

**CARACTERIZAÇÃO DO INFILTRADO
INFLAMATÓRIO EM CÂNCER DE MAMA
TRIPLO-NEGATIVO E SUA IMPLICAÇÃO
PROGNÓSTICA**

DANIEL GASPAR GONÇALVES

**Dissertação de mestrado apresentada a Fundação
Antônio Prudente para a obtenção do título de
Mestre em Ciências**

Área de concentração: Oncologia

Orientador: Dr. Israel Tojal da Silva

Coorientador: Dra. Dirce Maria Carraro

São Paulo

2019

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Gonçalves, Daniel Gaspar

Caracterização do infiltrado inflamatório em câncer de mama triplo-negativo e sua implicação prognóstica / Daniel Gaspar Gonçalves – São Paulo, 2019.

50p.

Dissertação (Mestrado)-Fundação Antônio Prudente.

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

Orientador: Israel Tojal da Silva

Descritores: 1. Neoplasias de Mama Triplo Negativas/diagnóstico/Triple Negative Breast Neoplasms/diagnosis. 2. Expressão Gênica/Gene Expression. 3. Bioinformática/Bioinformatic. 4. Transcriptoma/Transcriptome. 5. Técnicas de Apoio para a Decisão/Decision Support Techniques

Devemos aceitar a decepção finita, mas nunca perder a esperança infinita"

Martin Luther King Jr.

*Este trabalho é dedicado aos meus pais e meus irmãos
Que sempre me deram suporte.*

AGRADECIMENTOS

À Deus por ter me dado boas oportunidades e colocado pessoas importantes no meu caminho.

À minha família por todo o suporte, confiança e conselhos nas minhas decisões.

Ao meu orientador, Dr Israel Tojal da Silva, pela confiança, atenção e disponibilidade sempre que necessário.

Ao time do Laboratório de Bioinformática e Biologia Computacional, ou seja, Alexandre Defelicibus, Bruno Moda, Felipe Azevedo, Irina Bobrovnitchaia, Jaqueline Ramalho, Monize Provisor, Renan Valieris, Rodrigo Borges e Rodrigo Drummond que, por ser formado de bons profissionais, pode me auxiliar neste trabalho.

À Dra. Dirce Maria Carraro, Dr. Joao Paulo da Silveira Nogueira Lima, Dr. Vladimir Cláudio Cordeiro de Lima e Dr. Kenneth John Gollob, Kivvi Duarte de Mello, Rafael Canfield Brianese, que contribuíram de diferentes formas para o desenvolvimento deste trabalho.

À Fundação Antônio Prudente por todo suporte e ensino.

À CNPq, pelo fornecimento da bolsa de estudos (processo nº 131766/2017-1)

RESUMO

Gonçalves DG. **Caracterização do infiltrado inflamatório em câncer de mama triplo-negativo e sua implicação prognóstica.** São Paulo; 2019. [Dissertação de mestrado - Fundação Antônio Prudente]

O câncer de mama triplo-negativo (**CMTN**) é uma doença heterogênea tratada basicamente com quimioterapia e, possui o pior prognóstico entre os cânceres de mama. Embora haja dados mostrando que o perfil de infiltrado inflamatório (**INF**) em tumores sólidos apresenta uma função importante, seu papel como fator prognóstico ainda não é completamente estabelecido no **CMTN**. Neste trabalho, foi desenvolvida uma abordagem supervisionada para detecção de uma assinatura multigênica com 50 genes capaz de identificar pacientes de alto (**PP**) e baixo (**GP**) risco de óbito em até 3 anos. A validação da assinatura ocorreu em duas coortes distintas (tanto para pacientes com os estadiamentos I, II e III, como para os pacientes com estadiamentos I e II) e suas classificações foram independentes de outras variáveis clínicas. Adicionalmente, foram analisados os perfis imunológicos nos grupos classificados, onde um dos achados foi a maior proporção de macrófagos **M1** em pacientes do grupo **GP**. O presente trabalho auxiliou na estratificação de risco das pacientes em conjunto com a caracterização dos aspectos moleculares e imunológicos associados ao prognóstico de pacientes com **CMTN**, o que pode permitir a elaboração de um tratamento mais personalizado.

Descritores: Neoplasias de Mama Triplo Negativas/diagnóstico. Expressão Gênica. Bioinformática. Transcriptoma. Técnicas de Apoio para a Decisão.

SUMMARY

Gonçalves DG. **Caracterização do infiltrado inflamatório em câncer de mama triplo-negativo e sua implicação prognóstica**. São Paulo; 2019. [Dissertação de mestrado - Fundação Antônio Prudente]

Triple-negative breast cancer (**TNBC**) is a heterogeneous disease that is basically treated with chemotherapy and has the worst prognosis among breast cancers. Although there are indications that the inflammatory infiltrate (**INF**) profile in these tumors plays an important role, its role as a prognostic factor still poorly established in **TNBC**. In this work, a supervised approach was developed to detect a multigenic signature with 50 genes capable of identifying patients with high (**PP**) and low (**GP**) risk of death until 3 years. The signature was validated in two distinct cohorts (both for patients with Stages I, II and III, and for patients with Stages I and II) and their classifications were independent of other clinical variables. In addition, the immunological profiles were analyzed in the classified groups, where one of the findings was the highest proportion of **M1** macrophages in **GP** patients. The present study assisted in the risk stratification of patients together with the characterization of the molecular and immunological aspects associated with prognosis in **TNBC** patients, which may allow the elaboration of a more personalized treatment.

Key-words: Triple Negative Breast Neoplasms/diagnosis. Gene Expression. Bioinformatic. Transcriptome. Decision Support Techniques

LISTA DE FIGURAS

Figura 1	Estágios em que ocorre o desenvolvimento do câncer na maioria das vezes.	2
Figura 2	Incidência de câncer no Brasil e no Mundo	3
Figura 3	Distribuição das taxas de incidência e mortalidade de câncer.....	3
Figura 4	Subtipos moleculares de câncer de mama.....	5
Figura 5	Sobrevida global de pacientes com câncer de mama triplo- negativo quando comparada a pacientes com outros subtipos	7
Figura 6	Taxa de recorrência em câncer de mama triplo-negativo.....	8
Figura 7	Classificação molecular de amostras de câncer de mama das coortes do TCGA, METABRIC e A.C.Camargo Cancer Center.....	16
Figura 8	Etapas de elaboração de assinatura multigênica.....	20
Figura 9	Etapas para a classificação utilizando a assinatura multigênica.....	21
Figura 10	Sobrevida global dos subtipos moleculares das pacientes do TCGA..	24
Figura 11	Sobrevida global das pacientes com CMTN a partir dos dados do TCGA	25
Figura 12	Sobrevida global dos subtipos moleculares das pacientes do METABRIC	26

Figura 13	Sobrevida global das pacientes com CMTN a partir dos dados do METABRIC	26
Figura 14	Sobrevida global das pacientes com CMTN nos dados do A.C.Camargo Cancer Center.....	27
Figura 15	Curva de sobrevida e o respectivo perfil de expressão gênica das pacientes classificadas como PP e GP de acordo com a assinatura com 50 genes (Estadiamento I)	31
Figura 16	Curva de sobrevida e o respectivo perfil de expressão gênica das pacientes classificadas como PP e GP de acordo com a assinatura com 50 genes (Estadiamentos I e II)	32
Figura 17	Curva de sobrevida e o respectivo perfil de expressão gênica das pacientes classificadas como PP e GP de acordo com a assinatura com 50 genes (Estadiamentos I, II e III)	34
Figura 18	Subdivisão das pacientes CMTN em 3 grupos (GP, PPB10 e PPO) ..	36
Figura 19	Análise de enriquecimento da assinatura de 50 genes.....	37
Figura 20	Comparação da proporção de 22 tipos de leucócitos entre os grupos PP e GP das coortes do TCGA e METABRIC.....	38
Figura 21	Comparação da proporção entre os 22 tipos de leucócitos nos grupos PP e GP das coortes do TCGA e METABRIC.....	40
Figura 22	Avaliação da sobrevida das pacientes com CMTN nas coortes TCGA, METABRIC e A.C.Camargo Cancer Center de acordo com os grupos definidos pelo software <i>TNBCtype</i>	42

LISTA DE TABELAS

Tabela 1 Dados clínicos de pacientes com CMTN das coortes TCGA, METABRIC e A.C.Camargo Cancer Center.. 28

Tabela 2 Dados clínicos de pacientes com CMTN das coortes TCGA, METABRIC e A.C.Camargo Cancer Center (estadiamentos I, II e III) associados à classificação utilizando a assinatura de 50 genes... 29

LISTA DE SIGLAS E ABREVIATURAS

CMTN	Câncer de Mama Triplo-Negativo
ER	Receptor de Estrogênio
GP	Bom Prognóstico
HER2	Receptor da proteína HER2
HR	Receptor Hormonal
IHC	Imuno-histoquímica
INF	Infiltrado Inflamatório
PP	Mau Prognóstico
PPB10	Pior Prognóstico
PPO	Prognóstico Intermediário
PR	Receptor de Progesterona
prop.test	Teste de proporção de óbito até 3 anos

SUMARIO

1	INTRODUÇÃO	1
2	OBJETIVOS	12
2.1	Objetivo geral.....	12
2.2	Objetivos específicos	12
3	METODOLOGIA	13
3.1	Dados moleculares e clínicos.....	13
3.1.1	Coorte pública.....	13
3.1.2	Coorte do A.C.Camargo Cancer Center.....	14
3.1.3	Classificação molecular das coortes	14
3.2	Análise estatística.....	16
3.2.1	Análise de proporção (prop.test).....	16
3.2.2	Regressão de Cox	17
3.3	Classificação dos tumores de pacientes com CMTN em 6 subtipos moleculares (TNBCtype).....	17
3.4	Análise de enriquecimento.....	17
3.5	Infiltrado inflamatório.....	18
3.6	Criação da assinatura multigênica.....	19
3.7	Classificação das pacientes	20
3.8	Análise de sobrevida (Curvas de Kaplan-Meier).....	21
3.9	Classificação das pacientes em 3 grupos	22
4	RESULTADOS	23
4.1	Análise de sobrevida global	23
4.2	Características clínicas das coortes (TCGA, METABRIC e A.C.Camargo Cancer Center)	27
4.3	Dados clínicos dos grupos GP e PP	28
4.4	Sobrevida global e perfil de expressão nos estadiamentos iniciais.....	29

4.4.1	Estadiamento I	30
4.4.2	Estadiamentos I e II.....	31
4.5	Sobrevida global e perfil de expressão nos estadiamentos I, II e III	33
4.6	Divisão das pacientes com CMTN em 3 grupos	35
4.7	Análise de enriquecimento.....	37
4.8	Infiltrado inflamatório.....	38
4.9	Classificação <i>TNBCtype</i>	41
5	DISCUSSÃO.....	44
6	CONCLUSÃO.....	46
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	47

1 INTRODUÇÃO

O câncer representa um conjunto de doenças caracterizadas pela proliferação anormal das células, com perda da diferenciação e aquisição de potencial invasivo e metastático (National Cancer Institute-NCI 2015).

Sabe-se que células danificadas tendem a sofrer um processo denominado apoptose, ou morte celular programada, tendo por consequência a substituição por novas células. No entanto, determinadas mutações gênicas são capazes de desencadear um crescimento celular anormal, sendo este crescimento desordenado um fator chave para o desenvolvimento do câncer (NCI 2015).

O processo de desenvolvimento do câncer ocorre, na maioria das vezes, em quatro estágios principais, embora estas etapas necessariamente não ocorram de maneira sequencial e ordenada (Figura 1): **(i)** Hiperplasia, caracterizado pelo aumento benigno de um tecido devido ao acréscimo na taxa de divisão celular, porém as células mantêm aparência normal sob o microscópio; **(ii)** Displasia, estágio onde ocorre o aumento na taxa de divisão celular, e as células passam a adquirir aparência anormal ao microscópio. Além disso, normalmente, existe uma relação direta entre as chances da formação do câncer e o nível de anormalidade das células ou tecido; **(iii)** câncer in situ, conhecido como o estágio no qual as células tumorais estão ainda limitadas pela lâmina basal e não se espalharam para outros órgãos e **(iv)** invasivo, um câncer caracterizado pelos deslocamento das células do sítio de desenvolvimento (tumor primário) para regiões não contíguas do mesmo órgão ou para órgãos à distância. Caso essas células se espalhem para outros órgãos ou tecidos através da rede linfática ou sanguínea, não será somente câncer invasivo, mas também metastático (NCI 2015).

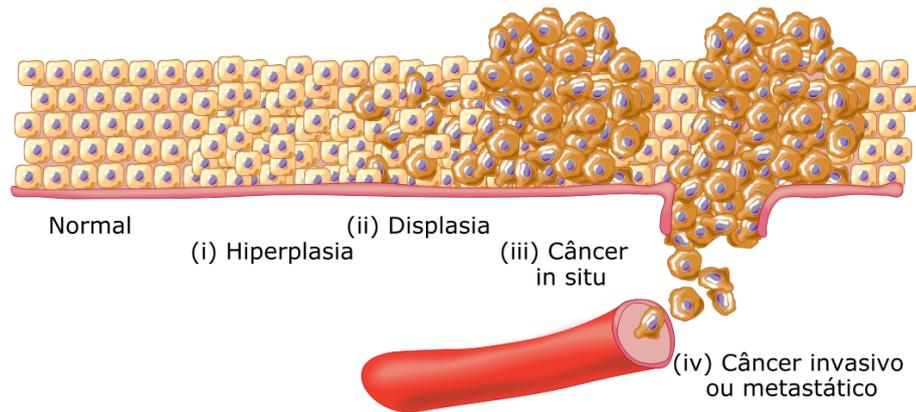
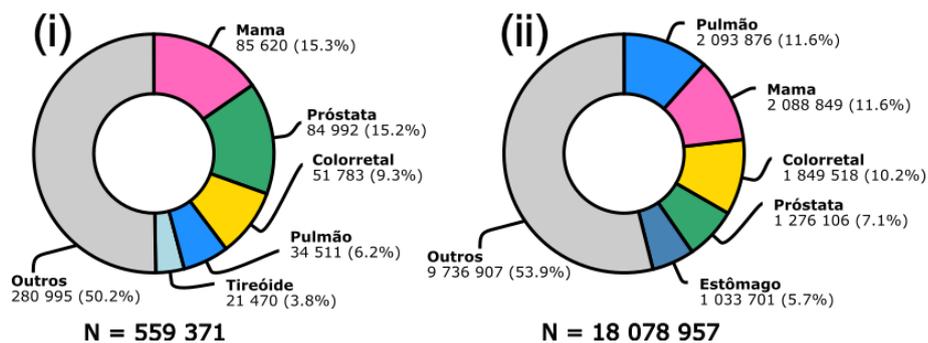


Figura 1 – Estágios em que ocorre o desenvolvimento do câncer na maioria das vezes. Estas etapas decorrem à partir de células normais, onde ocorre um aumento benigno de um tecido dado o acréscimo da divisão celular (i), seguido da anormalidade na aparência das células ante o microscópio (ii), ocasionalmente se desenvolvendo em um câncer restrito à lâmina basal (iii) e posteriormente se espalhando para regiões não contíguas do mesmo órgão ou para órgãos à distância.

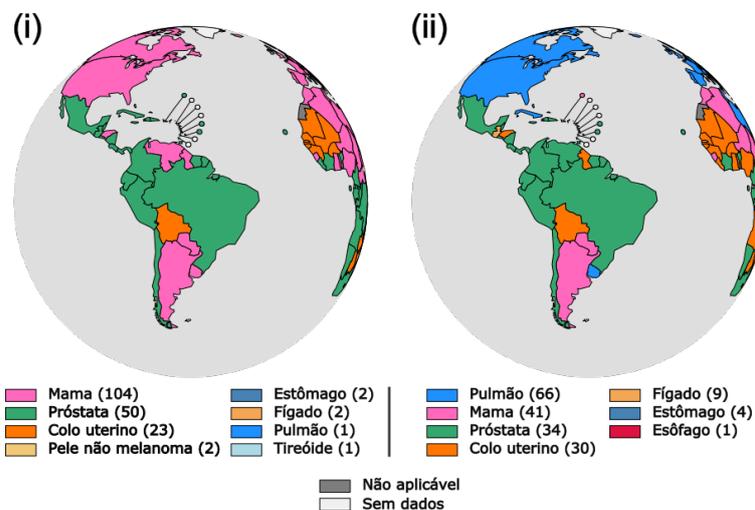
O câncer é a segunda maior causa de morte do mundo e, no ano de 2018, cerca de 9,6 milhões de pessoas morreram dessa doença (1 em 6 mortes). No Brasil, o câncer foi responsável por mais de 240 mil óbitos (World Health Organization-WHO 2018). A Figura 2 apresenta as taxas de incidência dos diferentes tipos de câncer no Brasil (i) e no mundo (ii). Em 2018, os cânceres de mama (15,3%), próstata (15,2%) e colorretal (9,3%) foram os mais incidentes no Brasil. No mesmo ano, o mais incidente no mundo foi o câncer de pulmão (11,58%), seguido do câncer de mama (11,55%) e câncer colorretal (10,2%) (GLOBOCAN 2018).



Fonte: GLOBOCAN (2018) - Adaptado.

Figura 2 – Incidência de câncer no Brasil (i) e no Mundo (ii).

De acordo com a Figura 3, ao avaliar os tipos de câncer com maior incidência e mortalidade por país, nota-se que o câncer de mama é o tipo mais incidente em 104 países, possuindo a maior taxa de mortalidade em 41 países (GLOBOCAN 2018).

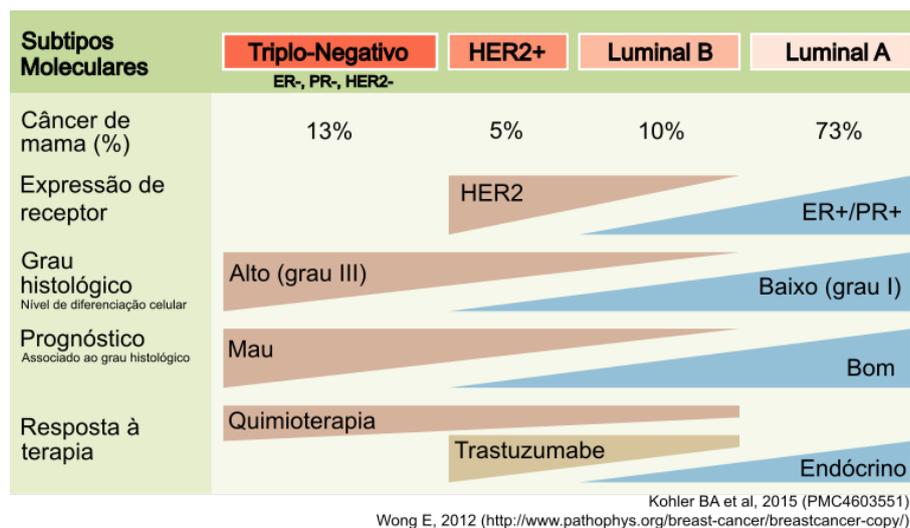


Fonte: GLOBOCAN (2018) - Adaptado.

Figura 3 – Distribuição das taxas de incidência (i) e mortalidade (ii) de câncer. Os números entre parênteses referem-se ao total de países com maior incidência ou mortalidade.

O câncer de mama é uma doença heterogênea com vários subtipos moleculares, possuindo uma grande variação em suas características biológicas e nos perfis genéticos (ANDERSON et al. 2014). A classificação molecular por imuno-histoquímica (IHC - Técnica para detectar moléculas de interesse

dentro dos tecidos utilizando anticorpos (NATURE-SUBJECTS-IHC 2019)) do câncer de mama baseia-se na análise de 4 proteínas: os receptores hormonais de estrogênio (**ER**) e progesterona (**PR**), o receptor **HER2** (receptor 2 do fator de crescimento epidérmico humano) e a proteína nuclear **KI-67**, utilizada como marcador de proliferação celular (GERDES et al. 1983; DOWSETT et al. 2011; CUYLEN et al. 2016). Na prática clínica, três destes receptores (**ER**, **PR** e **HER2**) são utilizados na classificação dos subtipos moleculares de câncer de mama (KOHLENER et al. 2015). Como pode-se notar na Figura 4, os subtipos são compostos pelas diferentes combinações entre receptores hormonais (**ER** e **PR**) e a proteína **HER2**. Para a positividade de receptores hormonais, basta que um deles (**ER** ou **PR**) seja positivo. Já para que um tumor seja considerado receptor hormonal negativo, ambos os receptores hormonais (**ER** e **PR**) devem ser negativos.



Fonte: Adaptado de WONG (2012).

Figura 4 – Subtipos moleculares de câncer de mama. A figura mostra a frequência estimada de cada subtipo de câncer de mama, a expressão esperada de cada marcador imuno-histoquímico, o grau histológico, o prognóstico esperado e a resposta habitual à terapia. A primeira linha é referente à incidência dos subtipos moleculares de câncer de mama no mundo (**triplo-negativo**=13%, **HER2+**=5%, **Luminal B**=10%, **Luminal A**=73%). A segunda linha é referente à expressão dos receptores em cada subtipo molecular (**triplo-negativo**: não possui expressão dos receptores hormonais e a proteína **HER2**; **HER2+**: não possui expressão dos receptores hormonais e possui a expressão da proteína **HER2**; **Luminal B**: possui expressão da proteína **HER2** e dos receptores hormonais; **Luminal A**: possui a expressão dos receptores hormonais e não possui a expressão da proteína **HER2**). A terceira linha é referente ao grau histológico dos subtipos moleculares, onde o subtipo triplo-negativo possui o maior grau histológico, enquanto o subtipo Luminal A possui o menor grau histológico. A quarta linha é referente ao prognóstico de cada subtipo molecular, onde o subtipo triplo-negativo possui o pior prognóstico e o subtipo Luminal possui o melhor prognóstico. A quinta linha é referente à terapia utilizada em cada subtipo molecular (**triplo-negativo**: quimioterapia; **HER2+**: quimioterapia e trastuzumabe; **Luminal B**: quimioterapia, trastuzumabe e terapia endócrina; **Luminal A**: terapia endócrina).

Conforme a Figura 4, o subtipo triplo-negativo (**CMTN**) é responsável por cerca de 13% dos casos de câncer de mama, embora não seja o mais representativo, apresenta o pior prognóstico, alto grau histológico e não possui tratamento molecular específico, sendo tratado com quimioterapia apenas, demonstrando maior sensibilidade a este tratamento do que os subtipos luminais (CAREY et al. 2007). Por outro lado, os subtipos com receptor **HER2** positivo podem ser tratados com terapia HER2-dirigida (trastuzumabe, pertuzumabe, TDM1, la-

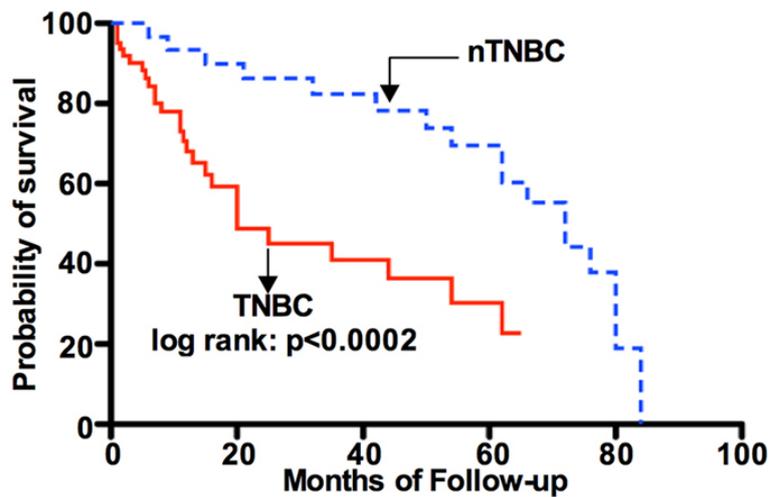
patinibe, neratinibe) (GENENTECH-PRESS-RELEASES 2006; BRANDAO et al. 2018). Adicionalmente, os subtipos hormônio positivo podem ser tratados com tratamentos endócrinos, como o tamoxifeno, que geralmente possui baixa toxicidade e reduzem a mortalidade por câncer de mama neste subgrupo (LUMACHI et al. 2011).

Os principais fatores de risco para o câncer de mama são: menarca precoce, menopausa tardia, terapia de reposição hormonal prolongada, mutação germinativa nos genes *BRCA1* e *BRCA2* e histórico familiar positivo, sendo os dois últimos considerados os mais relevantes (SINGLETARY 2003).

Com o avanço das técnicas de sequenciamento, foi possível estudar vários aspectos moleculares de diversos tipos de câncer, resultando em uma grande quantidade de dados a respeito de mutações somática e germinativa, expressão gênica, metilação, alteração do número de cópia de genes e outros eventos. A disponibilidade desses dados em repositórios públicos permite a investigação de questões adicionais. Um exemplo de repositório é o **TCGA** (*The Cancer Genome Atlas*), que possui dados foram gerados principalmente por dois estudos (KOBOLDT et al. 2012; CIRIELLO et al. 2015) que utilizaram expressão gênica, *microRNA*, metilação de *DNA*, entre outras abordagens para discriminar características clínicas e moleculares dos subtipos histológicos (**Carcinoma Lobular Invasivo** e **Carcinoma Ductal Invasivo**) (CIRIELLO et al. 2015) e moleculares (**CMTN**, **HER2+** e **Luminais A e B**) (KOBOLDT et al. 2012) de câncer de mama. Além disso, dados de cerca de 1.090 pacientes de câncer de mama estão disponíveis neste repositório, abrindo oportunidades para a comunidade científica investigar, por exemplo, os mecanismos moleculares associados à tumorigênese e desfecho clínico, dentre outros.

Uma análise extensa desse conjunto de dados (**TCGA**) foi o trabalho de Noman et al. (2016), que estudou uma coorte de 400 pacientes com câncer de mama disponíveis no **TCGA** (Figura 5). Neste estudo, foi observado que pacientes com **CMTN** apresentaram pior prognóstico quando comparados aos

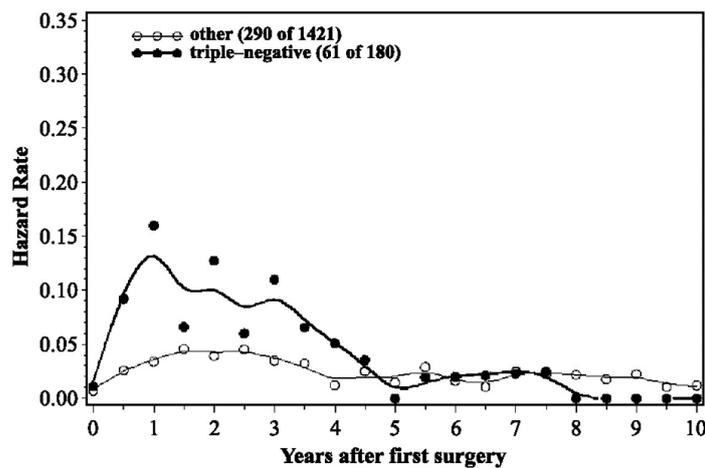
demais subtipos, onde a maior taxa de mortalidade foi vista nos primeiros meses, sendo a mediana de sobrevida das pacientes com CMTN igual a 20 meses (NOMAN et al. 2016).



Fonte: NOMAN et al. (2016).

Figura 5 – Sobrevida global de pacientes com câncer de mama triplo-negativo quando comparada a pacientes com outros subtipos.

Adicionalmente, alguns estudos indicam que 75% das recorrências ocorrem nos 3 primeiros anos e o risco de recorrência distante em CMTN atinge o seu pico próximo a 3 anos, reduzindo consideravelmente após este período (Figura 6) (DENT et al. 2007; ROOZENDAAL et al. 2016).



Fonte: DENT et al. (2007).

Figura 6 – Taxa de recorrência em câncer de mama triplo-negativo. O CMTN tende a apresentar maior taxa de recorrência, embora esse risco decline rapidamente após o terceiro ano de seguimento, ao contrário dos demais subtipos que, apesar de apresentarem um risco de recorrência menor, mantêm um risco residual de recorrência mesmo após 5 anos de seguimento.

Portanto, ao analisar o prognóstico das pacientes **CMTN**, existe um grupo de pacientes com menor taxa de sobrevivência e maior taxa de recorrência nos primeiros anos. Ou seja, existem ao menos dois grupos de pacientes distintos dentro do subtipo **CMTN** em relação ao prognóstico, o que demonstra a heterogeneidade deste subtipo.

Atualmente, a fim de direcionar o tratamento de pacientes com câncer de mama, alguns centros solicitam a análise de assinatura multigênica *MammaPrint* (BRANDAO et al. 2019) que, ao avaliar uma assinatura de 70 genes, é capaz de prever um escore associado a maior ou menor risco de recorrência de doença com menos de 5 anos. Esta abordagem, por sua vez, é sugerida nos guidelines da Sociedade Clínica Americana de Oncologia (**ASCO**). No entanto, se a paciente possuir câncer triplo-negativo, o resultado do *MammaPrint* é pouco útil para guiar decisões na aplicação de quimioterapia sistêmica adjuvante (KROP et al. 2017; VIEIRA e SCHMITT 2018).

Outra assinatura utilizada na clínica para câncer de mama é o *Oncoty-*

peDx, que utiliza 21 genes para distinguir pacientes entre alto ou baixo risco de recorrência quando tratados com tamoxifeno. Além disso, esta abordagem é também presente nos guidelines do Comitê Americano de Câncer (**AJCC**) e da **ASCO**. Novamente, a abordagem não deve ser utilizada na tomada de decisão quanto ao tratamento de **CMTN**, pois esta assinatura foi desenvolvida para câncer de mama hormônio positivo (HARRIS et al. 2016; VIEIRA e SCHMITT 2018). Dessa forma, apesar destas abordagens que utilizam assinatura multigênica possuírem aplicações em outros subtipos de câncer de mama, acabam por não contribuir significativamente na tomada de decisão quanto ao tratamento de **CMTN**.

Além destas assinaturas, outras duas utilizadas na prática clínica são a Prosigna e Endopredict (VIEIRA e SCHMITT 2018). A assinatura Prosigna possui 50 genes do **PAM50** (genes utilizados na classificação de subtipos moleculares de câncer de mama a partir de *RNA-Seq* (PARKER et al. 2009)), além de 8 genes *housekeeping* e serve como indicador para ausência de recorrência à distância em 10 anos em mulheres pós-menopáusicas com cancer de mama hormônio positivo a ser tratado com terapia endócrina adjuvante (PROSIGNA 2018). A assinatura Endopredict utiliza 12 genes e calcula um escore de risco, que pode ser utilizado juntamente com a informação do tamanho do tumor podendo guiar as decisões quanto a quimioterapia e terapia anti-hormonal (VIEIRA e SCHMITT 2018).

Outra abordagem na busca de um tratamento específico para o **CMTN** envolve o estudo do perfil molecular. Neste sentido, dada a heterogeneidade do subtipo **CMTN**, existem 3 abordagens principais (AHN et al. 2016) que buscam estratificar o subtipo **CMTN**: o subtipo do centro Unicancer na França, o subtipo da universidade de Baylor e o subtipo da universidade Vanderbilt. Esses 3 subtipos utilizaram dados de *mRNA* para um agrupamento hierárquico não supervisionado como etapa inicial, e, embora estes subtipos apresentem basicamente 4 perfis principais (basal, mesenquimal, receptor de andrógeno luminal e

imuno-modulador) seus painéis gênicos são distintos. A abordagem realizada por Lehmann e colaboradores (Vanderbilt) foi proposta em 2011 com o intuito de determinar perfis moleculares, onde foi analisado 587 amostras de **CMTN** por expressão gênica, subclassificando-as em 6 subtipos moleculares (LEHMANN et al. 2011). A partir deste estudo, foi disponibilizado uma ferramenta *online* (VANDERBILT 2018) para a classificação de amostras de **CMTN** através de dados de expressão gênica. Contudo, segundo Ahn et al. (2016), as diferenças na taxa de sobrevida global não foram suficientes para estratificar subtipos moleculares (AHN et al. 2016), apesar de alguns resultados indicarem seu potencial como preditor para sensibilidade à quimioterapia neoadjuvante (MASUDA et al. 2013).

Adicionalmente, devido a falta de tratamento molecular específico para pacientes com **CMTN**, existem esforços para tentar melhor compreender o papel do infiltrado inflamatório (**INF**) no microambiente tumoral. Nesse sentido, alguns estudos demonstram que altas quantidades de linfócitos T CD8+ no microambiente tumoral estão associadas à um melhor prognóstico em pacientes com **CMTN** (MATSUMOTO et al. 2015; SOTO-PEREZ-DE-CELIS et al. 2017). Em contrapartida, neutrófilos têm sido associados a um resultado desfavorável em diversos tumores, tais como câncer gástrico, colorretal, hepatocelular, cervical e carcinoma de células renais (SOTO-PEREZ-DE-CELIS et al. 2017). Adicionalmente, os linfócitos B também podem interferir negativamente através da promoção do crescimento tumoral pela produção de linfotoxinas que induzem angiogênese (YUEN et al. 2016).

Outras células possuem resultados diversos quanto à sua associação com o prognóstico de pacientes com câncer, sendo por vezes associados a um prognóstico favorável enquanto em outros estudos são associados a um pior resultado da paciente. Um tipo celular que apresenta este perfil é o linfócito T regulador, que é responsável pela regulação da resposta imune. Outra célula que atua de forma semelhante são os macrófagos, onde macrófagos M1 atuam no combate tumoral

através de sua atividade pró-inflamatória, promovendo respostas anti-tumorais, enquanto macrófagos M2 possuem atividade anti-inflamatória (YU e DI 2017).

Em suma, o câncer de mama triplamente negativo é uma doença heterogênea e ainda sem tratamento que leva em consideração as informações moleculares disponíveis. Esforços atuais são direcionados para se estratificar diferentes grupos de pacientes com baixo e alto risco a fim de direcionar tratamentos mais personalizados (LEHMANN et al. 2011). Neste cenário, abordagens que utilizam assinaturas gênicas e são capazes de detectar previamente pacientes de bom e mau prognóstico são promissoras e são o tema principal do presente trabalho. De maneira complementar, entender os aspectos imunológicos dessas pacientes representa uma oportunidade para compreender mecanismos subjacentes e possíveis marcadores envolvidos em respostas imunológicas auxiliando, assim, na escolha de pacientes que possam se beneficiar de tratamentos com imunoterápicos.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Identificar uma assinatura multigênica capaz de predizer pacientes com **CMTN** de baixo e alto risco de óbito em até 3 anos.

2.2 OBJETIVOS ESPECÍFICOS

- Estabelecer uma abordagem computacional para identificar genes que estratificam grupos de pacientes portadores de **CMTN** de acordo com a sobrevida global;
- Validar a assinatura multigênica identificada por este trabalho em coortes distintas;
- Verificar o papel funcional através da análise de enriquecimento da assinatura multigênica identificada;
- Estimar o diversidade do infiltrado inflamatório nos grupos identificados pelo classificador.

3 METODOLOGIA

3.1 DADOS MOLECULARES E CLÍNICOS

3.1.1 Coorte pública

Neste estudo utilizou-se dados moleculares e clínicos de pacientes diagnosticados com **CMTN** disponíveis nos bancos de dados públicos **TCGA** (GDC-v12) e **METABRIC**.

3.1.1.1 TCGA

As informações moleculares de expressão gênica e informações clínicas de pacientes com câncer de mama não metastáticos do repositório **TCGA** (*The Genome Cancer Atlas*) foram geradas principalmente por dois estudos (KOBOLDT et al. 2012; CIRIELLO et al. 2015). Resumidamente, a *pipeline* de bioinformática utilizada partiu dos dados de *RNA-Seq* no formato **fastQ** (contém sequências de nucleotídeos e seus respectivos índices de qualidade), seguido do alinhamento ao genoma utilizando o *software* **STAR2** (DOBIN et al. 2013), realizando a contagem da expressão dos genes utilizando HTSeq [Anders et al., 2015] e, finalmente, normalizando pela métrica **FPKM** (normaliza pelo tamanho da biblioteca e o tamanho dos genes) (TCGA-PIPELINE 2018).

3.1.1.2 METABRIC

Uma segunda coorte foi utilizada para validar a assinatura detectada na coorte **TCGA**. Assim, foram baixados os dados clínicos e de expressão gênica das pacientes com câncer de mama não metastáticos investigados no projeto **METABRIC** (LIU et al. 2018), disponível no site cBioPortal (METABRIC-

DADOS 2018).

3.1.2 Coorte do Hospital A.C.Camargo Cancer Center

As amostras sequenciadas pelo projeto temático do Hospital **A.C.Camargo Cancer Center (ACCCC)**, aprovado pela FAPESP (2014/22589-9) e pelo Comitê de Ética e Pesquisa da **A.C.Camargo Cancer Center** (1746-13C), sob coordenação da Dra. Dirce Maria Carraro, foram utilizadas como uma segunda coorte para validação. Ao total, foram sequenciadas 25 amostras tumorais de câncer de mama triplo-negativo não metastático pela plataforma Illumina-NextSeq 500, disponível no Núcleo de Genômica do Centro Internacional de Pesquisas (**CIPE**) do **A.C.Camargo Cancer Center**.

Após o sequenciamento, foi realizado um pré-processamento onde as sequências anotadas aos genes ribossomais foram removidas. As sequências remanescentes foram alinhadas ao genoma humano (**HG38**) pelo *software* **STAR2** (DOBIN et al. 2013). Somente sequências mapeadas unicamente aos genes anotados (ENSEMBL Genes build GRCh38.p12) foram mantidas. O resultado do alinhamento compreende um arquivo no formato **BAM** (*Binary Alignment Map*). Esses arquivos foram utilizados pelo programa **HTSeq** (ANDERS et al. 2015) para quantificar o total de sequências de cada gene. Ao final, esses valores foram normalizados por meio da métrica **FPKM** (*Fragments Per Kilobase of exon per Million reads mapped*).

3.1.3 Classificação molecular das coortes

Como os dados clínicos do **METABRIC** e do **TCGA** não possuem informações sobre a proteína nuclear *KI-67*, a abordagem utilizada para a definição dos subtipos moleculares das amostras baseou-se na análise por **IHC** de três

marcadores proteicos, sendo eles o receptor de estrogênio (**ER**), receptor de progesterona (**PR**) e **HER2**, validado com hibridização *in situ* da proteína **HER2**.

A abordagem utilizada resultou em 4 subtipos moleculares, incluindo a combinação de hormônio positivo (**HR+**) ou negativo (**HR-**) e **HER2** positivo (**HER2+**) e negativo (**HER2-**). Além disso, as pacientes que não possuíam informações quanto a positividade ou negatividade dos hormônios e da proteína **HER2** foram classificadas como **NA** (*Not Available*).

O total de amostras de cada subtipo molecular do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** estão disponíveis na Figura 7. Para a caracterização dos dados, foram inicialmente selecionadas pacientes com **HER2+** ou **HER2-** e depois verificado se cada paciente era **HR+** ou **HR-**.

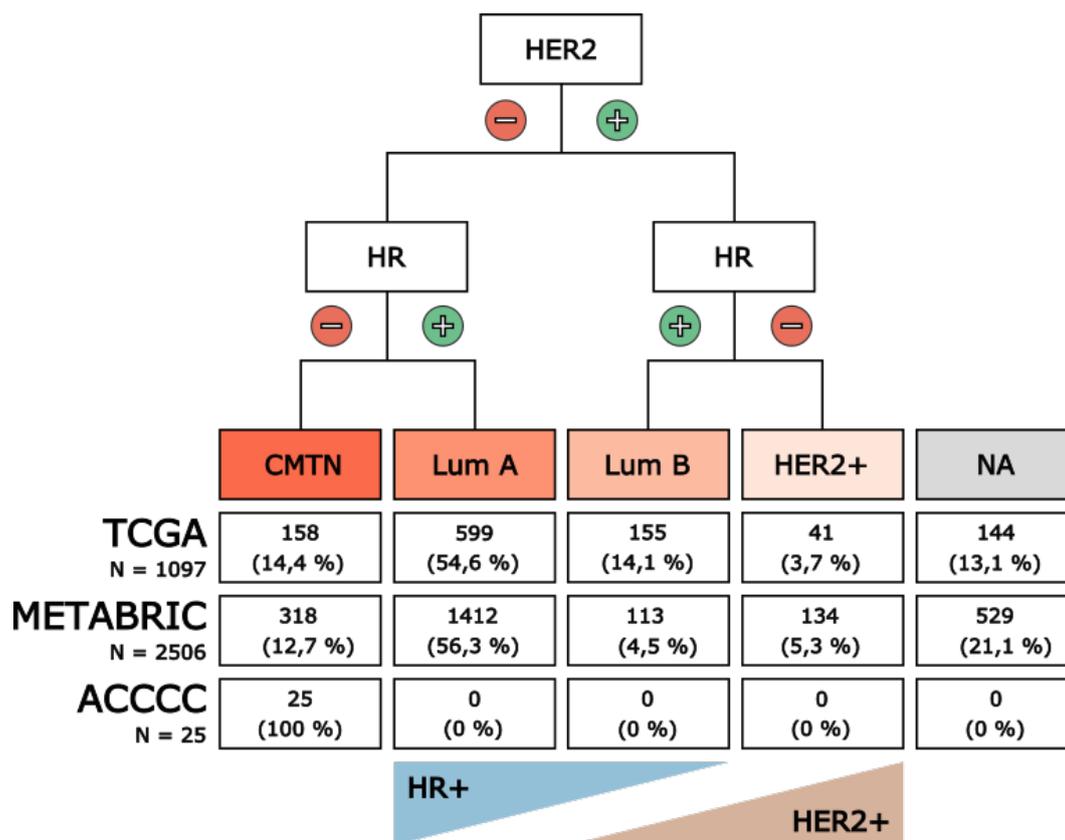


Figura 7 – Classificação molecular de amostras de câncer de mama das coortes do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center**. Através dos dados clínicos, foram verificados quais pacientes possuem expressão da proteína **HER2**, e posteriormente quais possuem expressão dos receptores hormonais (**ER** e **PR**). Aquelas pacientes que não possuem informação sobre a proteína **HER2** ou os receptores hormonais foram classificadas como **NA** (*Not Available*). Estas análises resultaram em 4 subtipos moleculares: **CMTN**; **Lum A** (**Luminal A**); **Lum B** (**Luminal B**) e **HER2+**.

3.2 ANÁLISE ESTATÍSTICA

3.2.1 Análise de proporção (prop.test)

As pacientes diagnosticadas com tumores triplo-negativos foram categorizadas em dois grupos de acordo com seu prognóstico e, descritos como segue: (i) **PP** (*Poor Prognosis*), grupo de pacientes que morreram em até 3 anos após o primeiro diagnóstico e, (ii) **GP** (*Good Prognosis*), pacientes com sobrevida superior a 3 anos. Para verificar a assertividade do método em classificar pacientes dentre esses grupos (**PP** e **GP**), foi utilizado o teste de proporção de *Fisher* (apresentado na curva de sobrevida *Kaplan-Meier*) para comparar quan-

tas pacientes foram classificadas corretamente, ou seja, verificar se o grupo **PP** estava enriquecido em pacientes que morreram em até 3 anos e se o grupo **GP** concentrava pacientes que viveram mais de 3 anos.

3.2.2 Regressão de Cox

Para avaliar se os grupos classificados como **PP** e **GP** identificam populações com prognóstico independente de outras variáveis clínicas, foi utilizado a análise multivariada de *Cox* (BENITEZ-PAREJO et al. 2011).

3.3 CLASSIFICAÇÃO DOS TUMORES DE PACIENTES COM CMTN EM 6 SUBTIPOS MOLECULARES (*TNBCTYPE*)

A classificação das pacientes **CMTN** do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** nos 6 subtipos moleculares propostos por (LEHMANN et al. 2011) foi feita através da ferramenta disponível *online* (VANDERBILT 2018). Esta ferramenta faz uso de uma matriz contendo dados de expressão gênica normalizados por *FPKM*. Baseado no conjunto de dados de expressão gênica desse estudo, foram propostos seis subtipos de **CMTN** descritos a seguir: dois subtipos basais (**BL1** e **BL2**), imunomodulador (**IM**), mesenquimal (**M**), mesenquimal-*like* (**MSL**) e luminal receptor de andrógeno (**LAR**).

3.4 ANÁLISE DE ENRIQUECIMENTO

O enriquecimento dos termos biológicos, associados aos genes das assinaturas, foi realizado pelo método *clusterProfiler* (YU et al. 2012), que calcula o teste de enriquecimento para os termos do *Gene Ontology* (**GO** - Uma iniciativa de bioinformática que fornece uma ontologia (vocabulário compartilhado) de termos definidos para representar propriedades específicas de genes (NATURE-SUBJECTS-GO 2019)) e vias do **KEGG** (*Kyoto Encyclopedia of Genes and Genomes*) com base na distribuição hipergeométrica. Para controlar as taxas de

falsos positivos (**FDR**, do inglês *False Discovery Rate*) em múltiplos testes, os valores **q** (Storey 2002) também foram estimados.

3.5 INFILTRADO INFLAMATÓRIO

A estimação da composição do **INF** foi realizada pelo *software* **CIBERSORT** (NEWMAN et al. 2015), disponível *online* (CIBERSORT 2019).

Os parâmetros utilizados na ferramenta **CIBERSORT** foram:

- Assinatura gênica capaz de estimar a proporção de 22 tipos de leucócitos (**LM22**);
- 1000 permutações;
- Escala em modo absoluto;
- Normalização por quartil desabilitada (recomendável para dados *RNA-Seq* utilizados no presente estudo).

Para estas análises foram utilizados os dados das coortes do **TCGA** e **METABRIC**, devido as suas classificações estratificarem em dois grupos com perfis de sobrevida distintos (prop.test: $p < 0,01$ em ambas as coortes).

Para as análises posteriores, foram consideradas amostras com significância na estimação ($p < 0,05$). Este valor indica o sucesso da quantificação das células na respectiva amostra.

Para avaliar a resposta imune presente em cada um dos grupos prognósticos, foram feitas duas análises:

A primeira análise é uma comparação (teste de *Mann-Whitney* com nível de significância $p < 0,05$) entre as proporções de cada um dos 22 tipos de leucócitos identificados pela ferramenta CIBERSORT nos grupos **PP** e **GP**. Exemplificando, é verificado se as proporções de macrófago M1 é superior no grupo **PP** em relação ao grupo **GP**.

A segunda análise é a comparação (teste de *Mann-Whitney* com nível de significância $p < 0,05$) entre os 22 tipos de leucócitos identificados pela ferramenta

CIBERSORT dentro de cada grupo (**PP** ou **GP**). Exemplificando, é analisado se a proporção de macrófago M1 é superior à proporção de macrófago M2 dentro do grupo **PP**.

3.6 CRIAÇÃO DA ASSINATURA MULTIGÊNICA

A criação da assinatura multigênica segue uma estrutura de filtros e testes estatísticos a fim de obter dois grupos de genes com perfis distintos capazes de identificar pacientes associadas a um alto (**PP**) e baixo risco (**GP**) de óbito em até 3 anos.

O ponto de corte para a divisão dos grupos foi em 3 anos devido à alta taxa de óbitos (Figura 5) e à alta taxa de recorrência (Figura 6) até este período.

A partir de uma análise supervisionada, considerando informações clínicas e de sobrevida global, foi realizada uma estratificação de dois grupos de pacientes denotadas como **GP** e **PP** (Figura 8). O grupo **PP** são as pacientes que morreram até 3 anos, enquanto **GP** são as pacientes que viveram mais que 3 anos (**i**). Em seguida, foram realizadas interações a fim de manter pacientes que compartilham perfis de expressão gênica mais similares dentro de cada grupo (**ii**). A partir desse conjunto de pacientes, o perfil de expressão de cada gene foi comparado entre os grupos **PP** e **GP** (**iii**). Finalmente, uma etapa adicional foi realizada a fim de remover genes com alta variabilidade (**iv**) por meio do coeficiente de *Gini* (O'HAGAN et al. 2018).

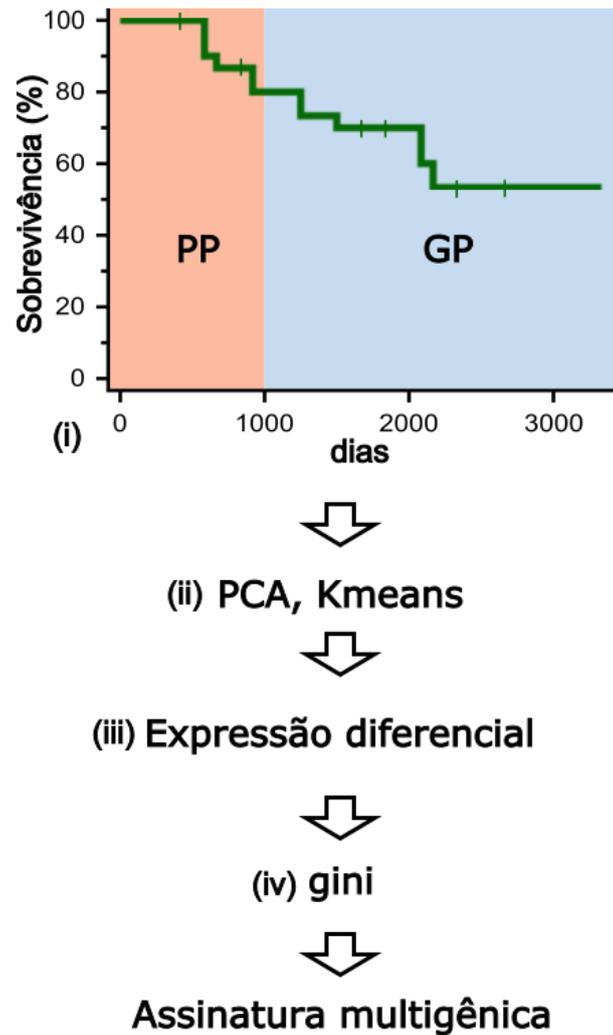


Figura 8 – Etapas de elaboração de assinatura multigênica. A criação da assinatura multigênica consiste em basicamente 4 etapas: (i) inicialmente, são definidos dois grupos de pacientes (**PP**: pacientes que morrem em até 3 anos; **GP**: pacientes que vivem mais do que 3 anos) utilizando os dados clínicos; (ii) em seguida, estes grupos de pacientes são filtrados utilizando PCA e *k-means*; (iii) após isso, é obtida a expressão diferencial utilizando o teste de *Mann-Whitney*; (iv) finalmente, são removidas pacientes que possuam alta variação nos genes utilizando o coeficiente de Gini.

3.7 CLASSIFICAÇÃO DAS PACIENTES

Para a classificação das pacientes, foram estabelecidas algumas etapas (Figura 9) que utilizam a assinatura multigênica gerada. A primeira etapa da classificação utiliza a assinatura multigênica para filtrar os dados de expressão gênica. Em seguida é feita uma normalização dos dados por *Z-Score*. Após isso, é feito a comparação por *Mann-Whitney* se o grupo de genes **highPP** (genes altamente expressos no grupo **PP** e pouco expressos no grupo **GP**) é

significativamente ($p < 0,05$) maior que o grupo de genes **lowPP** (genes altamente expressos no grupo **GP** e pouco expressos no grupo **PP**) e, neste caso, a paciente é classificada como **PP**, caso contrário, a paciente é classificada como **GP**.

Assinatura multigênica



Z-Score



Mann-whitney

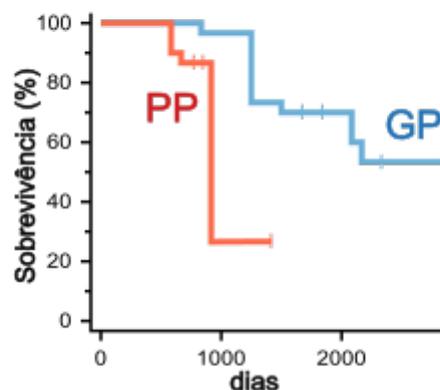


Figura 9 – Etapas para a classificação utilizando a assinatura multigênica. A classificação das pacientes utiliza a assinatura multigênica identificada (metodologia, seção 3.6), seguida da normalização por *Z-Score*. Em seguida é comparado por *Mann-Whitney* se os genes **highPP** possuem maior expressão ($p < 0,05$) do que o grupo de genes **lowPP**, caso sejam, então a paciente é classificada como **PP**, caso contrário é classificada como **GP**.

3.8 ANÁLISE DE SOBREVIDA (CURVAS DE *KAPLAN-MEIER*)

Desfechos clínicos como sobrevida global sobrevida livre de progressão e tempo para progressão são os eventos mais usados em oncologia. O estimador de *Kaplan-Meier* (GOEL et al. 2010) é um método não paramétrico baseado em dados quantitativos que gera uma função de distribuição do tempo até a

ocorrência desses eventos. No presente trabalho foi utilizado este estimador para avaliar a sobrevida global das pacientes previamente classificadas como **PP** ou **GP**.

3.9 CLASSIFICAÇÃO DAS PACIENTES EM 3 GRUPOS

A fim de se analisar se há subgrupos dentre as pacientes classificadas como **PP**, foi feita uma inferência utilizando-se os valores de **p** resultantes do teste de *Mann-Whitney*. Através destes valores de **p**, divide-se as pacientes **PP** em dois grupos, o primeiro grupo sendo formado de pacientes com valor de **p** contidos no primeiro decil (**PPB10**), ou seja, ordenam-se os valores de **p** do grupo **PP** e aquelas pacientes que possuírem os 10% menores valores de **p** são denominadas no grupo **PPB10**. O segundo grupo são as pacientes **PP** que não estão no grupo **PPB10** (**PPO**).

A mesma abordagem foi testada com o grupo **GP**, porém o teste *Log-rank* não resultou em diferenças significativas. Além disso, outras divisões de quartis foram testadas, porém, a melhor divisão foi no primeiro decil.

4 RESULTADOS

Uma vez que foram analisados dados de expressão gênica de duas coortes internacionais (**TCGA** e **METABRIC**) e uma coorte interna (**A.C.Camargo Cancer Center**), observamos que a anotação dos genes são diferentes. As coortes do **TCGA** e **A.C.Camargo Cancer Center** foram realizadas usando uma anotação (*build* GRCh38.p12) com 54.517 genes (*ENSEMBL ID*), enquanto para a coorte do **METABRIC** foram usados 24.369 genes (*Gene Symbol*). Desta forma, para a definição dos grupos **GP** e **PP**, 92 genes foram detectados na análise na coorte do **TCGA** e validado na coorte do **A.C.Camargo Cancer Center** e, destes, 50 genes também quantificados na coorte do **METABRIC** foram utilizados na classificação destas amostras. Assim, no que se segue, os resultados serão apresentados para a assinatura de 50 genes.

4.1 ANÁLISE DE SOBREVIDA GLOBAL

O estimador de *Kaplan-Meier* foi utilizado para avaliar o desfecho de sobrevida das pacientes classificadas nos grupos **GP** e **PP**. A Figura 10 apresenta a análise da coorte **TCGA** com 932 pacientes considerando o tempo de seguimento e desfechos, incluindo perda de seguimento e morte dos 4 subtipos estudados no **TCGA** (v12) e classificadas da maneira seguinte: **HR-HER2-** (N=155), **HR-HER2+** (N=40), **HR+HER2-** (N=584) e **HR+HER2+** (N=153).

Podemos observar que pacientes representadas pela curva vermelha (**HR+HER2+**) apresentam melhor sobrevida global, enquanto pacientes representadas pela curva verde (**HR-HER2-** ou com tumores triplo-negativos) apresentam pior sobrevida.

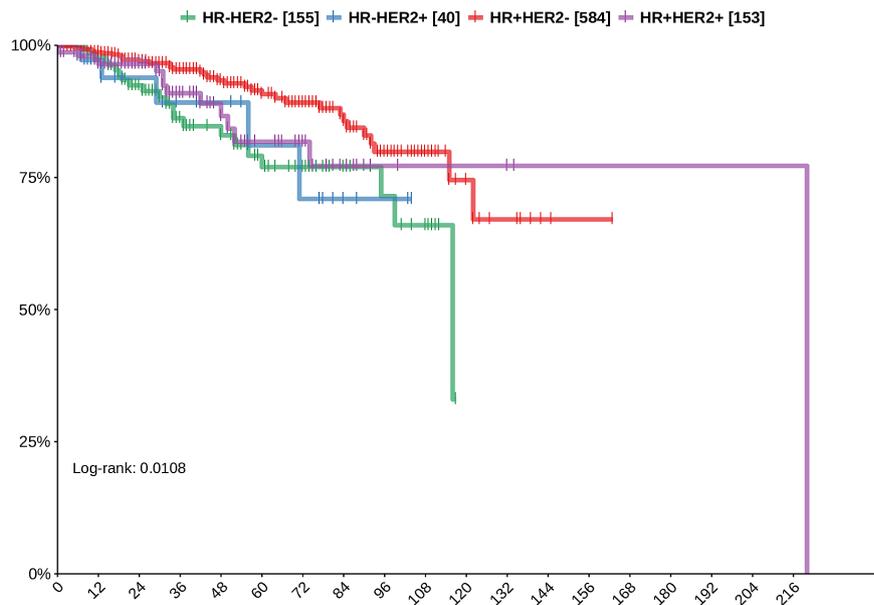


Figura 10 – Sobrevida global dos subtipos moleculares das pacientes do **TCGA**. Sobrevida global dos subtipos moleculares (**HR-HER2-**: triplo-negativo; **HR-HER2+**: HER2+; **HR+HER2-**: Luminal A; **HR+HER2+**: Luminal B) das pacientes com câncer de mama a partir dos dados do **TCGA**.

Ao se avaliar o grupo de pacientes com tumores triplo-negativos separadamente (Figura 11), podemos identificar uma característica em relação ao período de até 3 anos, onde ocorre a maioria dos eventos de mortes. Esses dados são consistentes com a literatura (DENT et al. 2007; ROOZENDAAL et al. 2016) e, assim, foi utilizado como primeiro critério para o classificador a fim de prever pacientes de alto (**PP**) baixo (**GP**) risco.

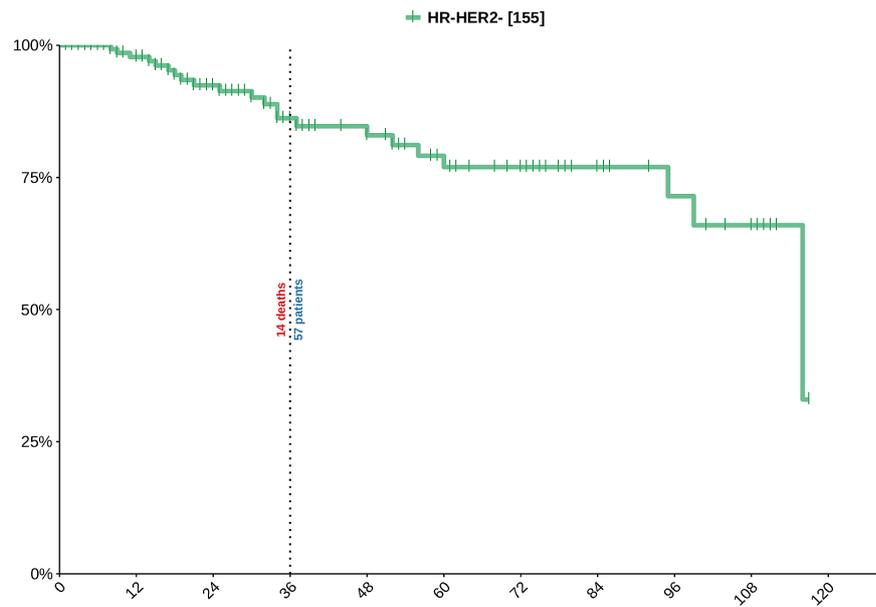


Figura 11 – Sobrevida global das pacientes com **CMTN** a partir dos dados do **TCGA**. Houve 14 óbitos em até 3 anos e 57 pacientes viveram mais do que 3 anos.

Ao se avaliar a sobrevida global da coorte do **METABRIC**, as pacientes com tumores triplo-negativos (linha verde) obtiveram a segunda pior sobrevida (Figura 12). Vale destacar que, no período desse trabalho (PEREIRA et al. 2016), as pacientes **HER2+** apresentaram pior prognóstico, isso decorreu possivelmente devido a algumas pacientes terem sido diagnosticadas antes de 2006, que foi quando o **FDA** (*Food and Drug Administration*) aprovou o primeiro tratamento específico para este subtipo (GENENTECH-PRESS-RELEASES 2006).

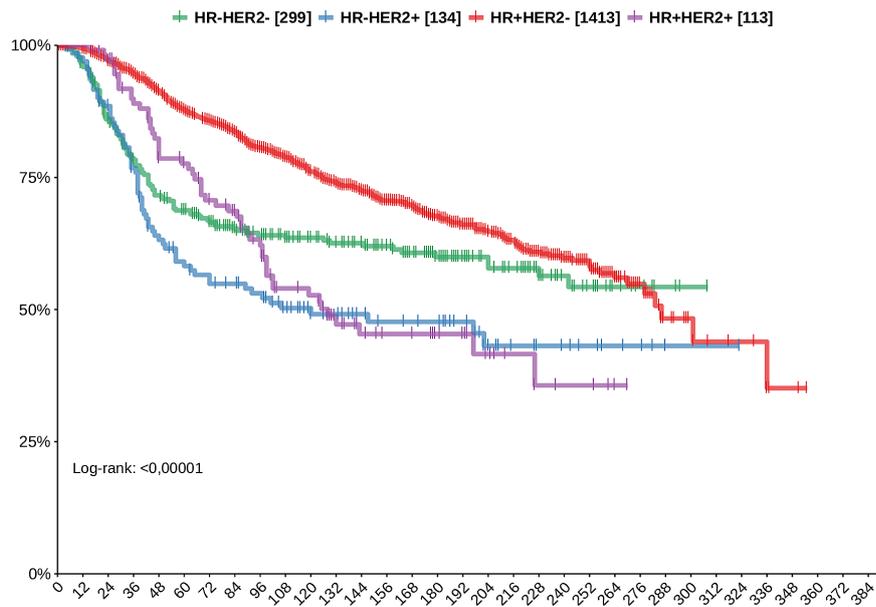


Figura 12 – Sobrevida global dos subtipos moleculares das pacientes do **METABRIC**. Sobrevida global das pacientes com câncer de mama em função dos subtipos moleculares (**HR-HER2-**: triplo-negativo; **HR-HER2+**: HER2+; **HR+HER2-**: Luminal A; **HR+HER2+**: Luminal B) a partir dos dados do **METABRIC**.

A Figura 13 ilustra somente as pacientes com **CMTN** e, novamente, observamos uma alta taxa de óbitos nos primeiros 3 anos.

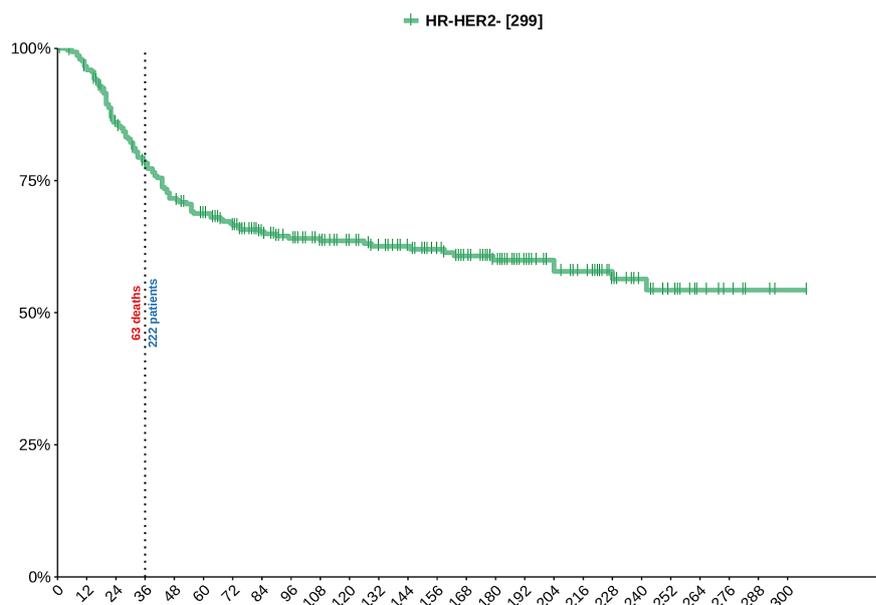


Figura 13 – Sobrevida global das pacientes com **CMTN** a partir dos dados do **METABRIC**. Houve 63 óbitos em até 3 anos e 222 pacientes viveram mais do que 3 anos.

Finalmente, a coorte do **A.C.Camargo Cancer Center** (Figura 14) reflete o mesmo perfil encontrado nos dados do **TCGA** e **METABRIC**, ou seja, uma alta taxa de mortalidade nos primeiros 3 anos.

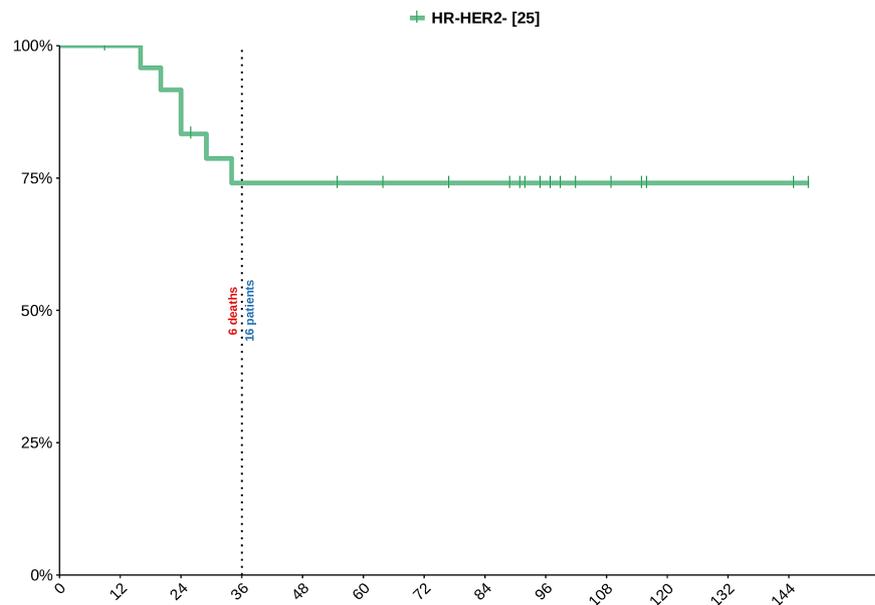


Figura 14 – Sobrevida global das pacientes com **CMTN** nos dados do **A.C.Camargo Cancer Center**. Houve 6 óbitos até 3 anos e 16 pacientes viveram mais do que 3 anos.

No cenário apresentado, podemos observar um padrão recorrente nas 3 coortes, fundamentando a escolha dos 3 anos como ponto de corte das pacientes **GP** e **PP**.

4.2 CARACTERÍSTICAS CLÍNICAS DAS COORTES (TCGA, METABRIC E ACCCC)

A Tabela 1 sumariza as características clínicas das coortes utilizadas para gerar a assinatura multigênica (**TCGA**) e para validação (**TCGA**, **METABRIC** e **A.C.Camargo Cancer Center**).

Tabela 1 – Dados clínicos de pacientes com **CMTN** das coortes **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center**.

LF: Perda de acompanhamento.

	TCGA	METABRIC	A.C.Camargo Cancer Center
Óbito (meses)			
<= 36	14	63	6
> 36	8	50	0
LF (meses)			
<= 36	84	14	3
> 36	49	172	16
Idade (anos)			
<= 30	1	6	3
31-50	57	110	22
51-70	79	142	0
> 70	18	41	0
BRCA			
BRCA1+	10	NA	8
BRCA2+	4	NA	0
Estadiamento			
I	29	62	1
II	99	130	11
III	24	25	11
NA	3	82	2
T			
T1	41	0	1
T2	96	0	17
T3	13	0	4
T4	5	0	3
NA	0	299	0
N			
N0	104	0	9
N1	31	0	9
N2	13	0	4
N3	7	0	3
NA	0	299	0

4.3 DADOS CLÍNICOS DOS GRUPOS GP E PP

Após a aplicação da abordagem proposta neste trabalho e sumarizada nas seções 3.6 e 3.7, o classificador com 50 genes foi capaz de identificar os grupos **GP** e **PP**. Os aspectos clínicos das respectivas pacientes estão apresentados na Tabela 2.

Uma característica relevante na abordagem para classificar pacientes de alto e baixo risco, é avaliar se o classificador por nós elaborado é capaz de identificar grupos com prognóstico distinto independentemente de outras variáveis clínicas. Para isso, foi utilizado a análise multivariada de *Cox*. Para os dados das pacientes com **CMTN** do **TCGA** e **A.C.Camargo Cancer Center** foram analisadas as seguintes variáveis: idade, TNM e estadiamento. Já, nos dados da coorte do **METABRIC** foram analisadas as seguintes variáveis: grau histológico,

terapia hormonal, estado de menopausa, idade e estadiamento. Ao considerar como limiar de significância o valor de $p < 0,05$, não foram encontradas variáveis que expliquem melhor os grupos **PP** e **GP** (classificação dos estadiamentos I,II e III) e, assim, a classificação realizada por nossa abordagem é independente de outras variáveis em todas as coortes avaliadas.

Tabela 2 – Dados clínicos de pacientes com **CMTN** das coortes **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** (estadiamentos I, II e III) associados à classificação utilizando a assinatura de 50 genes.

Teste exato de Fisher (óbito em até 3 anos comparado com sobrevida maior do que 3 anos nos grupos **PP** e **GP**) - **TCGA**: $p < 0,01$; **METABRIC**: $p < 0,01$; **A.C.Camargo Cancer Center**: $p = 0,1$.
LF: Perda de acompanhamento.

	TCGA		METABRIC		ACCCC	
	PP	GP	PP	GP	PP	GP
Óbito (meses)						
<= 36	13	1	22	41	3	3
> 36	1	7	10	40	0	0
LF (meses)						
<= 36	23	61	1	13	0	3
> 36	10	39	19	153	2	14
Idade (anos)						
<= 30	0	1	1	5	1	2
31-50	19	38	21	89	4	18
51-70	23	56	23	119	0	0
> 70	5	13	7	34	0	0
BRCA						
BRCA1+	1	9	NA	NA	0	8
BRCA2+	2	2	NA	NA	0	0
Estadiamento						
I	6	23	7	55	0	1
II	28	71	24	106	2	9
III	13	11	6	19	2	9
NA	0	3	15	67	1	1
T						
T1	9	32	0	0	0	1
T2	31	65	0	0	4	13
T3	4	9	0	0	0	4
T4	3	2	0	0	1	2
NA	0	0	52	247	0	0
N						
N0	31	73	0	0	1	8
N1	6	25	0	0	1	8
N2	5	8	0	0	2	2
N3	5	2	0	0	1	2
NA	0	0	52	247	0	0

4.4 SOBREVIDA GLOBAL E PERFIL DE EXPRESSÃO NOS ESTADIAMENTOS INICIAIS

A fim de avaliar se os genes detectados anteriormente são capazes de estratificar pacientes em estágio inicial, foram feitas duas análises, a primeira classificando pacientes com estadiamento I e a segunda classificando pacientes com

os estadiamentos I e II.

4.4.1 Estadiamento I

Para as classificações das pacientes com estadiamento I (Figura 15) foram analisadas pacientes das coortes do **TCGA** e **METABRIC** e não foram utilizadas pacientes da coorte do **A.C.Camargo Cancer Center**, pois nesta coorte há apenas uma paciente com estadiamento I. Na coorte do **TCGA**, não foi possível diferenciar as pacientes quanto à sobrevida, ao considerar os testes de *Log-rank* e de proporção de óbito até 3 anos (*Log-rank*: $p=1$, prop.test: $p=0,75$). Quanto as pacientes do **METABRIC**, houve diferença quanto aos perfis de sobrevida pelo teste de *Log-rank* ($p=0,01$), porém não houve diferença significativa nas proporções de óbitos até 3 anos ($p=0,43$) nas pacientes com alto (**PP**) e baixo risco (**GP**).

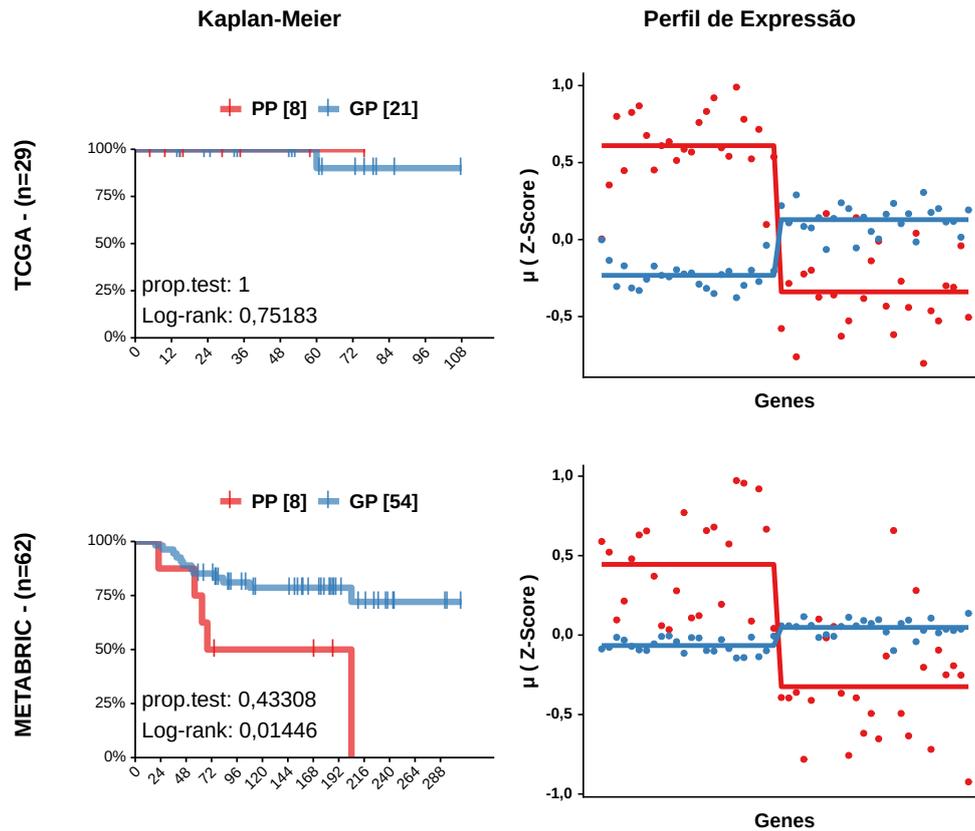


Figura 15 – Curva de sobrevivência e o respectivo perfil de expressão gênica das pacientes classificadas como **PP** (vermelho) e **GP** (azul) de acordo com a assinatura com 50 genes (Estadiamento I). As curvas de sobrevivência foram calculadas pelo método de *Kaplan-Meier* e os grupos comparados pelo teste de *Log-rank*. Os gráficos de expressão mostram dois conjuntos de genes: O primeiro conjunto de genes sendo altamente expresso nas pacientes classificadas como **PP** e pouco expresso nas pacientes classificadas como **GP** (conjunto de genes denominado como **highPP**); O segundo conjunto de genes sendo altamente expressos em pacientes classificadas como **GP** e pouco expressos em pacientes classificadas como **PP** (conjunto de genes denominado como **lowPP**).

4.4.2 Estadiamentos I e II

Conforme a Figura 16, a classificação de pacientes com estadiamentos I e II, utilizando-se 50 genes, foi capaz de identificar populações com sobrevivências distintas pelo teste de *Log-rank* e o teste de proporção de óbito até 3 anos nas coortes do **TCGA** (*Log-rank*: $p=0,04$, prop.test: $p=0,01$) e **METABRIC** (*Log-rank*: $p<0,01$, prop.test: $p=0,017$), porém, ao analisar a coorte do **A.C.Camargo Cancer Center**, não foi possível diferenciar as pacientes (*Log-rank*: $p=0,62$, prop.test: $p=1$).

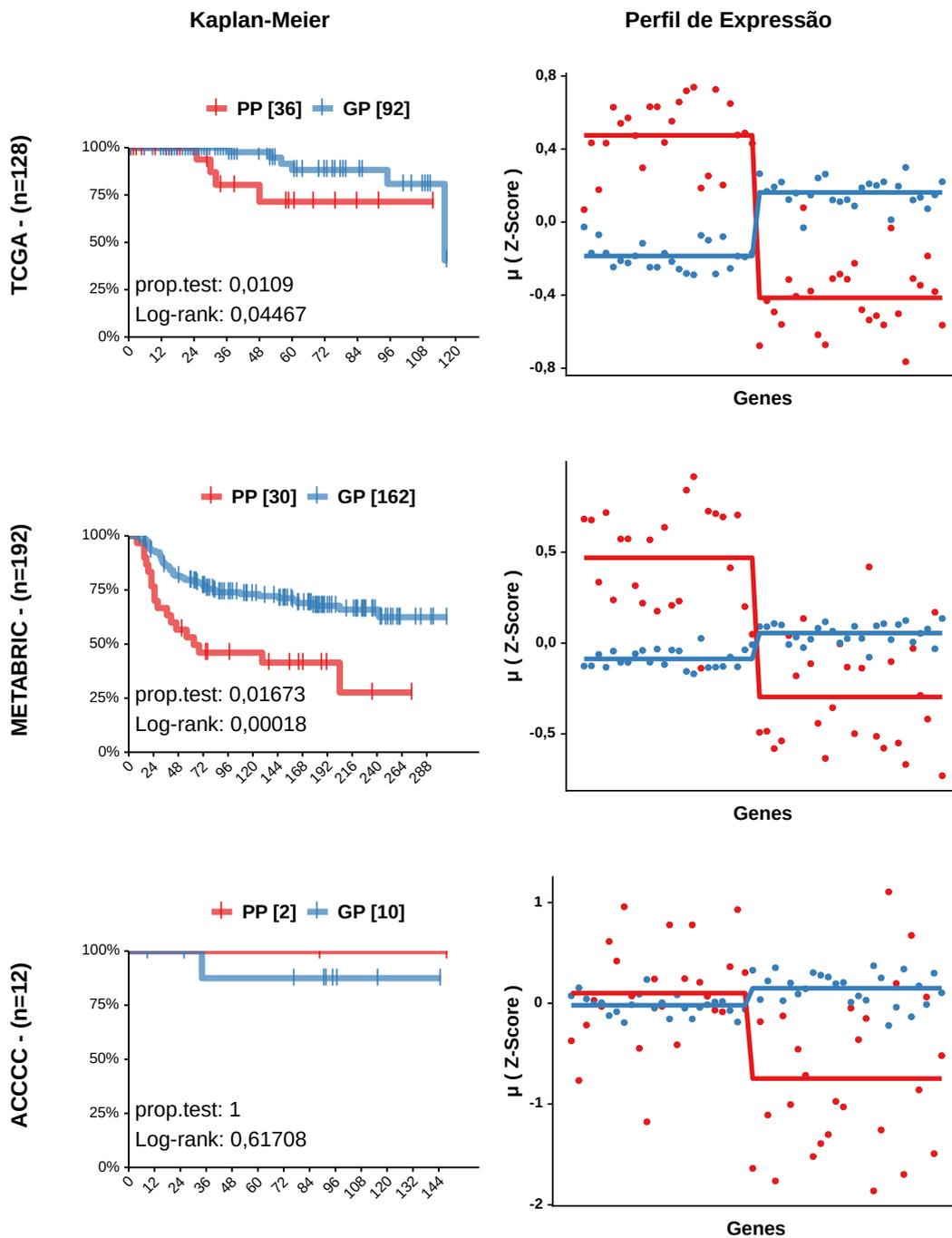


Figura 16 – Curva de sobrevida e o respectivo perfil de expressão gênica das pacientes classificadas como **PP** (vermelho) e **GP** (azul) de acordo com a assinatura com 50 genes (Estadiamentos I e II). As curvas de sobrevida foram calculadas pelo método de *Kaplan-Meier* e os grupos comparados pelo teste de *Log-rank*. Os gráficos de expressão mostram dois conjuntos de genes: O primeiro conjunto de genes sendo altamente expresso nas pacientes classificadas como **PP** e pouco expresso nas pacientes classificadas como **GP** (conjunto de genes denominado como **highPP**); O segundo conjunto de genes sendo altamente expressos em pacientes classificadas como **GP** e pouco expressos em pacientes classificadas como **PP** (conjunto de genes denominado como **lowPP**).

4.5 SOBREVIDA GLOBAL E PERFIL DE EXPRESSÃO NOS ESTADIAMENTOS I, II E III

Conforme a Figura 17, é possível observar que as coortes do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** apresentam um perfil de expressão gênica similar, onde, existem basicamente dois conjuntos de genes: o primeiro é um grupo de genes que é altamente expresso em **PP** e é pouco expresso em **GP** (conjunto de genes denominado como **highPP**); O segundo é um grupo de genes que é altamente expresso em **GP** e pouco expresso em **PP** (conjunto de genes denominado como **highPP**).

Em relação à sobrevida, nos dados do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center**, o teste *Log-rank* obteve significância ($p < 0,01$, $p < 0,01$, $p = 0,04$, respectivamente), enquanto no teste de proporção de óbito em até 3 anos, as coortes do **TCGA** e **METABRIC** obtiveram significância ($p < 0,01$, $p < 0,01$, respectivamente), porém, a coorte do **A.C.Camargo Cancer Center** não resultou em diferença significativa ($p = 0,1$). Ou seja, o método de classificação utilizando a assinatura de 50 genes foi capaz de identificar dois grupos com perfis de sobrevida distintos nas coortes do **TCGA** e **METABRIC**, porém, nas pacientes do **A.C.Camargo Cancer Center** não houve diferença significativa nas proporções de óbitos em até 3 anos nos grupos identificados (**PP** e **GP**), isso ocorre, provavelmente devido ao número pequeno de pacientes nesta coorte (25 amostras). Além disso, foi realizada a análise multivariada de *Cox* e as populações encontradas se mostraram independentes de outras variáveis clínicas nas coortes do **TCGA** e **A.C.Camargo Cancer Center** (idade, T, N e estadiamento) e na coorte do **METABRIC** (idade, grau histológico, terapia hormonal, estado de menopausa, estadiamento e tamanho do tumor).

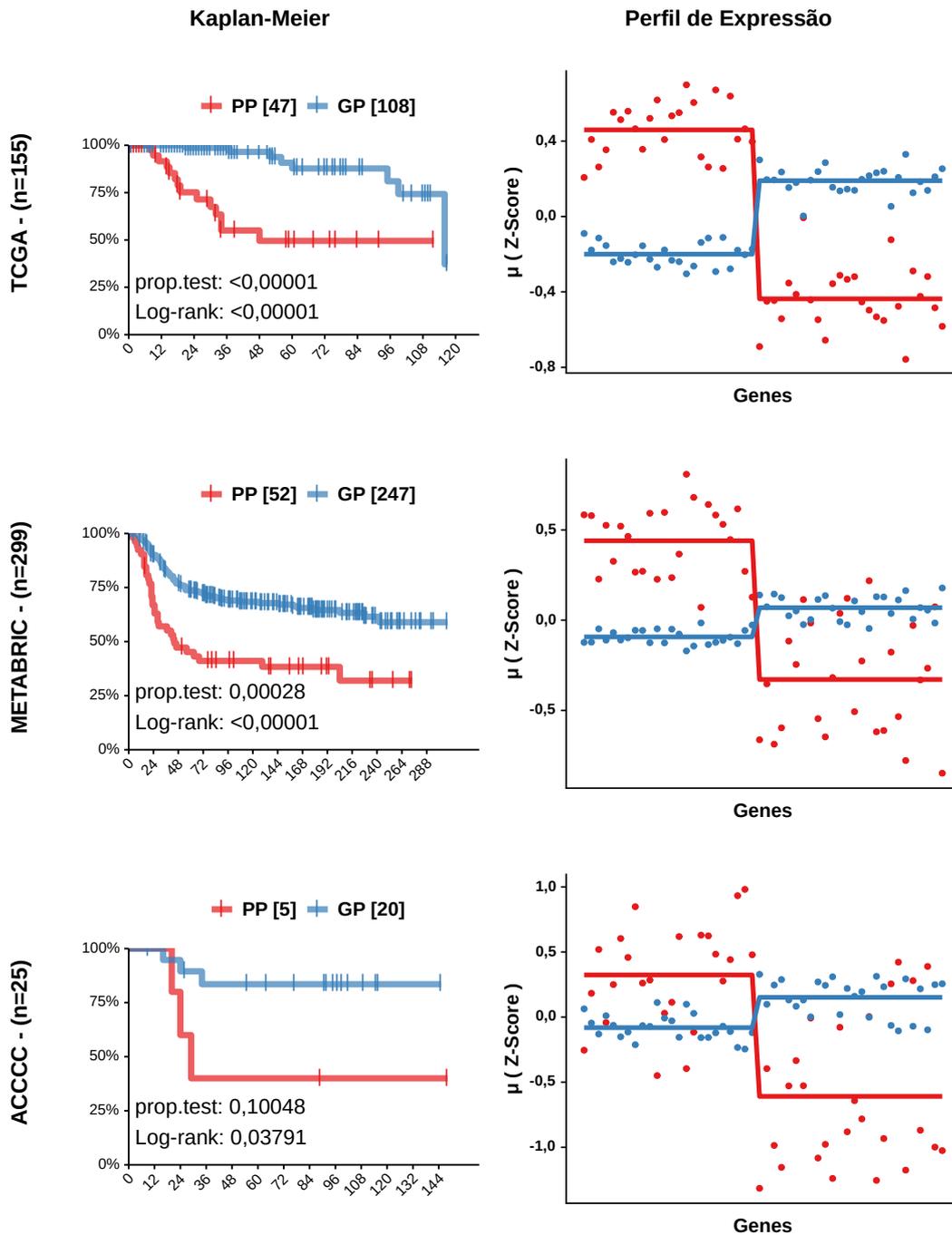


Figura 17 – Curva de sobrevivência e o respectivo perfil de expressão gênica das pacientes classificadas como **PP** (vermelho) e **GP** (azul) de acordo com a assinatura com 50 genes (Estadiamentos I, II e III). As curvas de sobrevivência foram calculadas pelo método de *Kaplan-Meier* e os grupos comparados pelo teste de *Log-rank*. Os gráficos de expressão mostram dois conjuntos de genes: O primeiro conjunto de genes sendo altamente expresso nas pacientes classificadas como **PP** e pouco expresso nas pacientes classificadas como **GP** (conjunto de genes denominado como **highPP**); O segundo conjunto de genes sendo altamente expressos em pacientes classificadas como **GP** e pouco expressos em pacientes classificadas como **PP** (conjunto de genes denominado como **lowPP**).

4.6 DIVISÃO DAS PACIENTES COM CMTN EM 3 GRUPOS

Nesta etapa, o subtipo triplo-negativo foi estratificado em 3 grupos (metodologia, seção 3.9) utilizando as coortes do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** classificadas utilizando a assinatura de 50 genes.

A partir dos dados classificados entre bom (**GP**) e mau prognóstico (**PP**) nas 3 coortes, é possível verificar que os grupos identificados em duas destas coortes (**TCGA** e **METABRIC**) apresentam dois perfis de sobrevida distintos. Na coorte do **A.C.Camargo Cancer Center** não apresentou diferença significativa (ou seja, valor de $p < 0,05$) mediante o teste de proporção de óbito até 3 anos entre os grupos **PP** e **GP**, e, ao analisar o teste *Log-rank*, existe diferença significativa ($p=0,04$) entre os grupos de pacientes identificados (**PP** e **GP**).

Quando analisado os subgrupos de **PP** (**PPB10** e **PPO**) identificados pelo método descrito na seção 3.9 juntamente com o grupo **GP**, ocorre diferença nos valores de p nos testes de *Log-rank* e no teste de proporção de óbito até 3 anos nos 2 conjuntos de dados com maior número de pacientes (**TCGA** e **METABRIC**), embora não tenha sido possível estratificar as pacientes quanto à sobrevida na coorte do **A.C.Camargo Cancer Center**, provavelmente devido ao número pequeno de pacientes.

Relembrando, o método de classificação compara, utilizando o teste de *Mann-Whitney*, se o grupo de genes **highPP** (genes com alta expressão nas pacientes **PP** e baixa expressão nas pacientes **GP**) é significativamente ($p < 0,05$) maior que o grupo de genes **lowPP** (genes com baixa expressão nas pacientes **GP** e alta expressão nas pacientes **PP**). Sendo assim, a partir dos valores de p resultantes do teste de *Mann-Whitney* na classificação, as pacientes **PP** com menores valores de p possuem a pior quando comparados às demais pacientes **PP**, que, por sua vez possuem pior sobrevida do que pacientes **GP**.

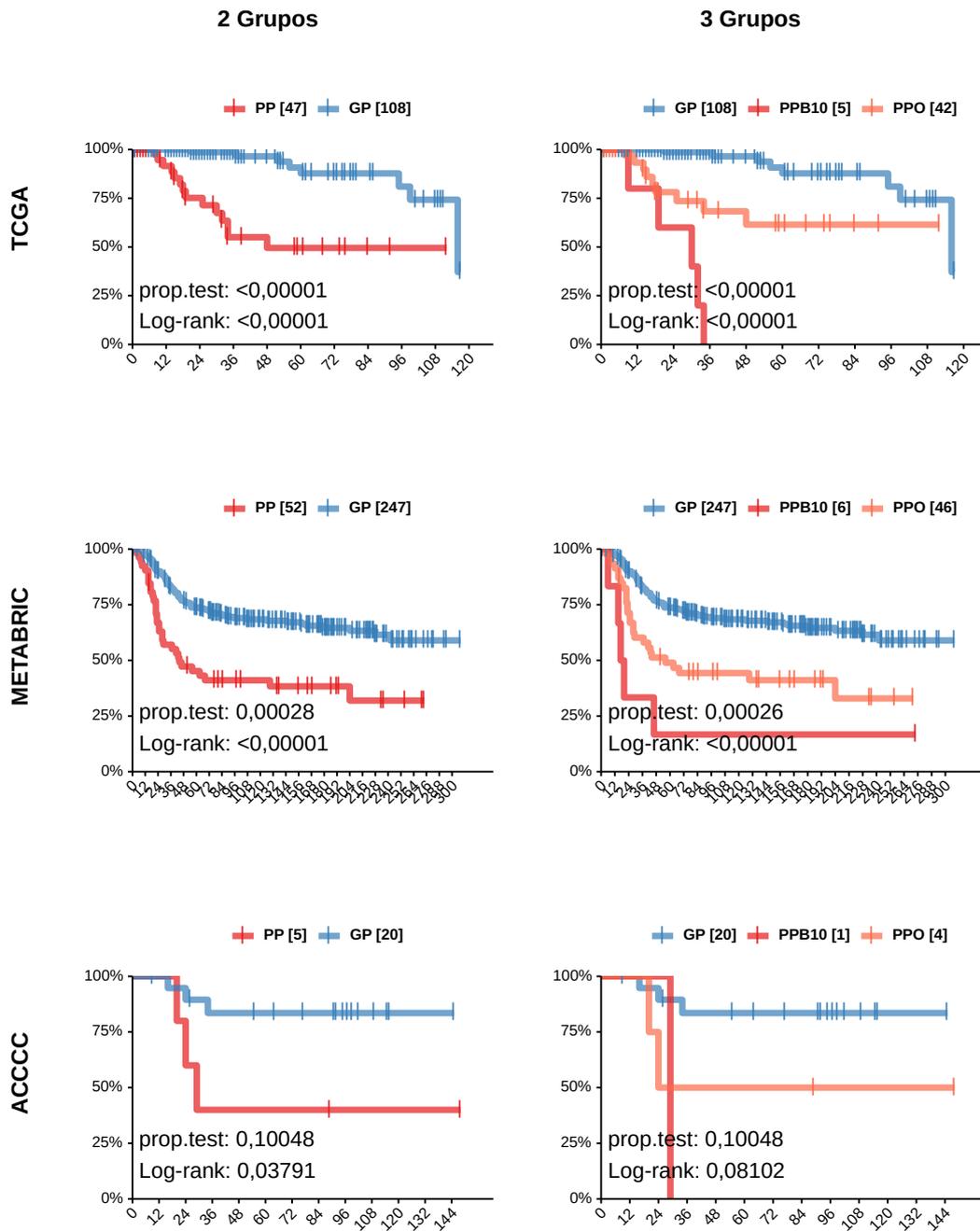


Figura 18 – Subdivisão das pacientes CMTN em 3 grupos (GP, PPB10 e PPO). As curvas de sobrevida foram calculadas pelo método de *Kaplan-Meier* e os grupos comparados pelo teste de *Log-rank* e o teste de proporção de óbito até 3 anos.

Podemos observar, na Figura 18, que existe um subgrupo de pacientes que vão pior dentro do grupo **PP**. Essa característica é observada nas coortes do **TCGA** e **METABRIC** quando classificadas pela mesma assinatura.

4.7 ANÁLISE DE ENRIQUECIMENTO

Uma vez que a análise proposta neste estudo identificou uma assinatura de 50 genes, uma etapa seguinte é a anotação funcional dos processos biológicos nos quais esses genes estão envolvidos, conforme Figura 19.

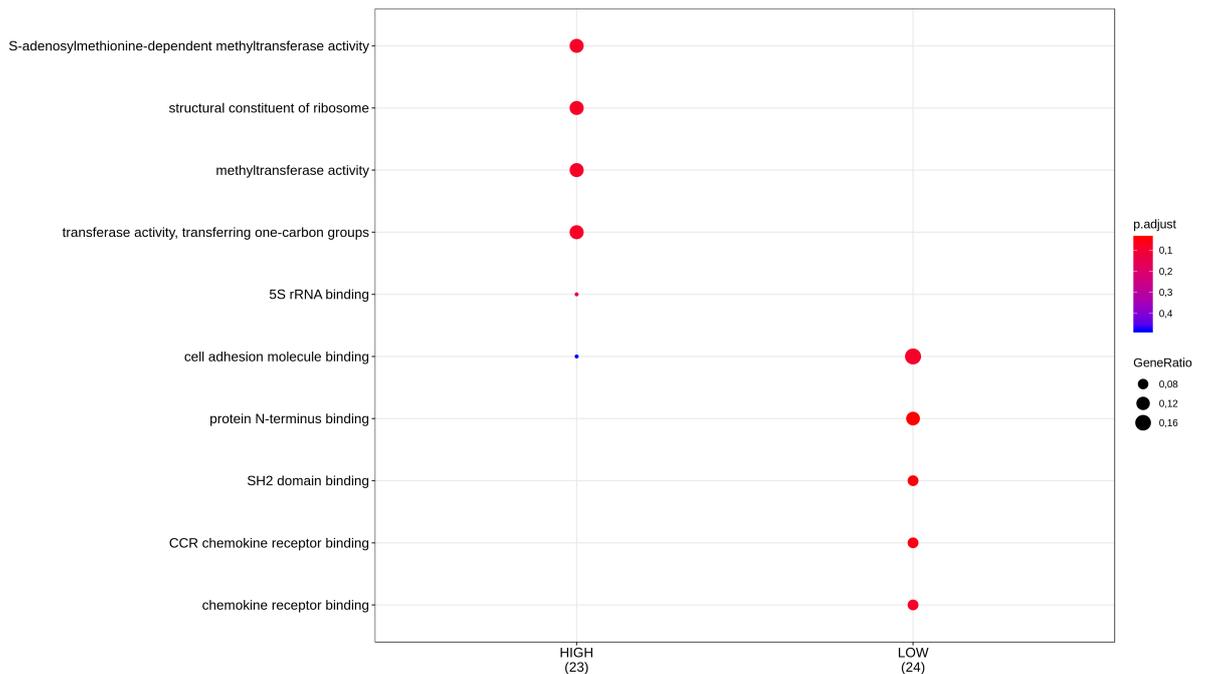


Figura 19 – Análise de enriquecimento da assinatura de 50 genes. Nas linhas estão os processos biológicos. Nas colunas estão dois grupos de genes: o primeiro é formado por genes altamente expressos em pacientes PP e pouco expressos em pacientes GP (**HIGH**); o segundo é referente aos genes altamente expressos em pacientes GP e pouco expressos em pacientes PP (**LOW**). A cor das bolas é referente ao valor de **p** ajustado e seu tamanho é referente a proporção de genes associados ao processo biológico.

A análise de enriquecimento da assinatura de 50 genes identificou características distintas para os dois grupos de genes (**highPP** e **lowPP**). Os genes **highPP** apresentaram as atividades catalítica na transferência de grupos de carbono, além da atividade de contribuição na integridade estrutural da célula. Ao passo que, os genes **lowPP** apresentaram atividades associadas à ligação proteica, entre elas, a ligação aos receptores de quimiocinas.

4.8 INFILTRADO INFLAMATÓRIO

Para avaliar o perfil de infiltrado inflamatório nos grupos classificados utilizando a assinatura de 50 genes, foram feitas duas análises.

A primeira análise (Figura 20) foi a comparação de cada um dos 22 tipos de leucócitos identificados pelo CIBERSORT em cada grupo (**PP** e **GP**) utilizando o teste de *Mann-Whitney*. Esta análise serve para analisar as diferenças de proporções de cada célula em cada grupo, por exemplo, se existe maior proporção de macrófago M1 no grupo **PP** ou **GP**.

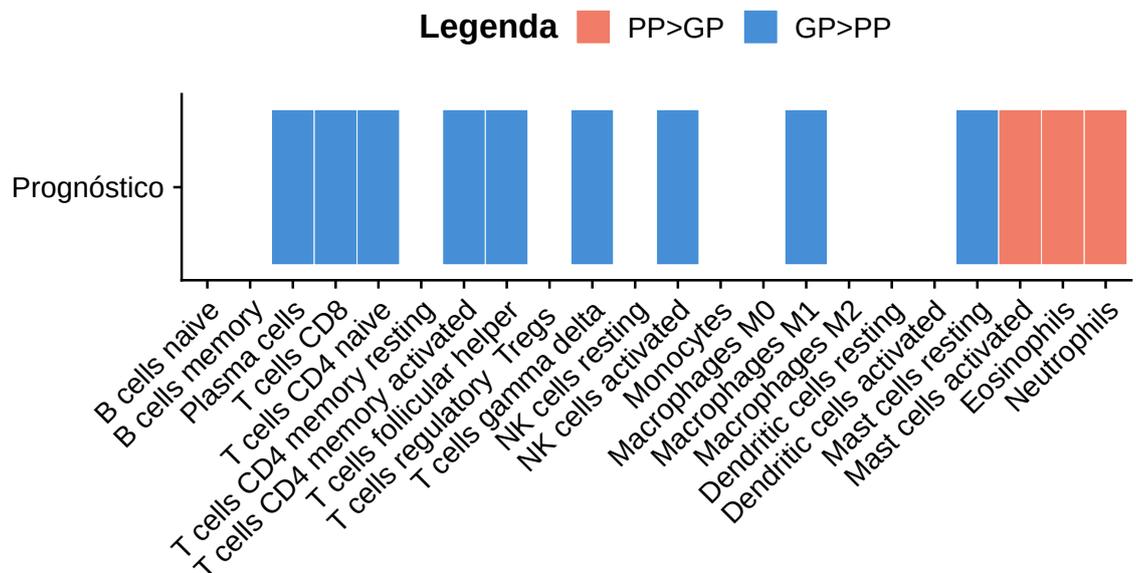


Figura 20 – Comparação da proporção de 22 tipos de leucócitos entre os grupos **PP** e **GP** das coortes do **TCGA** e **METABRIC**. A análise foi feita utilizando o teste de *Mann-Whitney* para analisar as proporções de cada leucócito entre os grupos **PP** e **GP** das coortes do **TCGA** e **METABRIC**. As barras em vermelho indicam os tipos de leucócitos com maiores proporções no grupo **PP**, enquanto as barras em azul indicam os tipos de leucócitos com maiores proporções no grupo **GP**.

A segunda análise (Figura 21) foi a comparação entre os 22 tipos de leucócitos dentro de cada grupo utilizando o teste de *Mann-Whitney* nos grupos **PP** e **GP** isoladamente. Esta análise serve para verificar quais tipos de leucócitos

estão mais presentes em cada grupo (**PP** ou **GP**), por exemplo, é comparado se há maior proporção de macrófago M1 ou macrófago M2 no grupo **PP**, porém não é comparado se há maior proporção de macrófago M1 no grupo **PP** em relação ao macrófago no grupo **GP**.

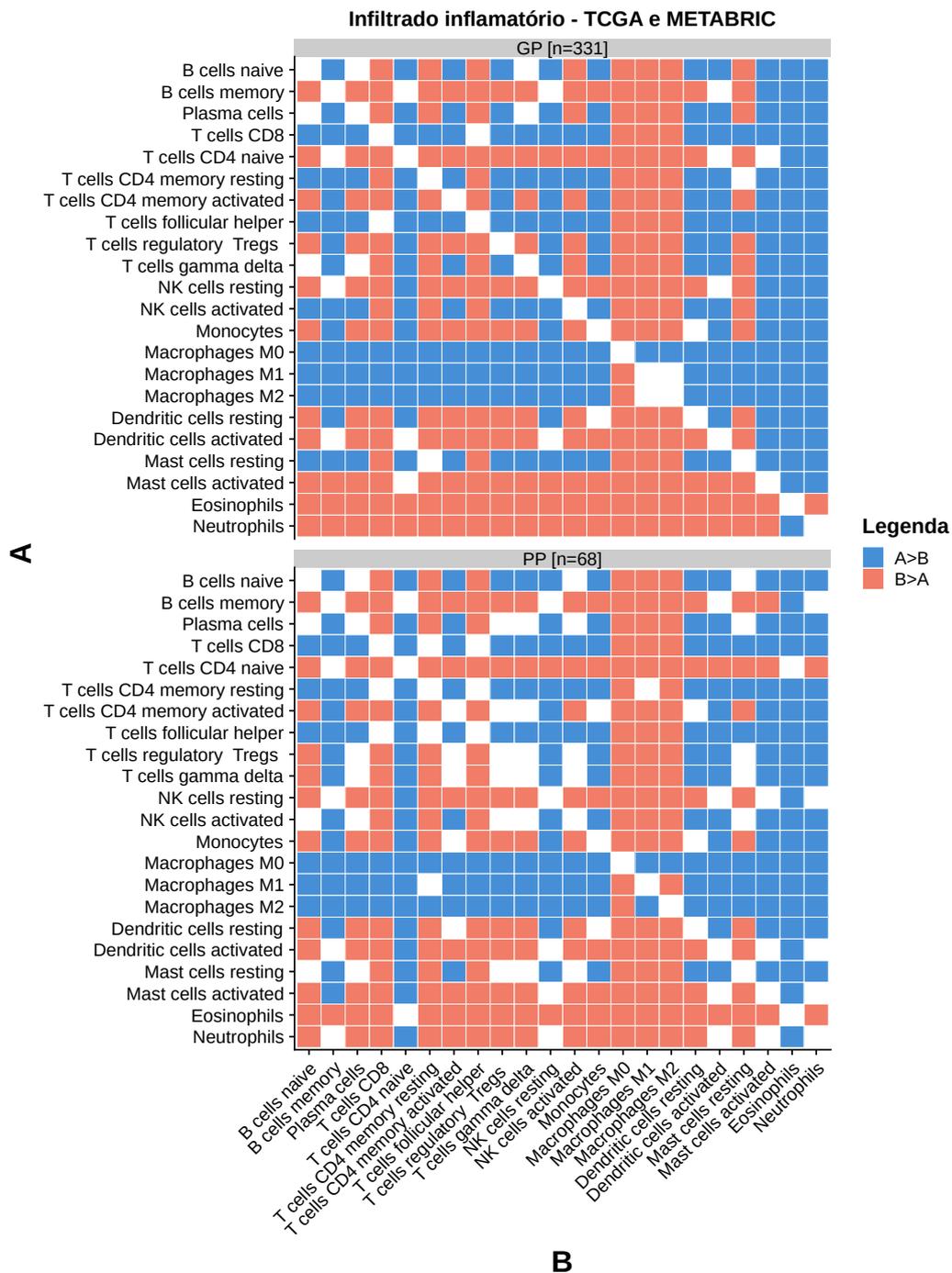


Figura 21 – Comparação da proporção entre os 22 tipos de leucócitos nos grupos **PP** e **GP** das coortes do **TCGA** e **METABRIC**. A análise foi feita utilizando o teste de *Mann-Whitney* para analisar as proporções de cada leucócito com os demais tipos de leucócitos em cada grupo (**PP** e **GP**) das coortes do **TCGA** e **METABRIC**. As barras em vermelho indicam os tipos de leucócitos com maiores proporções no grupo de leucócitos no eixo das abscissas quando comparados com os leucócitos no eixo das ordenadas (**B>A**), enquanto as barras em azul indicam o inverso (**A>B**).

À partir da primeira análise (Figura 20) é possível verificar que as pacien-

tes classificadas como **GP** possuem maiores proporções de células plasmáticas, T CD8+, T CD4+ naive, T CD4+ memória ativada, T follicular helper, T gamma delta, NK ativada, macrófago M1 e mastócitos em repouso, enquanto nas pacientes classificadas como **PP** foram identificadas maiores proporções de mastócitos ativados, eosinófilos e neutrófilos.

Na segunda análise (Figura 21), ao verificar somente as características divergentes nos dois grupos (**PP** e **GP**), as pacientes classificadas como **GP** apresentaram maiores proporções nas células NK ativadas, células T gamma delta, T CD4+ memória ativada, mastócitos em repouso e células plasmáticas em relação a células T reguladoras. Quando ao grupo **PP**, as proporções de macrófagos M2 foi superior as proporções de macrófagos M1.

4.9 CLASSIFICAÇÃO *TNBCTYPE*

A fim de avaliar se os subtipos propostos por Lehmann et al. (LEHMANN et al. 2011) são capazes de estratificar pacientes das coortes **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center**, foi utilizada a ferramenta *online* (VANDERBILT 2018). Como observado na Figura 22, os subtipos moleculares não foram capazes de separar as pacientes em populações com sobrevida global distintos.

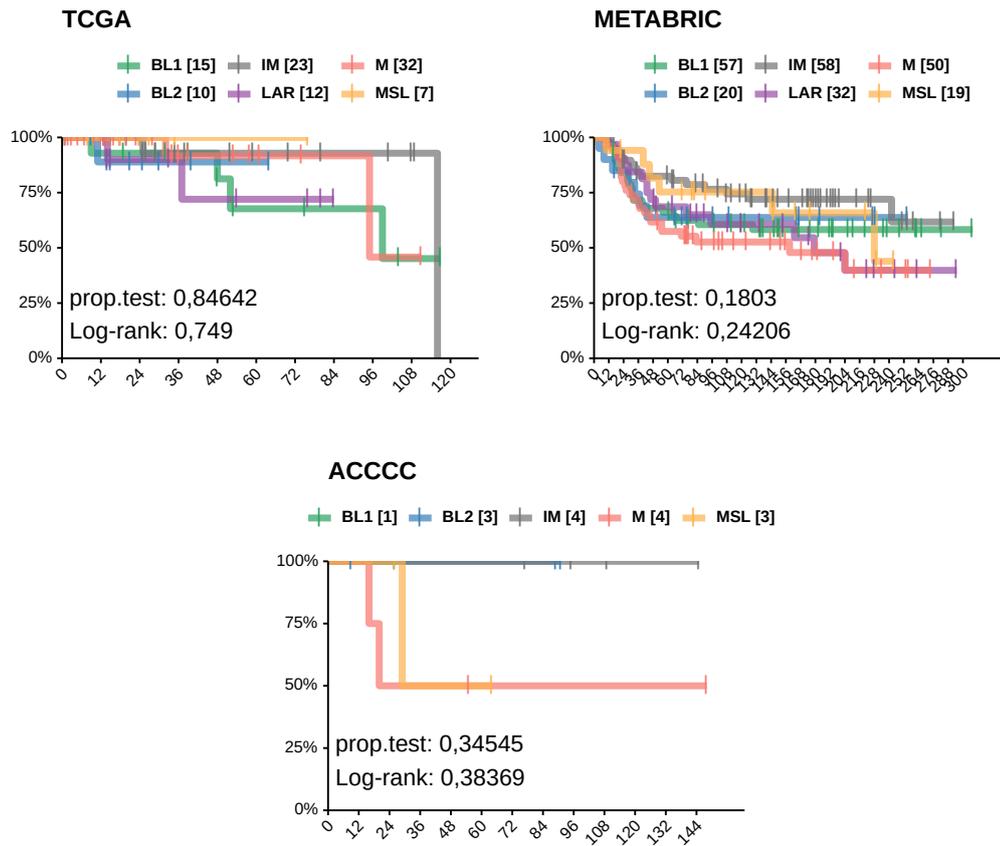


Figura 22 – Avaliação da sobrevida das pacientes com CMTN nas coortes **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** de acordo com os grupos definidos pelo software *TNBCtype*.

5 DISCUSSÃO

O câncer de mama triplo-negativo (**CMTN**) é uma doença heterogênea que não possui tratamento molecular específico, além de possuir o pior prognóstico entre os tipos de câncer de mama (NOMAN et al. 2016). No presente estudo, foram identificadas assinaturas multigênicas capazes de prever pacientes de baixo e alto risco em **CMTN**.

Uma abordagem que estratifica **CMTN** em 6 subtipos moleculares (*TNBCtype*) foi aplicada aos dados do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** e não houve diferença significativa quanto à sobrevida dos respectivos subtipos. Esses resultados implicam que o valor prognóstico do *TNBCtype* ainda é limitado, embora permita revelar alguns aspectos de tratamento.

A abordagem desenvolvida nesta dissertação gerou uma assinatura multigênica com 50 genes utilizando os dados da coorte do **TCGA**. O perfil das assinaturas utilizadas pelo classificador é formado por dois grupos de genes, o primeiro que possui alta expressão em pacientes com mau prognóstico e baixa expressão em pacientes com bom prognóstico (**highPP**) e outro conjunto de genes que possui baixa expressão em pacientes com mau prognóstico e alta expressão em pacientes com bom prognóstico (**lowPP**).

A partir deste conjunto de genes, o método desenvolvido foi capaz de classificar pacientes de bom (**GP**) e mau prognóstico (**PP**) nas coortes do **TCGA** e **METABRIC** (Figuras 17), utilizando a assinatura de 50 genes (estadiamentos I, II e III). Percebe-se o sucesso dessa classificação pela significância obtida nas comparações das sobrevidas globais dos dois grupos ($p < 0,05$). Embora a coorte do **A.C.Camargo Cancer Center** apresente um conjunto pequeno de pacientes ($N=25$), observamos uma tendência de separação dos dois grupos (Figura 17).

Adicionalmente, ao estimar o perfil de **INF** (metodologia, seção 3.5) nes-

tes grupos de pacientes, o grupo **GP** apresentou um perfil de **INF** com maiores proporções de linfócitos como células plasmáticas (linfócito B efetor), T CD8+, T CD4+ naive, T CD4+ memória ativada, T follicular helper, T gamma delta e outras células com perfil inflamatório como NK ativada, macrófago M1 e mastócitos em repouso em relação ao grupo **GP**. Por outro lado, as pacientes no grupo **PP** foram identificadas maiores proporções de células imunes inatas como mastócitos ativados, eosinófilos e neutrófilos em relação ao grupo **GP**. Além disso, ao analisar cada grupo isoladamente, as pacientes classificadas como **PP** possuem maior proporção de macrófagos M2 em relação à macrófagos M1, enquanto as pacientes classificadas como **GP** não possuem diferença significativa entre as proporções de macrófagos M1 e M2, o que pode indicar uma importância nos papéis desses leucócitos em pacientes com CMTN.

Além disso, para verificar a eficiência do classificador em estágios iniciais do tumor, foram feitas duas classificações das pacientes das coortes do **TCGA**, **METABRIC** e **A.C.Camargo Cancer Center** utilizando a assinatura de 50 genes, a primeira análise verificou as pacientes com estadiamento I (Figura 15) e a segunda análise verificou as pacientes com estadiamentos I e II (Figura 16).

Ao verificar a sobrevida na análise de pacientes com estadiamento I, não foi possível estratificar pacientes entre bom a mau prognóstico em nenhuma das coortes. Entretanto, ao analisar os estadiamentos I e II, foi possível estratificar as pacientes nas coortes do **TCGA** e **METABRIC** (prop.test: $p=0,01$ e $p=0,02$, respectivamente). Quanto aos dados da **A.C.Camargo Cancer Center**, não houve diferença significativa ($p<0,05$) no teste de proporção de óbito até 3 anos ($p=1$), possivelmente devido ao pequeno número de pacientes na análise. Ou seja, a assinatura de 50 genes foi capaz de classificar pacientes com **CMTN** nos estadiamentos I e II.

Em outra análise (metodologia, seção 3.9), foi utilizada a assinatura com 50 genes nos dados do **TCGA** e **METABRIC** que estratificou pacientes classificadas como mau prognóstico (**PP**) em dois subgrupos e, foi possível identificar

aqueles pacientes que possuem pior sobrevida dentro do grupo **PP**, sendo assim, ao analisar todas as pacientes **CMTN**, esta abordagem resultou em 3 subgrupos, sendo um com bom prognóstico (**GP**), outro com pior prognóstico dentro do grupo **PP** (**PPB10**) e as demais pacientes do grupo **PP** (**PPO**). E, conforme Figura 18, foi possível estratificar as pacientes das coortes do **TCGA** e **METABRIC** em 3 grupos (prop.test: $p < 0,01$ em ambas as coortes).

Neste trabalho, apresentamos uma proposta para a estratificação do **CMTN** considerando um perfil de expressão gênica que separa subgrupos prognósticos que, por sua vez, se correlacionam com o perfil imune. O entendimento do cenário imunológico somado ao classificador prognóstico pode apontar para novas abordagens terapêuticas (LIU et al. 2018).

Como perspectiva futura, será realizado um levantamento retrospectivo das amostras no biobanco do **A.C.Camargo Cancer Center** em conjunto com as informações do desfecho clínico a fim de levantar os casos para uma validação extensiva da assinatura. Pretendemos, assim, contribuir para o desenvolvimento de preditores de baixo e alto risco e torná-lo acessível às pacientes acometidas por essa doença.

6 CONCLUSÃO

No presente trabalho foi proposto uma abordagem para detectar uma assinatura multigênica capaz de identificar pacientes de baixo e alto risco. A assinatura foi validada em coortes independentes, mostrando a sensibilidade do método. Interessantemente, foi observado que o poder do classificador é independente de outras variáveis clínicas. Estes achados destacam as oportunidades para uma melhor estratificação das pacientes de **CMTN**. Ao observar o perfil imunológico nos grupos de pacientes, é possível propor terapias mais adequadas para os respectivos grupos. A análise funcional da assinatura revelou genes super expressos associados à proliferação celular em pacientes com alto risco (pior prognóstico). Diferentemente de outras classificações que estratificam o subtipo **CMTN** em ao menos 4 subtipos principais moleculares distintos (*TNBCtype*, por exemplo), a abordagem proposta por este estudo tem como intuito a estratificação entre pacientes entre altas e baixas taxas de óbito nos 3 primeiros anos. Assim, a contribuição deste trabalho permitiu identificar aspectos moleculares adicionais comuns no câncer **CMTN** auxiliando, assim, na estratificação de potenciais pacientes que possam se beneficiar de um tratamento mais personalizado.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- Ahn SG, Kim SJ, Kim C, et al. Molecular Classification of Triple-Negative Breast Cancer. **J Breast Cancer** 2016; 19:223–30.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. **Bioinformatics** 2015; 31:166–9.
- Anderson WF, Rosenberg PS, Prat A, et al. How many etiological subtypes of breast cancer: two, three, four, or more? **J Natl Cancer Inst** 2014; 106:dju165.
- Benitez-Parejo N, Rodriguez del Aguila MM, Perez-Vicente S. Survival analysis and Cox regression. **Allergol Immunopathol (Madr)** 2011; 39:362–73.
- Brandao M, Ponde N, Piccart-Gebhart M. Mammprint: a comprehensive review. **Future Oncol** 2019; 15:207–24.
- Brandao M, Ponde NF, Poggio F, et al. Combination therapies for the treatment of HER2-positive breast cancer: current and future prospects. **Expert Rev Anticancer Ther** 2018; 18:629–49.
- Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. **Clin Cancer Res** 2007; 13:2329–34.
- Soto-Perez-de Celis E, Chavarri-Guerra Y, Leon-Rodriguez E, et al. Tumor-Associated Neutrophils in Breast Cancer Subtypes. **Asian Pac J Cancer Prev** 2017; 18:2689–93.
- Cibersort. **Estimación do infiltrado inflamatório**. 2019. <URL: <https://cibersort.stanford.edu/>>. [2019 fev 6].
- Ciriello G, Gatza ML, Beck AH, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. **Cell** 2015; 163:506–19.
- Cuylen S, Blaukopf C, Politi AZ, et al. Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. **Nature** 2016; 535:308–12.
- Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. **Clin Cancer Res** 2007; 13:4429–34.
- Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. **Bioinformatics** 2013; 29:15–21.
- Dowsett M, Nielsen TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. **J Natl Cancer Inst** 2011; 103:1656–64.
- Genentech-Press-Releases. **Genentech - Press Releases**. 2006. <URL: <https://bit.ly/2ZX5UGu>>. [2019 fev 6].

Gerdes J, Schwab U, Lemke H, et al. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. **Int J Cancer** 1983; 31:13–20.

GLOBOCAN. **online-analysis**. 2018. <URL: <https://bit.ly/2GTyrVI>, <https://bit.ly/2UUFJAj>>. [2019 fev 6].

Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. **Int J Ayurveda Res** 2010; 1:274–78.

Harris LN, Ismaila N, McShane LM, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. **J Clin Oncol** 2016; 34:1134–50.

Koboldt DC, Fulton RS, McLellan MD, et al. Comprehensive molecular portraits of human breast tumours. **Nature** 2012; 490:61–70.

Kohler BA, Sherman RL, Howlader N, et al. Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State. **J Natl Cancer Inst** 2015; 107:djv048.

Krop I, Ismaila N, Andre F, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. **J Clin Oncol** 2017; 35:2838–47.

Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. **J Clin Invest** 2011; 121:2750–67.

Liu Z, Li M, Jiang Z, et al. A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer. **Transl Oncol** 2018; 11:311–29.

Lumachi F, Luisetto G, Basso SM, et al. Endocrine therapy of breast cancer. **Curr Med Chem** 2011; 18:513–22.

Masuda H, Baggerly KA, Wang Y, et al. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. **Clin Cancer Res** 2013; 19:5533–40.

Matsumoto H, Koo SL, Dent R, et al. role of inflammatory infiltrates in triple negative breast cancer. **J Clin Pathol** 2015; 68:506–10.

METABRIC-dados. **Dados clínicos e de expressão - METABRIC**. 2018. <URL: http://www.cbioportal.org/study?id=brca_metabric>. [2019 fev 6].

Nature-Subjects-GO. 2019. <URL: <https://go.nature.com/2ZNo5OK>>. [2019 fev 6].

Nature-Subjects-IHC. 2019. <URL: <https://go.nature.com/2DFvxls>>. [2019 fev 6].

[NCI] National Cancer Institute. **What is cancer**. 2015. <URL: <https://bit.ly/2no8FQy>, <https://bit.ly/2DtIW1N>>. [2019 fev 6].

Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. **Nat Methods** 2015; 12:453–57.

Noman AS, Uddin M, Rahman MZ, et al. Overexpression of sonic hedgehog in the triple negative breast cancer: clinicopathological characteristics of high burden breast cancer patients from Bangladesh. **Sci Rep** 2016; 6:18830.

O'Hagan S, Wright Muelas M, Day PJ, et al. GeneGini: Assessment via the Gini Coefficient of Reference "Housekeeping" Genes and Diverse Human Transporter Expression Profiles. **Cell Syst** 2018; 6:230–44.

Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. **J Clin Oncol** 2009; 27:1160–67.

Pereira B, Chin SF, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. **Nat Commun** 2016; 7:11479.

Prosigna. **Prosigna**. 2018. <URL: <https://bit.ly/2PHDtHN>>. [2019 fev 6].

van Roozendaal LM, Smit LHM, Duijsens GHNM, et al. Risk of regional recurrence in triple-negative breast cancer patients: a Dutch cohort study. **Breast Cancer Res Treat** 2016; 156:465–72.

Singletary SE. Rating the risk factors for breast cancer. **Ann Surg** 2003; 237:474–82.

Storey JD. A direct approach to false discovery rates. **J Royal Sta Soc: Series B (Statistical Methodology)** 2002; 64:479–98.

TCGA-pipeline. **Processamento dos dados de expressão - Bioinformática**. 2018. <URL: <https://bit.ly/2GWpKtT>>. [2019 fev 6].

Vanderbilt. **TNBCtype**. 2018. <URL: <http://cbc.mc.vanderbilt.edu/tnbc/>>. [2019 fev 6].

Vieira AF, Schmitt F. An Update on Breast Cancer Multigene Prognostic Tests-Emergent Clinical Biomarkers. **Front Med (Lausanne)** 2018; 5:248.

[WHO] World Health Organization. **Cancer**. 2018. <URL: <https://bit.ly/2CSR75X>>. [2019 fev 6].

Wong E. **breast cancer pathogenesis and histologic vs. molecular subtypes**. 2012. <URL: <https://bit.ly/2DKZMHU>>. [2019 fev 6].

Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. **OMICS** 2012; 16:284–7.

Yu T, Di G. Role of tumor microenvironment in triple-negative breast cancer and its prognostic significance. **Chin J Cancer Res** 2017; 29:237–52.

Yuen GJ, Demissie E, Pillai S. B lymphocytes and cancer: a love-hate relationship. **Trends Cancer** 2016; 2:747–57.