

## Interações realizadas por genes líderes do câncer de mama: uma abordagem computacional

Interactions performed by leading genes in breast cancer: a computational approach

Interacciones realizadas por genes líderes del cáncer de mama: un abordaje computacional

Daniela Oliveira Lima Magalhães<sup>1</sup>, Italo Rodrigues Magalhães<sup>2</sup>, Frederico Marques Andrade<sup>2</sup>, Lorena Roseli Rios Durães<sup>1</sup>, João Vítor da Silva Rodrigues<sup>2</sup>, Emerson Willian Santos de Almeida<sup>2\*</sup>, Orlene Veloso Dias<sup>2</sup>, Celio Pereira David<sup>1</sup>, Otávio Cardoso Filho<sup>3</sup>, Ludmilla Regina de Souza David<sup>1</sup>.

---

### RESUMO

**Objetivo:** Identificar os genes líderes e as interações realizadas por eles na rede do câncer de mama. **Métodos:** Trata-se de uma pesquisa exploratória de natureza computacional. A identificação dos genes envolvidos com o câncer de mama foi realizada na base de dados de genética humana *GeneCards*. Para isso, adotou-se o descritor *breast cancer*, selecionado na biblioteca de descritores de assuntos médicos *Medical Subject Headings*. Após a identificação dos genes, foi construído um mapa genômico com o auxílio do *software Search Tool for the Retrieval of Interacting Genes/Proteins* versão 10.0. **Resultados:** Foram identificados 8.610 genes, sendo selecionados 400 genes com as pontuações mais elevadas para a realização do mapa de interação genômico. Os genes *TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300*, *SRC* e *CDK2* foram considerados os genes líderes da rede câncer de mama, os quais interagiram com 90, 88, 64, 62, 60, 66, 58 e 52 genes, respectivamente. Interações entre os genes líderes e genes pertencentes a vias moleculares distintas foram realizadas. **Conclusão:** O mapeamento computacional dos genes mais relevantes para a condição câncer de mama revelou *TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300*, *SRC* e *CDK2* como genes líderes e as interações realizadas por eles dentro da rede em estudo.

**Palavras-chave:** Mapa genômico, Carcinoma de mama, Rede de interação gênica.

---

### ABSTRACT

**Objective:** To identify the leading genes and their interactions in the breast cancer network. **Methods:** This is an exploratory research of a computational nature. Identification of genes involved with breast cancer was performed in the *GeneCards* human genetics database. For this, the descriptor *breast cancer*, selected in the medical subject headings library, was used. After identifying the genes, a genomic map was constructed with the help of the *Software Search Tool for the Retrieval of Interacting Genes / Proteins* version 10.0. **Results:** 8,610 genes were identified and 400 genes with the highest scores were selected for genomic interaction mapping. The genes *TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300*, *SRC* and *CDK2* were considered the leading genes of the breast cancer network, which interacted with 90, 88, 64, 62, 60, 66, 58 and 52 genes respectively. Interactions between the leading genes and genes belonging to distinct molecular pathways were performed. **Conclusion:** Computational mapping of the genes most relevant to the breast cancer condition revealed *TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300*, *SRC* and *CDK2* as the leading genes and the interactions carried out by them within the study network.

**Key words:** Genomic map, Breast carcinoma, Gene interaction network.

---

<sup>1</sup> Faculdades Promove, Belo Horizonte – MG. \* E-mail: [emersonwillian@usp.br](mailto:emersonwillian@usp.br)

<sup>2</sup> Universidade Estadual de Montes Claros (UNIMONTES), Montes Claros - MG.

<sup>3</sup> Faculdade FACIONORTE especializada em odontologia, Montes Claros – MG.

## RESUMEN

**Objetivo:** Identificar los genes líderes y las interacciones realizadas por ellos en la red del cáncer de mama.

**Métodos:** Se trata de una investigación exploratoria de naturaleza computacional. La identificación de los genes involucrados con el cáncer de mama fue realizada en la base de datos de genética humana GeneCards. Para ello, se adoptó el descriptor de cáncer de mama, seleccionado en la biblioteca de descriptores de asuntos médicos Medical Subject Headings. Después de la identificación de los genes, se construyó un mapa genómico con la ayuda del software Search Tool para el Retrieval de Interacting Genes / Proteins versión 10.0. **Resultados:** Se identificaron 8.610 genes, siendo seleccionados 400 genes con las puntuaciones más elevadas para la realización del mapa de interacción genómica. Los genes TP53, PIK3CA, AKT1, EGFR, MAPK1, EP300, SRC y CDK2 se consideraron los genes líderes de la red de cáncer de mama, que interactuaron con 90, 88, 64, 62, 60, 66, 58 y 52 genes, respectivamente. Se realizaron interacciones entre los genes líderes y los genes pertenecientes a diferentes vías moleculares. **Conclusión:** El mapeo computacional de los genes más relevantes para la condición de cáncer de mama reveló TP53, PIK3CA, AKT1, EGFR, MAPK1, EP300, SRC y CDK2 como genes líderes y las interacciones realizadas por ellos dentro de la red en estudio.

**Palabras clave:** Mapa genómico, Carcinoma de mama, Red de interacción génica.

## INTRODUÇÃO

O câncer de mama é considerado o segundo tipo de câncer mais comum entre as mulheres, responsável por 25% de casos novos a cada ano (INCA, 2016). Compreende um grupo heterogêneo de doenças, com morfologia, biologia e prognóstico variáveis. Tem como agravante a disseminação precoce das células tumorais, que normalmente se disseminam através das vias linfáticas e se implantam em vértebras e arcos costais, podendo acometer também pulmão, fígado e cérebro (ORDING AG, 2016). Nesse tipo de tumor, a disseminação metastática é a principal causa da baixa sobrevida dos pacientes (RUITERKAMP R, et al., 2011).

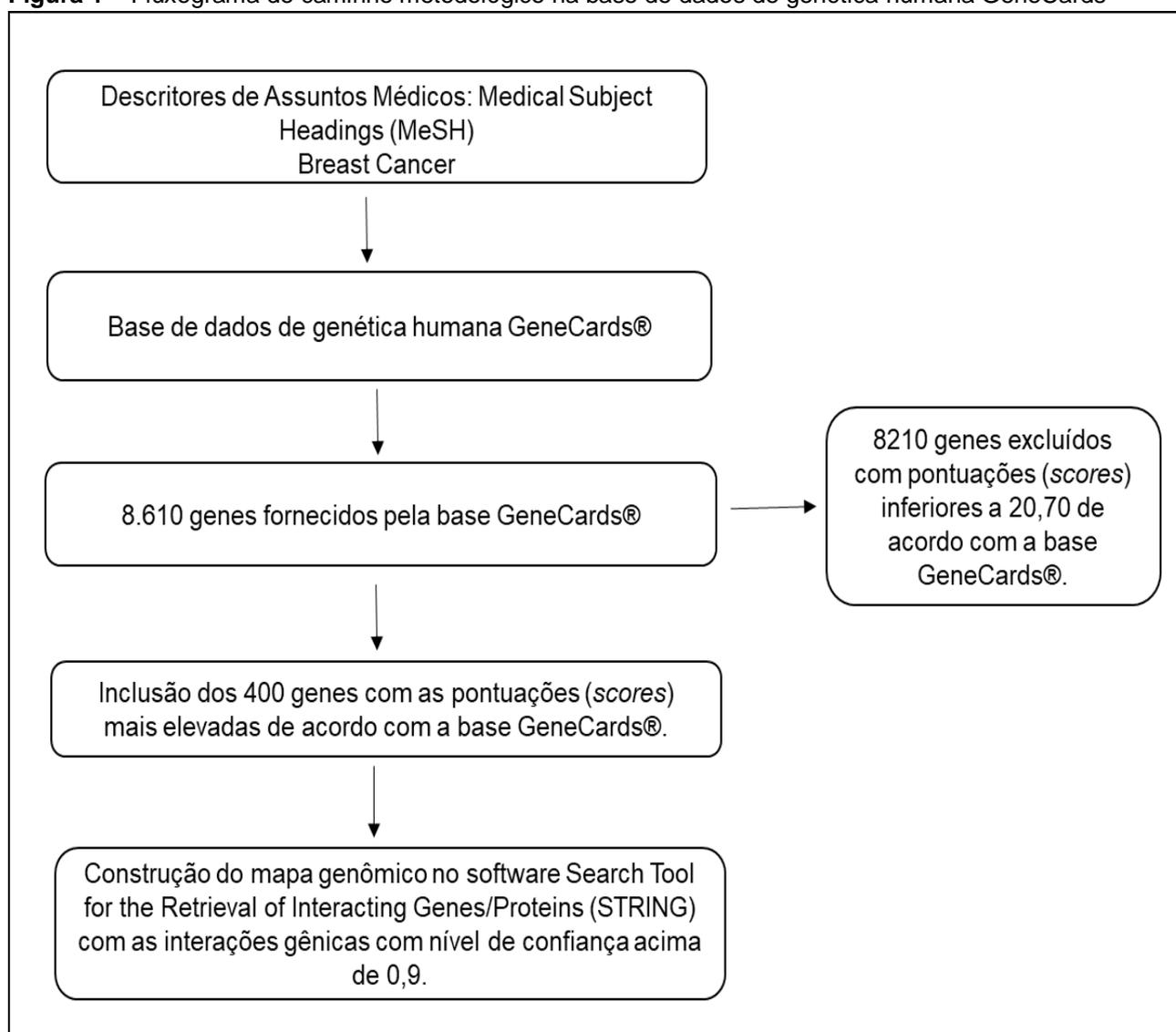
A avaliação de biomarcadores moleculares tem contribuído para o diagnóstico e a seleção da modalidade terapêutica empregada de acordo com características peculiares dos pacientes, que podem ser submetidos à cirurgia, radioterapia, quimioterapia ou terapia hormonal (AMOROSO V, et al., 2015). Os receptores hormonais de estrógeno e progesterona e o receptor de crescimento epidérmico humano-2 (PIZON M, et al., 2018; ILIE SM et al., 2018) são os fatores preditivos mais utilizados para a escolha do tratamento hormonal (WEIGEL e DOWSETT, 2010). A avaliação de mutações e polimorfismos na proteína supressora de tumor BRCA1 tem sido também bastante utilizada como ferramenta de prevenção da doença (SOUDEH GF, et al., 2018). O biomarcador CA 15-3, o qual detecta a forma solúvel da proteína MUC-1, é o marcador sérico mais amplamente usado em pacientes com câncer de mama, especialmente útil no monitoramento de pacientes com doença metastática (DUFFY MJ, et al., 2010). Apesar dos avanços das pesquisas e do reflexo delas no diagnóstico, prognóstico e conduta terapêutica, o câncer de mama ainda é a maior causa de morte por câncer no mundo entre o gênero feminino (INCA, 2016).

A partir dessas evidências, metodologias de investigação que contribuam para uma maior compreensão da biologia do câncer de mama se fazem necessárias. Estudos de bioinformática têm recebido bastante atenção no cenário mundial das ciências da saúde na interpretação e na compreensão de grandes volumes de dados (VERLI H, 2014; JAERGER S, et al., 2017). Em câncer de mama, estudos *in silico* identificaram 390 novos pares de fármacos anticâncer capazes de exibir efeitos sinérgicos antitumorais, muitos deles em fase de testes *in vitro* (JAERGER S, et al., 2017). Análise de bioinformática realizada a partir de bancos de dados relevou os genes diferencialmente expressos em cânceres de mama iniciais (CUI C, et al., 2018). Tendo em vista a necessidade de uma maior compreensão acerca da biologia do câncer de mama, em especial a partir do grande volume de informações genéticas disponíveis em bases de dados, a presente pesquisa teve por objetivo identificar os genes líderes e suas interações no mapa de interações genômicas do câncer de mama por meio de ferramentas computacionais.

## METODOLOGIA

O presente estudo caracteriza-se como uma pesquisa exploratória de natureza computacional. A identificação dos genes envolvidos com a condição câncer de mama ocorreu em janeiro de 2018 através de pesquisas na base de dados de genética humana GeneCards®, utilizando a nomenclatura oficial para genes humanos *Human Genoma Organization* (HUGO). Os descritores analisados foram utilizados em língua inglesa e definidos de acordo com a biblioteca de descritores de assuntos médicos *Medical Subject Headings* (MeSH). Como critério de inclusão para a condição câncer de mama, o termo selecionado a partir dos descritores foi aquele relativo exclusivamente à condição maligna: “Breast cancer”. A partir da base de dados estudada, foram obtidos 8.610 genes relacionados à condição maligna da mama, sendo selecionados para o corrente estudo 400 genes com as pontuações (*scores*) mais elevadas. Todos os genes estudados possuíram pontuação acima de 20,69 fornecido pela base GeneCards®.

**Figura 1** – Fluxograma do caminho metodológico na base de dados de genética humana GeneCards®



Fonte: Dados da pesquisa, 2018.

Após a identificação dos genes, foi construído um mapa genômico com o auxílio do software *Search Tool for the Retrieval of Interacting Genes/Proteins* (STRING) versão 10.0. Foram utilizadas todas as fontes de interações ativas na espécie *Homo sapiens* (mineração de texto, experimentos, base de dados, co-expressão, vizinhança, fusão de genes e co-ocorrência) que apresentavam grau de confiança acima de 0,9.

A análise do mapeamento da rede de genes na base de dados STRING resultou em um arquivo *Text Output* contendo as pontuações para todas as associações realizadas por cada gene. Assim, tais pontuações foram tabuladas e multiplicadas por 1000. Posteriormente foi realizada a soma da pontuação de interação de cada gene, com o intuito de encontrar a pontuação denominada número ponderado de ligações, *Weighted Number of Links* (WNL) (POSWAR FO, et al., 2015). O WNL refere-se ao cálculo da soma das associações de cada gene dentro da rede a que pertence (PEREIRA DVC, 2016).

Na presente pesquisa, também foi calculada a pontuação de interação global dos genes denominada *Significant Global Connectivity* (TIS), conectividade global significativa. Os valores de TIS foram estabelecidos somando-se as associações de cada gene a partir dos valores obtidos do arquivo de interações globais *Protein Network Data* gerado pelo software STRING (PEREIRA DVC, 2016).

A razão entre WNL/TIS representa os genes mais influentes dentro da rede em estudo. Genes que não realizaram interações com os genes selecionados da rede são chamados de genes órfãos (POSWAR FO, et al., 2015; PEREIRA DVC, 2016; SANTOS E, OTACILIO SH, 2016).

Os genes do mapa genômico foram agrupados em função dos seus valores de WNL por meio da análise de agrupamento realizada pelo algoritmo *K-means*, realizada no programa de estatística SPSS® PASS Statistic 18.0. Para avaliar as diferenças entre os grupos gerados pela análise de *K-means*, foi realizado o teste de normalidade Kolmogorov Smirnov ( $p = 0.000$ ) seguido da análise de Kruskal-Wallis. A significância estatística foi fixada em um  $p$  valor  $< 0,001$ . Os genes que apresentaram maiores valores de WNL foram considerados genes líderes (SANTOS E, OTACILIO SH, 2016). Os genes que interagem com cada gene líder da rede também foram avaliados.

## RESULTADOS

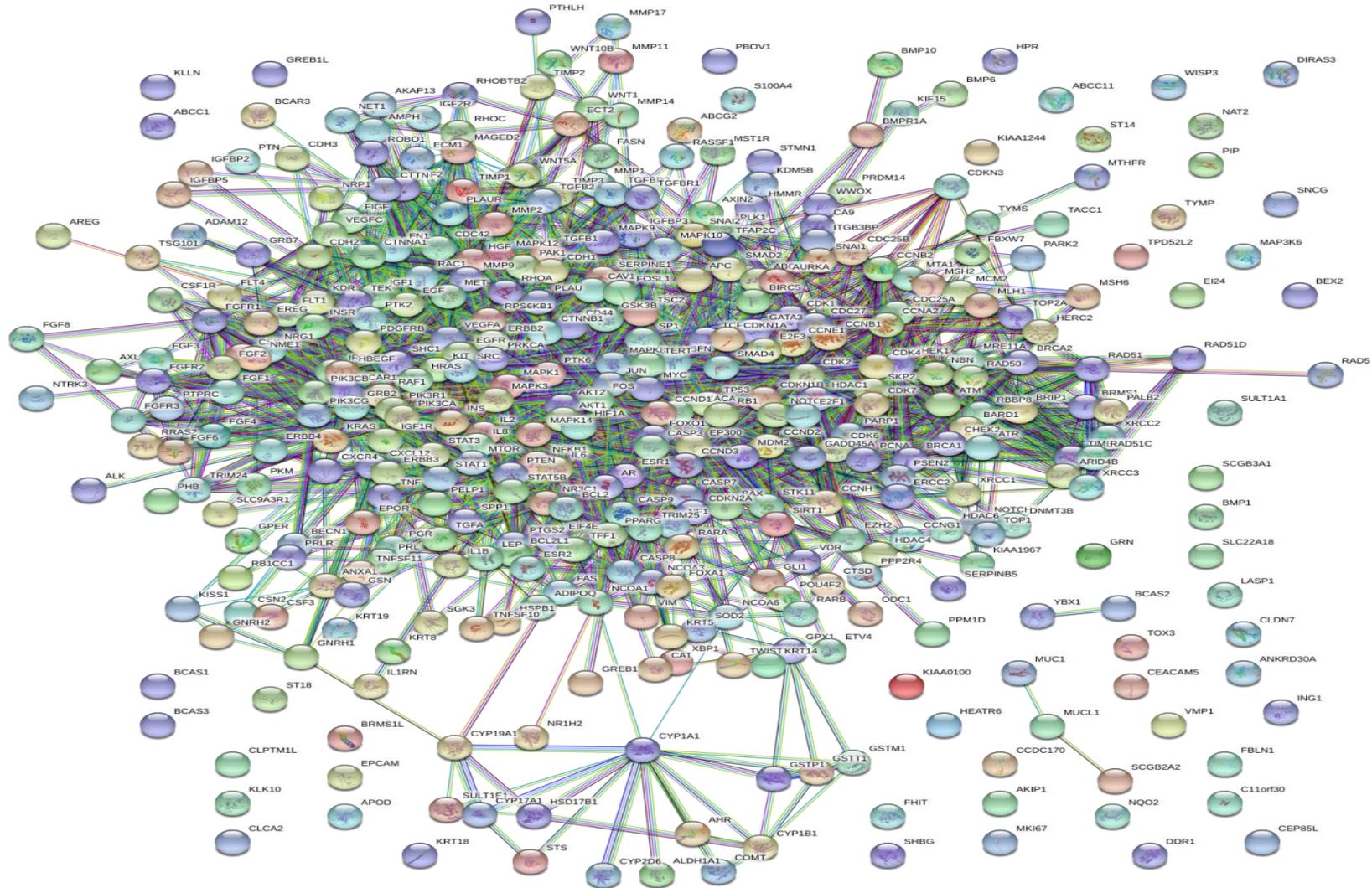
A pesquisa através da base de dados genética GeneCards® permitiu a identificação de um total de 8.610 genes para a condição estudada. Foram selecionados para o corrente estudo 400 genes com as pontuações (*scores*) mais elevadas. Destes, 396 genes foram inseridos no estudo por estarem presentes em *Homo sapiens*.

A **Figura 1** apresenta o mapa genômico da rede do câncer de mama, cujos nós da rede representam os genes enquanto as arestas representam as associações. As linhas de cores diferentes indicam os tipos de prova utilizadas nas associações, sejam elas por interações conhecidas ou previstas (SZKLARCZYK ML et al., 2017). Na rede do câncer de mama (CM) foram analisados 396 genes, sendo que 336 deles realizavam interação com outros genes da rede e 60 genes foram considerados órfãos.

A associação vermelha indica fusão de genes; verde: genes vizinhos; azul escuro: co-ocorrência de genes; roxa: homologia proteica; rosa: relação determinada experimentalmente; amarela: relação determinada por estudos de mineração de textos; azul claro: relação proveniente de um banco de dados com curadoria; preta: Co expressão gênica. Os genes órfãos não interagem com outros genes da rede.

O gene *TP53* foi o gene que realizou o maior número de interações na rede (90 interações) e com o maior número de genes líderes (*CDK2*, *EGFR*, *AKT1*, *MAPK1* e *SRC*). *PIK3CA*, o segundo gene líder do grupo A, realizou 88 interações na rede do câncer de mama, interagindo inclusive com os líderes *AKT1*, *EGFR* e *SRC*. *AKT1* realizou 64 interações na rede em estudo e interagiu com os genes líderes *PIK3CA*, *TP53*, *EP300* e *SRC*. Os genes líderes do grupo C, *EP300*, *EGFR*, *MAPK1*, *SRC* e *CDK2* realizaram 66, 62, 60, 58 e 52 interações, respectivamente, e ambos interagiram com genes líderes do grupo.

Figura 2 - Mapa de interação dos 396 genes envolvidos na rede do câncer de mama.



Fonte: Dados da pesquisa, 2018.

A análise do número ponderado de interações dentro da rede de câncer de mama (WNL) em função das interações realizadas na rede global (TIS) permitiu a distribuição dos genes por grupos e a identificação dos genes considerados líderes de cada condição. Os genes da rede em estudo foram distribuídos em 14 grupos (**Tabela 1**). Foram considerados genes líderes, os genes pertencentes aos grupos A, B e C, totalizando 8 genes. O gene *TP53* (Proteína supressora de tumor P53) apresentou maior valor de WNL na rede do câncer de mama, estando no grupo A juntamente com o gene *PIK3CA* (Subunidade catalítica de fosfatidilinositol-4,5-Bisfosfato 3-Quinase Alfa). O gene *AKT1* (Serina/treonina quinase 1) foi o único gene do grupo B, enquanto os genes líderes *EGFR* (Receptor do fator de crescimento epidérmico), *MAPK1* (Proteína quinase ativada por mitógeno 1), *EP300* (Proteína de Ligação E1A P300), *SRC* (Proto-oncogene SRC, tirosina quinase não receptora) e *CDK2* (Quinase dependente de ciclina 2) foram identificados no grupo C.

**Tabela 1** - Genes da rede do câncer de mama agrupados em função do número de interações realizadas.

Grupos	Genes
A	<i>TP53*</i> ; <i>PIK3CA*</i>
B	<i>AKT1*</i>
C	<i>SRC*</i> ; <i>MAPK1*</i> ; <i>EGFR*</i> ; <i>CDK2*</i> ; <i>PIK3R1*</i> ; <i>EP300*</i>
D	<i>EGF</i> ; <i>MYC</i> ; <i>HRAS</i>
E	<i>CTNNB1</i> ; <i>JUN</i> ; <i>CDK1</i> ; <i>PCNA</i> ; <i>VEGFA</i> ; <i>GRB2</i> ; <i>MAPK3</i>
F	<i>RHOA</i> ; <i>SHC1</i> ; <i>ESR1</i> ; <i>MAPK8</i> ; <i>CCND1</i> ; <i>PIK3CB</i>
G	<i>IGF1</i> ; <i>RAC1</i> ; <i>CDKN1A</i> ; <i>FOS</i> ; <i>CDC42</i> ; <i>MAPK14</i> ; <i>ERBB2</i> ; <i>CDK4</i> ; <i>CCNA2</i> ; <i>CCNB1</i> ; <i>RB1</i> ; <i>PIK3CG</i>
H	<i>CDKN1B</i> ; <i>ATM</i> ; <i>HDAC1</i> ; <i>HGF</i> ; <i>KRAS</i> ; <i>STAT3</i> ; <i>BRCA1</i> ; <i>CHEK1</i> ; <i>KDR</i> ; <i>E2F1</i> ; <i>SMAD4</i> ; <i>IGF1R</i> ; <i>SP1</i> ; <i>PDGFRB</i>
I	<i>ATR</i> ; <i>CDK6</i> ; <i>FGF2</i> ; <i>RPS6KB1</i> ; <i>STAT5B</i> ; <i>IRS1</i> ; <i>TGFB1</i> ; <i>CCNE1</i> ; <i>CDK7</i> ; <i>IL2</i> ; <i>MTOR</i> ; <i>NFKB1</i> ; <i>PTK2</i> ; <i>CCND3</i> ; <i>RAD50</i> ; <i>FN1</i> ; <i>CDH1</i> ; <i>INS</i> ; <i>RAD51</i> ; <i>CCND2</i> ; <i>MET</i> ; <i>KIT</i>
J	<i>BCAR1</i> ; <i>CCNH</i> ; <i>HIF1A</i> ; <i>AURKA</i> ; <i>FOXO1</i> ; <i>RAF1</i> ; <i>NBN</i> ; <i>FGF1</i> ; <i>MDM2</i> ; <i>IGF2</i> ; <i>CCNB2</i> ; <i>BCL2</i> ; <i>CASP3</i> ; <i>BARD1</i> ; <i>FLT1</i> ; <i>PRKACA</i> ; <i>PRKCA</i> ; <i>PAK1</i> ; <i>RBBP8</i> ; <i>CDC25A</i> ; <i>SERPINE1</i> ; <i>BRCA2</i> ; <i>AR</i> ; <i>TERT</i> ; <i>MMP2</i> ; <i>IL6</i> ; <i>AKT2</i> ; <i>E2F3</i> ; <i>NOTCH1</i> ; <i>GSK3B</i> ; <i>VEGFC</i> ; <i>STAT1</i> ; <i>CTNNA1</i> ; <i>FGFR1</i> ; <i>SKP2</i>
K	<i>CSK</i> ; <i>MMP9</i> ; <i>PLK1</i> ; <i>TEK</i> ; <i>RAD51C</i> ; <i>NR3C1</i> ; <i>TIMP1</i> ; <i>SFN</i> ; <i>FLT4</i> ; <i>INSR</i> ; <i>BRIP1</i> ; <i>HBEGF</i> ; <i>TGFB2</i> ; <i>NCOA1</i> ; <i>CDKN2A</i> ; <i>SMAD2</i> ; <i>PTEN</i> ; <i>NCOA3</i> ; <i>XRCC3</i> ; <i>PLAU</i> ; <i>PARP1</i> ; <i>CDC27</i> ; <i>CAV1</i> ; <i>CHEK2</i> ; <i>CASP8</i> ; <i>SIRT1</i> ; <i>PTK6</i> ; <i>XRCC2</i> ; <i>ERBB3</i> ; <i>FGF4</i> ; <i>CXCL12</i>
L	<i>TIMP3</i> ; <i>MAPK9</i> ; <i>ERBB4</i> ; <i>BCL2L1</i> ; <i>FOSL1</i> ; <i>CYP1A1</i> ; <i>NRG1</i> ; <i>BIRC5</i> ; <i>FGF6</i> ; <i>FGFR2</i> ; <i>CDH2</i> ; <i>ERCC2</i> ; <i>FGF3</i> ; <i>PALB2</i> ; <i>MCM2</i> ; <i>CDC25B</i> ; <i>BAX</i> ; <i>CASP9</i> ; <i>MAPK12</i> ; <i>EREG</i> ; <i>MLH1</i> ; <i>PELP1</i> ; <i>TOP2A</i> ; <i>CTTN</i> ; <i>CASP7</i> ; <i>TSC2</i> ; <i>IL1B</i> ; <i>GADD45A</i> ; <i>FGFR3</i> ; <i>MAPK10</i> ; <i>RARA</i> ; <i>MMP1</i> ; <i>TNF</i> ; <i>PLAUR</i> ; <i>TCF7L2</i> ; <i>TGFB1</i> ; <i>WNT5A</i> ; <i>CXCR4</i> ; <i>PPARG</i> ; <i>PRL</i> ; <i>LEP</i> ; <i>APC</i> ; <i>TIMELESS</i> ; <i>TGFBR2</i> ; <i>HERC2</i> ; <i>XRCC1</i> ; <i>ECM1</i> ; <i>MSH2</i> ; <i>PTGS2</i> ; <i>MAGED2</i> ; <i>CDKN3</i> ; <i>RHOC</i> ; <i>EZH2</i>
M	<i>CD44</i> ; <i>STK11</i> ; <i>TFF1</i> ; <i>EPOR</i> ; <i>FAS</i> ; <i>IGFBP3</i> ; <i>TRIM24</i> ; <i>BECN1</i> ; <i>ESR2</i> ; <i>EIF4E</i> ; <i>ANXA1</i> ; <i>CSF1R</i> ; <i>HDAC4</i> ; <i>CYP1B1</i> ; <i>RHOBTB2</i> ; <i>PGR</i> ; <i>NCOA6</i> ; <i>ECT2</i> ; <i>MSH6</i> ; <i>RAD51D</i> ; <i>ADIPOQ</i> ; <i>MMP14</i> ; <i>CYP19A1</i> ; <i>AXIN2</i> ; <i>GNRH1</i> ; <i>PRLR</i> ; <i>ITGB3BP</i> ; <i>TFAP2C</i> ; <i>GRB7</i> ; <i>NF1</i> ; <i>NRP1</i> ; <i>FOXA1</i> ; <i>TYMS</i> ; <i>GSTP1</i> ; <i>GSTM1</i> ; <i>GPX1</i> ; <i>SOD2</i> ; <i>TNFSF10</i> ; <i>HMMR</i> ; <i>GSTT1</i> ; <i>WNT1</i> ; <i>SNAI1</i> ; <i>NET1</i> ; <i>ROBO1</i> ; <i>TIMP2</i> ; <i>SNAI2</i> ; <i>IGF2R</i> ; <i>TRIM25</i> ; <i>GATA3</i> ; <i>HDAC6</i> ; <i>RARB</i> ; <i>FBXW7</i> ; <i>AKAP13</i> ; <i>PSEN2</i> ; <i>RRAS2</i> ; <i>BMPR1A</i> ; <i>CCNG1</i> ; <i>KISS1</i> ; <i>ABCB1</i> ; <i>GNRH2</i> ; <i>VIM</i> ; <i>AXL</i> ; <i>TGFA</i> ; <i>VDR</i> ; <i>AMPH</i> ; <i>FGF8</i> ; <i>CDH3</i> ; <i>CYP17A1</i> ; <i>HSPB1</i> ; <i>HSD17B1</i> ; <i>NME1</i>
N	<i>TNFSF11</i> ; <i>MTA1</i> ; <i>ARID4B</i> ; <i>DNMT3B</i> ; <i>CAT</i> ; <i>SPP1</i> ; <i>KRT5</i> ; <i>STMN1</i> ; <i>PHB</i> ; <i>SLC9A3R1</i> ; <i>GSN</i> ; <i>GLI1</i> ; <i>CTSD</i> ; <i>XBP1</i> ; <i>WNT10B</i> ; <i>TSG101</i> ; <i>KDM5B</i> ; <i>MMP11</i> ; <i>IGFBP2</i> ; <i>IGFBP5</i> ; <i>NOTCH4</i> ; <i>SULT1E1</i> ; <i>POU4F2</i> ; <i>MMIP17</i> ; <i>TOP1</i> ; <i>BRMS1</i> ; <i>MST1R</i> ; <i>SGK3</i> ; <i>RASSF1</i> ; <i>COMT</i> ; <i>KRT14</i> ; <i>RB1CC1</i> ; <i>ODC1</i> ; <i>CSN2</i> ; <i>PKM</i> ; <i>CA9</i> ; <i>ADAM12</i> ; <i>PTPRC</i> ; <i>MUCL1</i> ; <i>ETV4</i> ; <i>CSF3</i> ; <i>PTN</i> ; <i>WWOX</i> ; <i>STS</i> ; <i>CYP2D6</i> ; <i>PRDM14</i> ; <i>FASN</i> ; <i>AHR</i> ; <i>IL1RN</i> ; <i>BMIP6</i> ; <i>KIF15</i> ; <i>RAD54L</i> ; <i>TACC1</i> ; <i>MTHFR</i> ; <i>PPM1D</i> ; <i>BMP10</i> ; <i>SERPINB5</i> ; <i>NTRK3</i> ; <i>GREB1</i> ; <i>KRT8</i> ; <i>ALK</i> ; <i>SCGB2A2</i> ; <i>TWIST1</i> ; <i>ALDH1A1</i> ; <i>AREG</i> ; <i>PTHLH</i> ; <i>ABCG2</i> ; <i>MUC1</i> ; <i>BCAR3</i> ; <i>KRT19</i> ; <i>YBX1</i> ; <i>NR1H2</i> ; <i>BCAS2</i> ; <i>ABRAXAS1</i> ; <i>BRMS1L</i> ; <i>SNCG</i> ; <i>SLC22A18</i> ; <i>BCAR4</i> ; <i>KLLN</i> ; <i>MKI67</i> ; <i>BCAS1</i> ; <i>NQO2</i> ; <i>ANKRD30A</i> ; <i>BEX2</i> ; <i>TOX3</i> ; <i>GPER1</i> ; <i>NAT2</i> ; <i>MRE11</i> ; <i>SHBG</i> ; <i>BCAS3</i> ; <i>HEATR6</i> ; <i>FHIT</i> ; <i>CLCA2</i> ; <i>CEP85L</i> ; <i>TYMP</i> ; <i>PIP</i> ; <i>VEGFD</i> ; <i>ST14</i> ; <i>ING1</i> ; <i>LASP1</i> ; <i>KIAA0100</i> ; <i>EPCAM</i> ; <i>AKIP1</i> ; <i>CEACAM5</i> ; <i>EMSY</i> ; <i>ARFGEF3</i> ; <i>GRN</i> ; <i>CLDN7</i> ; <i>APOD</i> ; <i>DDR1</i> ; <i>KRT18</i> ; <i>ST18</i> ; <i>S100A4</i> ; <i>KLK10</i> ; <i>GREB1L</i> ; <i>SCGB3A1</i> ; <i>BMP1</i> ; <i>ABCC11</i> ; <i>FBLN1</i> ; <i>PTPA</i> ; <i>HPR</i> ; <i>PD52L2</i> ; <i>PBOV1</i> ; <i>CXCL8</i> ; <i>CCAR2</i> ; <i>MAP3K6</i> ; <i>ABCC1</i> ; <i>WISP3</i> ; <i>BRCA3</i> ; <i>MIR21</i> ; <i>BPIFA4P</i> ; <i>CLPTM1L</i> ; <i>EI24</i> ; <i>PRKN</i> ; <i>DIRAS3</i> ; <i>CCDC170</i> ; <i>SULT1A1</i>

\* Genes com asterisco são os considerados líderes

Fonte: Dados da pesquisa, 2018.

Todos os genes líderes da rede, com exceção do gene *CDK2*, realizam interações com genes biomarcadores do câncer de mama *ESR1* (receptor de estrógeno 1) e *ERBB2* (receptor de tirosina quinase 2). Ligações entre *BRCA1* e *BRCA2* foram observadas com os líderes *TP53* e *CDK2*. Catepsina D (CTSD) também foi um dos genes ligantes de *TP53*. Genes envolvidos em diversos mecanismos biológicos, como o controle da integridade do DNA, apoptose, angiogênese e pluripotência celular, podem ser observados nas interações gênicas da **tabela 2**.

**Tabela 2** - Interações realizadas pelos genes líderes na rede do câncer de mama.

Genes líderes	Número de interações	Genes que interagem com os genes líderes na rede de câncer de mama
<i>TP53</i>	90	<i>MDM2, BCL2, CDKN1A, CDK2*, CDKN2A, BAX, ATM, EGFR*, PTEN, MYC, RB1, CHEK2, PCNA, CHEK1, MAPK8, EP300, AKT1*, BRCA1, JUN, HDAC1, SIRT1, BIRC5, BCL2L1, CCNB1, HIF1A, GADD45A, MAPK9, MSH2, FOS, FAS, SP1, STK11, CDKN1B, AURKA, MLH1, MAPK14, CCNG1, MAPK1*, MMP2, IGFBP3, GSK3B, SNAI2, SFN, ATR, TSC2, SERPINE1, ABCB1, TERT, MAPK3, SERPINB5, CCND1, CAV1, MRE11A, TIMP3, CCNA2, HRAS, RAD50, MET, BARD1, SKP2, HGF, CCNE1, CASP3, AKT2, VDR, SPP1, CTSD, ERBB2, NBN, PARP1, TGFA, CASP8, VEGFA, MTOR, RBBP8, GPX1, BRCA2, ESR1, CDK1, KRAS, APC, BRIP1, SRC*, MAPK10, PTGS2, POU4F2, WWOX, CDH1, CD44, CDK4.</i>
<i>PIK3CA</i>	88	<i>AKT1*, PTEN, PIK3R1, RPS6KB1, IRS1, HRAS, RRAS2, IGF1R, EGFR*, SRC*, AKT2, ERBB2, VEGFA, CDC42, INSR, INS, PDGFRB, STAT3, EGF, IGF1, CCND1, ESR1, KRAS, PTK2, KDR, BCL2L1, GRB2, CDH1, RHOA, MYC, FLT1, BECN1, NFKB1, IL2, SHC1, RAF1, KIT, GSK3B, GSN, MET, FLT4, CDKN2A, AXL, CDH2, ERBB4, CTNNB1, FGF2, HGF, PIK3CA, FGFR1, RAC1, AR, FGFR3, PELP1, CSK, TEK, FGFR2, IL1B, HBEGF, MAPK1, SP1, CCND3, NR3C1, FGF1, BCAR1, CASP9, PLAU, VEGFC, MAPK3, GNRH1, FIGF, NRG1, PIK3CG, EREG, TRIM24, FGF4, ANXA1, ERBB3, FGF6, FOXO1, FGF3, CTNNA1, PIK3CB, KISS1, BCL2, ALK, GNRH2, PGR.</i>
<i>EP300</i>	66	<i>TP53*, HDAC1, NCOA3, SIRT1, NCOA1, FOS, CTNNB1, JUN, HIF1A, MYC, FOXO1, SMAD4, HDAC6, ESR1, SP1, AKT1*, TCF7L2, NR3C1, AR, RB1, MDM2, STAT3, CDKN1A, NFKB1, STAT1, HDAC4, VEGFA, MAPK8, SMAD2, NOTCH1, STAT5B, PCNA, PPARG, VDR, MAPK14, FOSL1, RARA, SKP2, NCOA6, EZH2, PELP1, BCL2, BAX, GATA3, GADD45A, PLK1, TERT, MMP1, CDK7, RARB, CCND2, CCNH, CA9, IL1B, LEP, IGFBP3, ADIPOQ, ODC1, CDC25A, ERCC2, TFAP2C, ABCB1, AURKA, XRCC1, KRT14, SFN.</i>
<i>AKT1</i>	64	<i>PIK3CB, PIK3CA*, PIK3CG, FOXO1, MTOR, PTEN, TSC2, INS, BCL2, ERBB3, CDKN1B, CTNNB1, MDM2, TP53*, CASP9, BCL2L1, CDKN1A, RHOA, ESR1, MYC, PIK3R1, VIM, HIF1A, NFKB1, TERT, CAV1, IL2, GRB2, PAK1, IRS1, GSK3B, RAF1, EP300*, RAC1, SHC1, CASP3, RB1, RPS6KB1, SRC*, CAT, SOD2, VEGFA, SFN, INRS, RARA, MAPK14, EGF, IGF1, CDC42, MST1R, PDGFRB, AKT2, BAX, PTK6, E2F1, AURKA, PELP1, STAT3, CCND1, NOTCH1, TNF, JUN, IL6, HSPB1.</i>
<i>EGFR</i>	62	<i>EGF, KRAS, GRB2, HRAS, TP53*, SHC1, TGFA, CTNNB1, PIK3CA*, PIK3CB, CDH1, STAT3, HBEGF, ERBB2, MTOR, CAV1, VEGFA, PIK3CG, NRG1, EREG, IL8, HIF1A, PIK3R1, INS, STAT5B, STAT1, PLAU, FGF2, IGF1, PLAUR, CDC42, MAPK1*, RHOA, ESR1, SRC, PRKCA, WNT5A, CTNNA1, CTTN, PTK6, TERT, FGF1, HGF, ERBB3, NOTCH1, IGF2R, FOS, TSG101, PDGFRB, ERBB4, TFF1, CSK, MDM2, RAC1, FN1, PTEN, AMPH, AREG, IRS1, CCND1, PTK2.</i>
<i>MAPK1</i>	60	<i>JUN, FOS, STAT3, GRB2, BCL2, MYC, CASP8, KRAS, SMAD2, HRAS, SP1, PTGS2, RHOA, IL6, TP53*, SHC1, RB1, IGF1, FGF2, IL2, MAPK3, TSC2, STAT1, NFKB1, IRS1, STAT5B, ESR1, RAF1, SRC, EGFR, PLAU, PGR, RPS6KB1, TGFB1, RAC1, HGF, FGF4, PRKCA, TNFSF10, SFN, FOXO1, PIK3CG, CDH1, PIK3CA*, CSK, PIK3CB, FGF1, MAPK8, IGF1R, CDK2*, CXCL12, MET, FGF2, FGF3, PLAUR, TERT, FN1, NRG1, ESR2, EGF.</i>
<i>SRC</i>	58	<i>BCAR1, EGF, STAT3, HRAS, PIK3CA, CTTN, CDC42, RHOA, CAV1, ESR1, CDH1, VEGFA, PIK3R1, AR, CTNNB1, PIK3CB, PIK3CG, PGR, IGF1, PELP1, HGF, MMP9, GRB2, PTK2, NFKB1, PLAUR, MAPK1*, AKT1*, FAZ, SHC1, EGFR, MMP2, ERBB2, PRKCA, HBEGF, PDFFRB, KDR, IGF1R, MET, MAPK8, GRB7, CXCL12, CSK, PLAU, FN1, MAPK14, RAC1, CDK1, PAK1, ESR2, CTNNA1, NRG1, EREG, PTN, FLT4, FLT1, TP53*, STAT5B.</i>
<i>CDK2</i>	52	<i>CCNB1, CDKN1B, CCND3, CCNE1, CCNB2, CCNA2, CCND1, PCNA, CDKN1A, CDC25A, CCND2, RB1, TP53*, MCM2, SKP2, CDC25B, ATR, CKN3, CDK7, CCNH, CDC27, E2F1, ATM, CDK6, MDM2, BRCA2, FOXO1, E2F3, RAD51, CDK4, BRCA1, MRE11A, CHEK1, RBBP8, BARD1, RAD50, RHOA, CDK2*, HDAC1, MAPK3, MAPK1*, CCNG1, NBN, SMAD4, MLH1, PTK6, ERCC2, SFN, CDKN2A, PRKACA, HERC2, MYC.</i>

Fonte: Dados da pesquisa, 2018. \* Genes com asterisco são os considerados líderes.

## DISCUSSÃO

O advento da bioinformática possibilitou à biologia do câncer identificar e validar marcadores biológicos, fornecendo importantes candidatos ao monitoramento da progressão de doenças e respostas a terapias (SANTOS E, OTACILIO SH, 2016; CUI et al., 2018). Diante disso, a proposta do nosso estudo foi investigar os genes que realizam os maiores números de interações na condição câncer de mama e descrever as interações gênicas realizadas por eles. A partir dos resultados obtidos, foram evidenciados 8 genes líderes (*TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300*, *SRC* e *CDK2*) entre os 396 genes analisados. Os genes líderes do câncer de mama, com exceção do gene *MAPK1*, apresentaram elevadas interações na rede em estudo e baixas interações nas redes globais. Além de uma elevada interação entre os líderes, uma importante interação entre eles e genes comumente utilizados como biomarcadores no câncer de mama foi observada. *TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300* e *SRC* interagem com o receptor de estrógeno *ESR1* e o receptor de tirosina quinase 2 (*ERBB2*). Ligações entre *BRCA1* e *BRCA2* foram também observadas com o gene líder *CDK2*. Catepsina D (*CTSD*), *BRCA1* e *BRCA2* também foram genes marcadores do câncer de mama que interagiram com *TP53*.

O gene *TP53* se destacou como o gene líder que mais realizou interações na rede do câncer de mama, seguido pelo *PIK3CA*. Conhecido como um guardião do genoma, o gene *TP53* codifica uma proteína chamada p53, que faz parte de uma família de proteínas com atividade supressora de tumor (MAXIMOV GK e MAXIMOV KG, 2008). Em condições normais, essa proteína induz o bloqueio do ciclo celular, senescência e apoptose em resposta aos eventos de estresse celular, como hipóxia e danos no DNA. Em câncer oral, *TP53* foi também apontado como gene líder em abordagem computacional e é comprovadamente associado ao prognóstico da doença em análises experimentais (SANTOS E e OTACILIO SH, 2016). *PIK3CA* é fundamental na via de sinalização fosfatidilinositol 3-quinase (PI3K), com implicações em diversos eventos celulares, tais como proliferação, diferenciação e migração. Em câncer de mama, cerca de 27% das mutações são detectadas em *PIK3CA* (MANGONE FR, et al., 2012). Evidências sugerem que mutações em *TP53* e *PIK3CA* podem ser considerados fatores chaves para a progressão desse tipo de câncer. *PIK3CA* e *TP53* são os genes mais comumente mutados em cânceres de mama positivos para o receptor de estrógeno e negativos para esse receptor, respectivamente (KIM MJ, et al., 2017). Estudos relatam associação entre a ocorrência de mutações do éxon 20 do gene *PIK3CA* com a presença de mutações do gene *TP53* (MANGONE FR et al., 2012). Em função disso, inibidores de *PI3K* já estão sendo utilizados no tratamento de pacientes com câncer de mama apresentando mutações em *PI3KCA* na triagem clínica (KIM MJ et al., 2017). A integridade funcional de *TP53* e *PI3KCA* são essenciais para o adequado funcionamento celular. Nossos estudos fornecem suporte adicional para pesquisas que buscam por alvos que inativem esses genes mutados e avaliem suas implicações moleculares e consequências para o prognóstico do câncer de mama.

A proteína Serina/Treonina Quinase 1 (*AKT1*) juntamente com PI3K participam de eventos celulares como proliferação, sobrevivência, crescimento tumoral, angiogênese e são necessários para a manutenção das características de células-tronco em uma população de células de câncer de mama (GARGINI R, et al., 2015). As células-tronco correspondem a uma pequena população de células tumorais responsáveis por desencadear o desenvolvimento do câncer devido à capacidade de auto-renovação e resistência à apoptose. Elevada incidência de mutações no gene *AKT1* relaciona-se à redução da sobrevida dessas células (GARGINI R et al. 2015). De maneira similar, *EGFR*, um dos genes líderes evidenciados no presente estudo, contribui com a expressão de marcadores de auto-renovação, ativação da proliferação, invasão e resistência aos tratamentos quimioterápicos (ABHOLD EL, et al., 2012; HAN W, et al., 2013). *EGFR* encontra-se superexpresso em 15-20% de todos os carcinomas mamários e em 50-70% dos cânceres de mama triplo negativo. Sua alta expressão associa-se a um maior volume tumoral, ocorrência de metástases e pior sobrevida (KOZLOVA N, et al., 2017). Análise de bioinformática realizada para o adenocarcinoma de esôfago revelou *TP53*, *EGFR* e *AKT1* como os três maiores candidatos a biomarcadores da doença (REZAEI-TAVIRANI M, et al., 2017). Percebe-se que diversas condições malignas compartilham genes comuns ao surgimento e evolução da doença, apesar de não serem utilizados como genes marcadores.

A proteína MAPK1, proteína quinase 1 ativada por mitógenos, também conhecida como ERKs (quinases reguladas por sinal extracelular), quando ativada por outras quinases, é transferida para o núcleo e fosforila alvos nucleares, participando assim em processos de regulação da transcrição, proliferação e diferenciação. MAPK1 é superativada em vários casos de câncer de mama, e parece interagir com receptores estrógeno, entre eles os receptores de andrógenos, os receptores de hormônio esteroide mais amplamente expressos em câncer de mama (AZZAM DG, et al., 2008). Em nosso estudo, *MAPK1* foi considerado um dos genes líderes do câncer de mama por interagir com um grande número de genes, entre eles o receptor de estrógeno *ESR1* e o receptor de tirosina quinase 2 (*ERBB2*). *MAPK1* também está envolvido na regulação do desenvolvimento do tumor de mama e quimioresistência (SI W, et al., 2018). Estudo realizado com 3592 indivíduos portadores de câncer de mama e 4183 indivíduos controle revelou que expressão gênica aumentada de *MAPK1*, em interação com dieta rica em agentes oxidantes, fibra e folato aumentam o risco de câncer de mama (SLATTERY ML, et al., 2015).

Um outro gene líder evidenciado no nosso estudo foi o EP300 (proteína P300 de ligação a E1A), o qual codifica uma acetiltransferase de histona, responsável por promover a acetilação de proteínas histonas e não-histonas no processo de ativação da transcrição gênica. Em função disso, essa proteína coordenada múltiplos eventos transcricionais, como proliferação, diferenciação, regulação do ciclo celular, resposta ao dano no DNA e apoptose. Sua alta expressão é observada em cânceres de mama e tem sido associada ao pior prognóstico (XIAO XS, et al., 2011; KOWALCZYK AE, et al., 2017). O fator transcricional p300 também promove o aumento da expressão de genes relacionados à migração em células de câncer de mama (HE H, et al., 2015). A ocorrência de uma mutação específica nesse gene tem sido apontada como um marcador para a tomada de decisões clínicas em câncer de mama triplo negativo (BEMANIAN V, et al., 2018).

Os genes líderes *SRC* e *CDK2* são essenciais nos mecanismos de regulação do ciclo celular e consequentemente na progressão de tumores, como o câncer de mama (VENTURA C, et al., 2017). *Src* e os membros dessa família compõem o maior grupo de tirosina quinases não receptoras. Têm apresentado papéis importantes para o crescimento celular, proliferação, angiogênese, invasão e disseminação metastática. Apresentam também função na ativação de osteoclastos e reabsorção óssea, características essas que são elevadas em metástases ósseas (FINN RS, 2008). *SRC* tem sido também crucial na resposta ao estrógeno. Esse hormônio é um dos mais importantes estímulos para induzir a proliferação de células mamárias, capaz de ativar a saída das células da fase de intervalo G1 para a fase de síntese (S) no ciclo celular. A ação do estrógeno é mediada por interação física entre *Src* e o receptor de estrógeno  $\alpha$ . *Src* fosforila p27 e impede que p27 exerça efeito inibitório sobre o complexo de progressão no ciclo celular ciclina E/CDK2. Dessa forma, a ativação de CDK2 em função do aumento de *SRC* promove o aumento da proliferação das células tumorais da mama (VENTURA C, et al., 2017).

A análise das interações realizadas pelos genes líderes na rede do câncer de mama revelou os genes que interagem com os genes líderes. *TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300* e *SRC* interagem com o receptor de estrógeno *ESR1* e o receptor de tirosina quinase 2 (*ERBB2*). 70% dos tumores de mama expressam receptores de estrógeno, que fazem deles usualmente sensíveis ao bloqueio hormonal (CLATOT F e AUGUSTO L, 2017). *ERBB2*, comumente referido com *HER2*, encontra-se superexpresso em 20-40% dos carcinomas de mama invasivos. Os cânceres de mama positivos para *ERBB2* são tratados como uma doença com comportamento mais agressivo (TOMASICH FDS, et al., 2018).

Catepsina D (*CTSD*), *BRCA1* e *BRCA2* também foram genes marcadores de câncer de mama que interagiram com *TP53*. A expressão da catepsina D tem sido avaliada em cânceres de mama visto que níveis elevados dessa proteína se associam ao pior prognóstico (PUJOL P, et al., 2012). Ligações entre *BRCA1* e *BRCA2* foram também observadas com o gene líder *CDK2*. *BRCA1* e *BRCA2* fazem parte de um complexo de proteínas envolvidas na manutenção da integridade do genoma, especialmente no mecanismo de reparo de DNA. Relação entre essas proteínas com a quinase *CDK2* faz parte dos eventos de controle da integridade e proliferação celular. A proliferação celular ativada por *CDK2* em tumores de mama tem sido relatada como provocada pelo estrógeno, o qual age como fator de transcrição desse gene (FOSTER JS, et al., 2018).

Por tudo isso, pode-se perceber que a partir dos resultados obtidos, muitas interpretações e interações gênicas podem ser exploradas e servirem de base para análises posteriores. Um dos principais desafios da biologia é desvendar as numerosas proteínas envolvidas nas condições biológicas e extrair informações úteis dos grandes conjuntos de dados gerados. No presente estudo, a análise do mapa de interações gênicas por bioinformática fornece previsões testáveis e novas informações sobre a rede do câncer de mama. Embora os genes líderes *TP53* e *PI3KCA* apresentem uma grande importância nos mecanismos moleculares da carcinogênese da mama, eles não podem ser inativados em função de suas diversas funções benéficas. Os demais genes líderes evidenciados no estudo, *AKT1*, *EGFR*, *MAPK1*, *EP300*, *SRC* e *CDK2*, também se destacam pelo elevado número de interações na rede do câncer de mama, cuja importância para a patogênese da doença é reforçada pelos estudos que demonstram a relação entre suas elevadas expressões gênicas com o pior prognóstico da doença. A inativação molecular ou medicamentosa desses genes mutados ou superexpressos pode ser uma ferramenta importante para impedir a progressão da malignidade. Apesar da ausência de estudos genômicos computacionais que apontem os genes líderes do câncer de mama e seu mapa de interações, diversas outras condições malignas demonstram a importância da biologia de sistemas computacionais na compreensão das interações entre componentes biológicos e na complementação dos estudos experimentais (SANTOS E, OTACILIO SH, 2016; REZAEI-TAVIRANI M, et al., 2017).

## CONCLUSÃO

As ferramentas de bioinformática possibilitaram identificar as redes de interações gênicas do câncer de mama, bem como a geração de hipóteses que possam ser avaliadas em futuros estudos moleculares. Particularmente, nossas análises sugerem muitos genes importantes para a patogênese do câncer de mama, assim como genes candidatos à inativação quando da sua super expressão, como *AKT1* e *EGFR*. Em suma, os genes *TP53*, *PIK3CA*, *AKT1*, *EGFR*, *MAPK1*, *EP300*, *SRC* e *CDK2* foram os genes líderes da rede do câncer de mama. Esses fizeram um total de 540 interações, sendo que o gene líder *TP53* foi o gene que fez mais interações na rede, 90 interações, e interagiu com um número maior de genes líderes da rede do câncer de mama, *CDK2*, *EGFR*, *AKT1*, *MAPK1* e *SRC*. O gene líder *PIK3CA* realizou 88 interações na rede do câncer de mama, interagindo com *AKT1*, *EGFR* e *SRC*. O gene *AKT1* realizou 64 interações na rede em estudo e interagiu com os genes líderes *PIK3CA*, *TP53*, *EP300* e *SRC*. O gene *EGFR* realizou 62 interações e fez interações com os líderes da rede *TP53*, *PIK3CA* e *MAPK1*. O gene *MAPK1* realizou 60 interações na rede do câncer de mama e interagiu com os seguintes genes líderes *TP53*, *PIK3CA* e *CDK2*. *EP300* realizou 66 interações e interagiu com apenas dois genes líderes em estudo, *TP53* e *AKT1*. O gene *SRC* realizou 58 interações na rede do câncer de mama, interagindo com os líderes *MAPK1*, *AKT1* e *TP53*. O gene que menos realizou interações foi o *CDK2*, 52 interações, e interagiu com os líderes *TP53*, *CDK2* e *MAPK1*. Todos os líderes apontados realizaram interações com genes envolvidos em diferentes vias moleculares, inclusive com marcadores de diagnóstico e prognóstico do câncer de mama. Em conclusão, a identificação de genes cruciais e suas interações não somente contribuem para elucidar a patogênese do câncer de mama, mas também fornece potenciais marcadores prognósticos e alvos terapêuticos.

## REFERÊNCIAS

1. ABHOLD EL, et al. EGFR kinase promotes acquisition of stem cell-like properties: a potential therapeutic target in head and neck squamous cell carcinoma stem cells. *PLoS One*, 2012; 7(2).
2. AMOROSO V, et al. International expert consensus on primary systemic therapy in the management of early breast cancer: highlights of the fifth symposium on primary systemic therapy in the management of operable breast cancer, Cremona, Italy (2013). *JNCI Monographs*, 2015; 51:90–6. Disponível em: doi: 10.1093 / jncimonographs / lgv023.
3. AZZAM DG, et al. ERK/MAPK regulation of the androgen responsiveness of breast cancer cells. *Adv Exp Med Biol*, 2008; 617:429-35.
4. BEMANIAN V, et al. Somatic EP300-G211S mutations are associated with overall somatic mutational patterns and breast cancer specific survival in triple-negative breast cancer. *Breast Cancer Res Treat*, 2018; 172(2):339-351.
5. CLATOT F, AUGUSTO L. ESR1 mutations in breast cancer. *Aging (Albany NY)*, 2017; 9(1): 3–4. Disponível em: doi: 10.18632/aging.101165
6. CUI C, LI L, ZHEN J. Bioinformatic analysis reveals the key pathways and genes in early-onset breast cancer. *Med Oncol*, 2018; 11;35(5):67.
7. DUFFY MJ, EVOY D, MCDERMOTT EW. CA 15-3: uses and limitation as a biomarker for breast cancer. *Clin Chim Acta*, 2010; 14;411(23-24):1869-74.

8. FINN RS . Targeting Src in breast cancer. *Ann Oncol*, 2008; 19(8):1379-86.
9. FOSTER JS, et al. Multifaceted regulation of cell cycle progression by estrogen: regulation of CDK inhibitors and cdc25A independent of cyclin D1-cdk4 function. *Mol Cell Biol*, 2001; 21:794–810.
10. GARGINI R, et al. Cancer stem cell-like phenotype and survival are coordinately regulated by Akt/FoxO/Bim pathway. *Stem Cells*, 2015; 33: 646–660.
11. HAN W, et al. STAT1 gene expression is enhanced by nuclear EGFR and HER2 via cooperation with STAT3. *Carcinogenesis Molecular*, 2012; 52 (12): 959-969.
12. HE H, et al. Transcriptional factors p300 and MRTF-A synergistically enhance the expression of migration-related genes in MCF-7 breast cancer cells. *Biochem Biophys Res Commun*, 2015; 467(4):813-20.
13. ILIE SM, et al. Potential clinically useful prognostic biomarkers in triple-negative breast cancer: preliminary results of a retrospective analysis. *Breast Cancer - Targets and Therapy*, 2018; 10.
14. INSTITUTO NACIONAL DE CÂNCER-INCA. Estimativa 2016: Incidência de Câncer no Brasil. Rio de Janeiro, 2016.
15. JAEGER S, et al. Quantification of pathway crosstalk reveals novel synergistic drug combinations for breast cancer. *Cancer Research*, 2017;77(2): 459-469.
16. KIM MJ, et al. Correlation of ALDH1 and Notch3 Expression: Clinical implication in Ovarian Carcinomas. *Journal of Cancer*, 2017; 8 (16): 3331–3342.
17. KOWALCZYK AE, et al. Expression of the EP300, TP53 and BAX genes in colorectal cancer: Correlations with clinicopathological parameters and survival. *Oncol Rep*, 2017; 38(1):201-210.
18. KOZLOVA N, et al. Hypoxia-inducible factor prolyl hydroxylase 2 (PHD2) is a direct regulator of epidermal growth factor receptor (EGFR) signaling in breast cancer. *Oncotarget*, 2017; 8 (6): 9885–9898.
19. MANGONE FR, et al. PIK3CA Exon 20 Mutations Are Associated with Poor Prognosis in Breast Cancer Patients. *Clinics*, 2012; 67(11): 1285–1290.
20. MAXIMOV GK, MAXIMOV KG. "The Role of p53 Tumor-Suppressor Protein in Apoptosis and Cancerogenesis." *Biotechnology & Biotechnological Equipment*, 2008; 22(2): 664-668.
21. ORDING AG, et al. Site of metastasis and breast cancer mortality: a Danish nationwide registry-based cohort study. *Clin Exp Metastasis*, 2017; 34:93-101.
22. PEREIRA DVC. Identificação e análise da rede de genes envolvidos no câncer de pele de melanoma através da mineração de texto. 2016. 128f. Dissertação (Mestrado em Tecnologia da Informação Aplicada à Biologia Computacional) - Faculdades Infórum de Tecnologia. Belo Horizonte, 2016.
23. PERRI F, et al. "P53 mutations and cancer: a tight linkage." *Ann Transl Med*, 2016; 4(24): 522.
24. PIZON M, et al. Influence of endocrine therapy on the ratio of androgen receptor (AR) to estrogen receptor (ER) positive circulating epithelial tumor cells (CETCs) in breast cancer. *J Transl Med*, 2018; 14;16(1):356.
25. POSWAR, Fabiano de Oliveira, et al. Bioinformatics, Interaction Network Analysis, and Neural Networks to Characterize Gene Expression of Radicular Cystand Periapical Granuloma. *Journal of Endodontics*, 2015; 41 (6): 877–883.
26. PUJOL P, et al. A prospective study of the prognostic value of cathepsin D levels in breast cancer cytosol. *Cancer*. 1993; 71(6):2006-2012.
27. REZAEI-TAVIRANI M, et al. Protein-Protein Interaction Network Analysis for a Biomarker Panel Related to Human Esophageal Adenocarcinoma. *Asian Pac J Cancer Prev*, 2017; 18(12): 3357–3363. Disponível em: doi: 10.22034/APJCP.2017.18.12.3357
28. REZAEI-TAVIRANI M, et al. Protein-Protein Interaction Network Analysis for a Biomarker Panel Related to Human Esophageal Adenocarcinoma. *Asian Pac J Cancer Prev*, 2017; 18(12): 3357–3363. Disponível em: doi: 10.22034/APJCP.2017.18.12.3357
29. RUITERKAMP J, et al. Improved survival of patients with primary distant metastatic breast cancer in the period of 1995–2008. A nationwide population-based study in the Netherlands. *Breast Cancer Res Treat*, 2011; 128: 495-503.
30. SANTOS E, OTACILIO SH. Obesity-Related Genes and Oral Cancer: A Bioinformatics Approach and Systematic Review. *Journal of Applied Bioinformatics & Computational Biology*, 2016; 05 (10):4172-2329.
31. SI W, et al. A miR-20a/MAPK1/c-Myc regulatory feedback loop regulates breast carcinogenesis and chemoresistance. *Cell Death Differ*. 2018 Feb;25(2):406-420.
32. SLATTERY ML, et al. MAPK genes interact with diet and lifestyle factors to alter risk of breast cancer: The Breast Cancer Health Disparities Study. *Nutr Cancer*. 2015;67(2):292-304.
33. SOUDEH GF, et al. Meta-Analysis of BRCA1 Polymorphisms and Breast Cancer Susceptibility. *Klin Onkol*, 2018; 31(5):330-338. Disponível em: doi: 10.14735/amko2018330.
34. SZKLARCZYK K. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible *Nucleic Acids Research*, 2017; 45: 362 – 368.
35. TOMASICH FDS, et al. Marcadores tumorais CEA e CA 72-4 na avaliação do câncer gástrico. *Rev Acta Oncol Brasil*, 2001;21(1):211-15.
36. UIMARI O, et al. Genome-wide genetic analyses highlight mitogen-activated protein kinase (MAPK) signaling in the pathogenesis of endometriosis. *Human Reproduction*, 2017; 32(4): 780–793. Disponível em: <https://doi.org/10.1093/humrep/dex024>
37. VENTURA C, et al. Hexachlorobenzene alters cell cycle by regulating p27-cyclin E-CDK2 and c-Src-p27 protein complexes, 2017; 15(270) :72-79. Disponível em: doi: 10.1016/j.toxlet.2017.02.013. Epub 2017 Feb 12.
38. VERLI H. *Bioinformática da Biologia à Flexibilidade Molecular*. 1. ed. São Paulo: SBBq, 282p, 2014.
39. WEIGEL MT, DOWSETT M. Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocr Relat Cancer*, 2010; 17(4): 245 – 262.
40. XIAO XS, et al. High expression of p300 in human breast cancer correlates with tumor recurrence and predicts adverse prognosis. *Chin J Cancer Res*, 2011; 23(3):201-7.