protein J.* 2010, *29*, 99–102.
- [17] Maccarrone, G., Rewerts, C., Lebar, M., Turck, C. W., Martins-de-Souza, D., Proteome profiling of peripheral mononuclear cells from human blood. *Proteomics* 2013, *13*, 893–897.
- [18] Kim, D.-K., Kang, B., Kim, O. Y., Choi, D.-S. et al., EVpedia: an integrated database of high-throughput data for systemic analyses of extracellular vesicles. *J. Extracell. Vesicles* 2013, *2*, 20384.
- [19] Raposo, G., Stoorvogel, W., Extracellular vesicles: exosomes, microvesicles, and friends. *J. Cell Biol.* 2013, *200*, 373–383.
- [20] Witwer, K. W., Buzás, E. I., Bemis, L. T., Bora, A. et al., Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *J. Extracell. Vesicles* 2013, *2*, 20360.
- [21] Bunkenborg, J., García, G. E., Paz, M. I. P., Andersen, J. S., Molina, H. The minotaur proteome: avoiding cross-species identifications deriving from bovine serum in cell culture models. *Proteomics* 2010, *10*, 3040–3044.
- [22] Obermajer, N., Doljak, B., Kos, J., Cytokeratin 8 ectoplasmic domain binds urokinase-type plasminogen activator to breast tumor cells and modulates their adhesion, growth and invasiveness. *Mol. Cancer* 2009, *8*, 88.
- [23] Wang, W., Eddy, R., Condeelis, J., The cofilin pathway in breast cancer invasion and metastasis. *Nat. Rev. Cancer* 2007, *7*, 429–440.
- [24] Li, B., Antonyak, M. A., Zhang, J., Cerione, R. A., RhoA triggers a specific signaling pathway that generates transforming microvesicles in cancer cells. *Oncogene* 2012, *31*, 4740–4749.
- [25] Bourguignon, L. Y., Zhu, H., Chu, A., Iida, N. et al. Interaction between the adhesion receptor, CD44, and the oncogene product, p185HER2, promotes human ovarian tumor cell activation. *J. Biol. Chem.* 1997, *272*, 27913–27918.
- [26] Ponta, H., Sherman, L., Herrlich, P. A., CD44: from adhesion molecules to signalling regulators. *Nat. Rev. Mol. Cell Biol.* 2003, *4*, 33–45.
- [27] Günthert, U., Hofmann, M., Rudy, W., Reber, S. et al., A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. *Cell* 1991, *65*, 13–24.
- [28] Stoeck, A., Keller, S., Riedle, S., Sanderson, M. P. et al., A role for exosomes in the constitutive and stimulus-induced ectodomain cleavage of L1 and CD44. *Biochem. J.* 2006, *393*, 609–618.
- [29] Antonyak, M. A., Li, B., Boroughs, L. K., Johnson, J. L. et al., Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells. *Proc. Natl. Acad. Sci. USA* 2011, *108*, 4852–4857.
- [30] Ruoslahti, E., Fibronectin in cell adhesion and invasion. *Cancer Metastasis Rev.* 1984, *3*, 43–51.
- [31] Birkenfeld, J., Betz, H., Roth, D., Identification of cofilin and LIM-domain-containing protein kinase 1 as novel interaction partners of 14–3–3 zeta. *Biochem. J.* 2003, *369*, 45–54.
- [32] Normoyle, K. P. M., Brieher, W. M., Cyclase-associated protein (CAP) acts directly on F-actin to accelerate cofilin-mediated actin severing across the range of physiological pH. *J. Biol. Chem.* 2012, *287*, 35722–35732.
- [33] Hla, T., Lee, M. J., Ancellin, N., Thangada, S. et al., Sphingosine-1-phosphate signaling via the EDG-1 family of G-protein-coupled receptors. *Ann. N. Y. Acad. Sci.* 2000, *905*, 16–24.
- [34] Long, J. S., Fujiwara, Y., Edwards, J., Tannahill, C. L. et al., Sphingosine 1-phosphate receptor 4 uses HER2 (ERBB2) to regulate extracellular signal regulated kinase-1/2 in MDA-MB-453 breast cancer cells. *J. Biol. Chem.* 2010, *285*, 35957–35966.
- [35] Allison, K. H., Dintzis, S. M., Schmidt, R. A., Frequency of HER2 heterogeneity by fluorescence in situ hybridization according to CAP expert panel recommendations: time for a new look at how to report heterogeneity. *Am. J. Clin. Pathol.* 2011, *136*, 864–71.



**Anexo 5** – Perfil do *Bioanalyzer* de bibliotecas para sequenciamento, obtido a partir de RNAs de EVs e células das linhagens C5.2 e HB4a.



**Anexo 6 – Dados clínicos de todas as pacientes recrutadas para este estudo.**

| Grupo   | Caso    | Nome          | RGH               | Idade         | RE   | RP   | ki67 | HER2 | FISH (HER2/cr17)       | Subtipo | Linfonodos acometidos/total | TNM         | Estadiamento | Local metástase         |   |
|---------|---------|---------------|-------------------|---------------|------|------|------|------|------------------------|---------|-----------------------------|-------------|--------------|-------------------------|---|
| in situ | 1       | Mama 4        | ACCC 10709330     | 59            | pos  | pos  | -    | 3+   | -                      | lumB    | 0/2                         | pT1spN0M0   | 0            | -                       |   |
|         | 2       | Mama 35       | ACCC 11508850     | 34            | neg  | neg  | -    | 3+   | -                      | Her2    | 0/3                         | pT1spN0M0   | 0            | -                       |   |
|         | 3       | Mama 39       | ACCC 11644780     | 54            | neg  | 10%  | -    | 3+   | -                      | lumB    | -                           | pT1spN0M0   | 0            | -                       |   |
|         | 4       | Mama 40       | ACCC 11653880     | 55            | 90%  | 40%  | -    | 3+   | -                      | lumB    | -                           | pT1spN0M0   | 0            | -                       |   |
|         | 5       | Mama 43       | ACCC 11624630     | 41            | 10%  | neg  | -    | 3+   | -                      | lumB    | -                           | pT1spN0M0   | 0            | -                       |   |
|         | 6       | Mama 45       | ACCC 11718920     | 47            | 80%  | 70%  | -    | 3+   | -                      | lumB    | -                           | pT1spN0M0   | 0            | -                       |   |
|         | 7       | Mama 47       | ACCC 11848580     | 52            | neg  | 1%   | -    | 3+   | -                      | lumB    | -                           | pT1spN0M0   | 0            | -                       |   |
| NO      | 1       | Mama 2        | ACCC 10688950     | 61            | neg  | neg  | 90%  | 3+   | -                      | Her2    | 0/21                        | pT1cpN0M0   | I            | -                       |   |
|         | 2       | Mama 3        | ACCC 10719080     | 38            | 40%  | 70%  | 20%  | 2+   | 3                      | lumB    | 0/2                         | pT1cpN0M0   | I            | -                       |   |
|         | 3       | Mama 5        | ACCC 10764360     | 68            | neg  | neg  | 20%  | 3+   | -                      | Her2    | 0/2                         | pT1bpN0M0   | I            | -                       |   |
|         | 4       | Mama 13       | ACCC 10868820     | 60            | 95%  | 30%  | 25%  | 2+   | 1.98 biopsia/ 2.4 peça | lumB    | 0/2                         | pT1bpN0M0   | I            | -                       |   |
|         | 5       | Mama 19       | ACCC 10383670     | 54            | 95%  | 40%  | 30%  | 3+   | -                      | lumB    | 0/1                         | pT1bpN0M0   | I            | -                       |   |
|         | 6       | Mama 22       | ACCC 11130060     | 40            | 100% | 90%  | 15%  | 2+   | 4                      | lumB    | 0/4                         | pT1cpN0M0   | I            | -                       |   |
|         | 7       | Mama 30       | ACCC 11230500     | 48            | 5%   | 5%   | 20%  | 3+   | -                      | lumB    | 0/3                         | pT1micpN0M0 | I            | -                       |   |
|         | 8       | Mama 31       | ACCC 10715110     | 48            | 90%  | 80%  | 2%   | 3+   | -                      | lumB    | 0/2                         | pT1bpN0M0   | I            | -                       |   |
|         | 9       | Mama 32       | ACCC 11221460     | 44            | 90%  | 80%  | 60%  | 3+   | -                      | lumB    | 0/4                         | pT1cpN0M0   | I            | -                       |   |
|         | 10      | Mama 44       | ACCC 11746820     | 54            | neg  | 30%  | 30%  | 3+   | -                      | lumB    | 0/4                         | pT1cpN0M0   | I            | -                       |   |
|         | 11      | Mama 49       | ACCC 11808350     | 71            | 90%  | 50%  | 30%  | 3+   | -                      | lumB    | 0/2                         | pT1cpN0M0   | I            | -                       |   |
|         | 12      | Mama 33       | ACCC 11343070     | 36            | 90%  | 90%  | 30%  | 3+   | -                      | lumB    | 0/5                         | pT2pN0M0    | IIA          | -                       |   |
|         | 13      | Mama 17       | ACCC 11041380     | 30            | neg  | 5%   | 80%  | 3+   | -                      | Her2    | 0/2                         | pT2pN0M0    | IIA          | -                       |   |
|         |         |               |                   | ACCC 11041380 |      | 90%  | 60%  | 70%  | 3+                     | -       | lumB                        | 0/2         | pT2pN0M0     | IIA                     | - |
|         | 14      | Mama 14       | ACCC 5013518      | 46            | 80%  | 30%  | 40%  | 3+   | -                      | lumB    | 0/1                         | pT2pN0M0    | IIA          | -                       |   |
|         | 15      | Mama 15       | ACCC 10377280     | 41            | 5%   | neg  | 40%  | 3+   | -                      | lumB    | 0/10                        | pT2pN0M0    | IIA          | -                       |   |
|         | 16      | Mama 27       | ACCC 11157780     | 37            | 90%  | 95%  | 70%  | 3+   | -                      | lumB    | 0/2                         | pT2pN0M0    | IIA          | -                       |   |
|         | 17      | Mama 34       | ACCC 10126230     | 65            | neg  | neg  | 50%  | 3+   | -                      | Her2    | 0/4                         | pT2pN0M0    | IIA          | -                       |   |
|         | 18      | Mama 36       | ACCC 11575030     | 28            | 95%  | 20%  | 60%  | 3+   | -                      | lumB    | 0/3                         | pT2pN0M0    | IIA          | -                       |   |
|         | 19      | Mama 7        | ACCC 10808430     | 50            | 30%  | 70%  | 80%  | 3+   | -                      | lumB    | N0*                         | cT2cN0M0*   | IIA          | -                       |   |
| 20      | Mama 26 | ACCC 11172570 | 66                | neg           | 10%  | 70%  | 3+   | -    | lumB                   | N0      | cT2pN0M0*                   | IIA         | -            |                         |   |
| N       | 1       | Mama 9        | ACCC 10637770     | 59            | 10%  | 10%  | 20%  | 3+   | -                      | lumB    | 2/12                        | pT1apN1aM0  | IIA          | -                       |   |
|         | 2       | Mama 20       | ACCC 10948680     | 62            | 100% | 100% | 10%  | 2+   | 2.37                   | lumB    | 1/2                         | pT1cpN1aM0  | IIA          | -                       |   |
|         | 3       | Mama 10       | ACCC 10812900     | 40            | neg  | 50%  | 70%  | 3+   | -                      | lumB    | 1/13                        | pT2pN1aM0   | IIB          | -                       |   |
|         | 4       | Mama 16       | ACCC 10950140     | 46            | neg  | neg  | 20%  | 3+   | -                      | Her2    | 2/14                        | pT2pN1aM0   | IIB          | -                       |   |
|         | 5       | Mama 12       | ACCC 10863400     | 37            | 95%  | 80%  | 20%  | 3+   | -                      | lumB    | 2/22                        | pT2pN1aM0   | IIB          | -                       |   |
|         | 6       | Mama 38       | ACCC 11579820     | 40            | 90%  | 70%  | 60%  | 3+   | -                      | lumB    | 3/27                        | pT2pN1aM0   | IIB          | -                       |   |
|         | 7       | Mama 18       | ACCC 10073930     | 56            | neg  | 30%  | 60%  | 3+   | -                      | lumB    | N1                          | cT2pN1M0*   | IIB          | -                       |   |
|         | 8       | Mama 25       | ACCC 11174050     | 64            | 5%   | neg  | 30%  | 3+   | -                      | lumB    | N1*                         | cT2cN1M0*   | IIB          | -                       |   |
|         | 9       | Mama 42       | ACCC 11774970     | 37            | 80%  | 10%  | 70%  | 3+   | -                      | lumB    | N1*                         | cT2cN1M0*   | IIB          | -                       |   |
|         | 10      | Mama 24       | ACCC 11133010     | 66            | neg  | neg  | 50%  | 3+   | -                      | Her2    | N1*                         | cT3cN1M0*   | IIIA         | -                       |   |
|         | 11      | Mama 28       | ACCC 10868320     | 46            | 70%  | 5%   | 40%  | 3+   | -                      | lumB    | N1*                         | cT3cN1M0*   | IIIA         | -                       |   |
|         | 12      | Mama 29       | ACCC 11260750     | 42            | 95%  | 95%  | 40%  | 3+   | -                      | lumB    | N2*                         | cT3cN2M0*   | IIIA         | -                       |   |
|         | 13      | Mama 23       | ACCC 11076980     | 36            | 80%  | 80%  | 80%  | 3+   | -                      | lumB    | N1*                         | cT4bcN1M0*  | IIIB         | -                       |   |
|         |         |               |                   | ACCC 11076980 |      | 5%   | 5%   | 90%  | 3+                     | -       | lumB                        | N1*         | cT4bcN1M0*   | IIIB                    | - |
|         | 14      | Mama 6        | ACCC 10513430     | 45            | 90%  | 5%   | 30%  | 3+   | -                      | lumB    | N1*                         | cT4dcN1M0*  | IIIB         | -                       |   |
|         | 15      | Mama 1        | ACCC 10679750     | 58            | neg  | neg  | 80%  | 3+   | -                      | Her2    | N1*                         | cT4bcN1M0*  | IIIB         | -                       |   |
|         | 16      | Mama 8        | ACCC 10807190     | 54            | neg  | neg  | 90%  | 3+   | -                      | Her2    | N1*                         | cT4bcN1M0*  | IIIB         | -                       |   |
|         | 17      | Mama 41       | ACCC 11652310     | 42            | 50%  | neg  | 20%  | 3+   | -                      | lumB    | N1*                         | cT4cN1M0*   | IIIB         | -                       |   |
|         | 18      | Mama 48       | ACCC 11826250     | 53            | 60%  | 60%  | 80%  | 3+   | -                      | lumB    | 5/32                        | pT4bpN2M0   | IIIB         | -                       |   |
| 19      | Mama 11 | ACCC 9039510  | 74                | 100%          | 100% | 40%  | 2+   | 3    | lumB                   | 19/20   | pT2pN3aM0                   | IIIC        | -            |                         |   |
| M       | 1       | Mama 21       | ACCC 11070710     | 42            | 70%  | 70%  | 50%  | 3+   | -                      | lumB    | N1*                         | cT4bcN1M1*  | IV           | medula óssea, fígado    |   |
|         | 2       | Mama 37       | ACCC 11583020     | 53            | neg  | neg  | 60%  | 3+   | -                      | Her2    | N1*                         | cT4cN1M1*   | IV           | medula óssea, pulmão    |   |
|         | 3       | Mama 46       | ACCC 11856720     | 49            | 2%   | 20%  | 80%  | 3+   | -                      | lumB    | N2*                         | cT4bcN2M1*  | IV           | pulmão                  |   |
|         | 4       | Mama 50       | Barretos 13.4813  | 34            | 60%  | 30%  | 90%  | 3+   | -                      | lumB    | N2*                         | cT4bcN2M1*  | IV           | mediastino, T12, fígado |   |
|         | 5       | Mama 51       | Barretos 13.6366  | 42            | 100% | neg  | pos  | 3+   | 3.5                    | lumB    | N1*                         | cT4bcN1M1*  | IV           | pulmão                  |   |
|         | 6       | Mama 52       | Barretos 13.7703  | 34            | neg  | neg  | 80%  | 3+   | -                      | Her2    | N3*                         | cT4cN3M1*   | IV           | fígado                  |   |
|         | 7       | Mama 53       | Barretos 13.7128  | 51            | neg  | neg  | pos  | 3+   | -                      | Her2    | N2*                         | cT4bcN2M1*  | IV           | pulmão, fígado          |   |
|         | 8       | Mama 54       | Barretos 04.14380 | 51            | neg  | 40%  | 85%  | 3+   | -                      | lumB    | N1*                         | cT3cN1M1*   | IV           | medula óssea            |   |
|         | 9       | Mama 55       | Barretos 13.7974  | 55            | neg  | neg  | 90%  | 3+   | -                      | Her2    | N3*                         | cT4cN3M1*   | IV           | cérebro, medula óssea   |   |
|         | 10      | Mama 56       | Barretos 13.8819  | 48            | pos  | pos  | 80%  | 3+   | -                      | lumB    | N0*                         | cT2cN0M1*   | IV           | medula óssea            |   |
|         | 11      | Mama 57       | Barretos 14.1833  | 35            | 80%  | 80%  | 75%  | 3+   | -                      | lumB    | N0*                         | cT4bcN0M1*  | IV           | pulmão                  |   |
|         | 12      | Mama 58       | Barretos 14.1150  | 47            | neg  | neg  | 20%  | 3+   | -                      | Her2    | N1*                         | cT4dcN1M1*  | IV           | pulmão, fígado          |   |
|         | 13      | Mama 59       | Barretos 14.4466  | 56            | -    | -    | -    | -    | -                      | -       | N3*                         | cT4bcN3M1*  | IV           | -                       |   |

**Legenda** – As pacientes incluídas no sequenciamento estão indicadas em negrito. Estadiamento de acordo com TNM – Classificação dos Tumores Malignos – INCA, 7ª Edição. Todas as pacientes são HER2+, subtipo Her2: ER/PR- e subtipo lumB: ER+ e/ou PR+.

**Anexo 7 – Perfil do *Bioanalyzer* de bibliotecas para sequenciamento, obtido a partir de RNAs de EVs derivadas do plasma de pacientes**

