**FACULDADE DE MEDICINA
UNIVERSIDADE DE SÃO PAULO**

ANDRES GALINDO CÉSPEDES

Utilização de mineração de dados para avaliação do perfil de expressão e do potencial valor prognóstico de genes identificados em tumores de mama triplo negativos com expressão diferencial de SPARC (*Secreted Protein Acidic and Rich in Cysteine*).

SÃO PAULO

2020

ANDRÉS GALINDO CESPEDES

Utilização de mineração de dados para avaliação do perfil de expressão e do potencial valor prognóstico de genes identificados em tumores de mama triplo negativo com expressão diferencial de SPARC (*Secreted Protein Acidic and Rich in Cysteine*).

Dissertação apresentada a Faculdade de Medicina da para obtenção do título de Mestre em Ciências. Área de concentração: Oncologia

SÃO PAULO

2020

Andrés Galindo Céspedes

Utilização de mineração de dados para avaliação do perfil de expressão e do potencial valor prognóstico de genes identificados em tumores de mama triplo negativo com expressão diferencial de SPARC (*Secreted Acidic and Rich in Cysteine*)

Orientadora:

Profa. Dra. **MARIA APARECIDA NAGAI**

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Oncologia

**São Paulo
2020**

FICHA CATALORGRÁFICA

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Galindo-Cespedes, Andrés

Utilização de mineração de dados para avaliação do perfil de expressão e do potencial valor prognóstico de genes identificados em tumores de mama triplo negativos com expressão diferencial de SPARC (Secreted Protein Acidic and Rich in Cysteine) /

Andrés Galindo-Cespedes. -- São Paulo, 2020.

Dissertação (mestrado) -- Faculdade de Medicina da Universidade de São Paulo.

Programa de Oncologia.

Orientadora: Maria Aparecida Nagai.

Descritores: 1.Neoplasias da mama 2.Biologia computacional 3.Neoplasias de mama triplo negativas 4.Biomarcadores 5.Mineração de dados 6.Osteonectina

USP/FM/DBD-095/20

Responsável: Erinalva da Conceição Batista, CRB-8 6755

DEDICATÓRIA

À os meus amados pais

AGRADECIMENTOS

- À os meus pais
- À minha avó, madrinha, irma, tios e tias
- À Dra. Maria Aparecida Nagai
- À Flavia
- À o Renan
- À o grupo de pesquisa do laboratorio de Genetica Molecular
- À o profesor Alexandre, Leonardo, Alan
- À Roxana, Camila e Alexis
- À Analí
- À Izabella
- À os meus colegas de laboratório : Mercia, Daniel
- À Daniel, Leandro, Wilfredo
- À Ludmila, Bruna, Elizabeth
- À Mathew, Otavio, Igor, Jean
- À “The Brotherhood”
- À Dani
- À os amigos do Perú em São Paulo
- À o pessoal da bioinformatica
- À Dona Cristiane, Dona Raquel
- À o CTO, o São Paulo e o Brasil inteiro
- À CAPES

RESUMO

Galindo-Céspedes A. *Utilização de mineração de dados para avaliar perfis de expressão e potencial valor prognóstico de genes identificados em tumores de mama triplo negativo com expressão diferencial do SPARC (Secreted Protein Acidic Rich in Cysteine)* [dissertação]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2020.

Introdução e objetivo: De acordo com a classificação molecular do câncer de mama (BC), os tumores TNBC (câncer de mama triplo negativo) prevalecem como o subtipo mais agressivo, representando 10 a 15% de todos os tumores de mama diagnosticados. Além disso, o fato de os pacientes portadores de tumores de TNBC apresentarem respostas desfavoráveis em relação ao comportamento clínico e taxas de recorrência mais altas motivou um número crescente de pesquisas a direcionar a atenção da descoberta de biomarcadores para o contexto do TNBC. Infelizmente, muitos biomarcadores não conseguiram manejar a previsão da doença. Portanto, há um interesse crescente em explorar as capacidades de novas moléculas dentro das populações de TBNC. O presente trabalho tem como objetivo utilizar ferramentas in silico e bancos de dados públicos de câncer disponíveis para avaliar o papel biológico de um conjunto de genes previamente identificado por nosso grupo no subtipo TNBC.

Desenho experimental: Uma abordagem em três etapas foi desenvolvida para selecionar novos e potenciais biomarcadores de uma lista de 100 genes diferencialmente expressos, obtidos de um estudo anterior realizado por nosso grupo na comparação do TNBC com expressão diferencial de SPARC. Primeiro, todos os genes foram anotados manualmente usando os bancos de dados NCBI e a literatura científica. Em seguida, os genes restantes foram avaliados quanto à expressão do tumor versus expressão normal e intrínseca do subtipo. Depois disso, um novo conjunto de genes foi selecionado para avaliar melhor o perfil da expressão e seu significado clínico no TNBC usando dados de três bancos de dados diferentes. Finalmente, o conjunto de genes selecionado foi avaliado de acordo com padrões de expressão, valor prognóstico, alterações genéticas e interações de redes moleculares com ênfase no TNBC. Todos os dados foram obtidos de bancos de dados públicos de câncer e as análises foram realizadas com o R Statistical Package (www.r-project.org/). **Resultados e discussão:** Cinco de 100 genes foram previstos de acordo com seu desfecho clínico para pior prognóstico, quando regulados positivamente (MESP1, ZNF556, KLHL13), bem como com negativamente (TSPAN8,

RIMS2) em pacientes portadores de tumores do tipo basal. Características como desdiferenciação, progressão tumoral e metástase, troncicidade, desregulação no ciclo celular e na matriz extracelular foram encontradas em todos os genes, refletindo assim o estado dos tumores agressivos da mama. Usando esse subconjunto de genes, observamos diferenças significativas na sobrevida em pacientes com câncer de mama nas análises de Kaplan-Meier. **Conclusão:** Os cinco genes candidatos identificados podem ajudar a revelar os mecanismos moleculares subjacentes ao câncer de mama e fornecer orientações sobre a possível seleção personalizada de esquemas terapêuticos. A abordagem usada neste estudo fornece novos insights para entender melhor o TNBC e outros subtipos de câncer.

Descritores: Neoplasias da mama; Biologia computacional; Neoplasias de mama triplo negativas; Biomarcadores; Mineração de dados; Osteonectina.

ABSTRACT

Galindo-Cespedes A. *The Use Of Data Mining To Assess Expression Patterns And Potential Prognostic Value Of Genes Identified In Triple Negative Breast Tumors With Differential Expression Of Sparc (Secreted Protein Acidic And Rich In Cysteine)* [dissertation]. São Paulo: “Faculdade De Medicina, Universidade De São Paulo”; 2020.

Introduction and purpose: according to breast cancer (BC) molecular classification, TNBC (triple negative breast cancer) tumors have prevailed as the most aggressive subtype, accounting for 10–15% of all diagnosed breast tumors. else, the fact that patients carrying tnbc tumors bear disadvantaged responses regarding clinical behavior and alta recurrence rates has motivated a rising number of researches to drive biomarker discovery attention into tnbc context. Unfortunately, many biomarkers have failed to manage disease prediction. Therefore, there is a growing interest to explore the capacities of novel oncogenic molecules inside new tbnc populations. The present work aims to use in silico tools and public available cancer databases to evaluate biological role of a gene set previously identified by our group. **Experimental design:** a three-stage approach was developed to select new genes from a list of 100 top differentially expressed genes obtained from a previous study performed by our group. First, all genes were manually annotated by using ncbi-databases and scientific literature. Then, the remaining genes were assessed for tumor versus normal and intrinsic subtype expression. After that, a new gene set was selected to further assess the expression profiling and its clinical significance in triple negative breast cancer using data from three different databases. Finally, the selected gene-set was evaluated according to expression patterns, prognostic value, genetic alterations and molecular network interactions with emphasis on TNBC. All data were obtained from public cancer databases and analyses were performed with r statistical package (www.r-project.org/). **Results and discussion:** five out of 100 genes were predicted according to their significant outcome for worse prognosis as upregulated (*mesp1*, *znf556*, *klhl13*) as well as downregulado (*tspan8*, *rims2*) in patients carrying basal-like tumors. Characteristics such as de-differentiation, tumor progression and metastasis, stemness, dysregulation in the cell cycle and in the extracellular matrix were found in all genes, which reflect the state of aggressive breast tumors. By using this subset of genes, we observed significant differences in survival among breast cancer patients in kaplan-meier analyses. **Conclusion:** the five candidate genes identified can help reveal the underlying

molecular mechanisms of BC and provide guidance on possible personalized selection of therapeutic regimens. It does not escape our knowledge that the approach of this study may give a new insight to better understand TNBC and other types of cancer.

Descriptors: Breast neoplasms; Computational biology; Triple negative breast neoplasm; Biomarkers; Data mining; Osteonectin.

LISTA DE FIGURAS

Figura 1. Representação proporcional da sobreposição entre TNBC e Basal-like, e Classificação molecular do subtipo TNBC em acordo com expressão genica e possível relação com vias moleculares	11
Figura 2. Representação gráfica da estrutura e função de SPARC e sua participação em processos envolvendo remodelação da matriz extracelular (MEC).....	12
Figura 3. Fluxograma de trabalho visando obter candidatos potenciais a biomarcadores em câncer de mama com ênfase no subtipo TNBC a partir de um conjunto de 100 genes	21
Figura 4. Representação gráfica de dos subconjuntos de genes obtidos a partir de um estudo do nosso grupo	22
Figura 5. Perfil de Alteração genicas para os 05 genes selecionados, utilizando dados de dois estudos em câncer de mama.	23
Figura 6. Rede de interação genica a partir de 22 genes da comparação transcriptômica TNBC/SPARC+ versus TNBC/SPARC-.	24
Figura 7. Perfil de expressão do transcrito MESP1 em tumores mamários. A.- amostras tumorais de câncer de mama e de tecidos normais.....	27
Figura 8. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de MESP1 em três datasets diferentes.	28
Figura 9. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao MESP1.....	29
Figura 10. Alterações genéticas e perfil de metilação do MESP1	30
Figura 11. Representação da expressão de MESP1 associada à resposta a quimioterapia.....	31
Figura 12. Perfil de expressão do transcrito ZNF556 em tumores mamários.	33
Figura 13. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de ZNF556 em três datasets diferentes..	33
Figura 14. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao ZNF556.	35
Figura 15. Alterações genéticas e perfil de metilação do ZNF556.....	36
Figura 16. Representação da expressão de ZNF556 associada à resposta a quimioterapia.....	37
Figura 17. Perfil de expressão do transcrito TSPAN8 em tumores mamários..	39
Figura 18. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao TSPAN8.	40
Figura 19. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de TSPAN8 em três datasets diferentes.	40
Figura 20. Representação da expressão de TSPAN8 associada à resposta a quimioterapia.....	42
Figura 21. Alterações genéticas e perfil de metilação do TSPAN8.	42

Figura 22. Perfil de expressão do transcrito RIMS2 em tumores mamários.	44
Figura 23. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de RIMS2 em três datasets diferentes.	45
Figura 24. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo o RIMS2.....	46
Figura 25. Alterações genéticas e perfil de metilação do RIMS2.....	48
Figura 26. Representação da expressão de RIMS2 associada à resposta a quimioterapia.	49
Figura 27. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de KLHL13 em três datasets diferentes.	51
Figura 28. Perfil de expressão do transcrito KLHL13 em tumores mamários	52
Figura 29. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao KLHL13	53
Figura 30. Perfil de metilação do KLHL13 em duas bases de dados.....	54
Figura 31. Representação da expressão de KLHL13 associada à resposta a quimioterapia.....	55

LISTA DE TABELAS

Tabela 1. Análise de Sobrevida para 20 genes selecionados a partir de tres bases de dados visando avaliar um prognóstico desfavorável no subtipo basal.....	25
---	----

LISTA DE FIGURAS SUPLEMENTARES

Figura suplementar 1.....	84
Figura Suplementar 2.....	85
Figura Suplementar 3.....	86
Figura Suplementar 4.....	87
Figura Suplementar 5.....	88
Figura Suplementar 6.....	89
Figura Suplementar 7.....	90
Figura Suplementar 8.....	91
Figura Suplementar 9.....	92
Figura Suplementar 10.....	93
Figura Suplementar 11.....	94
Figura Suplementar 13.....	95
Figura Suplementar 14.....	96
Figura Suplementar 15.....	97
Figura Suplementar 16.....	98
Figura Suplementar 17.....	99
Figura Suplementar 18.....	100

LISTA DE TABELAS SUPLEMENTARES

Tabela suplementar 1	101
Tabela suplementar 2.	102
Tabela suplementar 3	106
Tabela suplementar 4	107
Tabela suplementar 5	108

LISTA DE ABREVIATURAS E SIGLAS

mRNA	RNA mensageiro
RNA	Ácido ribonucleico
WHO	Organização Mundial de Saúde (World Health Organization)
vs	versus
INCA	Instituto Nacional do Câncer
DNA	Ácido desoxirribonucleico
TNBC	Câncer de mama triplo negativo
BL	Tumores do tipo Basal

SUMÁRIO

SUMÁRIO	13
1. INTRODUÇÃO	1
1.1. FATORES DE RISCO	1
1.2. ESTRUTURA DA GLÂNDULA MAMÁRIA E PROCESSO DE INICIAÇÃO TUMORAL	3
1.3. CLASSIFICAÇÃO MOLECULAR DO CÂNCER DE MAMA	5
2. OBJETIVOS	13
2.1. OBJETIVO GERAL	13
2.2. OBJETIVOS ESPECÍFICOS	13
3. MATERIAIS E METODOS	14
3.1. ANOTAÇÃO E SELEÇÃO DE GENES DIFERENCIALMENTE EXPRESSOS NA COMPARAÇÃO ENTRE TNBC COM ALTA EXPRESSÃO DE SPARC (TNBC/SPARC+) VERSUS TNBC COM BAIXA EXPRESSÃO DE SPARC (TNBC/SPARC-)	14
3.2. COLETA E DADOS A PARTIR DE REPOSITÓRIOS TCGA E GEO	14
3.3. SELEÇÃO DE GENES POR ANÁLISE DE SOBREVIVÊNCIA A PARTIR DE TRÊS CONJUNTOS DE DADOS	15
3.4. ANÁLISES DE EXPRESSÃO UTILIZANDO O BANCO DE DADOS UALCAN	15
3.5. ANÁLISES DE EXPRESSÃO UTILIZANDO OS BANCOS DE DADOS GOBO E GEO	15
3.6. PREDIÇÃO E ANÁLISE DE SOBREVIVÊNCIA DE MICRORNAs	16
3.7. PERFIS DE ALTERAÇÕES GÊNICAS E ANÁLISES DE CO-EXPRESSÃO	16
3.8. REDES DE INTERAÇÕES GÊNICAS	17
3.9. ANÁLISE DE CURVAS ROC	17
3.10. VISUALIZAÇÃO DE DADOS E ANÁLISES ESTATÍSTICAS	17
4. RESULTADOS	18
4.1. CLASSIFICAÇÃO FUNCIONAL DE GENES DIFERENCIALMENTE EXPRESSOS NA COMPARAÇÃO ENTRE TNBC COM ALTA EXPRESSÃO DE SPARC (TNBC/SPARC+) VERSUS TNBC COM BAIXA EXPRESSÃO DE SPARC (TNBC/SPARC-)	18
4.2. ANOTAÇÃO GÊNICA E SELEÇÃO DE UM CONJUNTO DE GENES IDENTIFICADOS COMO DIFERENCIALMENTE EXPRESSOS NA COMPARAÇÃO TNBC/SPARC+ VERSUS TNBC/SPARC-	19
4.3. ANÁLISES DE SOBREVIVÊNCIA	20
4.4. ANÁLISES DE REDES DE INTERAÇÕES GÊNICAS	20
4.5. MESP1	26
4.5.1. Análise de perfil de expressão dos transcritos MESP1 em pacientes com câncer de mama	26
4.5.2. Análise de sobrevivência em pacientes com câncer de mama triplo negativo, estratificados pela expressão de MESP1	26
4.5.3. Análise do impacto de microRNAs tendo como alvo o transcrito MESP1 na sobrevivência de pacientes com câncer de mama triplo negativo	28
4.5.4. Análise de metilação e alterações genéticas no gene MESP1 em pacientes com câncer de mama	28
4.5.5. Análise de curvas ROC do transcrito MESP1 em pacientes TNBC tratados com quimioterapia	30
4.6. ZNF556	32
4.6.1. Análise de perfis de expressão dos transcritos ZNF556 em pacientes com câncer de mama	32
4.6.2. Análise de sobrevivência em pacientes com câncer de mama triplo negativo, estratificados pela expressão de ZNF556	34
4.6.3. Análise do impacto de microRNAs tendo como alvo o transcrito ZNF556 na sobrevivência de pacientes com câncer de mama triplo negativo	34
4.6.4. Análise de metilação e alterações genéticas no gene ZNF556 em pacientes com câncer mamário	35
4.6.5. Análise de curvas ROC do transcrito ZNF556 em pacientes TNBC tratados com quimioterapia	36
4.7. TSPAN8	38
4.7.1. Análise de perfis de expressão dos transcritos TSPAN8 em pacientes com câncer de mama	38
4.7.2. Análise de sobrevivência em pacientes com câncer de mama triplo negativo, estratificados pela expressão de TSPAN8	38
4.7.3. Análise do impacto de microRNAs tendo como alvo o transcrito TSPAN8 na sobrevivência de pacientes com câncer de mama triplo negativo	41
4.7.4. Análise de metilação e alterações genéticas no gene TSPAN8 em pacientes com câncer de mama	41
4.7.5. Análise de curvas ROC do transcrito TSPAN8 em pacientes TNBC tratados com quimioterapia	41

4.8.	RIMS2	43
4.8.1.	Análise de perfis de expressão dos transcritos RIMS2 em pacientes com câncer mamário	43
4.8.2.	Análises de sobrevida em pacientes com câncer mamário triplo negativo, estratificados pela expressão de RIMS2	43
4.8.3.	Análise do impacto de microRNA tendo como alvo o transcrito RIMS2 na sobrevida de pacientes com câncer de mama triplo negativo	45
4.8.4.	Análise de metilação e alterações genicas no gene RIMS2 em pacientes com câncer de mama	46
4.8.5.	Análise de curvas ROC dos transcritos RIMS2 em pacientes TNBC tratados com quimioterapia	47
4.9.	KLHL13	50
4.9.1.	Análise de perfis de expressão dos transcritos KLHL13 em pacientes com câncer de mama	50
4.9.2.	Análise de sobrevida em pacientes com câncer mamário triplo negativo, estratificados pela expressão de KLHL13	50
4.9.3.	Análise do impacto de microRNAs tendo como alvo o transcrito KLHL13 na sobrevida de pacientes com câncer de mama triplo negativo.	51
4.9.4.	Análise de metilação e alterações genicas no gene KLHL13 em pacientes com câncer de mama	53
4.9.5.	Análise de curvas ROC do transcrito KLHL13 em pacientes TNBC tratados com quimioterapia	54
5.	DISCUSSÃO	56
5.1.	MESP1	56
5.2.	ZNF556	58
5.3.	TSPAN8	61
5.4.	RIMS2	64
5.5.	KLHL13	67
6.	CONCLUSÕES	70
7.	REFERENCIAS	72
8.	MATERIAL SUPLEMENTAR	84

1. INTRODUÇÃO

O câncer de mama apresentou uma incidência de 1,8 milhões de casos em todo o mundo em 2013, foi a segunda causa principal de morte por câncer em mulheres no mundo (464.000) e a principal causa de morbidade. Do total de óbitos, 63% ocorreram em países em desenvolvimento e 37% nos países desenvolvidos (DESANTIS et al., 2015; FITZMAURICE et al., 2015). No ano 2016, para os Estados Unidos o câncer de mama foi à doença mais comum entre as mulheres, com uma incidência estimada de 246.660 novos casos (29% de todos os cânceres locais) e 40.450 mortes estimadas (14% de todos os cânceres locais) (SIEGEL; MILLER; JEMAL, 2016). Para o presente ano 2018, segundo as estimativas da Agência Internacional para Pesquisa IARC (<http://gco.iarc.fr/>), há uma estimativa global de 18,1 milhões de novos casos de câncer e 9,6 milhões de mortes por câncer, sendo em mulheres a frequência e mortalidade de câncer de mama de 11,6% e 9,7 % respectivamente (BRAY et al., 2018). De acordo com INCA, para o Brasil, para cada ano do biênio 2018-2019, foram esperados 59,700 novos casos de câncer de mama, com um risco estimado de 56,33 casos a cada 100 mil mulheres, sendo as regiões do Sul, Sudeste, Centro-Oeste e Nordeste as mais incidentes (INCA/MS, 2018). No Brasil, apenas 20% dos casos de câncer de mama são diagnosticados nos estados precoces, sugerindo a importância do rastreamento ativo no câncer como forma de diagnóstico, prevenção da doença e redução dos índices de mortalidade. (GOSS et al., 2013; RENCK et al., 2014).

1.1. Fatores de risco

Conhecer os fatores de risco para o câncer de mama pode ajudar aos clínicos a tomarem medidas preventivas para reduzir a incidência da doença, sendo os principais fatores de risco: história familiar, idade, menarca precoce, menopausa tardia, nuliparidade, o uso prolongado de contraceptivos orais, terapia de reposição hormonal, obesidade, consumo de álcool e tabaco pode aumentar a predisposição de desenvolver o câncer de mama (CECILIO et al., 2015; ZHANG et al., 2012).

A herança genética é o principal fator de risco não modificável para o câncer de mama. Aproximadamente 5 a 10% dos tumores de mama são do tipo familiar sendo BRCA1 e BRCA2 os genes melhor caracterizados e mais frequentemente associados com câncer do tipo familiar. Estes genes desempenham funções relacionadas com diferentes processos celulares,

incluindo ativação e regulação transcricional, o reparo de lesões no DNA, controle do ciclo celular, proliferação e diferenciação celular. Os pacientes com mutações nesses genes estão mais propensos a desenvolver tumores (ATCHLEY et al., 2008; PINILLA et al., 2006). Também existem mutações atribuídas a diversas síndromes como TP53 e PTEN respectivamente. Por outra parte mutações em outros genes como ATM, BRIP1, PALB2 e a variante alélica 1100delC do gene CHEK2, são consideradas de penetrância moderada e conferem risco menor para o desenvolvimento do câncer de mama (AFGHAHI ANOSHEH, TELLI L. MELINDA, 2016; COUCH et al., 2016).

Mutações em genes de alta penetrância conferem um risco estimado de desenvolvimento da doença em até 80 anos, de 72% para mulheres com mutação BRCA1 e 69% para mulheres com mutação BRCA2 (KUCHENBAECKER et al., 2017). As mutações somáticas são mais frequentes e constituem 90% dos casos dos tumores mamários e surgem como resultado de lesões provocadas pela interação de agentes físicos, químicos e biológicos com o DNA (PAZ et al., 2018).

O estrogênio um hormônio feminino produzido principalmente nos ovários é importante para o desenvolvimento sexual e está envolvido na regulação da proliferação, diferenciação, metabolismo e homeostase dos tecidos-alvo, como tecidos reprodutivos. Na glândula mamária desempenha um papel principalmente proliferativo, porém uma maior exposição ao estrogênio ao longo da vida pode aumentar o risco de desenvolver o câncer de mama, tal risco está associado à uma menarca precoce ou à uma menopausa tardia (RUSSO; RUSSO, 2004). Nestes dois grupos de risco a atividade deste hormônio pode aumentar no tecido mamário promovendo a proliferação das células e tornando-as suscetíveis a mutações malignas que se não houvesse um sistema de reparo eficiente, cada diminuição de um ano na idade na menarca o risco de câncer de mama pode aumentar em até 5% (COLDITZ; BOHLKE, 2014).

A idade da primeira gravidez também tem sido relacionada com o câncer de mama. Ela induz alterações hormonais de longo prazo, com a redução dos níveis de prolactina e de estrógenos e aumento dos níveis de globulina transportadora de hormônios sexuais, o que pode proporcionar uma redução do risco contra o câncer de mama (COLDITZ; BOHLKE, 2014). Portanto, quanto maior o intervalo entre a menarca e a primeira gravidez, maior o risco de desenvolver câncer de mama.

Dependendo da população a obesidade é considerada o fator de risco modificável mais importante para o câncer de mama, além da atividade física, do tratamento hormonal e da

ingestão de álcool e está associada ao aumento do risco da doença entre as mulheres pós-menopausadas (GUNTER et al., 2015). Key e colaboradores mostraram que o aumento do risco de câncer de mama está envolvido com o aumento do IMC (Índice de Massa Corpórea) em mulheres pós-menopausadas e em grande parte é resultado do aumento associado ao tempo de exposição aos estrógenos, particularmente o estradiol (KEY et al., 2003). Além disso, o trabalho de Petridou e colaboradores (2018) identificou vários fatores de risco relacionados ao estilo de vida, incluindo obesidade, consumo de álcool e inatividade física que aumentam o risco de câncer de mama, especialmente em mulheres na pós-menopausa em até 40% (PETRIDOU; GEORGAKIS; ANTONOPOULOS, 2018).

Dados de dois grandes estudos de coortes dos EUA publicados por Cao e colegas (2015) descobriram que o consumo de álcool leve e moderado em mulheres estava associado a um aumento do risco total de câncer de mama (CAO et al., 2015). Acredita-se que as ações do álcool sejam moduladas por fatores genéticos, particularmente polimorfismos que afetam o metabolismo do álcool, metabolismo de folato e metionina e reparo do DNA (PRAUD et al., 2016). Mecanismos para o câncer de mama parecem interferir o metabolismo do estrogênio, e mesmo beber moderadamente aumenta os níveis circulantes de hormônios sexuais que ativam proliferação celular. (SCOCCIANTI et al., 2014).

1.2. Estrutura da glândula mamária e processo de iniciação tumoral

A glândula mamária sofre mudanças dinâmicas ao longo da vida sendo que ao contrário da maioria dos órgãos epiteliais, desenvolve-se no período pós-natal e sofre remodelação epitelial complexa durante a puberdade, gravidez, lactação e desmame (SELDIN; LE GUELTE; MACARA, 2017).

A glândula mamária possui uma estrutura epitelial ramificada composta por ductos e alvéolos, os quais estão feitos de um epitélio estratificado, de dois tipos de células epiteliais: as células luminais que revestem o lúmen interno e respondem ao estrogênio e produzem o leite, e as células basais/mioepiteliais são localizadas embaixo das células luminais e em contato com a membrana basal, estas células fornecem força contrátil para expelir o leite durante a lactação. (GUSTERSON; STEIN, 2012; MCNALLY; STEIN, 2017).

A glândula mamária já adulta contém uma hierarquia de tipos de células com vários potenciais de automanutenção e diferenciação. Estes incluem células capazes de produzir glândulas mamárias completas e funcionais in vivo e que por sua vez contêm células filhas

com o mesmo potencial regenerativo notável, bem como células com atividade clonogênica mais limitada *in vitro* (MAKAREM et al., 2013). Tem-se descrito que ambos os tipos de células, tanto ductais como alveolares, originam-se a partir de uma célula tronco comum a qual é carente da expressão do receptor de estrogênio (MAKAREM et al., 2013; MOLYNEUX; SMALLEY, 2011), sendo que as células luminais com expressão desse receptor em grande parte correspondem a células diferenciadas (SLEEMAN et al., 2007).

As células luminal e basal do epitélio mamário adulto distinguem-se umas das outras não apenas pelas suas diferentes localizações na glândula, mas também pela sua expressão diferencial de várias proteínas. Assim, tanto as células epiteliais da mama como de outros tecidos estratificados o simples tem expressão diferencial de citoqueratinas tipo I e II, dessa maneira as células luminais expressam as citoqueratinas CK7 e CK8 do tipo I e as CK18 e CK19 do tipo II. Por outra parte as células basais (e/ou mioepiteliais) expressam as citoqueratinas CK5 e CK17 do tipo II e a CK14 do tipo I (MOLL; DIVO; LANGBEIN, 2008).

Dois modelos hipotéticos têm sido propostos para explicar a evolução clonal dos subtipos de câncer de mama. O primeiro modelo, o modelo linear, a célula de origem é a mesma para diferentes subtipos de tumor e, portanto, o subtipo de tumor é determinado por eventos genéticos e epigenéticos adquiridos. O segundo modelo, o modelo não linear, a célula de origem é diferente (presumivelmente uma célula-tronco ou uma célula progenitora) para os diferentes subtipos (CAMPBELL; POLYAK, 2007; POLYAK, 2007). No entanto, Jeselsohn forneceu evidências para sugerir que existem duas populações de células progenitoras, uma originando os cânceres de mama tipo luminal e a outra para os cânceres de mama tipo basal (JESELSON et al., 2010).

O estágio denominado carcinoma ductal *in situ* representa uma condição não invasiva devido que as células cancerígenas estão confinadas ao sistema ductal da mama, sem evidências de invasão através da membrana basal no estroma circundante (LEONARD; SWAIN, 2004; SIZIOPIKOU, 2013). À medida que o tumor progride as células adquirem novas mutações que lhes conferem características próprias do câncer como capacidade proliferativa, resistência à apoptose, sobrevivência, invasão e metástase (HANAHAN; WEINBERG, 2011). Posteriormente as células mioepiteliais transformadas migram ao estroma onde em conjunto degradam a membrana basal resultando na progressão de carcinoma *in situ* a invasivo. Neste estágio as células podem invadir tecidos adjacentes e migrar a órgãos distantes eventualmente desenvolvendo metástase (POLYAK, 2007).

A complexidade do câncer de mama tem promovido à busca de novas e melhores opções terapêuticas. O perfil de expressão diferencial de genes tem sido uma das ferramentas mais utilizadas em investigação para este fim permitindo tanto a classificação e prognóstico de certos tipos de câncer assim como prever prospectivamente a capacidade de resposta a tratamentos terapêuticos (CAMPBELL; POLYAK, 2007; POLYAK, 2007). O câncer de mama é um dos poucos tipos de tumores em que o perfil de expressão gênica tem sido usado com sucesso para o desenho de terapias individualizadas, levando a melhorias significativas na sobrevivência específica da doença.

1.3. Classificação molecular do câncer de mama

Ao nível fenotípico os tumores de mama são classificados pelos patologistas com base as características morfológicas descritivas do tumor e das células neoplásicas (WEIGELT; REIS-FILHO, 2009). A histologia do carcinoma é determinada pela localização de células tumorais, e podem ser ductais ou lobulares, infiltrativo ou in situ, além disso, permite determinar o grau do tumor (estado de diferenciação das células tumorais), presença de invasão linfocitária, presença de necroses, mitoses e no caso dos linfonodos a existência de metástases (WEIGELT; REIS-FILHO, 2009).

O desenvolvimento da técnica de microarranjos de DNA permitiu desenvolver estudos que classificassem os tumores de câncer de mama segundo os seus perfis de expressão gênica. Assim, Perou e colaboradores (2000) caracterizaram variações no perfil de expressão gênica em um conjunto de 65 espécimes cirúrgicos de tumores de mama humanos de 42 indivíduos diferentes, utilizando microarranjos de DNA complementares representando 8.102 genes humanos, sendo que as 65 amostras foram analisadas em dois momentos, antes e após o tratamento quimioterápico neoadjuvante. Nesse estudo os tumores de mama foram agrupados por semelhança no perfil de expressão em dois ramos: receptores estrogênicos positivos ou negativos. Aqueles que expressaram receptores de estrogênio também expressaram genes específicos para células epiteliais (receptores de progesterona, CK8, CK18, BCL-2) e, por essa razão, foram chamados de luminais (constituíram 58%). No ramo do receptor de estrogênio negativo apareceram três subgrupos: basal com o (13% do total), positivo para HER2 (11%) e *normal-like* com o (18%). Os Tumores *normal-like* expressaram genes próprios de células mioepiteliais de mama normal e de células adiposas. A maioria dos tumores do *basal-like* não expressaram nem receptores de estrogênio nem HER2 (triplo

negativo) e expressaram genes específicos para células mioepiteliais, como CK5/6, CK17, integrina $\beta 4$ e laminina. Os tumores HER2+ expressaram altos níveis da oncoproteína HER2 (PEROU et al., 2000).

Posteriormente, o conjunto de genes intrínsecos (reduzido a 456 genes) foi analisado em 78 tumores. Os tumores foram agrupados novamente em dois subgrupos claramente diferenciados em acordo com os receptores de estrogênio. O ramo receptor de estrogênio positivo foi subdividido em três: luminal A (38%), luminal B (6%) e luminal C (12%). O ramo de receptores de estrogênio negativo foi subdividido nos mesmos três subgrupos observados no primeiro estudo: *basal-like* (16%), HER2 positivo (13%) e *normal-like* (15%). O comportamento biológico também foi estudado e observou-se que os tumores *basal-like*, HER2 positivo e luminal C apresentaram menor sobrevida global e menor tempo livre de doença (SORLIE et al., 2001a). Em um terceiro estudo do mesmo grupo, 534 genes (uma versão estendida do conjunto de genes intrínsecos) foram estudados em 115 amostras de tumores, encontrando os mesmos 5 subgrupos: luminal A (36%), luminal B (14%), HER2 (14%), *basal-like* (24%) e *normal-like* (12%). Foram incluídos 18 tumores de pacientes portadoras da mutação BRCA1, todos classificados como basais (SORLIE et al., 2003).

A classificação molecular dos tumores de mama tem sido confirmada por outros grupos de investigação em diferentes populações de pacientes com câncer de mama (BERTUCCI et al., 2008; DI COSIMO; BASELGA, 2010; SORLIE et al., 2003). A importância desta classificação na clínica revela-se no trabalho de Sorlie e colaboradores (2001), no qual se demonstra que essa classificação molecular baseada nos perfis de expressão gênica, correlaciona-se com a sobrevida das pacientes (SORLIE et al., 2001b). Outros grupos de pesquisa usaram esses mesmos genes em sua própria série de tumores, encontrando resultados semelhantes (CALZA et al., 2006; HU et al., 2006).

De acordo com a classificação molecular do câncer de mama, os tumores TNBC (câncer de mama triplo negativo) prevalecem como o subtipo mais agressivo de câncer de mama representando 10 a 15% de todos os tumores de mama diagnosticados (PEROU et al., 2000). Tais tumores ocorrem com frequência em mulheres jovens e são caracterizados por baixa diferenciação e altas taxas proliferativas em comparação com os subtipos Luminal (tumores de receptores hormonais positivos, HR+) e HER2+. Assim, pacientes com tumores TNBC mostram altas taxas de recorrência, alta incidência de metástases distantes e baixo resultado de sobrevida a curto prazo (DIAZ CASAS et al., 2019; REDDY et al., 2018). A maioria dos

triplo-negativo são do tipo carcinoma ductal, o menos comum inclui o medular, adenoide cístico e metaplásico (FOULKES; SMITH; REIS-FILHO, 2010).

Cerca de 65% para 85% dos triplo-negativo são do tipo basal, o que inclui uma menor diferenciação e alta proliferação. Como já antes mencionado, de acordo com a classificação de Perou e colegas, tal subtipo carece de ER, HER2, e apresenta citoqueratinas da célula mioepitelial/basal e EGFR (NIELSEN et al., 2004; RAKHA; GREEN, 2016). Posteriormente essa classificação tem sido confirmada por outros grupos de pesquisa em diferentes populações de pacientes com câncer de mama.

Desta maneira, um estudo avaliando o perfil de expressão genica de 97 tumores de mama triplo-negativo, definidos pela ausência do RE, RP, HER2 revelou que 91% desses tumores foram classificados como tumores “*Basal-Like*” (KREIKE et al., 2007). Outros estudos realizados em câncer de mama por microarranjos de expressão genica, tem confirmado que os TNBC formaram um grupo que se sobrepõe com os tumores BL num 70-80% dos casos definidos no subgrupo *Basal-like* e entre 20-30% dos BL não correspondem num fenótipo TN (BERTUCCI et al., 2008; KREIKE et al., 2007; LIVASY et al., 2006). Por tudo, os tumores basais e TNBC não se consideram sinônimos (RAKHA et al., 2007; REIS-FILHO; TUTT, 2008), entretanto em torno de 23 a 30% do subtipo basal pode não ser o triplo-negativo, sendo considerado como não-basal (BERTUCCI et al., 2008).

Uma forma recente de identificar os tumores “*Basal-Like*”, baseia-se num perfil de expressão por RT-PCR de 50 genes denominado PAM50 (PARKER et al., 2009), esta estratégia avalia os níveis de expressão de 50 genes que permitem diferenciar os subgrupos moleculares descritos por Perou e cols. Esses genes derivam da lista inicial de Sorlie e cols., usada para classificar aos tumores de mama em cada subgrupo (PARKER et al., 2009). A importância de definir esses subgrupos moleculares mediante tal estratégia relaciona-se com selecionar tumores de mama que permitam ajudar ao prognóstico das pacientes e predizer um maior benefício aos quimioterápicos ou terapias alvo-dirigidas.

O triplo-negativo não é uma única doença, existe uma classificação segundo Lehmann e cols. (2011), que analisando perfis de expressão genica de 21 conjuntos de dados de mama, categorizou os tumores TBNC em seis subtipos moleculares: Basal like-1 (BL-1); Basal like-2 (BL-2); Mesenquimal (M); Mesenquimal Sem Cell (MSL); Imunomodulador (IM); Receptor de andrógeno luminal (LAR) e instável (UNS). BL-1/2 apresentaram maior expressão de genes envolvidos na regulação do ciclo celular e na resposta a danos no DNA, enquanto M e

MSL estavam enriquecidos com genes associados a transição epitelial-mesenquimal o subtipo LAR foi caracterizado por maior expressão de genes envolvidos na sinalização do receptor de andrógeno. Interessantemente, os pesquisadores identificaram linhagens celulares de câncer de mama representativas de cada um desses subtipos moleculares as quais mostraram diferentes sensibilidades a várias terapias direcionadas atualmente sob investigação clínica, fornecendo uma plataforma atraente para desenvolvimento futuro de medicamentos no TNBC. Eles concluíram que cada subtipo representa uma nova entidade de câncer que apresenta diferentes desfechos clínicos da doença e na resposta à quimioterapia (LEHMANN et al., 2011).

Essa conquista aprimorou a busca de novas abordagens para conhecer melhor a diversidade dos TBNC, apesar disso, a alta agressividade dos tumores desse subtipo associada à heterogeneidade na expressão gênica variando em acordo com muitos fatores tornou a detecção, o prognóstico e tratamento ambíguos, levando à identificação e caracterização de novos biomarcadores (SLEDGE et al., 2014).

Diversos estudos de larga escala têm levado a um grande avanço na compreensão do câncer de mama. Muitos desses estudos encontram-se depositados em bancos de dados públicos para o livre uso da comunidade científica seja para reanálise desses dados ou para o teste de novas hipóteses. Por exemplo, Curtis e colaboradores (2012) integraram a análise do genoma e transcriptoma de 2000 tumores de mama, identificando 10 subgrupos a partir dos subtipos moleculares de mama já conhecidos. Deste modo, os resultados sugerem uma nova estratificação molecular da população de câncer de mama, que segundo os autores, derivam do impacto de CNAs (*Copy Number Alteration*) somáticas no transcriptoma. A elucidação desses subgrupos de câncer de mama e seus fatores moleculares foi possível a partir de números representativos de pacientes (CURTIS et al., 2012).

Posteriormente o estudo de Curtis foi analisado por diferentes grupos que tem facilitado a identificação de novos subtipos intrínsecos de câncer de mama, a predição de sobrevida e a resposta às drogas. Joe e Hojung Nam (2016) correlacionaram 4 módulos de expressão com prognóstico em câncer de mama e o grupo de Fertig (2016) caracterizou a expressão de 5 moléculas associadas com metástases nos 5 subtipos moleculares de câncer de mama (FERTIG et al., 2015; JOE; NAM, 2016). Liu e colaboradores (2016), usando dados depositados no TCGA com 1000 casos de câncer de mama, identificaram o RNA não codificante de cadeia longa LINC00657 (lncRNAs), como um biomarcador (LIU et al., 2016). O grupo de Fernandez e colaboradores (2012) analisaram o perfil de metilação de 1628

amostras normais, tumorais e não tumorais de diferentes tipos de tecidos, dentre eles o de mama. Nesse trabalho, foram encontradas alterações no padrão de metilação das células cancerígenas e em outras doenças humanas analisadas. Esses dados estão depositados na plataforma GEO-ncbi para o livre uso e estudos adicionais (FERNANDEZ et al., 2012).

O nosso grupo tem como linha de pesquisa principal a identificação e caracterização de biomarcadores para direcionamento de prognóstico e das melhores opções terapêuticas para as pacientes com câncer de mama. Estudos prévios de nosso grupo utilizando análise serial de expressão gênica (SAGE), visando determinar o perfil de expressão gênica em células epiteliais de mama com diferencia de expressão ERBB2 (HB4a e C5.2) antes e após tratamento com o Docetaxel, permitiu a identificação de diversos genes associados ao processo de transformação maligna como proliferação celular, apoptose e organização do citoesqueleto e via de sinalização do HER2, dentre eles o SPARC (*Secreted Protein Acidic Rich in Cystine*) (**Figura 1**), que pode representar um possível potencial marcador prognóstico e de sensibilidade ao tratamento quimioterápico (DOS SANTOS et al., 2006, 2009). Posteriormente, utilizando análise imuno-histoquímica em *tissue-microarrays* contendo espécimes tumorais de mama de pacientes avaliou-se o potencial valor prognóstico da proteína SPARC. O status de expressão de SPARC foi correlacionado com variáveis prognósticas e com a sobrevida clínica das pacientes. As pacientes expressando níveis reduzidos de SPARC tiveram pior sobrevida livre de doença ($p = 0.001$) e sobrevida global ($p = 0.001$) comparados àqueles que expressaram altos níveis da proteína. Além disso, quando avaliados apenas os tumores triplo-negativos segundo a expressão de SPARC, observamos pior evolução da doença em pacientes com expressão reduzida da proteína (NAGAI et al., 2011).

Um estudo recente do nosso grupo determinou o perfil de expressão gênica de tumores de mama triplo-negativos com diferenças de expressão da proteína SPARC e no comportamento clínico da doença. A partir da técnica de *cDNA microarrays* foi realizada a comparação entre TNBC/SPARC+ e TNBC/SPARC-, que permitiu identificar 775 transcritos, dos quais 338 apresentaram aumento e 437 diminuições da expressão. A anotação gênica nas bases de dados públicas e literatura científica foi utilizada para a seleção 03 de genes para validação em uma população de TNBC: SOHLH2 (*spermatogenesis and oogenesis specific basic helix-loop-helix 2*, FC:24,20), DNAJC12 (*DnaJ heat shock protein family (Hsp40) member C12*, FC: 3,5) e LIM1 (*LIM homeobox 1* FC:-22,51). Utilizando imunohistoquímica em TMAs contendo 51 casos de TNBC foi observada que a expressão positiva de SOHLH2 está

associada com eventos de recorrência, a expressão positiva de DNAJC12 foi associada com um pior prognóstico e que a alta expressão de LIM1 está associada um comportamento biológico mais agressivo desses tumores. A principal limitação de tal estudo foi o número reduzido de pacientes TNBC provenientes de uma única instituição. Assim para testar futuras hipóteses validando genes diferencialmente expressos (identificados neste estudo) é necessário desenvolver estudos mais precisos com métodos de validação mais robustos usando coortes de grande escala (DE ALCANTARA FILHO et al., 2018).

Frente ao exposto, o presente trabalho visa utilizar ferramentas *in silico* e bases de dados públicos para avaliar o perfil de expressão e o valor prognóstico de genes identificados pela técnica *cDNA-microarray* de em TNBC com expressão diferencial de SPARC (**Figura 2**).

Figura 1. Representação proporcional da sobreposição entre TNBC e Basal-like, e Classificação molecular do subtipo TNBC em acordo com expressão genica e possível relação com vias moleculares A.- A maioria de tumores é do tipo Basal e vice-versa, sendo que ~75% dos tumores são tanto TNBC quanto *Basal-like*. B.- Estratificação em acordo com perfis de expressão genica em seis subclasses cada uma dessas subclasses mostra taxas variadas de resposta completa patológica (pCR) após quimioterapia neoadjuvante padrão. BL1: Basal-like-1; BL2: Basal-like-2; MSL: Mesechymal Stem-like; IM: Imunomodulatorio; M: Mesenchymal; LAR: Luminal Androgen Receptor; GF: Growth factors; EGFR: Epidermal growth Factor Receptor; EMT: Epitelial Mesenchymal Transition. Adaptado de Mehanna e cols (2019) e Omarini e cols, (2018).

Figura 2. Representação gráfica da estrutura e função de SPARC e sua participação em processos envolvendo remodelação da matriz extracelular (MEC)
Estrutura do SPARC em acordo com três domínios principais (Acidificado, tipo Follistatina, e Ligante de Ca²⁺ extracelular) e funções participando como componente da matriz extracelular durante processos normais como a involução da mama a partir da glândula lactante e patológicos como câncer de mama. MMP: Metaloproteínase; PAI-1: Inibidor do ativador de plasminogênio tipo 1; FN: Fibronectina; TSP1: Thrombospondin 1. Adaptado de Oskarsson, 2013 e Neveen Said, 2016.

2. OBJETIVOS

2.1. Objetivo geral

Avaliar o perfil de expressão e potencial valor prognóstico de genes identificados em tumores de mama triplo negativo com expressão diferencial de SPARC utilizando mineração de dados.

2.2. Objetivos específicos

- Utilizar bases de dados de câncer de mama para avaliar perfil de expressão de genes identificados em tumores TNBC com expressão diferencial de SPARC.
- Utilizar bases de dados contendo coortes robustas de vários subtipos tumorais mamários para avaliar o potencial valor prognóstico de genes identificados em tumores TNBC com expressão diferencial de SPARC.
- Utilizar bases de dados e ferramentas *in silico* para construir perfis de interações gênicas de genes identificados em tumores TNBC com expressão diferencial de SPARC.

3. MATERIAIS E METODOS

3.1. Anotação e seleção de genes diferencialmente expressos na comparação entre TNBC com alta expressão de SPARC (TNBC/SPARC+) versus TNBC com baixa expressão de SPARC (TNBC/SPARC-)

Bancos de dados, públicos e dados de literatura científica foram utilizados para a anotação gênica detalhada dos genes identificados como diferencialmente expressos na comparação entre TNBC/SPARC + versus TNBC/SPARC- (50 com aumento de expressão e 50 com diminuição de expressão) (DE ALCANTARA FILHO et al., 2018). Adicionalmente, foi utilizada a ferramenta Gene ontology ([//geneontology.org/docs/go-enrichment-analysis/](http://geneontology.org/docs/go-enrichment-analysis/)) para realizar uma análise de enriquecimento de conjuntos de genes visando classificar os nossos genes em acordo com diferentes processos biológicos. Posteriormente, utilizando ferramentas como *gene* e *pubmed* do NCBI com base em informações como função gênica, localização cromossômica, processos biológicos e associação com o processo de tumorigênese, um subgrupo genes foi selecionado para uma segunda etapa realizando uma análise mais abrangente utilizando bancos de dados públicos como representado no fluxograma da (Figura 3).

3.2. Coleta e dados a partir de repositórios TCGA e GEO

Para este estudo, os dados de RNA-seq nível 3 de 1205 pacientes do banco de dados TCGA foram baixados do estudo Breast Invasive Carcinoma (TCGA-provisional), sendo 112 amostras de tecido normal e 1093 amostras de tecido tumoral. Os dados foram mesclados para expressão de transcritos primários (RNA-Seq), micro-RNA (miRSeq) e desfecho clínico. Todos os dados foram fornecidos pelo Firehose Pipeline do Broad Institute, baixados do The Cancer Genome Atlas (www.cancer.gov/tcga), que disponibiliza uma vasta quantidade de dados de múltiplos estudos em diferentes tipos de câncer para extração, análise e visualização de dados de mRNA, miRNA, mutação, metilação, expressão de proteína e dados clínicos relacionados a sobrevida. Para todas as amostras o método de quantificação empregado foi RSEM gerando dados em “expression_log2” e “z.score”. Adicionalmente, seguindo as mesmas diretrizes foram obtidos dados do tipo RNA-Seq e desfecho clínico a partir do *gene-set* GSE81450 (n = 3678) da plataforma GEO (ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse81540). Foi empregado o software R e

pacotes do repositório Bioconductor (<http://bioconductor.org/>) para obtenção e processamento de todos os dados.

3.3. Seleção de genes por análise de sobrevida a partir de três conjuntos de dados

O valor prognóstico dos genes selecionados foi avaliado usando a plataforma KMPlotter para câncer de mama (www.kmplot.com) que contém a assinatura de expressão gênica pela plataforma Affymetrix (sondas de 20129 genes) para 3421 pacientes (GYÖRFFY et al., 2010), e dados coletados da base de dados TCGA- e o dataset GSE817450 contendo 3175 pacientes da base de dados GEO - Gene Expression Omnibus um repositório de dados de expressão gênica de alto rendimento e matrizes de hibridação, chips e microarrays. As sondas disponíveis para análise de sobrevida global e sobrevida livre de recidiva de acordo com a plataforma KMPlotter encontram-se na **Tabela 1**. As pacientes foram estratificadas por alta e baixa expressão dos genes-alvo para o subtipo Basal-like/Triplo negativo em acordo com a sobrevida global, selecionando o melhor ponto de corte e 10 anos de seguimento total (para KMPlotter e TCGA) e 06 anos para GEO. Foram selecionados apenas genes que conseguiram estratificar os pacientes e apresentaram coincidência nas três bases de dados independente do valor logrank p. Adicionalmente, apenas para a base de dados KMPlotter tanto análises de sobrevida global quanto sobrevida livre de recorrência foram realizadas de acordo com o subtipo intrínseco: Basal-like, Luminal A, Luminal B e Her2-positivo, levando em consideração o valor prognóstico e seu impacto para um resultado desfavorável.

3.4. Análises de expressão utilizando o banco de dados UALCAN

Os padrões de expressão de genes de acordo com o tipo de amostra (tumor versus normal), subtipo intrínseco (Luminal, Her2+ e Triplo-negativo) e subtipos moleculares do subtipo triplo negativo do TCGA (BL1, BL2, M, MSL, IM, LAR) de acordo com a classificação de Lehmann (CHANDRASHEKAR et al., 2017).- foram obtidos usando o banco de dados UALCAN, e a ferramenta ‘gene expression analysis’ (<http://ualcan.path.uab.edu>).

3.5. Análises de expressão utilizando os bancos de dados GOBO e GEO

O nível de expressão de mRNA foi analisado usando a ferramenta GOBO (*Gene expression-based Outcome for Breast cancer Online*: <http://co.bmc.lu.se/gobo>), que fornece uma lista completa de genes e sondas Affymetrix de onze conjuntos de dados diferentes de 1881 amostras de tumor de mama (RINGNÉR et al., 2011). Para cada gene, usamos a ferramenta

tumor com con (Figurações padrão para executar a análise de expressão de genes baseado no status do ER, na classificação PAM50 e no grau histológico (NPI), seguindo as mesmas diretrizes foram obtidos dados a partir da plataforma GEO (GSE81450).

3.6. Predição e análise de sobrevida de microRNAs

Deteção e análise de microRNAs que têm como alvo o subconjunto de genes selecionados foi realizada usando o banco de dados mirTarBase, (mirtarbase.mbc.nctu.edu.tw) que contem miRNAs validados experimentalmente por ensaio repórter, western blot, microarray e experimentos de sequenciamento de nova geração. Os microRNAs validados por um método de validação com “evidencia forte” foram escolhidos para cada gene do nosso conjunto. Adicionalmente, a plataforma miRWalk (zmf.umm.uni-heidelberg.de/), a qual é uma plataforma de código aberto que fornece uma interface intuitiva que gera sitios de ligação a miRNA preditos de genes humanos conhecidos, foi empregada para selecionar miRNAs preditos até por doze bases de dados diferentes (**Tabela suplementar 5**), alvejando a região 3'UTR de cada gene, sendo que para todos os algoritmos foi utilizado um *p-value* menor a 0,0005. O valor prognóstico de microRNAs selecionados foi determinado usando a plataforma TCGA, assim todos os microRNAs foram avaliados para sobrevida global no subconjunto TCGA-TNBC.

3.7. Perfis de alterações genica e análises de co-expressão

Utilizando a plataforma cBioPortal (cbioportal.org), e a ferramenta: “QUERY MUTATIONS” – foi examinado o número de mutações tipo missense, inframe, e truncating a partir de dados depositados em dois estudos: TCGA-provisional (n =963) e METABRIC (n =2173). Os quais contém informações de mutações e CNA de pacientes com câncer de mama. Da mesma maneira, utilizando a ferramenta “query co-expression” com informação de mRNA e proteínas de 963 pacientes foi realizado o análise de co-expressão visando encontrar genes correlacionados com o nosso subgrupo de genes, o teste estatístico de Pearson ($P \geq 0.5$) foi aplicado para seleção usando a opção “show regression line” e “Log-Scale”. Adicionalmente, a análise de metilação da região promotora do gene de cada um dos genes selecionados versus o nível de mRNA do transcrito completo foi executado usando a ferramenta “PLOTS”, tal análise foi obtido da plataforma HM450 (Array Methylation Express) tendo em conjunto 885 casos comuns para ambos os dados.

3.8. Redes de interações genicas

Redes de interações biológicas de nossos genes foram geradas a partir de um subconjunto de câncer de mama triplo negativo usando o software Ingenuity Pathways analysis (IPA; www.ingenuity.com), que permite os padrões de expressão gênica usando um banco de dados baseado em literatura científica.

3.9. Análise de Curvas ROC

A plataforma ROC Plotter (<http://rocplot.org/>), que pode vincular a expressão gênica e a resposta à terapia usando dados no nível de transcriptoma de 3.104 pacientes com câncer de mama e 2.369 pacientes com câncer de ovário, foi utilizada para avaliar curvas ROC visando prever a resposta a quimioterapia para Taxane, Anthracycline, Ixabepilone, CMF (Cyclophosphamide, Methotrexate, Fluorouracil), FAC (Fluorouracil, Adriamycin, Cytosine), FEC (Fluorouracil, Epirubicin, Cytosine) com base na sobrevida livre de recidiva em cinco anos em pacientes TNBC. A melhor sonda JetSet foi usada para representar cada gene com o “*optimal cutoff*” como ponto de corte. Os parâmetros aceitos para uma boa predição foram $AUC \geq 6$, intermediária AUC entre 5 e 6, e baixa $AUC \leq 4$.

3.10. Visualização de dados e análises estatísticas

A maioria dos dados obtidos diretamente dos diferentes bancos de dados foram plotados como gráficos boxplots e curvas de sobrevida usando pacotes R como ggplot2, ggpubr e survminer. Os testes estatísticos foram realizados usando os testes de Wilcoxon e Kruskal-Wallis no pacote estatístico R, versão 3.6.1.

4. RESULTADOS

No presente estudo, procuramos determinar o perfil de expressão e o potencial valor prognóstico de genes previamente identificados por nosso grupo em amostras de tumores de mama do subtipo triplo-negativo (TNBC) com diferença na expressão de SPARC. As análises *in silico* foram realizadas utilizando bancos de dados públicos e ferramentas de mineração de dados, visando identificar potenciais novos biomarcadores em câncer de mama. Em uma primeira etapa, realizamos a anotação dos 100 genes mais diferencialmente expressos (50 *up*-regulados e 50 *down*-regulados) utilizando bancos de dados e dados da literatura científica, posteriormente nosso interesse foi continuar na caracterização *in silico* os genes selecionados em diferentes coortes de câncer de mama com relação ao impacto dos genes selecionados na sobrevida das pacientes. Adicionalmente, quatro genes pertencentes à família CEACAM (*Carcinoembryonic antigen-related cell adhesion molecule*), os quais se destacaram dentro do subconjunto dos 100 genes mais diferencialmente expressos e presentes nas análises de interações gênicas foram também investigados quanto ao perfil de expressão e potencial valor prognóstico em câncer de mama. Os dados dos membros da família CEACAM são apresentados na forma de manuscrito (**Manuscrito em preparação**).

4.1. Classificação funcional de genes diferencialmente expressos na comparação entre TNBC com alta expressão de SPARC (TNBC/SPARC+) versus TNBC com baixa expressão de SPARC (TNBC/SPARC-)

O conjunto de 100 genes foi enriquecido usando a ferramenta *Gene Ontology* (GO), via a plataforma DAVID (*The Database for Annotation, Visualization and Integrated Discovery* v6.8) (HUANG; SHERMAN; LEMPICKI, 2009), a qual fornece um sistema para classificar hierarquicamente genes ou produtos gênicos em termos organizados em uma estrutura gráfica (ou uma ontologia). Os termos são agrupados em três categorias: função molecular (descreve a atividade molecular de um gene), processo biológico (descreve o principal papel celular ou fisiológico desempenhado por um gene, coordenado com outros genes) e componente celular (descreve a localização na célula onde o produto do gene executa sua função). Cada gene pode ser descrito (anotado) com vários termos. A ferramenta GO é usada ativamente para classificar genes de humanos e de diferentes espécies.

Em acordo com os nossos resultados, 49 de 100 genes foram devidamente agrupados para os três termos GO: processos biológicos, compartimento celular e função molecular (**Tabela**

suplementar 1). Para processos biológicos apenas foram enriquecidos 15,3% de todos os genes, e para compartimento celular como para função molecular 18,5% e 16,1% respectivamente (**Figura 4A, Figura suplementar 1**). De acordo com os processos biológicos celulares: 7 (0.07%) genes estiveram associados à diferenciação celular, 5 (0.05%) à resposta em defesa a bactérias, no entanto os processos menos frequentes foram: processos sinápticos, desenvolvimento da epiderme e diferenciação do queratinócito 3 (0.03%), e para regulação negativa da atividade peptidase, estabelecimento de barreira cutânea e especificação de padrão embrionário, 2 (0.02%) respectivamente (**Figura 4A**).

4.2. Anotação gênica e seleção de um conjunto de genes identificados como diferencialmente expressos na comparação TNBC/SPARC+ versus TNBC/SPARC-

Utilizando ferramentas de mineração de dados do *NCBI (Gene e Pubmed)*, foram estabelecidos dois critérios de seleção, o primeiro a partir de informações a respeito de funções biológicas e bioquímicas, tanto em estados normais e patológicos, dentre elas câncer. E o segundo baseado em informações de eventos cromossômicos relacionados aos loci dos 100 genes identificados como mais diferencialmente expressos na comparação TNBC/SPARC+ versus TBNC/SPARC- (**Tabela suplementar 3**), ambos os critérios foram mutuamente não exclusivos. No primeiro, foram identificados genes relacionados a doenças cardiovasculares, neuronais, genéticas e câncer (supressores tumorais e oncogenes). Além disso, os eventos cromossômicos mais frequentes, tanto em câncer de mama como em outros cânceres foram: amplificações, deleções, SNP (*single nucleotide polymorphism*), CNV (*copy number variants*) e CNA (*copy numbers alterations*), rearranjos e processos relacionados com alteração da metilação, todos associados à perda e/ou ganho de função. Apenas 22 genes que se enquadraram em ambos os critérios (**Figura 4B**), foram avaliados para uma segunda etapa utilizando a base de dados TCGA, a partir de um terceiro critério visando analisar a expressão de cada um deles em amostras tumorais de mama em comparação a tecidos normais, assim como a expressão para diferentes subtipos tumorais de mama. Após dessa etapa, foram selecionados apenas 20 genes (**Tabela suplementar 4**).

4.3. Análises de sobrevida

Dos 20 genes avaliados para sobrevida global por três bancos de dados públicos, cinco foram selecionados para análises posteriores. A alta expressão de três deles: MESP1 (KMPlotter: logrank $p=5.6e-05$; TCGA: logrank $p=0.058$; GEO: logrank $p=1269$), ZNF556 (KMPlotter: logrank $p=0.0044$; TCGA: logrank $p=0.047$; GEO: logrank $p=0.0334$) e KLHL13 (KMPlotter: logrank $p=0.014$; TCGA: logrank $p=0.0036$; GEO: logrank $p=0.1226$) e a baixa expressão de TSPAN8 (KMPlotter: logrank $p=0.03$; TCGA: logrank $p=0.019$; GEO: logrank $p=0.0045$) e RIMS2 (KMPlotter: logrank $p=0.021$; TCGA: logrank $p<0.0001$; GEO: logrank $p=0.4141$) indicaram um prognóstico desfavorável no subtipo triplo-negativo.

4.4. Análises de redes de interações gênicas

Foram geradas nove redes de interações genicas a partir de duas combinações, a primeira tendo como entrada (*entry*) 22 genes (13 top downregulados/ 9 top upregulados; *cutoff* = 3) contra duas planilhas (*seed*) de 35 e 70 genes. A segunda tendo como entrada 50 genes, (25 top downregulados/25 top upregulados; *cutoff* = 3) contra as mesmas planilhas de 35 e 70 genes. O conjunto de genes utilizados como *entry* foram provenientes do estudo do nosso grupo. Os genes melhor enriquecidos foram CTNBB1, BRCA1, MYC, AR, NOTCH1, IL4, FOS, ERBB2, e ESR1 representando sinalizações associadas a proliferação, regulação da transcrição, eventos hormonais, metástase, interação célula-célula e modificações da matriz extracelular (**Figura6**).

Figura 3. Fluxograma de trabalho visando obter candidatos potenciais a biomarcadores em câncer de mama com ênfase no subtipo TNBC a partir de um conjunto de 100 genes

A

B

Figura 4. Representação gráfica de dos subconjuntos de genes obtidos a partir de um estudo do nosso grupo. A.-Enriquecimento no termo PROCESSOS BIOLÓGICOS para 100 genes identificados na análise de transcriptômica: TNBC/SPARC+ versus TNBC/SPARC-, utilizando a ferramenta GENE ONTOLOGY. **B.-**Enquadramento de 22 genes como resultado do fluxograma de trabalho a partir de Log2_expression da comparação transcriptômica entre tumores triplo negativo com expressão diferencial do SPARC, utilizando o software R.

Figura 5. Perfil de Alteração genicas para os 05 genes selecionados, utilizando dados de dois estudos em câncer de mama. A: METABRIC (n =1988); B: TCGA (n = 963). Indicadores: verde: mutação missense; cinza: mutação truncada; vermelho: amplificação; azul: deleção. Dados obtidos da plataforma Cbioportal, utilizando a ferramenta OncoPrint.

Figura 6. Rede de interação genica apartir de 22 genes da comparação transcriptômica TNBC/SPARC+ versus TNBC/SPARC-. Genes ou os produtos gênicos são representados como nós, e a relação biológica entre dois nós é representada como uma aresta. A intensidade da cor do nó indica o grau de regulação. Dados obtidos utilizando o programa IPA-Ingenuity.

Tabela 1. Análise de Sobrevida para 20 genes selecionados a partir de tres bases de dados visando avaliar um prognóstico desfavorável no subtipo basal.

SOBREVIDA GLOBAL PARA PIOR PROGNOSTICO								
gene	sonda	KMPlotter (Microarray)				TCGA (RNA-seq)		GSE815540 (RNA-seq)
		OS_ALL(n=1402)	RFS_ALL(n=3955)	Basal_OS(n=241)	Basal_RFS(n=879)	OS_ALL(n=1093)	TBNC_OS (n=115)	BASAL (n < 700)
MESP1	224476_at	up(0.00034)	up(0.0032)	up(5.6e-05)	down(0.00036)	down(0.12)	up(0.058)	up(0.1269)
ZNF556	220709_at	-	down(0.0058)	up(0.0044)	down(0.087)	up(0.079)	up(0.047)	up(0.0334)
CAPN8	220930_at	down(0.0022)	down(1.1e-14)	up(0.015)	down(0.016)	down(0.00037)	down(0.022)	down(0.0071)
SCGB2A2	206378_at	-	down(1.3e-07)	up(0.00083)	up(0.32)	-	up(0.094)	down(0.0626)
SCGB2A1	205979_at	down(0.024)	-	up(0.026)	up(0.0082)	down(0.042)	-	down(0.0336)
SLAIN1	203184_at	-	down(3e-05)	down(0.017)	down(0.001)	down(0.0068)	up(<0.0001)	up(0.0115)
FBN2	203184_at	-	up(5.4e-05)	down(0.007)	down(0.016)	up(0.16)	up(0.17)	down(0.0138)
TSPAN8	203824_at	down(0.001)	down(3.1e-10)	down(0.03)	down(0.0001)	up(0.016)	down(0.019)	down(0.0045)
RIMS2	229823_at	up(0.045)	up(0.012)	down(0.021)	down(0.046)	up(0.0099)	down(<0.0001)	down(0.4141)
MSMB	210297_s_at	up(0.24)	down(0.008)	up(0.022)	down(0.16)	up(0.00086)	down(0.2)	down(0.0005)
CHP2	206149_at	up(0.22)	down(4.2e-10)	down(0.047)	down(0.0004)	up(0.026)	up(0.034)	up(0.2524)
KLHL13	218990_s_at	-	down(0.0013)	up(0.081)	up(0.22)	-	up(0.213)	up(0.1226)
TOX3	214774_x_at	up(1.6e-5)	up(1.1e-5)	up(0.081)	up(0.014)	down(0.064)	up(0.0036)	down(0.0196)
SLURP1	214536_at	up(0.038)	-	down(0.091)	down(0.03)	up(0.21)	down(0.013)	up(0.0128)
DAPL1	229290_at	down(0.15)	down(0.0063)	down(0.11)	up(0.2)	down(<0.0001)	up(0.15)	up(0.3443)
SIM2	206558_at	up(0.001)	down(0.11)	up(0.25)	down(0.35)	down(0.0026)	up(0.028)	up(0.0002)
HMP19	218623_at	down(0.27)	down(0.0037)	up(0.27)	down(0.003)	up(0.098)	up(0.25)	-
CAMP	210244_at	up(0.0002)	up(3.8e-7)	down(0.32)	down(0.0058)	down(0.005)	down(0.13)	down(0.1691)
PKNOX2	222171_s_at	down(0.21)	down(4.1e-5)	down(0.24)	down(0.0085)	down(0.22)	up(0.05)	up(0.0856)
DDTL	202929_s_at	up(0.078)	up(2.5e-7)	-	up(0.0017)	down(0.0013)	up(0.12)	up(0.0062)

Melhor ponto de corte usado como *cut-off*. Up: regulado; Down: downregulado. Entre parênteses (Log-Rank). OS: sobrevida global, RFS: Sobrevida livre de recidiva. ALL: para todos os subtipos. Representados apenas genes que conseguiram estratificar pacientes, independentemente do *P-value*. Dados obtidos de três plataformas diferentes envolvendo duas técnicas para expressão genica: KMPlotter (microarray), TCGA e GSE (RNA-seq). Em vermelhos genes selecionados segundo critério de sobrevida.

4.5. MESP1

4.5.1. Análise de perfil de expressão dos transcritos MESP1 em pacientes com câncer de mama

A análise dos níveis de expressão do mRNA de MESP1 em tecidos normais e tumorais de mama foi realizada utilizando dados depositados em três bases de dados; TCGA, GEO (GSE81540) e GOBO. Nos dados TCGA, os transcritos de MESP1 encontram-se downregulados nas amostras de tumor de mama quando comparado com amostras de tecido mamário normal (TCGA, $p=NS$); (**Figura 7A**). Na comparação entre os subtipos moleculares, embora não significativo, os transcritos de MESP1 apresentaram menor expressão nos subtipos luminais, e maior expressão nos subtipos HER2+ e TNBC. Dados semelhantes foram obtidos em acordo com a classificação molecular PAM50 GEO (GSE81540, $p = 0.001$); (**Figura 7B, Figura suplementar 4A**). Além disso, foi avaliada a expressão por status de receptor de estrogênio, onde foi observado que níveis elevados de MESP1 estão significativamente relacionados com pacientes apresentando negatividade para ER (GSE81540, logrank $p = 0.0050$); (**Figura suplementar 4B**). Quando avaliada a classificação molecular para triplo negativo, MESP1 foi mais abundante nos subtipos MSL, BL1, BL2 (**Figura 7C**).

4.5.2. Análise de sobrevida em pacientes com câncer de mama triplo negativo, estratificados pela expressão de MESP1

Em termos gerais a alta expressão do transcrito esteve associada a um prognóstico desfavorável quando avaliado para sobrevida global e livre de recorrência para todos os subtipos de câncer de mama (**Figura suplementar 2A e B**). Altos níveis de expressão do MESP1 em pacientes com tumores de mama do subtipo triplo negativo foram relacionados com sobrevida global e livre de recorrência (logrank $p=5.5e-05$; logrank $p=3.3e-05$) curta para dados depositados na plataforma KMPlotter (**Figura 8A, Figura suplementar 3H**). Da mesma maneira resultados semelhantes embora não significativos foram achados na base de dados TCGA para subtipo TNBC onde a alta expressão do MESP1 esteve associada com um prognóstico desfavorável (logrank $p=0.058$). Por outro lado, no dataset GSE81540, não houve associação estatisticamente significativa entre a expressão do gene *MESP1* e a sobrevida global em pacientes do subtipo Basal (logrank $p=0.1269$); (**Figuras 8B e C**).

Figura 7. Perfil de expressão do transcrito MESP1 em tumores mamários. A.- amostras tumorais de câncer de mama e de tecidos normais. B.- diferentes subtipos de câncer de mama (Luminal, HER2+ e Triplo negativo) C.- subtipos moleculares para triplo negativo. *P-value* para testes T-test e ANOVA $<0,05=*$; $<0,005=**$; $<0,0005=***$.

4.5.3. Análise do impacto de microRNAs tendo como alvo o transcrito MESP1 na sobrevida de pacientes com câncer de mama triplo negativo

Não foram obtidos resultados com a base dados miRTarBase. Em acordo com resultados da base de dados mirWALK, utilizando o máximo número de algoritmos de predição (06), foram preditos nove microRNAs que tem como alvo região 3' UTR do MESP1, dos quais a baixa expressão de dois deles (hsa-miR-34b-5p, $p = 0.12$); (hsa-miR-1229-3p, $p = 0.31$) e alta expressão de (hsa-miR-18a-5p, $p = 0.069$) e (hsa-miR-18b-5p, $p = 0.044$) esteve associada com uma sobrevida global curta em pacientes com subtipo triplo negativo (**Figuras 9A-D**).

Figura 8. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de MESP1 em três datasets diferentes. A.- KM PLOTTER. B.-TCGA. C.-GSE81540. O melhor ponto de corte foi empregado para todas as análises.

4.5.4. Análise de metilação e alterações genéticas no gene MESP1 em pacientes com câncer de mama

Encontrou-se uma correlação inversa entre o nível de metilação e expressão dos transcritos MESP1 para todos os pacientes com câncer de mama ($Pearson = -0,22$); (**Figura 10A**). A frequência de alterações a partir da base de dados do TCGA provisional, a qual contém 963 amostras de câncer de mama, relatou apenas uma porcentagem de 4% de alterações como ‘amplificações’, entretanto não foi encontrado nenhum tipo de mutações somáticas. Para a base de dados METABRIC a porcentagem de amplificações foi estimada em 1.7% (**Figuras 5B e A**). Alterações no número de cópia (CNA) também foram investigadas para o gene MESP1 usando o conjunto de dados do estudo METABRIC com 2173 pacientes de câncer de mama o qual mostrou que há aumento na expressão de mRNA apenas quando ganhos, deleções homozigotas (*deep deletions*), deleções heterozigotas (*ShalBaixa deletions*) a respeito de

amplificações. Resultados semelhantes foram encontrados para a base de dados TCGA (Figuras 10B e C).

Figura 9. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao MESP1. A.-hsa-miR-18b-5p, B.- hsa-miR-18a-5p; C.- hsa-miR-1229-3p, D.-hsa-miR-34b-5p. Dados obtidos da plataforma TCGA.

Figura 10. Alterações genéticas e perfil de metilação do MESP1 A.- Correlação entre expressão e metilação. Perfis de expressão versus alterações no numero de cópias CNA a partir de: B.-TCGA. C.-METABRIC.Dados obtidos da plataforma CBIportal.

4.5.5. Análise de curvas ROC do transcrito MESP1 em pacientes TNBC tratados com quimioterapia

A curva característica do operador receptor (ROC) foi usada para avaliar os papéis preditivos dos genes selecionados. MESP1 mostrou estar downregulado nos tumores do grupo de pacientes que responderam a quimioterapia, como já descrita na parte metodológica, em relação ao grupo de pacientes não responderam. Entretanto, a curva

abaixo da área apresentou uma tendência a identificar melhor os pacientes responsivos mostrando uma melhor sensibilidade e especificidade (**Figuras 11A e B**).

Figura 11. Representação da expressão de MESP1 associada à resposta a quimioterapia. A.- boxplots entre pacientes responsivos versus não responsivos; B.- Curva ROC exibindo AUC. Relação sensibilidade/especificidade, dados obtidos da plataforma ROCPLLOT.

4.6. ZNF556

4.6.1. Análise de perfis de expressão dos transcritos ZNF556 em pacientes com câncer de mama

Em acordo os dados da base de dados TCGA, transcritos ZNF556 encontraram-se downregulados nas amostras de tumor de mama quando comparados com amostras de tecido mamário normal (TCGA, $p = 0.008$). Não houve diferenças significativas na expressão de ZNF556 entre os subtipos moleculares de câncer de mama, embora se observe uma tendência de menores níveis de expressão desse gene nos subtipos triplo-negativo e HER2+ (**Figuras 12A e B**). Dados semelhantes foram obtidos para expressão de ZNF556 em relação a classificação molecular PAM50 da plataforma GEO (GSE81540, $p = 0.001$), mas diferiram da base de dados GOBO onde ZNF556 esteve mais expresso no subtipo HER2+, expressão intermediária no subtipo Basal e expressão diminuída nos subtipos Luminais ($p = 0.0068$) (**Figuras suplementares 8A e D**). Além disso, foi avaliada a expressão por status de ER, onde foi achado que os níveis elevados de ZNF556 estão significativamente relacionados com pacientes apresentando negatividade para ER (GSE81540, $p < 0.0001$; GOBO, $p = 0.419$) (**Figuras suplementares 8B e E**). Quando avaliada expressão do ZNF556 em acordo com a classificação molecular intrínseca para triplo negativo, houve mais abundância do transcrito de ZNF556 nos subtipos M, MSL, BL1, BL2 (**Figura 12C**).

Figura 12. Perfil de expressão do transcrito ZNF556 em tumores mamários. A.- amostras tumorais de câncer de mama e de tecidos normais. B.- diferentes subtipos de câncer de mama (Luminal, HER2+ e Triplo negativo) C.- subtipos moleculares para triplo negativo. *P-value* para testes T-test e ANOVA <0,05=*; <0,005 = **; <0,0005 = ***.

Figura 13. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de ZNF556 em três datasets diferentes. A.- KM PLOTTER. B.-TCGA. C.-GSE81540. O melhor ponto de corte foi empregado para todas as análises.

4.6.2. Análise de sobrevida em pacientes com câncer de mama triplo negativo, estratificados pela expressão de ZNF556

Segundo a base de dados KMPlotter, pacientes cujos tumores apresentaram alta expressão do ZNF556 exibiram um prognóstico desfavorável quando avaliado para sobrevida global (logrank $p=0.001$), não entanto os pacientes que tiveram uma diminuição do transcrito mostraram uma sobrevida livre de recorrência curta (logrank $p=3.1e-10$), ambas as sobrevidas para todos os subtipos de câncer de mama (**Figura suplementar 6A e B**). Entretanto, altos e baixos níveis do ZNF556 em pacientes com tumores de mama do subtipo triplo negativo foram relacionados com sobrevida global e livre de recorrência (logrank $p=0.0044$; logrank $p=0.087$) curta respectivamente para dados depositados na plataforma KMPlotter (**Figura 13A, Figura suplementar 7H**). Da mesma maneira, resultados semelhantes foram achados na base de dados TCGA para subtipo TNBC onde a alta expressão do transcrito esteve associada com um prognóstico desfavorável (logrank $p=0.047$). Semelhantemente, num terceiro conjunto de dados GSE81540, altos níveis de expressão do transcrito foram correlacionados com diminuição da sobrevida global em pacientes do subtipo Basal (logrank $p=0.0334$) (**Figuras 13B e C**).

4.6.3. Análise do impacto de microRNAs tendo como alvo o transcrito ZNF556 na sobrevida de pacientes com câncer de mama triplo negativo

A plataforma miRTarBase previu 114 microRNAs dos quais 11 apresentaram valor prognóstico em pacientes TNBC, sendo que apenas quatro deles exibiram um logrank significativo, três de eles com alta expressão associada a pior prognóstico: hsa-miR-550a-5p, ($p = 0.007$), hsa-miR-410-3p ($p = 0.0062$), hsa-miR-181d-5p ($p < 0.0001$); e um deles com expressão reduzida hsa-miR-214-5p ($p = 0.0098$); (**Figuras 14A-D**). No entanto, a base de dados mirWALK, utilizando o máximo número de algoritmos de predição (07), previu 03 microRNAs alvejando a região 3' UTR do ZNF556, cuja alta expressão de todos eles (hsa-miR-20b-5p, $p = 0.18$); (hsa-miR-93-5p, $p = 0.064$) e (hsa-miR-17-5p, $p = 0.16$) esteve associada com uma sobrevida global curta em pacientes triplo negativo. Apenas serão representados miRNAs com *p-value* significativo.

Figura 14. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao ZNF556. A.-hsa-miR-550a-5p, B.- hsa-miR-410-3p; C.- hsa-miR-181d-5p, D.-hsa-miR-214-5p. Dados obtidos da plataforma TCGA.

4.6.4. Análise de metilação e alterações genicas no gene ZNF556 em pacientes com câncer mamário

Encontrou-se uma correlação inversa entre o nível de metilação e expressão do transcrito ZNF556 para todos os pacientes com câncer de mama ($Pearson = -0,64$) (**Figura 15A**). A frequência de alterações a partir da base de dados do TCGA provisional, a qual contem 963 amostras de câncer de mama, relatou apenas uma porcentagem de 1.7 % de alterações entre ‘amplificações’, ‘deleções’, e ‘mutações’, sendo que destas últimas foram encontradas duas mutações missense (A282T, R445K) em duas pacientes. Na base de dados METABRIC a porcentagem de amplificações foi estimada em 0.3% (**Figuras 5B e A**). Alterações no número de cópia (CNA) também foram investigadas para o gene ZNF556 usando os conjuntos de dados dos estudos TCGA e METABRIC o quais mostraram que há

aumento na expressão de mRNA apenas quando há ganhos, deleções heterozigotas (*Shallow deletions*) (Figura 15B e C).

4.6.5. Análise de curvas ROC do transcrito ZNF556 em pacientes TNBC tratados com quimioterapia

Na curva característica do operador receptor (ROC), ZNF556 mostrou estar levemente mais expresso em pacientes que não responderam a quimioterapia a respeito daqueles que responderam. Entretanto, a curva abaixo da área apresentou certa tendência a identificar melhor as pacientes responsivas AUC= 0.503 (Figuras 16A e B).

Figura 15. Alterações genéticas e perfil de metilação do ZNF556 A.- Correlação entre expressão e metilação. Perfis de expressão versus alterações no número de cópias CNA a partir de: B.-TCGA. C.-METABRIC. Dados obtidos da plataforma CBioportal.

Figura 16. Representação da expressão de ZNF556 associada à resposta a quimioterapia. A.- boxplots entre pacientes responsivos versus não responsivos; B.- Curva ROC exibindo AUC. Relação sensibilidade/especificidade, dados obtidos da plataforma ROCPLLOT.

4.7. TSPAN8

4.7.1. Análise de perfis de expressão dos transcritos TSPAN8 em pacientes com câncer de mama

Segundo os dados TCGA, embora não significativos níveis de *TSPAN8* encontraram-se downregulados nas amostras de tumor quando comparado com amostras de tecido mamário normal (**Figura 17A**). Não houve diferença significativa na expressão dos transcritos de TSPAN8 em relação aos diferentes subtipos de câncer de mama (**Figura 17B**). Dados similares foram obtidos em acordo com a classificação molecular PAM50 tanto para GEO (GSE81540, logrank $p < 0.00$) quanto para GOBO (logrank $p < 0.00$) mostram maior expressão dos transcritos de TSPAN8 nos tumores do subtipo HER2+ e menor expressão nos tumores do subtipo Luminal B (**Figuras suplementares 11A e D**). Além disso, foi avaliada a expressão por status de receptor de estrogênio, onde foi achado que os níveis elevados de TSPAN8 estão significativamente relacionados com pacientes apresentando negatividade para ER (GSE81540, logrank $p < 0.001$; GOBO, logrank $p < 0.394$) (**Figuras suplementares 11B e E**). Segundo a classificação molecular a partir do subtipo triplo negativo, TSPAN8 apresentou maior expressão no subtipo LAR (**Figura 17C**).

4.7.2. Análise de sobrevida em pacientes com câncer de mama triplo negativo, estratificados pela expressão de TSPAN8

Em termos gerais, a baixa expressão dos transcritos TSPAN8 esteve associada a um prognóstico desfavorável quando avaliado para sobrevida global (logrank $p = 0.001$) e livre de recorrência (logrank $p = 3.1e-10$) para todos os subtipos de câncer de mama. (**Figuras suplementares 9A e B**). Níveis reduzidos de expressão do TSPAN8 em pacientes com tumores de mama triplo negativo foram relacionados com sobrevida global e livre de recorrência (logrank $p = 0.03$; logrank $p = 0.00014$) curta segundo a plataforma KMPlotter (**Figura 18A, Figura suplementar 10H**). Em simultâneo, resultados significativos foram achados na base de dados TCGA-TNBC (logrank $p = 0.049$) e GEO (logrank $p = 0.0045$) onde a baixa expressão do TSPAN8 esteve associada com um prognóstico desfavorável para sobrevida geral. (**Figuras 18B e C**).

Figura 17. Perfil de expressão do transcrito TSPAN8 em tumores mamários. A.- amostras tumorais de câncer de mama e de tecidos normais. B.- diferentes subtipos de câncer de mama (Luminal, HER2+ e Triplo negativo) C.- subtipos moleculares para triplo negativo. *P-value* para testes T-test e ANOVA <0,05=*; <0,005 = **; <0,0005 = ***.

Figura 19. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de TSPAN8 em três datasets diferentes. A.- KMPlotTER. B.-TCGA. C.-GSE81540. O melhor ponto de corte foi empregado para todas as análises.

Figura 18. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao TSPAN8. miRNAs preditos por miRTarBase A.-hsa-miR-125b-5p; por miRWalk: B.- hsa-miR-330-3p. Dados obtidos da plataforma TCGA.

4.7.3. Análise do impacto de microRNAs tendo como alvo o transcrito TSPAN8 na sobrevida de pacientes com câncer de mama triplo negativo.

A plataforma miRTarBase identificou apenas 01 microRNA apresentando baixa expressão associada a pior prognóstico: hsa-miR-125b-5p ($p = 0.01$); (**Figura 19A**). No entanto, a base de dados mirWALK, utilizando o máximo número de algoritmos de predição (09), previu 01 microRNA: hsa-miR-330-3p ($p = 0.083$) alvejando a região 3' UTR do TSPAN8, cuja alta expressão esteve associada com uma sobrevida global curta em pacientes triplo negativo (**Figura 19B**).

4.7.4. Análise de metilação e alterações genéticas no gene TSPAN8 em pacientes com câncer de mama

Encontrou-se uma correlação inversa entre o nível de metilação e expressão do transcrito TSPAN8 para todos os pacientes com câncer de mama ($Pearson = -0,26$); (**Figura 20A**). A frequência de alterações a partir da base de dados do TCGA provisional, relatou apenas uma porcentagem de 2.9% de alterações como 'amplificações', entretanto não foi encontrado nenhum tipo de mutações somáticas. De maneira semelhante, para a base de dados METABRIC a porcentagem de amplificações foi estimada em 2.3% (**Figuras 5B e A**). Alterações no número de cópia (CNA) também foram investigadas para o gene TSPAN8 usando conjunto de dados do estudo METABRIC com 2173 pacientes de câncer de mama o qual mostrou que há aumento na expressão de mRNA apenas quando há ganhos, deleções homozigotas (*deep deletions*), e amplificações a respeito de deleções heterozigotas (*Shallows deletions*) (**Figuras 20B e C**).

4.7.5. Análise de curvas ROC do transcrito TSPAN8 em pacientes TNBC tratados com quimioterapia

Na curva característica do operador receptor (ROC), TSPAN8 mostrou estar levemente mais expresso em pacientes que responderam a quimioterapia a respeito daqueles que não responderam. Entretanto, a curva abaixo da área não apresentou uma tendência a identificar melhor as pacientes responsivas $AUC = 0.503$ (**Figuras 21A e B**).

Figura 21. Alterações genéticas e perfil de metilação do TSPAN8 A.- Correlação entre expressão e metilação. Perfis de expressão versus alterações no numero de cópias CNA a partir de: B.-TCGA. C.-METABRIC. Dados obtidos da plataforma CBioportal.

Figura 20. Representação da expressão de TSPAN8 associada à resposta a quimioterapia. A.- boxplots entre pacientes responsivos versus não responsivos; B.- Curva ROC exibindo AUC. Relação sensibilidade/especificidade, dados obtidos da plataforma ROCPLLOT.

4.8. RIMS2

4.8.1. Análise de perfis de expressão dos transcritos RIMS2 em pacientes com câncer mamário

Segundo dados TCGA, níveis de expressão de *RIMS2* estiveram downregulados nas amostras de tumor de mama quando comparado com os controles (TCGA, $p < 0.0001$) (**Figura 22A**). Nos perfis de expressão molecular, embora não significativo, os transcritos de RIMS2 apresentaram menor expressão nos tumores do subtipo luminal, quando comparado aos subtipos HER2+ e TNBC (**Figura 22B**). No entanto, foram obtidos dados discordantes segundo a classificação molecular PAM50 GEO (GSE81540, logrank $p < 0.001$) e GOBO (logrank $p < 0.001$) (**Figuras suplementares 14A e D**). Ademais, foi avaliada a expressão por status de receptor de estrogênio, onde foi relatado que os níveis elevados de TSPAN8 estão significativamente relacionados em pacientes com tumores apresentando negatividade par ER (GSE81540, logrank $p < 0.001$) e GOBO (logrank $p = 0.394$) (**Figuras suplementares 14B e E**). Finalmente, quando avaliada a classificação molecular para triplo negativo, RIMS2 esteve mais distribuído nos subtipos BL1, IM e LAR (**Figura 22C**).

4.8.2. Análises de sobrevida em pacientes com câncer mamário triplo negativo, estratificados pela expressão de RIMS2

Em termos gerais a alta expressão dos transcritos de RIMS2 esteve associada a um prognóstico desfavorável quando avaliado para sobrevida global e livre de recorrência (logrank $p = 0.045$; logrank $p = 0.012$) para todos os subtipos de câncer de mama (**Figuras suplementares 12A e B**). Entretanto, quando avaliados apenas os pacientes com tumores de mama do subtipo triplo negativo, níveis reduzidos de RIMS2 foram relacionados com sobrevida global e livre de recorrência (logrank $p = 0.021$; logrank $p = 0.046$) curta para dados depositados na plataforma KMPlotter ((**Figuras 23A, Figura suplementar 13H**)). Da mesma maneira, resultados significativos foram achados na base de dados TCGA para subtipo TNBC onde a baixa expressão do RIMS2 esteve associada com um prognóstico desfavorável (logrank $p < 0.001$). Por outro lado, no dataset GSE81540, baixos níveis de expressão do gene RIMS2 foram correlacionados com diminuição da sobrevida global em pacientes do subtipo BASAL (logrank $p = 0.0045$) (**Figuras 23B e C**).

Figura 22. Perfil de expressão do transcrito RIMS2 em tumores mamários. A.- amostras tumorais de câncer de mama e de tecidos normais. B.- diferentes subtipos de câncer de mama (Luminal, HER2+ e Triplo negativo). C.- subtipos moleculares para triplo negativo. *P-value* para testes T-test e ANOVA <0,05=*, <0,005 = **, <0,0005 = ***.

Figura 23. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de RIMS2 em três datasets diferentes. A.- KM PLOTTER. B.-TCGA. C.-GSE81540. O melhor ponto de corte foi empregado para todas as análises.

4.8.3. Análise do impacto de microRNA tendo como alvo o transcrito RIMS2 na sobrevida de pacientes com câncer de mama triplo negativo

A plataforma miRTarBase identificou apenas 01 microRNA apresentando baixa expressão com tendência a estar associado a pior prognóstico: hsa-miR-1-3p ($p = 0.16$); **Figura 24A**. No entanto, a base de dados mirWALK, utilizando o máximo número de algoritmos de predição (09), previu 29 microRNAs, sendo que apenas 05 apresentaram associações com o prognóstico das pacientes: hsa-miR-142-5p ($p = 0.066$), hsa-miR-200c-3p ($p = 0.04$), hsa-miR-429 ($p = 0.035$), hsa-miR-450b-5p ($p = 0.0078$), hsa-miR-539-5p ($p = 0.035$) todos tendo como alvo a região 3'UTR do RIMS2, sendo que a baixa expressão de todos eles esteve associada com uma sobrevida global curta em pacientes triplo-negativo (**Figuras 24B-F**).

Figura 24. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo o RIMS2. miRNAs preditos por miRTarBase A.- hsa-miR-1-3p, p = 0.16, por miRWalk: B.- hsa-miR-142-5p, C.- hsa-miR-200c-3p, D.- hsa-miR-429, E.- hsa-miR-450b-5p, F.- hsa-miR-539-5p. Dados obtidos da plataforma TCGA.

4.8.4. Análise de metilação e alterações genicas no gene RIMS2 em pacientes com câncer de mama

Encontrou-se uma correlação inversa entre o nível de metilação e expressão do transcrito RIMS2 para todos os pacientes com câncer de mama (Pearson = -0,23) (**Figura 25A**). A frequência de alterações a partir da base de dados do TCGA provisional, mostrou uma porcentagem de 18% (n=178) de alterações entre ‘amplificações’ (n=166), e ‘mutações’ (n=12), sendo 11 missense (L854F, I919T, R587C, T712A, V72I, R825Q, S79L, S316N,

C949F, W810G, F846L) e 01 truncada (E823). Para a base de dados METABRIC a porcentagem de amplificações foi estimada em 22% (**Figuras 5B e A**). Alterações no número de cópia (CNA) também foram investigadas para o gene RIMS2 usando conjunto de dados do estudo METABRIC o qual mostrou que há aumento na expressão de mRNA apenas quando há amplificações, ganhos, deleções heterozigotas (*Shallow deletions*) a respeito de deleções homozigotas (*deep deletions*) (**Figura 25B**). Resultados semelhantes foram encontrados para a base de dados TCGA (**Figura 25C**).

4.8.5. Análise de curvas ROC dos transcritos RIMS2 em pacientes TNBC tratados com quimioterapia

Na curva característica do operador receptor (ROC), RIMS2 mostrou estar mais expresso em pacientes que responderam a quimioterapia a respeito daqueles que não responderam. Entretanto, a curva abaixo da área não apresentou uma tendência a identificar melhor as pacientes responsivas (**Figuras 26A e B**).

Figura 25. Alterações genéticas e perfil de metilação do RIMS2 A.- Correlação entre expressão e metilação. Perfis de expressão versus alterações no numero de cópias CNA a partir de: B.-TCGA. C.-METABRIC. Dados obtidos da plataforma CBioportal.

Figura 26. Representação da expressão de RIMS2 associada à resposta a quimioterapia. A.- boxplots entre pacientes responsivos versus não responsivos; B.- Curva ROC exibindo AUC. Relação sensibilidade/especificidade, dados obtidos da plataforma ROCPLLOT.

4.9. KLHL13

4.9.1. Análise de perfis de expressão dos transcritos KLHL13 em pacientes com câncer de mama

Nos dados TCGA, os transcritos KLHL13 encontram-se downregulados nas amostras de tumor de mama quando comparado com amostras de tecido mamário normal ($p < 0.0005$); (**Figura 27A**). Nos subtipos moleculares, a expressão de KLHL13 foi baixa nos tumores dos subtipos luminais em comparação com os tumores dos subtipos HER2+ e TNBC (**Figura 27B**). Dados semelhantes foram obtidos para expressão KLHL13 de acordo com a classificação molecular PAM50 para GEO (GSE81540, logrank $p = 0.001$); (**Figura suplementar 17A**). Além disso, foi avaliada a expressão por status de receptor de estrogênio, onde foi achado que os níveis elevados de KLHL13 estão significativamente relacionados com pacientes apresentando negatividade par ER (GSE81540, logrank $p = 0.0050$) (**Figura suplementar 17B**). Quando avaliada a classificação molecular para triplo negativo, KLHL13 foi mais abundante nos subtipos MSL, BL1 e BL2 (**Figura 27C**).

4.9.2. Análise de sobrevida em pacientes com câncer mamário triplo negativo, estratificados pela expressão de KLHL13

Em termos gerais a alta expressão do transcrito esteve associada a um prognóstico desfavorável quando avaliado para sobrevida global e livre de recorrência para todos os subtipos de câncer de mama (**Figuras suplementares 15A e B**). Altos níveis de expressão do KLHL13 em pacientes com tumores de mama do subtipo triplo negativo foram relacionados com sobrevida global e livre de recorrência (logrank $p = 5.5 \times 10^{-5}$; logrank $p = 3.3 \times 10^{-5}$) curta para dados depositados na plataforma KMPlotter (**Figuras 28A, Figura suplementar 16H**). Da mesma maneira resultados semelhantes embora não significativos foram achados na base de dados TCGA para subtipo TNBC onde a alta expressão do KLHL13 esteve associada com um prognóstico desfavorável (logrank $p = 0.058$). Por outro lado, no dataset GSE81540, baixos níveis de expressão do gene KLHL13 foram correlacionados com diminuição da sobrevida

global em pacientes do subtipo BASAL (logrank $p=0.1269$) (**Figuras 28B e C**).

Figura 27. Sobrevida global em pacientes do subtipo Basal-like/Triplo negativo estratificada pela expressão de KLHL13 em três datasets diferentes. A.- KMPlotter. B.-TCGA. C.-GSE81540. O melhor ponto de corte foi empregado para todas as análises.

4.9.3. Análise do impacto de microRNAs tendo como alvo o transcrito KLHL13 na sobrevida de pacientes com câncer de mama triplo negativo.

A plataforma miRTarBase identificou apenas 01 microRNA apresentando baixa expressão com tendência a estar associada a pior prognóstico: hsa-miR-128-5p ($p = 0.11$); (**Figura 29A**). No entanto a base de dados mirWALK, utilizando o máximo número de algoritmos de predição (09), previu 13 microRNA, sendo que a expressão de 4 deles exibiram associações estatisticamente significativas com prognóstico: hsa-let-7e-5p ($p = 0.00097$), hsa-let-7c-5p ($p = 0.0062$), alvejando a região 3' UTR do KLHL13, cuja alta expressão de todos eles esteve associada com um sobrevida global curta em pacientes triplo negativo (**Figuras 29B-E**).

Figura 28. Perfil de expressão do transcrito KLHL13 em tumores mamários A.- amostras tumorais de câncer de mama e de tecidos normais. B.- diferentes subtipos de câncer de mama (Luminal, HER2+ e Triplo negativo) C.- subtipos moleculares para triplo negativo. *P-value* para testes T-test e ANOVA <0,05=*; <0,005 = **; <0,0005 = ***.

Figura 29. Sobrevida global em pacientes TNBC estratificados pela expressão de miRNAs tendo como alvo ao KLHL13. miRNAs preditos por miRTarBase A.- hsa-miR-128-5p; por miRWalk: B.- hsa-miR-7e-5p, C.- hsa-miR-7c-5p, D.- hsa-miR-7c-5i, E.- hsa-miR-625-5p. Dados obtidos da plataforma TCGA.

4.9.4. Análise de metilação e alterações genéticas no gene KLHL13 em pacientes com câncer de mama

Não foi relatado nível de metilação e expressão dos transcritos de KLHL13, para pacientes com câncer de mama. A frequência de alterações a partir da base de dados do TCGA provisional, relatou apenas uma porcentagem de 1.7% de alterações em 16 de 963 pacientes como ‘amplificações’, ‘deleções’ e ‘mutações’ do tipo missense (*A519V*, *E210K*, *D162G*, *P510R*, *F119L*, *T568P*). Para a base de dados METABRIC essas alterações foram encontradas apenas 19 pacientes

(~1%); (**Figuras 5B e A**). Alterações no número de cópia (CNA) também foram investigadas usando conjunto de dados do estudo METABRIC, que mostrou que há aumento na expressão de mRNA apenas quando há ganhos e ampliações a respeito de deleções homozigotas (*deep deletions*), deleções heterozigotas (*Shallow deletions*), da mesma maneira, padrões semelhantes para cada tipo de alteração foram encontrados para a base de dados TCGA (**Figuras 30A e B**).

Adicionalmente, foi possível estabelecer correlações entre a expressão de KLHL13 e cinco genes que cumpriram nossos requerimentos: CACHD1, spearman: 0.56; DMD, spearman: 0.52; PAK5, spearman: 0.56; SYNM, spearman: 0.52; SFRP1, spearman: 0.53 (**Figuras suplementares 18A-E**).

Figura 30. Perfil de metilação do KLHL13 em duas bases de dados A.- Correlação entre expressão e alterações no numero de copias CNA a partir de: A.-TCGA. B.-METABRIC. Dados obtidos da plataforma CBioportal.

4.9.5. Análise de curvas ROC do transcrito KLHL13 em pacientes TNBC tratados com quimioterapia

Na curva característica do operador receptor (ROC), KLHL13 mostrou estar mais expresso em pacientes que não responderam a quimioterapia a respeito daqueles que responderam. Entretanto, a curva abaixo da área (AUC) não apresentou uma tendência a identificar melhor as pacientes responsivas (**Figuras 31A e B**).

Figura 31. Representação da expressão de KLHL13 associada à resposta a quimioterapia. A.- boxplots entre pacientes responsivos versus não responsivos; B.- Curva ROC exibindo AUC. Relação sensibilidade/especificidade, dados obtidos da plataforma ROCPLLOT.

5. DISCUSSÃO

5.1. MESP1

O primeiro dos genes a ser descrito é *Mesp1* (*Mesoderm posterior-1*), pois apresentou uma correlação significativa clínica e fenotípica em pacientes com câncer de mama, com destaque para pacientes com tumores do subtipo basal/triplo negativo.

De acordo com os nossos dados, níveis elevados do mRNA de MESP1 indicaram significativamente que o excesso de transcrito em pacientes com tumores basais ou triplo negativo estaria associado com curta sobrevida global e livre de recorrência para dados depositados em duas bases. Por outro lado, os padrões de expressão para os níveis de mRNA indicam que uma abundância maior do MESP1 poderia estar associada com um comportamento agressivo em tumores pouco diferenciados, o que seria confirmado pela presença do transcrito em subtipos moleculares TBNC de origem basal: como BL1 (Basal-Like1), BL2 (Basal-like2) e também no subtipo MSL (Mesenchymal-like), MESP1 poderia estar envolvido em processos de transição epitelial mesenquimal (EMT) e angiogênese. Mesmo quando o nível de metilação indicou uma correlação inversa a respeito da expressão, MESP1 apresentou uma correlação significativa direta com o seu ortólogo MESP2, em pacientes com tumores mamários, isso poderia indicar que *Mesp1*, e 2 estariam agindo conjuntamente em diferentes processos cancerígenos.

Outro aspecto importante seria a desregulação do MESP1 por dois *microRNA* em pacientes de mama triplo negativo, a qual poderia significar futuros alvos para terapia. Ademais, de acordo com os nossos achados de curva ROC, MESP1 poderia ser um potencial indicador de resposta a quimioterapia em pacientes com câncer.

Até agora, não tem se reportado estudos descrevendo *Mesp1*, que pertence a família de fatores de transcrição *helix-loop-helix* em nenhum tipo de câncer, no entanto é estabelecido como regulador hierárquico do sistema cardiovascular. De acordo com os estudos de mapeamento do destino celular, o *Mesp1* aparece pela primeira vez no mesoderma nascente e extraembrionário na faixa primitiva (SAGA et al., 1996). Posteriormente, sua expressão foi detectada nos progenitores comuns do miocárdio, endocárdio e sistema vascular sendo agora aceito como o primeiro marcador molecular para linhagens cardíacas (BONDUE et al., 2008; BONDUE; BLANPAIN, 2010; BRADE et al., 2013; DAVID et al., 2008; LINDSLEY et al., 2008).

Os camundongos nocauteados para *Mesp1* exibem uma anormalidade morfológica do coração, cárdia bífida, devido à migração tardia das células do progenitor cardíaco. Entre

tanto, a inibição inicial da migração celular é resgatada pela expressão dos alelos ortólogos do *Mesp1/2*. Contudo, os embriões de duplo nocaute para *Mesp1* e *Mesp2* morrem sem desenvolver o coração. O acúmulo anormal de células é observado na sequência primitiva de embriões de duplo nocaute porque as células mesodérmicas não podem migrar desde sequência primitiva. Esses dados indicam que *Mesp1* e *Mesp2* agem em conjunto durante a grutulação precoce e são essenciais para o desenvolvimento do mesoderma (KITAJIMA et al., 2000; SAGA; KITAJIMA; MIYAGAWA-TOMITA, 2000).

O *MESP1* é necessário para a saída do estado pluripotente e a indução do programa de expressão gênica cardiovascular (LESCROART et al., 2018). A transcrição da *MESP1* é amplamente controlada pela via WNT canônica e fatores de transcrição como *BRAQUYURY* e *EOMES* (COSTELLO et al., 2011; DAVID et al., 2011) que conduz com sucesso a formação de mesendoderme e subsequente diferenciação cardiovascular (SOIBAM et al., 2015). Estudos mostraram que o *MESP1* acelera a expressão de marcadores vasculares endoteliais como *CD31*, *CD34* e *KDR* (também conhecido como Receptor 2 do Fator de Crescimento Endotelial Vascular 2; *VEGFR2* ou *FLK1*) (BONDUE et al., 2008; DEN HARTOGH et al., 2016; WU, 2008), porém ele poderia desempenhar um papel na angiogênese tumoral devido que tais marcadores estão associados com a via Hedgehog em câncer de mama (DI MAURO et al., 2017).

Contudo, *MESP1* regula fatores de transcrição importantes para EMT e o comprometimento cardiovascular, incluindo *SNAI1*, *SIP1*, *SLUG* e *TWIST1*. Esses mesmos fatores de transcrição participam na EMT em câncer reprimindo promotores de marcadores epiteliais e ativando marcadores mesenquimais, resultando em alterações morfológicas como invasão e migração celular (KELLEHER; FENNELLY; RAFFERTY, 2006; LINDSLEY et al., 2008). A capacidade altamente invasiva quanto transcricional de *Mesp1*, e a participação em vias regulatórias como Notch, Hedgehog e WNT corroboram a hipótese que essa proteína possa estar associada a processos de EMT em tumores mamários agressivos como os do subtipo triplo-negativo.

5.2. ZNF556

O gene ZNF556 (*Zinc finger protein 556*) codifica um membro da família de fatores de transcrição (FT) com motivos de dedos de zinco. Ele pertence à família das proteínas dos dedos de zinco do tipo Krüppel, cujo motivo canônico (C2H2) compreende 28 a 30 resíduos de aminoácidos e sua estrutura é estabilizada por um íon de zinco coordenado por quatro resíduos altamente conservados, duas cisteínas e duas histidinas. A estrutura dobrada de forma estável consiste em duas a três fitas beta e uma hélice alfa. A hélice alfa medeia a ligação ao DNA através de interações não covalentes entre três de seus resíduos de aminoácidos e três bases adjacentes dentro da curvatura principal do DNA. Além disso, ZNF556 possui o domínio KRAB (Krüppel-associated Box), localizado no extremo amino terminal, o qual medeia interação proteína-proteína e pode controlar a localização subcelular, a união ao DNA, e a expressão gênica regulando a união seletiva de fatores de transcrição entre si ou com outros componentes celulares (CORSINOTTI et al., 2013; ECCO et al., 2016; EMERSON; THOMAS, 2009; JEN; WANG, 2016).

Atualmente dados na literatura são escassos, pois até agora não há estudos descrevendo ZNF556 em nenhum processo biológico seja normal ou patogênico. Não entanto, ZNF556 está localizado na região 19p13.3, uma região no cromossomo humano conhecida como locus de susceptibilidade ao câncer colorretal (TROJAN et al., 2000), de mama (OESTERREICH et al., 2001; YANG et al., 2004), de ovário (WANG et al., 1999), e mais recentemente Linfoma de Hodgkin (COZEN et al., 2014).

Segundo os nossos achados, encontramos que o nível transcrito ZNF556 esteve reduzido a respeito das contrapartes normais. Dados similares para o gene ZNF340 em linhagens mamarias corroboram nossos dados (LAN et al., 2016). Tal desregulação poderia ser resultado de mecanismos epigenéticos devido que ZNF556 apresentou uma correlação inversa entre níveis de metilação e expressão. Isso está em acordo com outros trabalhos da literatura, como o trabalho realizado por Farman e colaboradores (2018), que avaliaram o nível de metilação de E2F4 - um fator de transcrição envolvido no controle do ciclo celular - em 100 pacientes com câncer de mama. Eles relataram uma correlação inversa $r = -33$ sendo que quando E2F4 estava mais metilado era menos expresso, embora tivesse havido um aumento do transcrito de duas vezes em comparação a seus respectivos controles $p = 0.022$ (FARMAN et al., 2018).

Por outro lado, embora não significativa, a alta distribuição do ZNF556 no subtipo Basal e nos tumores ER- estaria sugerindo um fenótipo agressivo em tumores pouco diferenciados o

que pode ser refletido em acordo com quatro subtipos da classificação molecular para o subtipo triplo negativo: M (Mesenquimal) e MSL (Mesenquimal associado à tronccidade) os quais participam na transição epitélio mesenquimal e BL1-2 (*Basal-like* 1 e 2) os quais estão envolvidos na regulação do ciclo celular. A respeito disso, Yang e colaboradores confirmaram que ZKSCAN3, um FT membro da família Krüppel-C2H2, ativa transcricionalmente a integrina- β 4 e o fator de crescimento endotelial vascular, ambos envolvidos na tumorigênese colorretal (YANG et al., 2008). Posteriormente, os pesquisadores determinaram que a sobre-expressão de ZKSCAN3, promove a proliferação a traves de ativação transcricional da ciclina D2 em mieloma múltiplo (YANG et al., 2011). De fato, devido à presença do domínio KRAB, que exerce repressão transcricional gene-específica com interação de fatores remodeladores da cromatina, a maioria dos membros da família KRAB-ZFPs tem um papel na regulação do desenvolvimento embrionário, diferenciação celular, proliferação celular, apoptose, transformação neoplásica e regulação do ciclo celular (URRUTIA, 2003).

Em termos de sobrevida, nossos achados mostram uma forte correlação entre o excesso de transcrito e a sobrevida curta de pacientes em geral e do subtipo basal/triplo negativo confirmado por três *datasets*. Resultados semelhantes foram obtidos a partir do trabalho *in silico* realizado por Machnik e colegas, no qual a alta expressão de 05 TFs do tipo KRAB-domain, (ZNF273, ZNF320, ZNF485, ZNF525, ZNF643) esteve associada a um prognóstico desfavorável em pacientes com tumores mamários (MACHNIK et al., 2019). Quanto ao subtipo triplo negativo, no estudo de Liu e colaboradores, foram obtidos resultados discordantes dado que a baixa expressão de GGNBP2, uma proteína do tipo dedo de zinco com apenas um motivo C2H2, esteve correlacionada com um prognostico desfavorável em 138 pacientes TNBC ($p=0.029$), adicionalmente tal expressão mostrou ser um fator independente entre ambos os grupos GGNBP2Baixa versus GGNBP2Alta; HR= 5.4 (2.73-10.6) (LIU et al., 2019). Uma possível explicação que daria suporte ao efeito do ZNF556 em pacientes TNBC, poderia ser na regulação através do microRNA ‘hsa-miR-214-5p’ o qual demonstrou um comportamento repressivo na sobrevida global para os mesmos pacientes aonde ZNF556 foi avaliado. Tal microRNA, se liga na região 3’ UTR do gene alvo e mesmo sem evidencia experimental em câncer mamário, está relacionado a processos subjacentes a várias doenças do tipo neoplásicas, cardiomiopaticas e metabólicas (CHI et al., 2009; KARGINOV; HANNON, 2013).

Interessantemente, quando avaliada a sobrevida livre de recidiva encontrou-se que é a falta do transcrito o que estaria representando um risco de recorrência em pacientes com tumores não apenas do tipo basal, mas também de todos os subtipos moleculares. Sendo que RFS está

fortemente relacionado com o tipo tumoral, grau histológico, cirurgia e tratamento. Nesse último aspecto, nosso gene, poderia ser um possível preditor de resposta associada a quimioterapia, pois mostrou que poderia diferenciar pacientes que sobreviveram por mais de cinco anos versus aqueles que tiveram recidiva depois cinco anos após o início da terapia.

O papel funcional das proteínas KRAB-C2H2-ZF depende da organização dos diferentes módulos de proteína, isto é, o domínio repressor KRAB localizado no resíduo amino-terminal NH₂ e os domínios de ligação ao DNA presentes na região carboxilo-terminal COOH. Tal organização poderia explicar a diferença na expressão gênica, pois variações nas estruturas dos elementos reguladores, através de duplicações, amplificações ou deleções dos domínios do dedo de zinco, contribuem para a diversificação funcional. Nos nossos achados, a quantidade de alterações genéticas encontrada nas nossas coortes foi mínima de maneira que não é possível concluir que as mutações estariam causando a desregulação da proteína no câncer mamário.

Finalmente, outra característica das proteínas com motivos dedos de zinco e caixa de Kruppel é sua capacidade de não apenas ligar DNA, mas também RNA. Como relatado segundo nossa análise de interação molecular de redes, ZNF556 foi vinculado a uma associação com TRMT12 uma metiltransferase envolvida na metilação e repressão do DNA, sendo que já foi relatada estar amplificada em câncer mamário (RODRIGUEZ et al., 2007). Ao mesmo tempo ZNF556 estabeleceu interação com ELAVL1, uma proteína de ligação ao RNA que se liga à região 3'-UTR dos mRNAs e aumenta sua estabilidade. No TNBC, ELAVL1 age estabilizando o mRNA do ZEB1 conferindo assim uma vantagem para a progressão da doença (LUO et al., 2020). Com tudo, o papel de ZNF556 em processos neoplásicos parece ser dependente do contexto celular e da interação com parceiros celulares-específicos.

5.3. TSPAN8

A proteína codificada por esse gene é um membro das tetraspaninas, também conhecidas como proteínas da superfamília transmembranar 4 (TM4SF), elas têm quatro hélices alfa transmembranares e dois domínios extracelulares, um curto (chamado domínio ou laço extracelular pequeno, SED / SEL ou EC1) e um mais longo, tipicamente de 100 resíduos de aminoácidos (chamado domínio extracelular grande / alça, LED / LEL ou EC2). Embora várias famílias de proteínas tenham quatro hélices alfa transmembranares, as tetraspaninas são definidas por sequências de aminoácidos conservadas, incluindo quatro ou mais resíduos de cisteína no domínio EC2, com dois em um motivo 'CCG' altamente conservado. Pensa-se que as tetraspaninas atuam como proteínas de andaimes, ancorando várias proteínas em uma área da membrana celular (HEMLER, 2005, 2014).

Existem 34 tetraspaninas em mamíferos, 33 das quais também foram identificadas em seres humanos. Elas agem como glicoproteínas de membrana contribuindo para o arranjo de proteínas na membrana celular, organizando os chamados microdomínios enriquecidos com tetraspaninas (tetraspanina em rede). Ao organizar tais complexos multimoleculares de membrana estas proteínas medeiam os eventos de transdução de sinal desempenhando um papel na regulação do desenvolvimento, ativação, crescimento e motilidade celular (CHARRIN et al., 2014; HEMLER, 2005; ZÖLLER, 2009).

TSPAN8 codifica uma glicoproteína da superfície celular que é conhecida por formar complexos com integrinas (HEILER; WANG; ZÖLLER, 2016) e junto a outras tetraspaninas humanas como Tspan12, CD9, CD37, CD63, CD81, CD82 e CD151 desempenham um papel na progressão do câncer (HEMLER, 2014; ZÖLLER, 2009). A desregulação dessas tetraspaninas nos tumores tem sido correlacionada com um prognóstico bom ou ruim em diferentes tipos de câncer. Historicamente, várias tetraspaninas foram identificadas por estudos que compararam tecidos normais e malignos. Tspan8 foi originalmente identificado como um antígeno associado ao tumor por um anticorpo (CO-029) (SZALA et al., 1990). Evidências crescentes sugerem que o TSPAN8 promove a migração, invasão e metástase de células tumorais em vários tipos de câncer humano, incluindo câncer colorretal, ovariano e gástrico, hepatocarcinoma, adenocarcinoma pancreático e glioma (GESIERICH et al., 2005; PAN et al., 2015; PARK et al., 2016; VOGLSTAETTER et al., 2019).

A respeito de nossos resultados, embora não significativos TSPAN8 exibiu uma diminuição do transcrito primário nas amostras tumorais em comparação com tecido normal. Isto evidencia contraste com o trabalho do Voglstaetter e colaboradores que avaliaram a expressão

de mRNA de TSPAN8 por RT-qPCR em 15 amostras de câncer mamário sendo 01 tecido normal, 07 tumores primários e 07 metástases de pulmão e linfonodos, relatando uma expressão aumentada nos espécimes metastáticos, intermediária em tumores primários e mínima no tecido normal (VOGLSTAETTER et al., 2019). Adicionalmente, isto também foi dissemelhante para câncer colorretal e para de ovário, dado que foram encontrados níveis elevados de transcrito primário e proteína nas amostras tumorais em comparação com os respectivos controles (PARK et al., 2016; ZHANG et al., 2020).

No entanto TSPAN8 apresentou-se mais distribuído no subtipo HER2+ e no grupo dos pacientes cujos tumores apresentaram negatividade para ER. Por outro lado, a diminuição da expressão de TSPAN8 exibiu um impacto desfavorável na sobrevida nos pacientes em termos gerais e nos pacientes com subtipos triplo negativo, luminal A e luminal B. Isto difere significativamente do trabalho de Zhu e colegas que relataram uma alta expressão do TSPAN8 em pacientes com subtipo triplo negativo a respeito de luminais ($p < 0.05$) e Her2+ ($p = \text{NS}$). Ademais, de maneira geral quando a sobrevida foi avaliada, um aumento da expressão do TSPAN8 esteve correlacionado com uma sobrevida global curta ($n=90$; $p=0.023$) (ZHU et al., 2019). Da mesma maneira, resultados opostos na sobrevida global foram confirmados para câncer de ovário, onde a alta expressão do TSPAN8 esteve associada com pior prognóstico ($n=74$; $p=0.03$) (PARK et al., 2016).

Em nossas análises de interação de redes gênicas, o receptor nuclear AR (receptor de andrógeno) interagiu diretamente com TSPAN8, corroborando os dados obtidos na análise de expressão nos subtipos triplo-negativo, onde TSPAN8 esteve mais expresso no subtipo LAR (*Luminal Androgen Receptor*) em comparação aos outros subtipos. Por tanto isso nós leva a hipotetizar que TSPAN8 poderia estar participando da sinalização do AR, tornando-se um possível alvo para o bloqueio de androgênios e inibidores do PIK3 (OMARINI et al., 2018). Mais outras três proteínas principais interagiram diretamente com TSPAN8, entre elas CTNNB, cuja proteína β -catenina ancora o citoesqueleto de actina e é responsável pela transmissão do sinal de inibição de contato que faz com que as células parem de se dividir quando a camada epitelial estiver completa (MCGINN et al., 2020).

A interação entre TSPAN8 e β -catenina já foi recentemente relatada como um 'loop' na progressão e invasão no melanoma e em promover a troncicidade no câncer colorretal (EL KHARBILI et al., 2019; ZHAN et al., 2019), entretanto em câncer de mama Tspan8 foi identificado num subconjunto de células-tronco mamárias altamente quiescentes (Lgr5+

Tspan8^{hi}), localizado dentro da árvore ductal proximal e cujo transcriptoma se assemelha a carcinoma da mama com baixo nível expressão de claudinas, o que pode refletir em parte o subtipo *basal-like* (FU et al., 2017). Posteriores estudos demonstraram que, em câncer mamário Tspan8 poderia estar atuando na sinalização de E-cadherina/ β -catenina acompanhado pelo aumento de vesículas extracelulares circulantes, regulando assim a proliferação e a adesão célula-célula (VOGLSTAETTER et al., 2019).

Por outro lado, SMARCA4 que exerceu ação sobre BRCA1 e TSPAN8, é uma proteína codificada membro da família de proteínas SWI/SNF os quais têm atividades helicase e ATPase e regulam a transcrição de certos genes alterando a estrutura da cromatina em torno deles, assim BRCA1 e SMARCA4 formam um complexo cuja perda levaria a um estado de desdiferenciação em células progenitoras epiteliais mamarias como causa do dando ao DNA (WANG et al., 2016). Além disso, a sobre-expressão de SMARCA4 foi correlacionada com um prognóstico desfavorável em pacientes com câncer mamário tanto para sobrevida global como livre de recidiva, o que poderia indicar uma possível regulação negativa para TSPAN8 em acordo com nossos achados (DO et al., 2016).

Em quanto a interação com microRNAs, metilação e resposta a drogas os nossos dados não foram os suficientes para obter conclusões consistentes a respeito do papel de Tspan8 no câncer de mama em geral e no triplo negativo, no entanto, reforçamos o papel de TSPAN8 como um ‘downstream gene’ em vias relacionadas a proliferação, autorregulação, troncidade e um possível mediador (mas não um alvo) da resposta a drogas em tumores mamários.

5.4. RIMS2

A proteína codificada pelo gene RIMS2 pertence a um complexo de proteínas que interagem com Rab3 (RIMs, *Rab3 interacting molecules*) na zona ativa pré-sináptica desempenhando uma função essencial na liberação de neurotransmissores (LONART, 2002). No estado normal, RIMS2 age como proteína de andaime e é ativado pela proteína dependente de GTP, RAB3 que por sua vez atua ancorando outras proteínas como Munc13, ELKS, proteínas de ligação à RIM (RIM-BPs), α -liprinas e sinaptotagminas-1, para assim direcionar as vesículas sinápticas para aos canais de cálcio (CLARKE; CHEN; NISHIMUNE, 2012; DRESBACH et al., 2001).

A estrutura do RIMS2 contém um domínio FYVE de dedo Zn^{2+} na região N-terminal, um domínio PDZ e dois domínios C2. O domínio FYVE demonstrou ligar dois íons de zinco e possui oito posições potenciais de cisteínas coordenadoras de zinco. Domínios FYVE se ligam especificamente ao fosfatidilinositol 3-fosfato e são frequentemente encontrados em proteínas direcionadas às membranas lipídicas envolvidas na regulação do tráfego de membranas (SANKARAN et al., 2001; TIBBETTS et al., 2004). Domínios PDZ desempenham um papel fundamental na ancoragem de proteínas receptoras na membrana para componentes citoesqueléticos. As proteínas com esses domínios ajudam a unir e organizar complexos de sinalização nas membranas celulares. Esses domínios desempenham um papel fundamental na formação e função de complexos de transdução de sinal (KOUSHIKA et al., 2001). Finalmente, os domínios C2 são módulos de membrana dependentes de Ca^{2+} envolvidos na transdução de sinal e tráfego de membranas (FARAH; SOSSIN, 2012).

Segundo nossos achados, o nível de expressão de RIMS2 foi significativamente downregulado a respeito dos controles, e apresentou uma maior distribuição no subtipo triplo negativo e nos tumores negativos para receptor de estrogênio, que foi confirmado por mais duas bases de dados. Dado que o subgrupo TBNC carece de receptores ER\PR e HER2, e a ausência ou baixa expressão de ER é bem conhecida por ser frequente em tecidos indiferenciados, essas características tornam tais tumores resistentes à terapia hormonal, ademais padrões de expressão do RIMS2 para subtipos BL1, BL2 e padrões de metilação podem implicar que a presença de moléculas RIMS2 estaria associada a um comportamento agressivo em tumores triplo negativos (KUMAR; AGGARWAL, 2016; MAYER et al., 2014). Ademais, altos níveis de RIMS2 foram encontrados em subtipos IM, LAR dando a

entender que poderia estar envolvido com processos subjacentes a indução transcricional de genes envolvidos na inflamação, resposta imune, resposta de dano ao DNA, alteração na expressão de marcadores mioepiteliais e fatores de crescimento, mitógenos controladores na regulação do ciclo celular, e sinalização dependente de AR.

Quanto a processos oncológicos, em câncer de mama RIMS2 apenas foi identificada em um único estudo como uma das proteínas de interação com o domínio PDZ de Claudina-2, o qual é necessário para progressão independente de ancoragem e requerido para metástase ao fígado lamentavelmente o grande tamanho da proteína (1349 aa) representou um impedimento para continuar com a pesquisa (TABARIÈS et al., 2019). Assim, isso pode supor uma possível participação do RIMS2 em interações célula-célula na progressão tumoral.

Nossos resultados de interações de redes podem corroborar alguns dos processos nos quais RIMS2 estaria envolvido. RIMS2 interage com FYN uma oncoproteína relacionada à família tirosina quinases SCR que tem sido implicado no controle do crescimento celular, sendo que em tumores mamários de tipo basal, afeta vários processos como movimento celular e a apoptose, transdução de sinal da matriz extracelular e regulação da atividade de integrinas (LEE et al., 2018; MI et al., 2018). MYC também interage de maneira direta com RIMS2. MYC é um proto-oncogene e codifica uma fosfoproteína nuclear que desempenha um papel na progressão do ciclo celular, apoptose e transformação celular. A amplificação desse gene é frequentemente observada em vários cânceres humanos, em TNBC c-Myc já foi demonstrado mediar a troncidade e a transição epitelial mesenquimal (KATSUTA et al., 2019; YIN et al., 2017).

Por outro lado, as análises de sobrevida indicaram que níveis diminuídos do RIMS2 correlacionaram significativamente com uma sobrevida curta para pacientes TNBC, resultados semelhantes foram encontrados por Fernandez-Nogueira e colegas que usando mineração de dados relataram a sobre-expressão de conjunto de 07 genes neurais em câncer mamário tipo basal sendo que a baixa expressão de três de eles (APP, NGRF, ERNB-1) esteve associada com prognóstico desfavorável (FERNÁNDEZ-NOGUEIRA et al., 2016).

Outro dado importante do presente trabalho foi à frequência de amplificações (18%) que está diretamente associada com a expressão de RIMS2, porém quando avaliada a sobrevida global e livre de doença os pacientes que apresentaram a alteração para o gene relataram um pior prognóstico a respeito dos que não (www.cbioportal.org/). Além disso, a porcentagem de mutações somáticas apenas cobriu 1,2%, e segundo o portal de mutações POLYPHEN ([/genetics.bwh.harvard.edu/](http://genetics.bwh.harvard.edu/)), apenas 01 delas aparentou ser possivelmente deletéria.

Quanto à resposta a drogas, a expressão de RIMS2 foi mais alta nas pacientes que não responderam a tratamento quimioterápico em relação às pacientes que responderam. Igualmente, a expressão desse transcrito não apresentou uma predição robusta para diferenciar os dois tipos de pacientes, o qual indica que RIMS2 não seria um gene preditor para uso de quimioterápicos em pacientes TNBC. Não obstante, RIMS2 poderia participar como mediador no bloqueio da sinalização androgênicos (AR) e crescimento celular (MYC).

Evidências sugerem que o próprio sistema nervoso, assim como neurotransmissores e neuropeptídeos presentes no microambiente tumoral, desempenha um papel na orquestração da progressão tumoral. Os tumores induzem a formação de novos vasos sanguíneos (angiogênese) e vasos linfáticos (linfangiogênese) secretando diferentes fatores, dessa maneira eles também podem induzir a formação de novas terminações nervosas (neoneurogênese) secretando fatores neurotróficos e moléculas de orientação axonal (DUMITRU et al., 2018; FERNÁNDEZ-NOGUEIRA et al., 2016; TAN et al., 2020; VAN SWEARINGEN; SIEGEL; ANDERS, 2014).

Sugerimos que o RIMS2 como proteína reguladora do transporte intracelular, poderia ter um papel importante na progressão e metástase do tumor, promovendo o remodelamento da matriz extracelular. Ademais, atuando a nível tanto de amplificação quanto expressão, e por estar envolvido em diferentes rotas ligadas a oncogênese, pode ser um gene candidato associado ao risco de câncer de mama triplo-negativo. No entanto, estudos adicionais, abordando tanto genes da engrenagem neural como quanto genes relacionados diretamente com câncer mamário são necessários para confirmar esses achados.

5.5. KLHL13

KLHL13 (*Kelch Like Family Member 13*) é uma proteína adaptadora que junto com Cullin-3 e outras proteínas que tendo como substrato a AURORA forma o complexo ubiquitina-proteína ligase à base de Cullin3 E3. Esse complexo é necessário para a segregação cromossômica adequada e a conclusão da citocinese. As proteínas do tipo Kelch (KLHL) geralmente possuem um domínio BTB/POZ, um domínio BACK e cinco a seis motivos Kelch. Os domínios BTB facilitam a ligação e dimerização de proteínas. O domínio BACK não tem função conhecida e ainda é de importância funcional, pois mutações nesse domínio estão associadas à doença (STOGIOS; PRIVÉ, 2004). Os domínios Kelch formam uma estrutura terciária de hélices β que têm um papel nas funções extracelulares, morfologia e ligação a outras proteínas (DHANOA et al., 2013; GUPTA; BEGGS, 2014; SHI et al., 2019).

Conforme nossos resultados, níveis diminuídos dos transcritos de KLHL13 foram encontrados em amostras tumorais em comparação com as amostras de tecido normais, e apresentou uma maior distribuição no subtipo triplo negativo e nos tumores negativos para receptor de estrogênio. A alta expressão do KLHL13 no subtipo TN, assim como outros genes no presente estudo, indicaria uma possível associação com desdiferenciação e agressividade a respeito dos subtipos menos agressivos. Ora nossos dados também relataram uma associação significativa entre o subtipo luminal e o TN (KUMAR; AGGARWAL, 2016; MAYER et al., 2014). Resultados similares foram evidenciados por Zhang e colegas na proteína KLHL19 que estava mais expressa nas amostras normais a respeito das tumorais, não obstante os níveis de KLHL7 para *Basal-like* foram encontrados estar significativamente correlacionados com o subtipo HER2+ segundo Kurozumi e colaboradores (KUROSUMI et al., 2018; ZHANG et al., 2016).

Entretanto, níveis de KLHL13 foram encontrados em subtipos TNBC, como BL1(Basal-like 1) e M (Mesenquimal), sugerindo que KLHL13 poderia estar envolvido com processos subjacentes a indução transcricional de genes envolvidos na resposta de dano ao DNA, alteração na expressão de marcadores mioepiteliais e regulação da transição epitelial mesenquimal. Adicionalmente, algumas dessas associações podem ser corroboradas pela nossa análise de redes, pois KLHL13 foi alvo de ESR1, entanto que HNF4A estabeleceu uma interação indireta com KLHL28, uma possível explicação e que KLHL13 esteja interagindo com ESR1 através do seu domínio BTB implicado na formação de dímeros, ou também pode

ser alvo do ER pela ativação transcricional de elementos ERE na região promotora. Por outro lado HNF4 é um fator de transcrição nuclear que está correlacionado com invasão, metástase e transição epitelial mesenquimal (XIANG et al., 2015). Com tudo, os processos envolvendo ubiquitinação influenciam as propriedades semelhantes às células-tronco do câncer no câncer de mama, e as ligases E3 podem ser ativas no crescimento de células-tronco do câncer (GALLO; KO; DONOGHUE, 2017).

A respeito das análises de sobrevida, a pesar de que foram não significativas, a expressão do KLHL13 foi consistente em três bases dados *Basal-like*/Triplo-Negativo. Atualmente, não há dados mostrando o papel de KLHL13 nem em processos patológicos nem em câncer, no entanto outros membros da família de proteínas tipo Kelch, como KEAP-1, cuja alta expressão esteve associada com bom prognóstico em pacientes com câncer mamário, demonstram discordância com os nossos resultados (ZHANG et al., 2016). Complementarmente, o trabalho de Kurozumi e colegas, evidenciaram a alta expressão do transcrito ($p < 0.0001$) e proteína ($p = 0.0025$) de KLHL7, uma proteína do tipo Kelch contendo os mesmos domínios que KLHL13, estar significativamente associada com bom prognóstico em pacientes com câncer mamário, isso apresenta certa semelhança com os nossos resultados (KUROZUMI et al., 2018).

Entre os fatores que podem explicar o comportamento dos níveis de expressão de KLHL13 em acordo com sobrevida das pacientes com subtipos triplo negativo, estão três microRNAs, que como relatado nos nossos achados a alta expressão deles estaria associada com sobrevida global alta. Até agora, não se sabe o papel funcional desses microRNA, no entanto, no presente trabalho eles são relatados como novos possíveis alvos em câncer de mama triplo negativo. Por outro lado, KLHL13 esteve correlacionado de maneira direta com cinco genes nos pacientes com câncer mamário (www.cbioportal.org/): CACHD1 (proteína contendo o domínio do canal Ca^{2+} e receptor de quimiotaxia (cache) - 1), DMD (Distrofina), PACK5 (P21 (RAC1) Quinase Ativada-5), SYNM (Sinemina), SFRP1 (Proteína secretada relacionada à FRIZZLED-1). Esses genes estão altamente envolvidos na manutenção do citoesqueleto, remodelação da matriz extracelular, estresse mecânico e sinalização dependente de GTPases e Tirosina-Quinases. Além disso, a porcentagem de mutações somáticas cobriu apenas 0,7% do total de 963 pacientes, e segundo o portal de mutações POLYPHEN ([/genetics.bwh.harvard.edu/](http://genetics.bwh.harvard.edu/)), duas delas aparentaram ser benignas, e 04 delas foram previstas como ‘provavelmente daninhas’. Entretanto a frequência de ampliações foi 1%

(www.cbioportal.org/) sendo que o aumento do transcrito esteve associado com alterações como ganho e ampliações a respeito de deleções homozigotas. Resultados similares foram obtidos para o gene KLHL7 em câncer mamário (KUROSUMI et al., 2018).

Finalmente, embora KLHL13 não seja um bom preditor de resposta a drogas, sendo que é dependente de AURORA pode ser proposto com um futuro alvo no manejo quimioterápico em pacientes TNBC. Tudo nos leva a hipotetizar um possível envolvimento de KLHL13 em câncer de mama, no entanto são necessários estudos funcionais que possam corroborar nossos achados.

6. CONCLUSÕES

Os avanços da genômica levaram à identificação de cinco subtipos moleculares de câncer de mama com base em padrões de expressão gênica. A partir desse momento, numerosos genes foram encontrados estarem envolvidos na tumorigênese da mama, os quais podem atuar como biomarcadores para o diagnóstico precoce e posterior aplicação clínica. Embora dezenas de genes relacionados tenham sido encontradas. Eles são insuficientes para elucidar a tumorigênese do BC, a menos que genes mais relevantes sejam identificados. Portanto, descobrir novos genes candidatos é uma tarefa extremamente crucial.

É demorado e dispendioso descobrir genes relacionados a doenças apenas por experimentos, devido que o espaço da pesquisa é muito grande. Assim abordagens bioinformáticas são métodos alternativos que podem ajudar os pesquisadores a lidar com vários problemas biológicos, como analisar padrões de expressão e redes biológicas complexas com o intuito de identificar novos genes.

Nosso método foi limitado por subtipo de BC, especificamente TNBC. Ele envolveu uma abordagem extensa analisando detalhadamente cada gene, destacando características importantes em quanto à função biológica, estrutural e níveis de expressão, visando encontrar o impacto dessas características a respeito do valor prognóstico em diferentes cortes de pacientes com TNBC. No entanto, é necessário o uso de recursos computacionais mais robustos para o desenvolvimento de preditores moleculares mais eficazes que superem a variabilidade do subtipo BC, bem como a identificação de novos genes.

Os genes previstos partiram de um conjunto de 100 genes, de um estudo levado a cabo pelo nosso grupo, como resultado da comparação transcriptômica entre pacientes TNBC/SPARC+ vs TNBC/SPARC-. Tais genes atendem características como desdiferenciação, progressão tumoral e metástase, troncidade, desregulação no ciclo celular e na matriz extracelular, as quais refletem o estado de tumores mamários agressivos.

Embora esses achados, com base nos níveis de expressão de mRNA e características clínicas como sobrevida, mostrem um conjunto coerente de alterações associadas ao câncer de mama triplo negativo, deve-se levar em conta que os níveis de expressão de mRNA nem sempre estão diretamente correlacionados com os níveis de expressão de proteínas e que nossos resultados no nível de mRNA podem ser apenas correlativos, em vez de causativos. Portanto, são necessários mais estudos para medir os níveis de expressão de mRNA e as concentrações de proteínas para estabelecer as correlações existentes entre alterações nos dois tipos de moléculas.

Além das informações sobre minerações de dados referentes à expressão e sobrevida neste estudo, as informações de nível epigenético e de mutações também desempenham um papel importante na progressão de BC. Num trabalho futuro, usando conjuntos de dados multidimensionais pretendemos identificar mutações ‘driver’ que tenham efeito direto no crescimento seletivo como vantagem das células BC.

No presente trabalho foi desenvolvido um método utilizando ferramentas *in silico* para predizer genes candidatos a biomarcadores em câncer de mama com base na expressão e no valor prognóstico. Após a aplicação das ferramentas para mineração de dados obtivemos cinco genes relacionados a câncer mamário com ênfase no subtipo triplo negativo. Utilizando esse subconjunto de genes observamos diferenças significativas na sobrevida entre pacientes com câncer mamário em análises de Kaplan-Meier.

Em resumo, os cinco genes candidatos identificados podem ajudar a revelar os mecanismos moleculares subjacentes do BC e fornecer orientação sobre a possível seleção personalizada de regimes terapêuticos. Não escapa a nosso conhecimento de que a abordagem deste estudo pode fornecer também novas informações para um melhor entendimento dos diferentes subtipos de câncer de mama.

7. REFERENCIAS

- AFGHAHI ANOSHEH, TELLI L. MELINDA, K. W. A. Genetics of triple-negative breast cancer: Implications for patient care. **Current Problems in Cancer**, v. 40, n. 2–4, p. 130–140, 1 mar. 2016.
- ATCHLEY, D. P. et al. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 26, n. 26, p. 4282–8, 10 set. 2008.
- BERTUCCI, F. et al. How basal are triple-negative breast cancers? **International Journal of Cancer**, v. 123, n. 1, p. 236–240, 1 jul. 2008.
- BONDUE, A. et al. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. **Cell Stem Cell**, v. 3, n. 1, p. 69–84, 3 jul. 2008.
- BONDUE, A.; BLANPAIN, C. **Mesp1, 2010Circulation Research**American Heart Association, Inc., , 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21148448>>. Acesso em: 22 out. 2019
- BRADY, T. et al. **Embryonic heart progenitors and cardiogenesis**Cold Spring Harbor **Perspectives in Medicine**Cold Spring Harbor Laboratory Press, , 2013.
- BRAY, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, 12 set. 2018.
- CALZA, S. et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. **Breast Cancer Research**, v. 8, n. 4, p. R34, 17 ago. 2006.
- CAMPBELL, L. L.; POLYAK, K. Breast Tumor Heterogeneity: Cancer Stem Cells or Clonal Evolution? **Cell Cycle**, v. 6, n. 19, p. 2332–2338, 27 out. 2007.
- CAO, Y. et al. Light to moderate intake of alcohol, drinking patterns, and risk of cancer: results from two prospective US cohort studies. **BMJ**, v. 351, 2015.
- CECILIO, A. P. et al. Breast cancer in Brazil: epidemiology and treatment challenges. **Breast cancer (Dove Medical Press)**, v. 7, p. 43–9, 2015.
- CHANDRASHEKAR, D. S. et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. **Neoplasia**, v. 19, n. 8, p. 649–658, 1 ago. 2017.
- CHARRIN, S. et al. Tetraspanins at a glance. **Journal of Cell Science**, v. 127, n. 17, p. 3641–3648, 2014.
- CHI, S. W. et al. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.

Nature, v. 460, n. 7254, p. 479–486, 23 jul. 2009.

CLARKE, G. L.; CHEN, J.; NISHIMUNE, H. **Presynaptic active zone density during development and synaptic plasticity** *Frontiers in Molecular Neuroscience* *Frontiers*, , 30 jan. 2012.

COLDITZ, G. A.; BOHLKE, K. Priorities for the primary prevention of breast cancer. **CA: A Cancer Journal for Clinicians**, v. 64, n. 3, p. 186–194, maio 2014.

CORSINOTTI, A. et al. Global and Stage Specific Patterns of Krüppel-Associated-Box Zinc Finger Protein Gene Expression in Murine Early Embryonic Cells. **PLoS ONE**, v. 8, n. 2, p. e56721, 22 fev. 2013.

COSTELLO, I. et al. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. **Nature Cell Biology**, v. 13, n. 9, p. 1084–1091, 7 set. 2011.

COUCH, F. J. et al. Risks of triple negative breast cancer associated with cancer predisposition gene mutations. **Journal of Clinical Oncology**, v. 34, n. 15_suppl, p. 1513, 20 maio 2016.

COZEN, W. et al. A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. **Nature Communications**, v. 5, n. 1, p. 1–10, 12 jun. 2014.

CURTIS, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. **Nature**, v. 486, n. 7403, p. 346–352, 18 jun. 2012.

DAVID, R. et al. MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. **Nature Cell Biology**, v. 10, n. 3, p. 338–345, mar. 2008.

DAVID, R. et al. Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell. **Cardiovascular Research**, v. 92, n. 1, p. 115–122, 1 out. 2011.

DE ALCANTARA FILHO, P. R. et al. Gene expression profiling of triple-negative breast tumors with different expression of secreted protein acidic and cysteine rich (SPARC). **Breast Cancer Management**, p. bmt-2017-0019, 11 set. 2018.

DEN HARTOGH, S. C. et al. A comprehensive gene expression analysis at sequential stages of in vitro cardiac differentiation from isolated MESP1-expressing-mesoderm progenitors. **Scientific Reports**, v. 6, n. 1, p. 19386, 19 maio 2016.

DESANTIS, C. E. et al. International Variation in Female Breast Cancer Incidence and Mortality Rates. **Cancer Epidemiology and Prevention Biomarkers**, v. 24, n. 10, 2015.

DHANOA, B. S. et al. Update on the Kelch-like (KLHL) gene family. **Human Genomics**, v. 7, n. 1, p. 13, 15 maio 2013.

- DI COSIMO, S.; BASELGA, J. Management of breast cancer with targeted agents: importance of heterogeneity. **Nature Reviews Clinical Oncology**, v. 7, n. 3, p. 139–147, 2 mar. 2010.
- DI MAURO, C. et al. Hedgehog signalling pathway orchestrates angiogenesis in triple-negative breast cancers. **British Journal of Cancer**, v. 116, n. 11, p. 1425–1435, 25 maio 2017.
- DIAZ CASAS, S. et al. Clinical Behavior of Triple Negative Breast Cancer in a Cohort of Latin American Women. **Cureus**, 21 jun. 2019.
- DO, S.-I. et al. Increased Brahma-related Gene 1 Expression Predicts Distant Metastasis and Shorter Survival in Patients with Invasive Ductal Carcinoma of the Breast. **Anticancer research**, v. 36, n. 9, p. 4873–82, 1 set. 2016.
- DOS SANTOS, M. et al. Transcriptome characterization of human mammary cell lines expressing different levels of ERBB2 by serial analysis of gene expression. **International Journal of Oncology**, v. 28, n. 6, p. 1441–1461, 1 jun. 2006.
- DOS SANTOS, M. L. et al. Transcriptome changes induced by docetaxel in human mammary cell lines expressing different levels of ERBB2. **International Journal of Molecular Medicine**, v. 23, n. 6, p. 733–743, 5 maio 2009.
- DRESBACH, T. et al. **The presynaptic cytomatrix of brain synapses** Cellular and Molecular Life Sciences Birkhauser Verlag Basel, , 2001.
- DUMITRU, A. et al. Alterations in Calcium Signaling Pathways in Breast Cancer. In: **Calcium and Signal Transduction**. [s.l.] InTech, 2018.
- ECCO, G. et al. Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. **Developmental Cell**, v. 36, n. 6, p. 611–623, 21 mar. 2016.
- EL KHARBILI, M. et al. Tspan8- β -catenin positive feedback loop promotes melanoma invasion. **Oncogene**, v. 38, n. 20, p. 3781–3793, 16 maio 2019.
- EMERSON, R. O.; THOMAS, J. H. Adaptive Evolution in Zinc Finger Transcription Factors. **PLoS Genetics**, v. 5, n. 1, p. e1000325, 2 jan. 2009.
- FARAH, C. A.; SOSSIN, W. S. The role of C2 domains in PKC signaling. **Advances in Experimental Medicine and Biology**, v. 740, p. 663–683, 2012.
- FARMAN, F. U. et al. Aberrant promoter methylation status is associated with upregulation of the E2F4 gene in breast cancer. **Oncology Letters**, v. 15, n. 6, p. 8461–8469, 1 jun. 2018.
- FERNÁNDEZ-NOGUEIRA, P. et al. Differential expression of neurogenes among breast cancer subtypes identifies high risk patients. **Oncotarget**, v. 7, n. 5, p. 5313–5326, 10 dez. 2016.

- FERNANDEZ, A. F. et al. A DNA methylation fingerprint of 1628 human samples. **Genome research**, v. 22, n. 2, p. 407–19, 1 fev. 2012.
- FERTIG, E. J. et al. Analysis of gene expression of secreted factors associated with breast cancer metastases in breast cancer subtypes. **Scientific Reports**, v. 5, n. 1, p. 12133, 15 dez. 2015.
- FITZMAURICE, C. et al. The Global Burden of Cancer 2013. **JAMA Oncology**, v. 1, n. 4, p. 505, 1 jul. 2015.
- FOULKES, W. D.; SMITH, I. E.; REIS-FILHO, J. S. Triple-Negative Breast Cancer. **New England Journal of Medicine**, v. 363, n. 20, p. 1938–1948, 11 nov. 2010.
- FU, N. Y. et al. Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. **Nature Cell Biology**, v. 19, n. 3, p. 164–176, 1 mar. 2017.
- GALLO, L. H.; KO, J.; DONOGHUE, D. J. **The importance of regulatory ubiquitination in cancer and metastasis**Cell CycleTaylor and Francis Inc., , 3 abr. 2017.
- GESIERICH, S. et al. Colocalization of the tetraspanins, CO-029 and CD151, with integrins in human pancreatic adenocarcinoma: Impact on cell motility. **Clinical Cancer Research**, v. 11, n. 8, p. 2840–2852, 15 abr. 2005.
- GOSS, P. E. et al. Planning cancer control in Latin America and the Caribbean. **The Lancet Oncology**, v. 14, n. 5, p. 391–436, abr. 2013.
- GUNTER, M. J. et al. Breast Cancer Risk in Metabolically Healthy but Overweight Postmenopausal Women. **Cancer Research**, v. 75, n. 2, p. 270–274, 15 jan. 2015.
- GUPTA, V. A.; BEGGS, A. H. **Kelch proteins: Emerging roles in skeletal muscle development and diseases**Skeletal MuscleBioMed Central Ltd., , 1 jun. 2014.
- GUSTERSON, B. A.; STEIN, T. Human breast development. **Seminars in Cell & Developmental Biology**, v. 23, n. 5, p. 567–573, 1 jul. 2012.
- GYÖRFFY, B. et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. **Breast Cancer Research and Treatment**, v. 123, n. 3, p. 725–731, 18 out. 2010.
- HANAHAN, D.; WEINBERG, R. A. Hallmarks of Cancer: The Next Generation. **Cell**, v. 144, n. 5, p. 646–674, 4 mar. 2011.
- HEILER, S.; WANG, Z.; ZÖLLER, M. Pancreatic cancer stem cell markers and exosomes - the incentive push. **World Journal of Gastroenterology**, v. 22, n. 26, p. 5971, 14 jul. 2016.
- HEMLER, M. E. **Tetraspanin functions and associated microdomains**Nature Reviews Molecular Cell Biology, nov. 2005.
- HEMLER, M. E. **Tetraspanin proteins promote multiple cancer stages**Nature Reviews

Cancer, jan. 2014.

HU, Z. et al. The molecular portraits of breast tumors are conserved across microarray platforms. **BMC Genomics**, v. 7, n. 1, p. 96, 27 abr. 2006.

INCA/MS, 2018. **INCA - Instituto Nacional de Câncer - Estimativa 2018**. Disponível em: <<http://www.inca.gov.br/estimativa/2018/>>. Acesso em: 9 ago. 2018.

JEN, J.; WANG, Y.-C. Zinc finger proteins in cancer progression. **Journal of Biomedical Science**, v. 23, n. 1, p. 53, 13 dez. 2016.

JESELSON, R. et al. Cyclin D1 Kinase Activity Is Required for the Self-Renewal of Mammary Stem and Progenitor Cells that Are Targets of MMTV-ErbB2 Tumorigenesis. **Cancer Cell**, v. 17, n. 1, p. 65–76, 19 jan. 2010.

JOE, S.; NAM, H. Prognostic factor analysis for breast cancer using gene expression profiles. **BMC Medical Informatics and Decision Making**, v. 16, n. S1, p. 56, 18 jul. 2016.

KARGINOV, F. V.; HANNON, G. J. Remodeling of Ago2-mRNA interactions upon cellular stress reflects miRNA complementarity and correlates with altered translation rates. **Genes and Development**, v. 27, n. 14, p. 1624–1632, 15 jul. 2013.

KATSUTA, E. et al. High MYC mRNA Expression Is More Clinically Relevant than MYC DNA Amplification in Triple-Negative Breast Cancer. **International Journal of Molecular Sciences**, v. 21, n. 1, p. 217, 28 dez. 2019.

KELLEHER, F. C.; FENNELLY, D.; RAFFERTY, M. Common critical pathways in embryogenesis and cancer. **Acta Oncologica**, v. 45, n. 4, p. 375–388, 8 jan. 2006.

KEY, T. J. et al. Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. **Journal of the National Cancer Institute**, v. 95, n. 16, p. 1218–26, 20 ago. 2003.

KITAJIMA, S. et al. MesP1 and MesP2 are essential for the development of cardiac mesoderm. **Development**, v. 127, n. 15, 2000.

KOUSHIKA, S. P. et al. A post-docking role for active zone protein Rim. **Nature Neuroscience**, v. 4, n. 10, p. 997–1005, 17 set. 2001.

KREIKE, B. et al. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. **Breast Cancer Research**, v. 9, n. 5, p. R65, 2 out. 2007.

KUCHENBAECKER, K. B. et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for *BRCA1* and *BRCA2* Mutation Carriers. **JAMA**, v. 317, n. 23, p. 2402, 20 jun. 2017.

KUMAR, P.; AGGARWAL, R. An overview of triple-negative breast cancer. **Archives of Gynecology and Obstetrics**, v. 293, n. 2, p. 247–269, 4 fev. 2016.

- KUROZUMI, S. et al. Clinical and biological roles of Kelch-like family member 7 in breast cancer: a marker of poor prognosis. **Breast Cancer Research and Treatment**, v. 170, n. 3, p. 525–533, 1 ago. 2018.
- LAN, Z. J. et al. GGNBP2 acts as a tumor suppressor by inhibiting estrogen receptor α activity in breast cancer cells. **Breast Cancer Research and Treatment**, v. 158, n. 2, p. 263–276, 1 jul. 2016.
- LEE, G. H. et al. FYN promotes mesenchymal phenotypes of basal type breast cancer cells through STAT5/NOTCH2 signaling node. **Oncogene**, v. 37, n. 14, p. 1857–1868, 1 abr. 2018.
- LEHMANN, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. **The Journal of clinical investigation**, v. 121, n. 7, p. 2750–67, 1 jul. 2011.
- LEONARD, G. D.; SWAIN, S. M. Ductal carcinoma in situ, complexities and challenges. **Journal of the National Cancer Institute**, v. 96, n. 12, p. 906–20, 16 jun. 2004.
- LESCROART, F. et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. **Science (New York, N.Y.)**, v. 359, n. 6380, p. 1177–1181, 9 mar. 2018.
- LIANG, X. et al. Targeted next-generation sequencing identifies clinically relevant somatic mutations in a large cohort of inflammatory breast cancer. **Breast Cancer Research**, v. 20, n. 1, p. 88, 7 ago. 2018.
- LINDSLEY, R. C. et al. Mesp1 Coordinately Regulates Cardiovascular Fate Restriction and Epithelial-Mesenchymal Transition in Differentiating ESCs. **Cell Stem Cell**, v. 3, n. 1, p. 55–68, 3 jul. 2008.
- LIU, H. et al. Long non-coding RNAs as prognostic markers in human breast cancer. **Oncotarget**, v. 7, n. 15, p. 20584–20596, 12 abr. 2016.
- LIU, J. et al. GGNBP2 suppresses triple-negative breast cancer aggressiveness through inhibition of IL-6/STAT3 signaling activation. **Breast Cancer Research and Treatment**, v. 174, n. 1, p. 65–78, 28 fev. 2019.
- LIVASY, C. A. et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. **Modern Pathology**, v. 19, n. 2, p. 264–271, 2 fev. 2006.
- LONART, G. **RIM1: An edge for presynaptic plasticity** *Trends in Neurosciences* Elsevier Current Trends, , 1 jul. 2002.
- LUO, N. et al. ZEB1 induced- upregulation of long noncoding RNA ZEB1- AS1 facilitates the progression of triple negative breast cancer by binding with ELAVL1 to maintain the stability of ZEB1 mRNA. **Journal of Cellular Biochemistry**, p. jcb.29572, 10 jan. 2020.
- MACHNIK, M. et al. The expression signature of cancer-associated KRAB-ZNF factors

- identified in TCGA pan-cancer transcriptomic data. **Molecular Oncology**, v. 13, n. 4, p. 701–724, 1 abr. 2019.
- MAKAREM, M. et al. Stem Cells and the Developing Mammary Gland. **Journal of Mammary Gland Biology and Neoplasia**, v. 18, n. 2, p. 209–219, 27 jun. 2013.
- MAYER, I. A. et al. New strategies for triple-negative breast cancer--deciphering the heterogeneity. **Clinical cancer research : an official journal of the American Association for Cancer Research**, v. 20, n. 4, p. 782–90, 15 fev. 2014.
- MCGINN, O. et al. Cytokeratin 5 alters β -catenin dynamics in breast cancer cells. **Oncogene**, p. 1–15, 27 jan. 2020.
- MCNALLY, S.; STEIN, T. Overview of Mammary Gland Development: A Comparison of Mouse and Human. In: [s.l.] Humana Press, New York, NY, 2017. p. 1–17.
- MI, H. et al. miR-381 induces sensitivity of breast cancer cells to doxorubicin by inactivation of MAPK signaling via FYN. **European Journal of Pharmacology**, v. 839, p. 66–75, 15 nov. 2018.
- MOLL, R.; DIVO, M.; LANGBEIN, L. The human keratins: biology and pathology. **Histochemistry and Cell Biology**, v. 129, n. 6, p. 705–733, 7 jun. 2008.
- MOLYNEUX, G.; SMALLEY, M. J. The Cell of Origin of BRCA1 Mutation-associated Breast Cancer: A Cautionary Tale of Gene Expression Profiling. **Journal of Mammary Gland Biology and Neoplasia**, v. 16, n. 1, p. 51–55, 19 abr. 2011.
- NAGAI, M. A. et al. Prognostic value of NDRG1 and SPARC protein expression in breast cancer patients. **Breast Cancer Research and Treatment**, v. 126, n. 1, p. 1–14, 6 fev. 2011.
- NIELSEN, T. O. et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. **Clinical cancer research : an official journal of the American Association for Cancer Research**, v. 10, n. 16, p. 5367–74, 15 ago. 2004.
- OESTERREICH, S. et al. High rates of loss of heterozygosity on chromosome 19p13 in human breast cancer. **British Journal of Cancer**, v. 84, n. 4, p. 493–498, 13 fev. 2001.
- OMARINI, C. et al. **Neoadjuvant treatments in triple-negative breast cancer patients: Where we are now and where we are going** *Cancer Management and Research* Dove Medical Press Ltd, , 15 jan. 2018.
- PAN, S. J. et al. Over-expression of tetraspanin 8 in malignant glioma regulates tumor cell progression. **Biochemical and Biophysical Research Communications**, v. 458, n. 3, p. 476–482, 13 mar. 2015.
- PARK, C. S. et al. Therapeutic targeting of tetraspanin8 in epithelial ovarian cancer invasion and metastasis. **Oncogene**, v. 35, n. 34, p. 4540–4548, 25 ago. 2016.

- PARKER, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 27, n. 8, p. 1160–7, 10 mar. 2009.
- PAZ, M. F. C. J. et al. Correlations between Risk Factors for Breast Cancer and Genetic Instability in Cancer Patients—A Clinical Perspective Study. **Frontiers in Genetics**, v. 8, p. 236, 16 fev. 2018.
- PEROU, C. M. et al. Molecular portraits of human breast tumours. **Nature**, v. 406, n. 6797, p. 747–752, 17 ago. 2000.
- PETRIDOU, E. T.; GEORGAKIS, M. K.; ANTONOPOULOS, C. N. Effect of Oestrogen Exposure, Obesity, Exercise and Diet on Breast Cancer Risk. In: **Breast Cancer Management for Surgeons**. Cham: Springer International Publishing, 2018. p. 31–42.
- PINILLA, S. M. R. et al. Caveolin-1 expression is associated with a basal-like phenotype in sporadic and hereditary breast cancer. **Breast Cancer Research and Treatment**, v. 99, n. 1, p. 85–90, 16 set. 2006.
- POLYAK, K. Breast cancer: origins and evolution. **The Journal of clinical investigation**, v. 117, n. 11, p. 3155–63, 1 nov. 2007.
- PRAUD, D. et al. Cancer incidence and mortality attributable to alcohol consumption. **International Journal of Cancer**, v. 138, n. 6, p. 1380–1387, 15 mar. 2016.
- RAKHA, E. A. et al. Are triple-negative tumours and basal-like breast cancer synonymous? **Breast Cancer Research**, v. 9, n. 6, p. 404, 30 dez. 2007.
- RAKHA, E. A.; GREEN, A. R. Molecular classification of breast cancer: what the pathologist needs to know. **Pathology**, 28 dez. 2016.
- REDDY, S. M. et al. Long-term survival outcomes of triple-receptor negative breast cancer survivors who are disease free at 5 years and relationship with low hormone receptor positivity. **British Journal of Cancer**, v. 118, n. 1, p. 17–23, 12 jan. 2018.
- REIS-FILHO, J. S.; TUTT, A. N. J. **Triple negative tumours: A critical review** *Histopathology* John Wiley & Sons, Ltd, , 13 jan. 2008. Disponível em: <<http://doi.wiley.com/10.1111/j.1365-2559.2007.02889.x>>. Acesso em: 1 mar. 2020
- RENCK, D. V. et al. Equidade no acesso ao rastreamento mamográfico do câncer de mama com intervenção de mamógrafo móvel no sul do Rio Grande do Sul, Brasil. **Cadernos de Saúde Pública**, v. 30, n. 1, p. 88–96, jan. 2014.
- RINGNÉR, M. et al. GOBO: Gene Expression-Based Outcome for Breast Cancer Online. **PLoS ONE**, v. 6, n. 3, p. e17911, 21 mar. 2011.
- RODRIGUEZ, V. et al. Chromosome 8 BAC array comparative genomic hybridization and

- expression analysis identify amplification and overexpression of *TRMT12* in breast cancer. **Genes, Chromosomes and Cancer**, v. 46, n. 7, p. 694–707, 1 jul. 2007.
- RUSSO, J.; RUSSO, I. H. Molecular Basis of Breast Cancer. **Springer Berlin Heidelberg**, v. 53, n. 9, p. XIV, 448, 2004.
- SAGA, Y. et al. MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. **Development**, v. 122, n. 9, 1996.
- SAGA, Y.; KITAJIMA, S.; MIYAGAWA-TOMITA, S. **Mesp1 expression is the earliest sign of cardiovascular development** *Trends in Cardiovascular Medicine*, 2000.
- SANKARAN, V. G. et al. High-affinity binding of a FYVE domain to phosphatidylinositol 3-phosphate requires intact phospholipid but not FYVE domain oligomerization. **Biochemistry**, v. 40, n. 29, p. 8581–8587, 24 jul. 2001.
- SCOCCIANI, C. et al. Female breast cancer and alcohol consumption: a review of the literature. **American journal of preventive medicine**, v. 46, n. 3 Suppl 1, p. S16-25, mar. 2014.
- SELDIN, L.; LE GUELTE, A.; MACARA, I. G. Epithelial plasticity in the mammary gland. **Current Opinion in Cell Biology**, v. 49, p. 59–63, 1 dez. 2017.
- SHI, X. et al. **Kelch-like proteins: Physiological functions and relationships with diseases** *Pharmacological Research* Academic Press, , 1 out. 2019.
- SIEGEL, R. L.; MILLER, K. D.; JEMAL, A. Cancer statistics, 2016. **CA: A Cancer Journal for Clinicians**, v. 66, n. 1, p. 7–30, jan. 2016.
- SIZIOPIKOU, K. P. Ductal Carcinoma In Situ of the Breast: Current Concepts and Future Directions. **Archives of Pathology & Laboratory Medicine**, v. 137, n. 4, p. 462–466, 1 abr. 2013.
- SLEDGE, G. W. et al. Past, present, and future challenges in breast cancer treatment. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 32, n. 19, p. 1979–86, 1 jul. 2014.
- SLEEMAN, K. E. et al. Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. **The Journal of cell biology**, v. 176, n. 1, p. 19–26, 1 jan. 2007.
- SOIBAM, B. et al. Genome-Wide Identification of MESP1 Targets Demonstrates Primary Regulation Over Mesendoderm Gene Activity. **STEM CELLS**, v. 33, n. 11, p. 3254–3265, 1 nov. 2015.
- SORLIE, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. **Proceedings of the National Academy of Sciences**, v. 98, n. 19,

p. 10869–10874, 11 set. 2001a.

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

SORLIE, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 14, p. 8418–23, 8 jul. 2003.

STOGIOS, P. J.; PRIVÉ, G. G. **The BACK domain in BTB-kelch proteins** *Trends in Biochemical Sciences* Elsevier Ltd, , 1 dez. 2004.

SZALA, S. et al. Molecular cloning of cDNA for the human tumor-associated antigen CO-029 and identification of related transmembrane antigens. **Proceedings of the National Academy of Sciences**, v. 87, n. 17, 1990.

TABARIÈS, S. et al. Afadin cooperates with claudin-2 to promote breast cancer metastasis. **Genes and Development**, v. 33, n. 3–4, p. 180–193, 1 fev. 2019.

TAN, R. et al. Neural Functions Play Different Roles in Triple Negative Breast Cancer (TNBC) and non-TNBC. **Scientific Reports**, v. 10, n. 1, p. 3065, 2020.

The Spectrum of Genetic Mutations in Breast Cancer. **Asian Pacific Journal of Cancer Prevention**, v. 16, n. 6, p. 2177–2185, 2015.

TIBBETTS, M. D. et al. Crystal structure of a FYVE-type zinc finger domain from the caspase regulator CARP2. **Structure**, v. 12, n. 12, p. 2257–2263, 1 dez. 2004.

TROJAN, J. et al. 5'-CpG island methylation of the LKB1/STK11 promoter and allelic loss at chromosome 19p13.3 in sporadic colorectal cancer. **Gut**, v. 47, n. 2, p. 272–276, 1 ago. 2000.

URRUTIA, R. KRAB-containing zinc-finger repressor proteins. **Genome Biology**, v. 4, n. 10, p. 231, 23 set. 2003.

VAN SWEARINGEN, A. E. D.; SIEGEL, M. B.; ANDERS, C. K. Breast cancer brain metastases: Evidence for neuronal-like adaptation in a “breast-to-brain” transition? **Breast Cancer Research**, v. 16, n. 3, p. 304, 6 maio 2014.

VOGLSTAETTER, M. et al. Tspan8 is expressed in breast cancer and regulates E-cadherin/catenin signalling and metastasis accompanied by increased circulating extracellular vesicles. **The Journal of Pathology**, v. 248, n. 4, p. 421–437, 18 ago. 2019.

WANG, H. et al. BRCA1/FANCD2/BRG1-Driven DNA Repair Stabilizes the Differentiation State of Human Mammary Epithelial Cells. **Molecular Cell**, v. 63, n. 2, p. 277–292, 21 jul. 2016.

WANG, Z. J. et al. Allele loss and mutation screen at the Peutz-Jeghers (LKB1) locus

(19p13.3) in sporadic ovarian tumours. **British Journal of Cancer**, v. 80, n. 1–2, p. 70–72, 26 mar. 1999.

WEIGELT, B.; REIS-FILHO, J. S. Histological and molecular types of breast cancer: is there a unifying taxonomy? **Nature Reviews Clinical Oncology**, v. 6, n. 12, p. 718–730, 1 dez. 2009.

WU, S. M. Mesp1 at the Heart of Mesoderm Lineage Specification. **Cell Stem Cell**, v. 3, n. 1, p. 1–2, 2008.

XIANG, X. et al. Hepatocyte nuclear factor 4 alpha promotes the invasion, metastasis and angiogenesis of neuroblastoma cells via targeting matrix metalloproteinase 14. **Cancer Letters**, v. 359, n. 2, p. 187–197, 10 abr. 2015.

YANG, L. et al. Unbiased screening for transcriptional targets of ZKSCAN3 identifies integrin $\beta 4$ and vascular endothelial growth factor as downstream targets. **Journal of Biological Chemistry**, v. 283, n. 50, p. 35295–35304, 12 dez. 2008.

YANG, L. et al. Evidence of a role for the novel zinc-finger transcription factor ZKSCAN3 in modulating Cyclin D2 expression in multiple myeloma. **Oncogene**, v. 30, n. 11, p. 1329–1340, 17 mar. 2011.

YANG, T.-L. et al. High-resolution 19p13.2-13.3 allelotyping of breast carcinomas demonstrates frequent loss of heterozygosity. **Genes, Chromosomes and Cancer**, v. 41, n. 3, p. 250–256, 1 nov. 2004.

YIN, S. et al. Myc mediates cancer stem-like cells and EMT changes in triple negative breast cancers cells. **PLOS ONE**, v. 12, n. 8, p. e0183578, 17 ago. 2017.

ZHAN, Z. et al. A positive tetraspanin 8 (TSPAN8)/b-catenin regulatory loop enhances the stemness of colorectal cancer cells. **Medical Science Monitor**, v. 25, p. 9594–9601, 15 dez. 2019.

ZHANG, H. S. et al. TSPAN8 promotes colorectal cancer cell growth and migration in LSD1-dependent manner. **Life Sciences**, v. 241, p. 117114, 15 jan. 2020.

ZHANG, L. et al. Expression and clinical significance of Kelch-like epichlorohydrin-associated protein 1 in breast cancer. **Genetics and Molecular Research**, v. 15, n. 2, 13 maio 2016.

ZHANG, X. et al. Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor target gene activation. **Proceedings of the National Academy of Sciences**, v. 109, n. 33, p. 13331–13336, 14 ago. 2012.

ZHU, R. et al. TSPAN8 promotes cancer cell stemness via activation of sonic Hedgehog signaling. **Nature Communications**, v. 10, n. 1, 1 dez. 2019.

ZÖLLER, M. **Tetraspanins: Push and pull in suppressing and promoting metastasis** *Nature Reviews Cancer*, jan. 2009.

8. MATERIAL SUPLEMENTAR

A

Componente Celular

B

Função Molecular

Figura suplementar 1. Termos Gene Ontology (GO) para classificação de 100 genes obtidos da comparação transcricional SPARC+ vs SPARC-. A.- Componente celular B.- Função molecular

Figura Suplementar 2. Análise de sobrevida para pacientes com câncer mamário estratificados pela expressão de MESP1 A.- global e B.- livre de recorrência para MESP1 em todos os subtipos de câncer de mama. Dados obtidos da plataforma KMPlotter.

Figura Suplementar 3. Sobrevida global e Livre de recaída para MESP1. O melhor ponto de corte foi empregado como cutoff. A-D: sobrevida global; E-H: Sobrevida livre de recaída. A, E: luminal A; B, F: luminal B; C, G: Her2; D, H: Basal. Dados obtidos de plataforma KMPlotter. Dados obtidos de plataforma KMPlotter.

Figura Suplementar 4. Análise de expressão do transcrito *MESP1* de acordo com A. Subtipo molecular PAM50; B. status de Receptor de Estrogênio. C. Grau histológico. Dados obtidos da plataforma GEO (GSE51840).

Figura Suplementar 5. Análise de co-expressão entre nível de mRNA de MESP1 e MESP2 em pacientes com cancer de mama. dados obtidos da plataforma CBIPORTAL para o estudo TCGA provisional.

Figura Suplementar 6. Análise de A.- sobrevida global e B.- livre de recidiva para ZNF556 em todos os subtipos de câncer de mama. Dados obtidos da plataforma KMPlotter.

Figura Supplementar 7. Sobrevida global e Livre de recidiva para ZNF556. O melhor ponto de corte foi empregado como cutoff. A-D: sobrevida global; E-H: Sobrevida livre de recaída. A, E: luminal A; B, F: luminal B; C, G: Her2; D, H: Basal. Dados obtidos de plataforma KMPlotter.

Figura Suplementar 8. Análise de expressão do transcrito ZNF556 de acordo com A,D. Subtipo molecular PAM50; B,E. status de Receptor de Estrogênio; C,F. Grau histológico. Dados obtidos das plataformas GEO (GSE51840): A-D, e GOBO: D-F.

Figura Suplementar 9. Análise de A.- sobrevida global e B.- livre de recidiva para TSPAN8 em todos os subtipos de câncer de mama. Dados obtidos da plataforma KMPlotter.

Figura Supplementar 10. Sobrevida global e Livre de recaída para TSPAN8. O melhor ponto de corte foi empregado como cutoff. A-D: sobrevida global; E-H: Sobrevida livre de recaída. A, E: luminal A; B, F: luminal B; C, G: Her2; D, H: Basal. Dados obtidos de plataforma KMPlotter. Dados obtidos de plataforma KMPlotter.

Figura Suplementar 11. Análise de expressão do transcrito TSPAN8 de acordo com A,D. Subtipo molecular PAM50; B,E. status de Receptor de Estrogênio; C,F. Grau histológico. Dados obtidos das plataformas GEO (GSE51840): A-D, e GOBO: D-F.

Figura Suplementar 12. Sobrevida global e Livre de recaída para RIMS2. O melhor ponto de corte foi empregado como cutoff. A-D: sobrevida global; E-H: Sobrevida livre de recaída. A, E: luminal A; B, F: luminal B; C, G: Her2; D, H: Basal. Dados obtidos de plataforma KMPlotter. Dados obtidos de plataforma KMPlotter.

Figura Suplementar 13. Análise de expressão do transcrito RIMS2 de acordo com A,D. Subtipo molecular PAM50; B,E. status de Receptor de Estrogênio; C,F. Grau histológico. Dados obtidos das plataformas GEO (GSE51840): A-D, e GOBO: D-F.

Figura Suplementar 14. Análise de A.- sobrevida global e B.- Livre de recidiva para KLHL13 em todos os subtipos de câncer de mama. Dados obtidos da plataforma KMPlotter.

Figura Suplementar 15. Sobrevida global e Livre de recaída para KLHL13. O melhor ponto de corte foi empregado como cutoff. A-D: sobrevida global; E-H: Sobrevida livre de recaída. A, E: luminal A; B, F: luminal B; C, G: Her2; D, H: Basal. Dados obtidos de plataforma KMPlotter. Dados obtidos de plataforma KMPlotter.

Figura Suplementar 16. Análise de expressão do transcrito KLHL13 de acordo com A. Subtipo molecular PAM50; B. status de Receptor de Estrogênio. C. Grau histológico. Dados obtidos da plataforma GEO (GSE51840).

Figura Suplementar 17. Análise de co-expressão entre nível de mRNA de KLHL13 e 05 genes em pacientes com cancer de mama. A: CACHD1; B: DMD; C: PAK5; D: SYNM; SFRP1. Dados obtidos da plataforma CBIPORTAL a partir do estudo TCGA provisional.

Tabela suplementar 1. Anotação funcional (Termos GO) de 100 transcritos diferencialmente expressos obtidos a partir da análise transcriptômica de TNBC/SPARC+ versus TNBC/SPARC-

ID	Termos GO	N de transcritos
	Processos Biológicos	
GO:0042742	Resposta em defesa a bactérias	5 (0,05%)
GO:0030154	Diferenciação celular	7 (0,07%)
GO:0007416	Montagem de sinapse	3 (0,03%)
GO:0030216	Diferenciação do Queratinócito	3 (0,03%)
GO:0008544	Desenvolvimento da Epiderme	3 (0,03%)
	Regulação Negativa da atividade	
GO:0010466	peptidase	2 (0,02%)
GO:0061436	Estabelecimento da barreira da pele	2 (0,02%)
GO:0009880	Especificação do Padrão embrionário	2 (0,02%)
	Componente Celular	
GO:0070062	Exosomo extracelular	27 (0,29%)
GO:0005615	Espaço extracelular	16 (0,17%)
GO:0005576	Região extracelular	16 (0,17%)
GO:0031225	Componente ancorado da membrana	4 (0,04%)
GO:0001533	Envelope cornificado	3 (0,03%)
GO:0005796	Lúmen do aparato de Golgi	3 (0,03%)
	Função Molecular	
GO:0005509	União ao Calcio	9 (0,09%)
GO:0031731	União ao receptor CCR6 de Citocinas	2 (0,02%)
GO:0005198	Atividade molecular estrutural	5 (0,05%)
	Atividade inibitoria de endopeptidase	
GO:0004867	tipo serina	3 (0,03%)

Dados obtidos da plataforma DAVID usando a ferramenta GO enrichment

Tabela suplementar 2. Descrição da função biológica de 22 genes identificados na comparação TNBC/SPARC+ versus TBNC/SPARC-

GENE NAME	GENE DESCRIPTION	F.C.	BIOLOGICAL FUNCTION (NCBI)	LOCUS
CAMP	(cathelicidin antimicrobial peptide)	- 17,7 5	This gene encodes a member of an antimicrobial peptide family, characterized by a Altaly conserved N-terminal signal peptide containing a cathelin domain and a structurally variable cationic antimicrobial peptide, which is produced by extracellular proteolysis from the C-terminus. In addition to its antibacterial, antifungal, and antiviral activities, the encoded protein functions in cell chemotaxis, immune mediator induction, and inflammatory response regulation.	3p21.31
CAPN8	calpain 8	-33	Calpains constitute a family of intracellular Ca (2+)-regulated cysteine proteases that are indispensable in the regulation of a wide variety of cellular functions. The improper activation of calpain causes lethality or various disorders, such as muscular dystrophies and tumor formation.	1q41
CHP2	(calcineurin like EF-hand protein 2)	- 17,9 9	This gene product is a small calcium-binding protein that regulates cell pH by controlling plasma membrane-type Na ⁺ /H ⁺ exchange activity. This protein shares sequence similarity with calcineurin B and can bind to and stimulate the protein phosphatase activity of calcineurin A (CnA) and functions in the calcineurin/NFAT (nuclear factor of activated T cells) signaling pathway. Another member of the CHP subfamily, Calcineurin B homologous protein 1, is located on Chromosome 15 and is an inhibitor of calcineurin activity and has a genetic phenotype associated with Parkinson's Disease (OMIM:606988). This gene was initially identified as a tumor-associated antigen and was previously referred to as Hepatocellular carcinoma-associated antigen 520.	16p12.2
DAPL1	(death associated protein like 1)	- 11,23	May play a role in the early stages of epithelial differentiation or in apoptosis and is normally expressed in hair follicle, corneal epithelium, epidermis and footpad epithelium (at protein level).	2q24.1
DDTL	(D-dopachrome tautomerase like)	- 21,6 0	D-dopachrome tautomerase (D-DT) is a newly described cytokine and a member of the macrophage migration inhibitory factor (MIF) protein superfamily. MIF is a broadly expressed pro-inflammatory cytokine that regulates both the innate and the adaptive immune response. MIF activates the MAP kinase cascade, modulates cell migration, and counter-acts the immunosuppressive effects of glucocorticoids. For many cell types, MIF also acts as an important survival or anti-apoptotic factor. The protein encoded by this gene is a component of connective tissue microfibrils and may be involved in elastic fiber assembly. Mutations in this gene cause congenital contractural arachnodactyly.	22q11.23
FBN2	(fibrillin 2)	11,53		5q23.3
HMP19	(neuronal vesicle trafficking associated 2)	-12,2	HMP19 encodes a Mr ~ 19 kDa protein that reportedly localizes to the Golgi apparatus in neural and neuroendocrine cells. Specifically expressed in neural and neuroendocrine tissues. Pituitary and less in adrenal gland and testis. Expressed in the hippocampus throughout development. Remains enriched in layer V cortical neurons during development.	5q35.2
KLHL13	(kelch like family	22,0	This gene encodes a BTB and kelch domain containing protein and belongs to the kelch repeat	Xq24

	member 13)	9	domain containing superfamily of proteins. The encoded protein functions as an adaptor protein that complexes with Cullin 3 and other proteins to form the Cullin 3-based E3 ubiquitin-protein ligase complex. This complex is necessary for proper chromosome segregation and completion of cytokinesis. Alternate splicing results in multiple transcript variants.	
MAGEA1	(MAGE family member A1)	14,4 9	This gene is a member of the MAGEA gene family. The members of this family encode proteins with 50 to 80% sequence identity to each other. The promoters and first exons of the MAGEA genes show considerable variability, suggesting that the existence of this gene family enables the same function to be expressed under different transcriptional controls. The MAGEA genes are clustered at chromosomal location Xq28. They have been implicated in some hereditary disorders, such as dyskeratosis congenita.	Xq28
MESP1	(mesoderm posterior bHLH transcription factor 1)	-21,2	The MESP1 gene encodes a member of the basic helix-loop-helix (bHLH) family of transcriptional regulatory proteins that may play a role in the development of nascent mesoderm.	15q26.1
MSMB	(microseminoprotein beta)	7,97	The protein encoded by this gene is a member of the immunoglobulin binding factor family. It is synthesized by the epithelial cells of the prostate gland and secreted into the seminal plasma. This protein has inhibin-like activity. It may have a role as an autocrine paracrine factor in uterine, breast and other female reproductive tissues. The expression of the encoded protein is found to be decreased in prostate cancer. Two alternatively spliced transcript variants encoding different isoforms are described for this gene. The use of alternate polyadenylation sites has been found for this gene.	10q11.22
PKNOX2	(PBX/knotted 1 homeobox 2)	- 18,7 9	Homeodomain proteins are sequence-specific transcription factors that share a highly conserved DNA-binding domain and play fundamental roles in cell proliferation, differentiation, and death. PKNOX2 belongs to the TALE (3-amino acid loop extension) class of homeodomain proteins characterized by a 3-amino acid extension between alpha helices 1 and 2 within the homeodomain.	11q24.2
RIMS2	(regulating synaptic membrane exocytosis 2)	6,67	The protein encoded by this gene is a presynaptic protein that interacts with RAB3, a protein important for normal neurotransmitter release. The encoded protein can also bind several other synaptic proteins, including UNC-13 homolog B, ELKS/Rab6-interacting/CAST family member 1, and synaptotagmin 1. This protein is involved in synaptic membrane exocytosis. Polymorphisms in this gene have been associated with degenerative lumbar scoliosis.	8q22.3
SCGB2A1	(secretoglobin family 2A member 1)	- 25,5 9	Mammaglobin A (secretoglobin, family 2A, member 2 – SCGB2A2, or mammaglobin-1) is a member of the secretoglobin superfamily, a group of small, secretory, rarely glycosylated, dimeric proteins, mainly expressed in mucosal tissues.	11q12.3

SCGB2A2	(secretoglobin family 2A member 2)	- 30,4 7	SCGB2A2 is a member of the superfamily of secretoglobins, a group of small dimeric secreted and sometimes glycosylated proteins. Expressed mainly in mucosa, secretoglobins seem to be involved in cell signalling, immune response, and chemotaxis, and may also serve as transporters for steroid hormones in human.	11q12.3
SIM2	(SIM bHLH transcription factor 2)	- 11,85	This gene represents a homolog of the Drosophila single-minded (sim) gene, which encodes a transcription factor that is a master regulator of neurogenesis. The encoded protein is ubiquitinated by RING-IBR-RING-type E3 ubiquitin ligases, including the parkin RBR E3 ubiquitin protein ligase. This gene maps within the so-called Down syndrome chromosomal region, and is thus thought to contribute to some specific Down syndrome phenotypes. Alternative splicing of this gene results in multiple transcript variants.	21q22.13
SLAIN1	(SLAIN motif family member 1)	13,4 4	SLAIN1 and its more widely expressed homologue SLAIN2 comprise a new family of structurally unique genes conserved throughout vertebrate evolution. Slain1 is expressed at the stem cell and epiblast stages of ESC differentiation and in the epiblast, nervous system, tailbud and somites of the developing mouse embryo **it codes for Microtubule plus-end tracking protein that might be involved in the regulation of cytoplasmic microtubule dynamics, microtubule organization and microtubule elongation.	13q22.3
SLURP1	(secreted LY6/PLAUR domain containing 1)	- 10,3 1	The protein encoded by this gene is a member of the Ly6/uPAR family but lacks a GPI-anchoring signal sequence. It is thought that this secreted protein contains antitumor activity. Mutations in this gene have been associated with Mal de Meleda, a rare autosomal recessive skin disorder. This gene maps to the same chromosomal region as several members of the Ly6/uPAR family of glycoprotein receptors.	8q24.3
SPRR3	(small proline rich protein 3)	- 31,5 4	The small proline-rich region (SPRR) protein 3 belongs to the SPRR family that in humans is represented by 11 different SPRR members encoded by two SPRR1 genes (SPRR1a and SPRR1b), 7 SPRR2 (A–G), one SPRR3 and one SPRR4. These genes are clustered tightly together with other genes involved in the formation of the cornified envelope (CE) and in the regulation of keratinocyte terminal differentiation, within the epidermal differentiation complex (EDC) on human chromosome 1q21 SPRR3 is a marker for terminal squamous cell differentiation. but also functions in non-squamous tissues	1q21.3
TOX3	(TOX Alta mobility group box family member 3)	12,5 2	The protein encoded by this gene contains an HMG-box, indicating that it may be involved in bending and unwinding of DNA and alteration of chromatin structure. The C-terminus of the encoded protein is glutamine-rich due to CAG repeats in the coding sequence. A minor allele of this gene has been implicated in an elevated risk of breast cancer. Two transcript variants encoding different isoforms have been found for this gene.	16q12.1
TSPAN8	(tetraspanin 8)	7,41	The protein encoded by this gene is a member of the transmembrane 4 superfamily, also known as the	12q21.1

			tetraspanin family. Most of these members are cell-surface proteins that are characterized by the presence of four hydrophobic domains. The proteins mediate signal transduction events that play a role in the regulation of cell development, activation, growth and motility. This encoded protein is a cell surface glycoprotein that is known to complex with integrins. This gene is expressed in different carcinomas. The use of alternate polyadenylation sites has been found for this gene.	
ZNF556	(zinc finger protein 556)	18,09	The Zinc Finger Protein 658B is a Pseudogene. Among its related pathways are Gene Expression and may be involved in transcriptional regulation.	19p13.3

Dados obtidos do banco NCBI, utilizando a ferramenta RefSeq. FC:Fold Change.

Tabela suplementar 3. Os 100 genes melhores diferencialmente expressos identificados em análise de transcriptômica: SPARC positive versus SPARC negative

Table 3. 50 top genes differentially expressed identified in transcriptome analysis: SPARC positive versus SPARC negative – negatively regulated.				Table 2. 50 top genes differentially expressed identified in transcriptome analysis: SPARC positive versus SPARC negative – positively regulated.			
Gene symbol	Description	Fold change	Associated with cancer	Gene symbol	Description	Fold change	Associated with cancer
KRT1	Keratin 1	-57.75	Yes ¹	SOHLH2	Spermatogenesis and oogenesis-specific basic helix-loop-helix 2	24.20	Yes ¹
CAPN8	Calpain 8	-33.00	No	KLHL13	Kelch-like family member 13	22.09	No
IVL	Involucrin	-32.16	Yes ¹	HPGD	Hydroxyprostaglandin dehydrogenase 15- β (HGD)	20.01	Yes ²
SPRR3	Small proline-rich protein 3	-31.54	Yes ¹	GPR158	G-protein-coupled receptor 158	18.73	Yes
SCGB2A2	Secretoglobulin, family 2A, member 2	-30.474	Yes ¹	FABP7	Fatty acid-binding protein 7, brain	18.30	Yes ²
SCGB2A1	Secretoglobulin, family 2A, member 1	-25.594	Yes ¹	CTAG1A	Cancer/testis antigen 1A	18.09	Yes
LNX1	LIM homeobox 1	-22.51	Yes	ZNF556	Zinc finger protein 556	17.34	No
CEACAM7	Carcinoembryonic antigen-related cell adhesion molecule 7	-21.611	Yes ¹	SLC26A4	Solute carrier family 26 (anion exchanger), member 4	15.66	Yes ¹
DDIT1	D-dopachrome tautomerase-like	-21.601	No	MAGEA1	Melanoma antigen family A, 1	14.49	Yes ²
MesP1	Mesoderm posterior basic helix-loop-helix transcription factor 1	-21.20	No	UGT2B11	UDP glucuronosyltransferase 2 family, polypeptide B11	14.13	Yes ¹
CEACAM6	Carcinoembryonic antigen-related cell adhesion molecule 6 (nonspecific cross-reacting antigen)	-19.95	Yes ²	SLAIN1	SLAIN motif family, member 1	13.44	Yes
FAM25A	Family with sequence similarity 25, member A	-19.13	No	C6orf105	Androgen-dependent TTP-regulating protein	13.05	No
PKNOX2	PKX/knotted 1 homeobox 2	-18.79	No	GSTM2P1	Glutathione S-transferase mu 2 (muscle) pseudogene 1	13.04	No
CNP2	Calcineurin-like EF-hand protein 2	-17.99	Yes	HS3ST4	Heparan sulfate (glucosamine) 3-O-sulfotransferase 4	12.58	No
CAMP	Cathelicidin antimicrobial peptide	-17.759	Yes ¹	CYP26A1	Cytochrome P450, family 26, subfamily A, polypeptide 1	12.56	Yes ¹
SERPINF4	Serpin peptidase inhibitor, clade B (ovalbumin), member 4	-17.099	Yes	TDX3	TDX high-mobility group box family member 3	12.52	Yes ¹
SERPINF3	Serpin peptidase inhibitor, clade B (ovalbumin), member 3	-16.55	Yes ¹	FBN2	Fibrillin 2	11.53	Yes ¹
CAP5	Calcyphosine	-14.89	Yes ¹	FSP1	Fibrous sheath-interacting protein 1	11.29	Yes ¹
SCGB1D2	Secretoglobulin, family 1D, member 2	-14.83	Yes ¹	CDH1	Cadherin 1, type 1, E-cadherin (epithelial)	10.64	Yes ²
HS6ST3	Heparan sulfate 6-O-sulfotransferase 3	-13.98	Yes ¹	ZNF385B	Zinc finger protein 385B	10.52	Yes ¹
PLA2G4D	Phospholipase A2, group IV/D (cytosolic)	-13.79	Yes	TDRO9	Tudor domain containing 9	9.71	No
KRT32	Keratin 32	-13.72	No	CHRD1	Chordin-like 2	9.67	Yes ¹
SPRX6	Serine peptidase inhibitor, Kazal type 6	-12.23	Yes	CELF4	CUGBP, Elav-like family member 4	9.10	Yes
HMP19	HMP19 protein	-12.20	Yes	TNFRSF17	Tumor necrosis factor receptor superfamily, member 17	8.63	Yes ¹
KRTDAP	Keratinocyte differentiation-associated protein	-11.97	Yes	DDC	Dopa decarboxylase (aromatic L-amino acid decarboxylase)	8.53	Yes
MIST2H3A	Histone cluster 2, H3a	-11.97	Yes	KARML3	Kin of RRE like 3 (Drosophila)	8.31	Yes
C6orf169	Cysteine-rich tail protein 1	-11.867	No	FBXW7B	F-box and WD repeat domain containing 10	8.20	Yes
SM2	Single-minded family bHLH transcription factor 2	-11.857	Yes ¹	GLYT1L2	Glycine-N-acyltransferase-like 2	8.13	Yes
DAPL1	Death-associated protein-like 1	-11.237	Yes	TRIM36	Tripartite motif containing 36	7.99	Yes ¹
FABP6	Fatty acid binding protein 6, ileal	-11.17	Yes	MSMB	Microsminoprotein, β -	7.97	Yes ¹
SNORA74A	Small nucleolar RNA, H/ACA box 74A	-11.16	Yes	ZNF204P	Zinc finger protein 204, pseudogene	7.63	No
SNORA53	Small nucleolar RNA, H/ACA box 53	-11.12	No	SEMA3D	Sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3D	7.47	Yes
CEACAM3	Carcinoembryonic antigen-related cell adhesion molecule 3	-11.06	Yes	TSPAN8	Tetraspanin 8	7.41	Yes
DEFB1	Defensin, β 1	-11.03	Yes	MUC20	Mucin 20, cell surface associated	7.37	Yes
SNORA28	Small nucleolar RNA, H/ACA box 28	-10.45	No	APOD	Apolipoprotein D	7.33	Yes ¹
S100A14	S100 calcium-binding protein A14	-10.39	Yes ²	C6orf46	Chromosome 6 open-reading frame 46	7.31	No
SLURP1	Secreted LY6/PLAUR domain containing 1	-10.31	Yes ²	ATP1A2	ATPase, Na ⁺ /K ⁺ transporting, α 2 polypeptide	7.26	Yes ¹
RG520	Regulator of G-protein signaling 20	-10.23	Yes ²	MYO12	Myosin 2	7.02	Yes ¹
CEACAM5	Carcinoembryonic antigen-related cell adhesion molecule 5	-10.21	Yes ²	FAM158	Family with sequence similarity 65, member 8	6.78	Yes
HMP12	HMP-like 2	-10.19	Yes	CLDN1	Claudin 1	6.69	Yes ²
TMEM45A	Transmembrane protein 45A	-10.14	Yes ¹	RIMS2	Regulating synaptic membrane exocytosis 2	6.67	Yes
RN7SK	RNA, 7SK small nuclear	-10.11	No	SCSD	Sterol-C5-desaturase	6.56	No
DEFB4A	Defensin, β 4A	-9.81	Yes ¹	PPMTN	Protein phosphatase, Mg ²⁺ /Mn ²⁺ dependent, 1H	6.51	Yes ¹
SNORD38-1	Small nucleolar RNA, C/D box 38-1	-9.76	No	MAGEC1	Melanoma antigen family C, 1	6.26	Yes ¹
CDSN	Corneodesmosin	-9.56	No	MSLN	Mesothelin	6.26	Yes ²
RAET1E	Retinoic acid early transcript 1E	-9.54	Yes ¹	EAF2	ELL-associated factor 2	6.14	Yes
PVALB	Parvalbumin	-9.49	Yes				

¹Associated with breast cancer.
²Associated with triple-negative breast cancer.

¹Associated with breast cancer.
²Associated with TNBC.
 TNBC: Triple-negative breast cancer.

Dados obtidos apartir do trabalho de Alcantara Filho e cols, 2018 Acesso (DOI: [10.2217/bmt-2017-0019](https://doi.org/10.2217/bmt-2017-0019))

Tabela suplementar 4. Distribuição da expressão para 20 genes em acordo com estado tumoral e subtipos moleculares

Gene	Normal versus Tumor	Subtype status		
		Luminal	Her2+	Triple Negative
CAMP	Upregulado (< 0,0001)	Intermediaria	Alta	Baixa
CAPN8	Downregulado (< 0,0001)	Alta	Intermediaria	Baixa
CHP2	Downregulado (0.03)	Baixa	Intermediaria	Alta
DAPL1	Downregulado (0.15)	Intermediaria	Baixa	Alta
DDTL	Upregulado (< 0,0001)	Intermediaria	Baixa	Alta
HMP19	Upregulado (<0.0003)	Alta	Intermediaria	Baixa
MESP1	Downregulado (0.14)	Baixa	Intermediaria	Alta
PKNOX2	Downregulado (< 0,0001)	Alta	Intermediaria	Baixa
SCGB2A1	Downregulado (0.14)	Intermediaria	Alta	Baixa
SCGB2A2	Downregulado (< 0,0001)	Alta	Intermediaria	Baixa
SIM2	Upregulado (< 0,0001)	Baixa	Intermediaria	Alta
SLURP1	Downregulado (< 0,0001)	Baixa	Alta	Intermediaria
KLHL13	Downregulado (< 0,0001)	Intermediaria	Baixa	Alta
ZNF556	Downregulado (0.008)	Alta	Intermediaria	Baixa
SLAIN1	Downregulado (< 0,0001)	Intermediaria	Baixa	Alta
TOX3	Upregulado (< 0,0001)	Alta	Intermediaria	Baixa
FBN2	Upregulado (< 0,0001)	Alta	Baixa	Intermediaria
MSMB	Upregulado (< 0,0001)	Intermediaria	Alta	Baixa
TSPAN8	Downregulado (0.052)	Baixa	Alta	Intermediaria
RIMS2	Downregulado (< 0,0001)	Baixa	Intermediaria	Alta

Dados obtidos do estudo TCGA provisional pela ferramenta web UALCAN. Nível de expressão normalizado a TPP, Transcritos por milhão. Diferencia na expressão baseada na comparação entre as medianas

Tabela suplementar 5. Características, vantagens e desvantagens das principais aplicações de algoritmos de predição de miRNAs

Algoritmo de Predição	Característica		Qualidade da predição	
	Pontuação final baseada em	Conservação entre espécies	Robustez	Desvantagens
miRanda	Complementariedade e ligação de energia livre	Filtro de conservação entre espécies	Benéfico para locais de previsão com ligação imperfeita na região das 'seeds'	Baixa precisão, muitos falsos positivos. Desempenho médio em alvos não canônicos
TargetScan	Correspondência de seeds, complementaridade 3' baseada	A aplicação pontua cada resultado por conservação	Muitos parâmetros incluídos na pontuação do alvo. A pontuação final se correlaciona com a regulação negativa da proteína	Sítios com baixo emparelhamento de seeds são omitidos. Baixo desempenho na previsão de alvos não canônicos
TargetScanS	Tipo de correspondência de seed	Somente sítios conservadores são considerados	Ferramenta simples para pesquisa de sites conservados com emparelhamento rigoroso de seeds	Subestimação miRNAs com vários sítios alvo
PicTar	Energia de ligação, complementariedade e conservação	Emparelhamento necessário em posições conservadas	miRNAs com múltiplos alinhamentos são favorecidos	Não prevê sítios não conservadores. Baixo desempenho em sites não canônicos
DIANA-microT	Energia livre de ligação e complementaridade de seeds	Conjunto de dados de UTRs conservados entre humanos e camundongos é usado	Probabilidade fornecida para cada site de destino. Possibilidade de usar sequência miRNA próprio como entrada	Baixa eficiência em comparação com outros algoritmos, principalmente na previsão de alvos não canônicos
PITTA	Energia de acessibilidade do sitio alvo	Definido pelo usuário	A estrutura secundária do 3'UTR é considerada para a interação do miRNA	Alguns miRNAs com vários sítios alvo podem ser omitidos
RNA22	Padrão de reconhecimento e energia de dobramento	Não disponível	Permite identificar sítios alvejados por miRNAs ainda não descobertos	Baixa eficiência comparada a outros algoritmos

Adaptado de Witkos et al, 2011