

Tamires Avila de Souza Clemente

EXPLORAÇÃO DE DADOS PÚBLICOS PARA INVESTIGAÇÃO *IN SILICO* DE GENES RELACIONADOS A EOSINÓFILOS NO CÂNCER DE MAMA



**Monografia apresentada ao Instituto de Microbiologia Paulo de Góes, da Universidade Federal do Rio de Janeiro, como pré-requisito para a obtenção do grau de Bacharel em Ciências Biológicas: Microbiologia e Imunologia.**

INSTITUTO DE MICROBIOLOGIA PAULO DE GÓES  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
RIO DE JANEIRO  
DEZEMBRO/2025

**Trabalho realizado no Instituto de Biofísica Carlos Chagas Filho, UFRJ, sob a orientação do Professor Bruno Lourenço Diaz, co-orientação de Fernanda Neves da Cunha e tutoria de Márcia Giambiagi de Marval.**

## FICHA CATALOGRÁFICA

### CIP - Catalogação na Publicação

A626e      Avila de Souza Clemente, Tamires  
Exploração de dados públicos para investigação in silico de genes relacionados a eosinófilos no câncer de mama. / Tamires Avila de Souza Clemente. -- Rio de Janeiro, 2025.  
56 f.

Orientador: Bruno Lourenço Diaz.

Coorientadora: Fernanda Neves da Cunha.

Trabalho de conclusão de curso (graduação) - Universidade Federal do Rio de Janeiro, Instituto de Microbiologia, Bacharel em Ciências Biológicas: Microbiologia e Imunologia, 2025.

1. câncer de mama. 2. eosinófilos. 3. análise in silico. 4. transcriptoma. 5. bioinformática. I. Lourenço Diaz, Bruno, orient. II. Neves da Cunha, Fernanda, coorient. III. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

**INSTITUTO DE MICROBIOLOGIA PAULO DE GÓES / UFRJ**  
**COORDENAÇÃO DE ENSINO DE GRADUAÇÃO**

**ATA DA APRESENTAÇÃO DE MONOGRAFIA PARA APROVAÇÃO NO RCS DE TRABALHO DE CONCLUSÃO DE CURSO, BACHARELADO EM CIÊNCIAS BIOLÓGICAS: MICROBIOLOGIA E IMUNOLOGIA**

**ALUNO(A): TAMIRES AVILA DE SOUZA CLEMENTE**

**DRE: 122074076**

**BANCA EXAMINADORA:**

Prof. Leticia de Albuquerque Maranhão Carneiro (Presidente)  
 Prof. Israel Diniz Lima  
 Prof. Roberta Soares Faccion  
 Prof. Fabianno Ferreira Dutra (Suplente)

**Título da Monografia: “Exploração de dados públicos para investigação *in silico* de genes relacionados a eosinófilos no câncer de mama”**

**Local:** Sala G1-030, Instituto de Biofísica Carlos Chagas Filho.

**Data e hora de início:** 08 de Dezembro de 2025

Em sessão pública, após exposição de cerca de 50 minutos, o aluno foi argüido pelos membros da Banca Examinadora, demonstrando suficiência de conhecimentos e capacidade de sistematização no tema de sua Monografia, tendo, então, obtido nota 9,0 neste requisito do RCS de **TRABALHO DE CONCLUSÃO DE CURSO**. Na forma regulamentar, foi lavrada a presente ata que é assinada pelos membros da banca examinadora, aluno, orientador (e/ou coorientador) e pelo coordenador do RCS.

Rio de Janeiro, 8 de dezembro de 2025.

**NOTA:**

9,0

9,0

9,0

Banca Examinadora:

Prof. Leticia de Albuquerque Maranhão Carneiro (Presidente)

Israel Diniz Lima

Prof. Israel Diniz Lima

Prof. Roberta Soares Faccion

Prof. Fabianno Ferreira Dutra (Suplente)

**Aluno(a):**

**Orientador(a)**

**(e coorientador):**

**Coordenadora de**

**TCC:**

Tamires Avila de Souza Clemente

Tamires Avila de Souza Clemente

Prof. Bruno Lourenço Díaz

Prof. Marinella Silva Laport

**Dedico este trabalho às mulheres que  
um dia sonharam com a biologia,  
com a ciência, mas não puderam  
perseguir seu sonho. À minha mãe,  
que me inspirou a estar no ramo da  
ciência.**

## AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me sustentado nessa trajetória e ter me permitido conhecer e concluir um curso que se dedica a fazer ciência, meu sonho de criança.

Agradeço à toda minha família, que mais do que meu apoio, foi minha fortaleza nesses 4 anos árduos, sempre me oferecendo encorajamento para seguir em frente e abraçar o que amo.

Em especial, agradeço à minha mãe, que dividiu o sonho de estudar Biologia na UFRJ comigo. Foi o sonho de uma jovem de família pobre, que teve que trabalhar cedo e estudar o que lhe oferecesse emprego, que hoje me fez estar aqui também. Foi a paixão de estudarmos juntas durante o período de provas do Ensino Médio que percebi como a ciência é linda. E com esse sonho, que um dia foi guardado, hoje concluo a graduação na Universidade Federal do Rio de Janeiro.

Ao meu pai, que sempre priorizou apoiar meus estudos e foi minha fonte de inspiração no quesito de gostar de buscar conhecimento. Àquele que não teve a oportunidade, mas hoje vê sua filha se formando, muito obrigada, eu jamais teria conseguido sem o senhor. Foram nossas conversas, nossos bate-papos, nosso código “pai, pode ir tomar um refri comigo?”, que me aliviaram nos dias mais difíceis. Foram esses momentos de aconchego, em que eu ouvia suas histórias de vida e de quartel, que me deram força de suportar dia após dia com mais leveza.

Às minhas irmãs, Thaís Helena e Ana Beatriz, toda a minha gratidão. Foram vocês que me acolheram quando eu estava ansiosa em relação às provas, trabalhos e apresentações. Vocês sempre estiveram ali como meu ponto de apoio, vindo me abraçar ao mínimo som de choro, fosse por ansiedade ou felicidade. Vocês fizeram meus dias mais felizes e leves ao chegar em casa e poder ouvir suas rotinas com tanta alegria. Vocês tornaram minha graduação mais leve, porque eu sabia que sempre poderia correr para o abraço das duas. Esses pequenos momentos terei guardado como grande parte da minha formação como bióloga.

Aos meus avós Oswaldo e Aparecida, obrigada pelo amor e admiração de vocês. Minha querida avó Cida, que não teve a oportunidade de ver minha ingressão na área da saúde, hoje sua neta está trilhando os caminhos que um dia você também trilhou. Você como enfermeira, eu como bióloga com enfoque na área de imunologia. Obrigada por cada conselho, cada risada, cada afago. Admito que foi difícil não poder compartilhar histórias de laboratório com você. E ao meu avô Oswaldo, obrigada por tanto. Por sempre perguntar o que faço na faculdade, por se gabar à rua inteira que sua neta vai ser doutora (amém!). Seu apoio e admiração, desde o início, pelo que faço e escolhi como carreira foram combustíveis para que eu chegasse até aqui. Obrigada por me olhar com esses olhos de quem vai longe ainda e que vai fazer muita coisa.

À minha avó Manoelina, obrigada por seu amor em forma de cuidado e admiração. O fato de todo dia perguntar se cheguei bem à faculdade ou em casa, a famosa preocupação de vó, me fez continuar a graduação com garra e determinação. Para minha avózinha Lina, que não teve a oportunidade de sequer ser alfabetizada, eu sou grata por cada momento por sempre se interessar na minha rotina, nas minhas apresentações, por me apoiar e por se interessar em tudo relacionado à minha formação acadêmica. Ter a senhora me acompanhando nesse processo fez com que eu levasse a graduação da melhor forma possível.

Ao meu namorado Nicholas, sou grata por seu amor e por seu carinho nesse caminho. Nos conhecemos pela UFRJ, e hoje você faz parte da minha vida. Obrigada por aparecer nas apresentações mais aleatórias do meu curso, somente porque eu tinha pedido para me dar maior conforto, ainda que no início do relacionamento. Você foi parte importante de tudo isso, e sou extremamente grata por isso.

Às minhas amigas do colégio, Rayanne, Maria Eduarda, Ana Carolina, Thayná e Mylene, que durante essa trajetória se certificaram de manter a amizade mais forte do que nunca, obrigada por tudo. Nossas conversas e encontros para matar a saudade e atualizar todas as fofocas dos meses me deram um conforto imensurável nos dias de graduanda. Obrigada, meninas.

Às minhas amigas de graduação, Maria Eduarda Galdino (Maria Duda), Isabelle Colchone e Ana Júlia Rangel, obrigada por tudo meninas! Vocês fizeram a graduação ser divertida com nossas fofocas, risadas, montagem de esferoides caóticas e companheirismo. Sou muito grata pelos pequenos aos grandes momentos que compartilhamos juntas. Sem vocês, a graduação não teria sido a mesma, de forma alguma.

Ao Laboratório de Inflamação, meus agradecimentos. Primeiramente, por ser aceita como aluna de Iniciação Científica pelo professor Bruno Diaz, por todo conhecimento compartilhado e adquirido, e em segundo pelo companheirismo e acolhimento da equipe constituinte do laboratório. Em especial, agradeço à minha co-orientadora, Fernanda Neves, pela infinita paciência, pela bondade e por sempre estar visando a melhor forma de me orientar e ensinar. Agradeço ao laboratório pela imensa contribuição na minha formação profissional.

## RESUMO

**TAMIRES AVILA DE SOUZA CLEMENTE**

### **EXPLORAÇÃO DE DADOS PÚBLICOS PARA INVESTIGAÇÃO *IN SILICO* DE GENES RELACIONADOS A EOSINÓFILOS NO CÂNCER DE MAMA**

**Orientador: Bruno Lourenço Diaz**

**Co-orientadora: Fernanda Neves da Cunha**

O câncer de mama é responsável pela segunda maior taxa de incidência e quarta de mortalidade (excluindo câncer de pele não melanoma) mundialmente, com as mulheres respondendo por cerca de 99% dos casos e com apenas 1% de homens, sendo a primeira causa de morte por câncer entre as mulheres. O carcinoma mamário é composto por um microambiente tumoral heterogêneo que compreende células do estroma e epiteliais, envolvendo células tumorais, fibroblastos associados ao tumor (CAFs), e células do sistema imunológico infiltradas. Dentre os leucócitos, os eosinófilos são granulócitos classicamente associados a quadros de alergia e infestação parasitária. Contudo, no contexto do câncer, seu papel é destacado de forma contraditória, em que a infiltração no tecido tumoral de mama é frequentemente correlacionada a um bom prognóstico. A análise *in silico* de dados de transcriptomas de repositórios públicos permite analisar a expressão gênica e entender a interação de genes relevantes ao metabolismo do organismo. Assim, nosso estudo tem como objetivo analisar a expressão gênica e a interação de genes associados à presença de eosinófilos no microambiente tumoral do câncer de mama por assinaturas gênicas que modulam a presença ou a função desses granulócitos no tumor. Para investigar alterações envolvidas na progressão tumoral pela presença, pelo recrutamento e manutenção de eosinófilos nesse tecido, examinamos dados de RNAseq obtidos de amostras do tecido adjacente ao tumor, do repositório público TCGA, utilizando ferramentas de bioinformática pela linguagem de programação Python, visualizando os dados resultantes pelas bibliotecas Matplotlib, Pandas, Seaborn e Lifelines. Os resultados apresentaram um perfil de mau prognóstico com relação a presença de eosinófilos, que é definida a partir da expressão dos genes avaliados Galectina 10 (CLC), EPX (EPO), IL5RA, PRG2 (MBP), pela maior expressão de CLC no subtipo mais agressivo, Basal-like, além de observarmos a presença do infiltrado eosinofílico em todos os tumores avaliados, apesar do grau de agressividade, por meio da expressão de EPX e PRG2. Para recrutamento, definido pela expressão dos genes CCL11 (Eotaxina 1), CCL24 (Eotaxina 2) e CCL26 (Eotaxina 3), observamos um perfil associado ao bom prognóstico, pela alta presença de CCL11, importante na formação da mama, além da maior distribuição nos subtipos de melhor desfecho clínico. Por sua vez, a manutenção de eosinófilos, avaliado pela expressão gênica de IL33, IL25, IL5 e TSLP, mostrou um perfil associado ao mau prognóstico, pela expressão aumentada de IL5 no subtipo de maior agressividade. Assim, apesar da disponibilidade de poucos estudos na literatura acerca da eosinofilia no contexto do câncer, nossos dados condizem com os trabalhos descritos na literatura, sugerindo uma contrariedade em relação ao papel dos eosinófilos no câncer de mama, ainda que os mecanismos referentes a essa relação não sejam completamente conhecidos.

Palavras-chave: câncer de mama, eosinófilos, análise *in silico*, transcriptoma, bioinformática, Python

**ABSTRACT****TAMIREZ AVILA DE SOUZA CLEMENTE****EXPLORATION OF PUBLIC DATA FOR *IN SILICO* INVESTIGATION OF EOSINOPHIL RELATED GENES IN BREAST CANCER****Orientador: Bruno Lourenço Diaz****Co-orientadora: Fernanda Neves da Cunha**

Breast cancer is responsible for the second highest incidence rate and the fourth highest mortality rate worldwide (excluding non-melanoma skin cancer), with women representing about 99% of all cases and men only 1%, being the leading cause of cancer-related death among women. Breast carcinoma consists of a heterogeneous tumor microenvironment that includes stromal and epithelial cells, involving tumor cells, cancer associated fibroblasts (CAFs), and infiltrating immune cells. Among leukocytes, eosinophils are granulocytes classically associated with allergic responses and parasitic infections. However, in the context of cancer, their role is associated in a contradictory way, but in breast cancer their infiltration into tumor tissue is frequently correlated with a favorable prognosis. In silico analysis of transcriptome data from public repositories enables the investigation of gene expression and the understanding of interactions between relevant genes. Thus, the aim of our study is to analyze gene expression and gene interactions associated with the presence of eosinophils in the tumor microenvironment of breast cancer through gene signatures that modulate the presence or function of these granulocytes within the tumor. To investigate alterations involved in tumor progression by eosinophils presence, recruitment and maintenance in this tissue, we examined RNA-seq data, obtained from tumor adjacent tissue, from the public repository TCGA, using bioinformatics tools through programming language Python and visualized the resulting data through the libraries Matplotlib, Pandas, Seaborn and Lifelines. The results presented a bad prognostics profile in regarding to eosinophil presence, defined from the expression of the evaluated genes Galectin 10 (CLC), EPX (EPO), IL5RA, PRG2 (MBP), by the highest CLC expression in the most aggressive subtype, *Basal-like*. Besides, we observed the presence of the eosinophilic infiltrated in all the tumors, despite the aggressiveness grade, from the EPX and PRG2 expression. About recruitment, defined by the expression of the genes CCL11 (Eotaxin 1), CCL24 (Eotaxin 2) e CCL26 (Eotaxin 3), we observed a profile associated with good prognosis, by the highest presence of CCL11, important on the development of the breast, in addition to greater distribution between the subtypes of better outcome. On the other hand, the eosinophils maintenance, evaluated by the gene expression of IL33, IL25, IL5 and TSLP, showed a profile associated with bad prognosis, for the increased expression of IL5 in the most aggressive subtype. Thus, despite the limited number of studies in the literature addressing eosinophilia in the context of cancer, our data match with the papers described in the literature, suggesting an opposition regarding to the role of eosinophils in breast cancer, even though the mechanisms relating to this relation are not fully understood.

**Keywords:** breast cancer, eosinophils, in silico analysis, transcriptome, bioinformatics, Python

**RESUMO PARA PESSOAS LEIGAS****TAMIRES AVILA DE SOUZA CLEMENTE****INVESTIGAÇÃO DE DADOS PÚBLICOS POR BIOINFORMÁTICA DE GENES  
RELACIONADOS A EOSINÓFILOS NO CÂNCER DE MAMA****Orientador: Bruno Lourenço Diaz****Co-orientadora: Fernanda Neves da Cunha**

**Resumo da Monografia apresentada no Instituto de Microbiologia Paulo de Góes da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para obtenção do título de Bacharel em Ciências Biológicas: Microbiologia e Imunologia e aprovação no RCS Trabalho de Conclusão de Curso.**

O câncer de mama é a segunda principal causa de morte por câncer no mundo, tendo como população mais afetada as mulheres. Ele se desenvolve em uma massa celular que inclui não apenas células tumorais, mas também células de suporte e do sistema imunológico, como os eosinófilos. Tradicionalmente associados a alergias e parasitas, os eosinófilos quando se acumulam no tumor parecem levar a um resultado mais favorável para o paciente. Para confirmar se estas células auxiliam no combate ao câncer, nosso estudo utilizou informações sobre o nível de expressão de diferentes genes, obtidas de bancos de dados públicos. A análise foi realizada com dados obtidos de amostras provenientes do tecido vizinho ao tumor, tecido tumoral, e tecido metastático, buscando identificar padrões de genes associados à presença de eosinófilos. A análise e interpretação dos níveis de expressão dos genes de interesse e a posterior visualização dos resultados foram feitos com programas de computador específicos. O objetivo foi identificar os padrões de genes que modulam a presença e a função dos eosinófilos no câncer de mama. No subtipo de menor agressividade, a presença de genes associados a presença, manutenção e resposta imunológica do tipo 2 — característica de alergias — e de acúmulo de eosinófilos estão associados a formas de menor agressividade e melhor prognóstico do câncer de mama.

## LISTA DE FIGURAS

|                                                                                                                                                                        |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1.</b> Microambiente tumoral do câncer de mama.....                                                                                                          | 7  |
| <b>Figura 2.</b> Eosinófilos infiltrados no tumor podem influenciar células do microambiente tumoral, levando a expressão de perfis anti-tumorais ou pró-tumorais..... | 10 |
| <b>Figura 3.</b> Site UCSC Xena Browser que armazena dados públicos de câncer.....                                                                                     | 17 |
| <b>Figura 4.</b> Fluxograma de amostras utilizadas de acordo com a filtragem para cada análise.....                                                                    | 18 |
| <b>Figura 5.</b> Dataset de expressão gênica RNAseq do cohort TCGA Breast Cancer (BRCA) para análise de dados.....                                                     | 19 |
| <b>Figura 6.</b> Dataset de expressão fenotípica de genes.....                                                                                                         | 21 |
| <b>Figura 7.</b> Dataset de dados curados de sobrevida das amostras, mostrando tempo e status de sobrevida.....                                                        | 22 |
| <b>Figura 8.</b> Gráfico dos tipos de amostras obtidas.....                                                                                                            | 25 |
| <b>Figura 9.</b> Distribuição de amostras entre os subtipos PAM50 de câncer de mama.....                                                                               | 26 |
| <b>Figura 10.</b> Comparação da expressão simultânea dos genes CLC, EPX, IL5RA, PRG2 contra as demais combinações de genes por subtipo.....                            | 27 |
| <b>Figura 11.</b> Comparação da expressão simultânea dos genes EPX e PRG2 contra a expressão de nenhum dos dois. ....                                                  | 28 |
| <b>Figura 12.</b> Frequência da expressão gênica de PRG2 por subtipos PAM50.....                                                                                       | 29 |

|                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 13.</b> Frequência da expressão gênica de EPX por subtipos PAM50.....                                                                                 | 29 |
| <b>Figura 14.</b> Frequência da expressão gênica de <i>IL5RA</i> por subtipos PAM50.....                                                                        | 30 |
| <b>Figura 15.</b> Frequência da expressão gênica de CLC por subtipos PAM50.....                                                                                 | 31 |
| <b>Figura 16.</b> Gráfico Kaplan-Meier de sobrevida de alta vs baixa expressão de genes de presença de eosinófilos.....                                         | 32 |
| <b>Figura 17.</b> Comparação da expressão dos genes <i>CCL11</i> , <i>CCL24</i> , <i>CCL26</i> contra as demais combinações de genes por subtipo.....           | 33 |
| <b>Figura 18.</b> Frequência da expressão gênica de <i>CCL11</i> por subtipos PAM50.....                                                                        | 34 |
| <b>Figura 19.</b> Frequência da expressão gênica de <i>CCL26</i> por subtipos PAM50.....                                                                        | 34 |
| <b>Figura 20.</b> Frequência da expressão gênica de <i>CCL24</i> por subtipos PAM50.....                                                                        | 35 |
| <b>Figura 21.</b> Gráfico Kaplan-Meier de sobrevida de alta vs baixa expressão de genes de recrutamento de eosinófilos.....                                     | 36 |
| <b>Figura 22.</b> Comparação da expressão dos genes <i>IL5</i> , <i>IL33</i> , <i>IL25</i> , <i>TSLP</i> contra as demais combinações de genes por subtipo..... | 37 |
| <b>Figura 23.</b> Frequência da expressão gênica de <i>IL33</i> por subtipos PAM50.....                                                                         | 38 |

|                                                                                         |    |
|-----------------------------------------------------------------------------------------|----|
| <b>Figura 24.</b> Frequência da expressão gênica de <i>IL25</i> por subtipos PAM50..... | 39 |
| <b>Figura 25.</b> Frequência da expressão gênica de <i>IL25</i> por subtipos PAM50..... | 40 |
| <b>Figura 26.</b> Frequência da expressão gênica de <i>IL5</i> por subtipos PAM50.....  | 41 |

## LISTA DE ABREVIATURAS E SIGLAS

AMPc – Monofosfato de Adenosina cíclico

BRCA1 – Gene do Câncer de Mama 1 do inglês *Breast Cancer gene 1*

BRCA2 – Gene do Câncer de Mama 2 do inglês *Breast Cancer gene 2*

CAFs – Cancer-associated fibroblasts (Fibroblastos associados ao câncer)

CCR3 – Receptor de quimiocina C-C tipo 3

CCL11 – Eotaxina-1, ligante de quimiocina C-C tipo 1, do inglês *C-C motif chemokine 11*

CCL24 - Eotaxina 2, ligante de quimiocina C-C tipo 2, do inglês *Chemokine (C-C motif) ligand 24*

CCL26 - Eotaxina 3, ligante de quimiocina C-C tipo 3, do inglês *Chemokine (C-C motif) ligand 26*

CLC- nomenclatura HUGO do gene Galectina 10

EDN - Neurotoxina derivada de eosinófilos, do inglês “*Eosinophil derived neurotoxin*”

EGFR – Receptor do fator de crescimento epidérmico do inglês *Epidermal Growth Factor Receptor*

EPO- Peroxidase de eosinófilo, do inglês *eosinophil peroxidase*

EPX- Nomenclatura HUGO do gene codificador da proteína EPO

ER – Receptor de estrogênio do inglês *Estrogen Receptor*

ER- $\alpha$  – Receptor de estrogênio alfa do inglês *Estrogen Receptor alfa*

ER- $\beta$  – Receptor de estrogênio beta do inglês *Estrogen Receptor beta*

ESR1 – Gene 1 do Receptor de Estrogênio do inglês *Estrogen Receptor 1 gene*

FABP7 – Proteína ligante de ácidos graxos 7 do inglês *Fatty Acid Binding Protein 7*

FGFR1 – Receptor de fator de crescimento de fibroblastos 1 do inglês *Fibroblasts Growth Factor Receptor 1*

GATA-1 – Proteína de ligação GATA 1 do inglês *GATA binding protein 1*

GATA-3 – Proteína de ligação GATA 3 do inglês *GATA binding protein 3*

GRB7 – Proteína 7 ligada ao receptor do fator de crescimento do inglês *Growth factor receptor-bound protein 7*

HER-1 – Fator de crescimento epidermal humano 1 do inglês *Human Epidermal growth factor Receptor 1*

HER-2 – Fator de crescimento epidermal humano 2 do inglês *Human Epidermal growth factor Receptor 2*

HIF1- $\alpha$  – Fator indutor de hipóxia 1 do inglês *Hypoxia-Inducible Factor 1-alfa*

IL-2 – Interleucina 2

IL-4 – Interleucina

IL5RA- receptor alfa de IL5, do inglês *IL5 receptor alpha*

IL-5 – Interleucina 5

IL-10 – Interleucina 10

IL-12 – Interleucina 12

IL-13 – Interleucina 13

LTC4 – Leucotrieno 4

MBP – Proteína básica principal 1 do inglês *Major Basic Protein*

MAPK – *Mitogen-Activated Protein Kinase*

MMP9 – Metaloproteinase de matriz 9

mPR $\alpha$  – Receptor Progesterona alfa de membrana do inglês *Membrane Progesterone Receptor alpha*

PGRMC1 – *Progesterone Receptor Membrane Component 1*

PI3K – Fosfoinosítideo 3-quinase do inglês *Phosphoinositide 3-Kinase*

PRG2- Nomenclatura HUGO do gene codificador da proteína MBP

RANKL – Ligante do receptor ativador de fator nuclear kappa-B do inglês *Receptor activator of nuclear factor kappa-B ligand*

TATE – Eosinofilia tecidual associada ao tumor do inglês *Tumor-Associated Tissue Eosinophilia*

TGF – Fato transformador de crescimento do inglês *Transforming Growth Factor*

TP53 – Proteína tumoral p53 do inglês *Tumor Protein p53*

XBP1 – Proteína ligante X-box 1 do inglês *X-box Binding Protein 1*

ZNF703 – Proteína dedo de Zinco 703 do inglês *Zinc Finger Protein 703*

## SUMÁRIO

|                                                                                 |      |
|---------------------------------------------------------------------------------|------|
| RESUMO .....                                                                    | vii  |
| ABSTRACT.....                                                                   | viii |
| RESUMO PARA PESSOAS LEIGAS.....                                                 | ix   |
| LISTA DE FIGURAS .....                                                          | x    |
| LISTA DE ABREVIATURAS E SIGLAS .....                                            | xiv  |
| 1. INTRODUÇÃO .....                                                             | 1    |
| 1.1. Epidemiologia .....                                                        | 1    |
| 1.2. Câncer de mama.....                                                        | 1    |
| 1.2.1. Classificação molecular.....                                             | 2    |
| 1.2.1.1. Luminal A.....                                                         | 2    |
| 1.2.1.2. Luminal B.....                                                         | 3    |
| 1.2.1.3. HER-2 .....                                                            | 4    |
| 1.2.1.4. <i>Basal like</i> .....                                                | 5    |
| 1.3. O microambiente tumoral do câncer de mama .....                            | 6    |
| 1.3.1. Os eosinófilos no microambiente tumoral.....                             | 8    |
| 1.4. Análise transcriptômica e bioinformática no câncer de mama.....            | 13   |
| 2. JUSTIFICATIVA .....                                                          | 14   |
| 3. OBJETIVOS .....                                                              | 16   |
| 3.1. Objetivo geral .....                                                       | 16   |
| 3.2. Objetivos específicos .....                                                | 16   |
| 4. METODOLOGIA .....                                                            | 17   |
| 4.1. Obtenção de amostras .....                                                 | 17   |
| 4.2. Processamento de dados .....                                               | 23   |
| 4.3. Análise dos dados .....                                                    | 23   |
| 4.3.1. Análise dos grupos de sobrevida global .....                             | 24   |
| 4.3.2. Análise da frequência de expressão gênica de acordo com os subtipos..... | 24   |
| 4.4. Análise estatística .....                                                  | 24   |
| 5. RESULTADOS .....                                                             | 25   |
| 5.1. Classificação de amostras analisadas .....                                 | 25   |

|                                                                                                                          |    |
|--------------------------------------------------------------------------------------------------------------------------|----|
| 5.2. Avaliação da expressão gênica da presença de eosinófilos no tumor de mama ....                                      | 26 |
| 5.2.1. Correlação dos subtipos PAM50 com a frequência de expressão de genes relacionados à presença de eosinófilos ..... | 26 |
| 5.2.2. Avaliação da expressão gênica em cada subtipo do câncer de mama.....                                              | 28 |
| 5.2.3. Correlação entre a expressão dos genes de eosinófilos e a sobrevida dos pacientes .....                           | 31 |
| 5.3. Avaliação do recrutamento de eosinófilos no tumor de mama .....                                                     | 32 |
| 5.3.1. Correlação dos subtipos PAM50 com a frequência de expressão de quimiocinas de recrutamento de eosinófilos.....    | 32 |
| 5.3.2. Avaliação da expressão de quimiocinas capazes de recrutar eosinófilos para o microambiente tumoral.....           | 33 |
| 5.3.3. Correlação da expressão de quimiocinas de recrutamento de eosinófilos com a sobrevida global dos pacientes.....   | 35 |
| 5.4. Avaliação da manutenção de eosinófilos no microambiente tumoral do câncer de mama .....                             | 36 |
| 5.4.1. Correlação dos subtipos PAM50 com a frequência de expressão de citocinas de manutenção de eosinófilos.....        | 36 |
| 5.4.2. Avaliação da expressão de citocinas capazes de realizar a manutenção de sobrevida dos eosinófilos.....            | 37 |
| 5.4.3. Correlação da expressão de citocinas de manutenção de eosinófilos com a sobrevida global dos pacientes .....      | 41 |
| 6. DISCUSSÃO .....                                                                                                       | 42 |
| 7. CONCLUSÃO .....                                                                                                       | 45 |
| 8. REFERÊNCIAS BIBLIOGRÁFICAS.....                                                                                       | 47 |

## 1. INTRODUÇÃO

### 1.1. Epidemiologia

O câncer de mama é o câncer de maior incidência nas mulheres no Brasil e no mundo, acometendo cerca de 2,3 milhões de indivíduos (WHO, 2025). As taxas de ocorrência revelam que 99% do sexo feminino são afetados e 0,5 a 1% do sexo masculino, levando 670.000 indivíduos a óbito no ano de 2022 (WHO, 2025).

No Brasil, estimou-se que em 2023 haveria 73.610 novos casos, considerando o aumento no decorrer dos anos. Entre 2020 e 2021, mais de 18.139 óbitos por câncer de mama foram registrados, dos quais 220 referiam-se a homens e 18.139 a mulheres (INCA, 2023).

A presença de fatores de risco como idade, sexo e a longa exposição à altos níveis do estrogênio favorecem o desenvolvimento do carcinoma mamário (Wilkinson, Louise, and Toral Gathani. 2022). Adicionalmente, nos países desenvolvidos o risco aumenta de acordo com o estilo de vida ocidental, do inglês “*Western lifestyle*”, que engloba o frequente consumo de alimentos industrializados e produtos de “*fast-food*”, frequente tabagismo, estresse excessivo e falta de atividades físicas (Wilkinson, Louise e Toral Gathani. 2022).

Além disso, há o fator de risco da predisposição hereditária à mutação de genes associados ao câncer de mama (*BRCA 1*, *BRCA 2* e *PALB2*), representada em 5 a 10% dos pacientes (Smolarz, Beata *et al.* 2022).

### 1.2. Câncer de mama

O câncer de mama é caracterizado como um conjunto de desordens que acarretam na desregulação do crescimento mamário (Kumar, V., Abbas, AK, e Aster, JC. 2017), sendo decorrente de fatores ambientais e genéticos que levam ao processo de transformação por hiperplasia, mudanças pré-malignas e carcinoma *in situ* (Nolan, Emma *et al.* 2023). Dentre esses fatores, destacam-se alterações genômicas e expressão de genes, em conjunto ao microambiente tumoral, que contribuem para a classificação de acordo com a histopatologia, as características clínicas e análise molecular avançada (Nolan, Emma *et al.* 2023).

Além disso, a partir dos parâmetros de número e características das células, local de secreção, arquitetura celular e localização no tecido mamário, o carcinoma é subdividido em ductal e lobular (Nolan, Emma *et al.* 2023). Os Carcinomas Invasivos Ductais Não Específicos, caracterizados pela morfologia com núcleo saliente e número elevado de mitoses, abrangem 40% a 70% dos pacientes, enquanto que o Carcinoma Lobular Invasivo apresenta

células tumorais com pequenas atipias e distribuídas de forma uniforme em um padrão concêntrico no estroma, é diagnosticado em 5-15% dos novos casos e afetando principalmente mulheres de idade avançada.

### **1.2.1. Classificação molecular**

Dentre as classificações de câncer de mama, os subtipos moleculares são baseados na expressão imunohistoquímica dos receptores hormonais de estrogênio e progesterona, e do receptor de fator de crescimento epidermal HER-2 (Orrantia-Borunda *et al.*, 2022).

Proposta por Charles M. Perou, a classificação molecular do câncer de mama surgiu da análise de microarranjos de DNA complementares, que avaliaram simultaneamente a expressão de milhares de genes (Zhang, Xinmin. 2023; Perou, C. M. *et al.* 2000).

Essa abordagem revelou padrões de expressão gênica distintos, permitindo a divisão dos tumores em subtipos com perfis moleculares específicos (Zhang, Xinmin. 2023). Por meio de análises de agrupamento hierárquico, foi possível identificar assinaturas gênicas características associadas a diferentes comportamentos clínicos e biológicos dos tumores (Perou C. M. *et al.* 2000; Sørlie T. *et al.* 2001).

Assim, de acordo com a expressão ou ausência dos receptores, foi proposta a divisão de 4 subtipos moleculares intrínsecos: Luminal A, Luminal B, HER-2 e Basal-like (Nolan, Emma *et al.* 2023; Perou C. M. *et al.* 2000).

#### **1.2.1.1. Luminal A**

O subtipo Luminal A é o perfil molecular de maior predominância entre as pacientes, sendo caracterizado como tumores de menor agressividade (baixo grau) (Orrantia-Borunda *et al.* 2022).

É caracterizado pela expressão de receptores de estrogênio e progesterona, expressos no epitélio luminal do ducto mamário, mas com ausência de expressão de HER-2 (Orrantia-Borunda *et al.* 2022), o que justifica sua baixa proliferação comparado aos demais subtipos.

Esse subtipo apresenta características morfológicas semelhantes ao tecido mamário normal com expressão de genes característicos luminais como ESR1, GATA-3, XBP1, FOXA1, Bcl-2, Citoqueratina CK8/18 e baixa expressão do marcador de proliferação Ki-67 (Nolan, Emma *et al.* 2023), proteína presente em todas as fases do ciclo celular, mas ausente em células em repouso, que se encontram na fase G0 (Li, Lian Tao *et al.* 2014) .

Estudos demonstraram que a expressão do marcador imunohistoquímico GATA-3 é maior no subtipo Luminal A, permitindo a diferenciação em relação aos outros grupos de câncer de mama (Eroles, P. *et al.* 2012).

O marcador FOXA1 é requerido para a expressão ótima de pelo menos 50% dos genes regulatórios de ER- $\alpha$  e da proliferação induzida por estrogênio. A dependência dos cânceres de mama ao ER- $\alpha$  para sobrevivência e proliferação celular podem, inclusive, ser relacionados com os níveis de FOXA1 (Badve, Sunil *et al.* 2007).

Cerca de 13% dos casos de Luminal A apresentam mutação em *TP53*, que codifica o fator de transcrição p53 (Sørlie, T *et al.* 2001), e sua baixa proliferação pode ser explicada pela menor expressão de genes e marcadores proliferativos, como o Ki-67 (Eroles, P. *et al.* 2011).

Pacientes com esse subtipo apresentam bom prognóstico, com recidiva em torno de 27%, geralmente em metástase para ossos (Orrantia-Borunda *et al.* 2022), melhores taxas de sobrevivência quando comparados aos demais tumores mamários (Nolan, Emma *et al.* 2023) e são altamente responsivos aos tratamentos hormonais (Orrantia-Borunda *et al.* 2022).

#### **1.2.1.2. Luminal B**

O subtipo Luminal B de câncer de mama é caracterizado por uma menor expressão dos receptores hormonais de estrogênio (ER) e progesterona (PR), quando comparado ao subtipo Luminal A e correspondendo a cerca de 40% dos casos de câncer de mama (Yersal, O. e Sabri, B. 2014).

Em contrapartida, apresenta uma alta expressão de genes relacionados à proliferação e progressão do ciclo celular. Sua assinatura gênica é caracterizada pela presença dos genes FGFR1, HER1, PI3K e Src (Nolan, Emma *et al.* 2023).

Além disso, há uma expressão aumentada de genes ligados a fatores de crescimento, e aproximadamente 20% dos tumores Luminal B também expressam o receptor HER-2 (Ades, F. *et al.* 2014).

Esse subtipo exibe uma instabilidade genômica mais acentuada, com maior frequência de mutações e aberrações cromossômicas (Ades, F. *et al.* 2014). Dentre os oncogenes mais expressos, destaca-se o ZNF703, cuja expressão está associada à indução de proliferação celular de forma independente da estimulação de estrogênio e a um pior prognóstico (Ades, F. *et al.* 2014).

Como forma de classificação, o grau de um tumor é definido pela avaliação histológica das células tumorais de acordo com seu estado de diferenciação, podendo ser agravados pela formação de túbulos desorganizados, maior pleomorfismo nuclear e maior frequência de células em mitose (Carriaga, M. T., e Henson, D. E. 1995). Sendo assim, os tumores Luminal B apresentam grau mais elevado que os Luminal A (Li, Zhi-Hua *et al.* 2016), sendo considerados mais agressivos.

Eles se distinguem por alterações no número de cópias de genes, padrões de metilação do DNA e mutações somáticas pontuais (Nolan, Emma *et al.* 2023). Destaca-se a mutação no gene GATA-3, associada à sensibilidade endócrina e à diminuição da expressão de Ki-67 (Nolan, Emma *et al.* 2023).

### **1.2.1.3. HER-2**

O subtipo HER-2-enriquecido de câncer de mama é caracterizado molecularmente pela alta expressão de genes localizados na região do amplicon ERBB2, no cromossomo 17 região q22–24, incluindo os genes ERBB2 e GRB7 (Sørlie, T *et al.* 2001).

Embora definido pela positividade para HER-2 e, em alguns casos, também para receptores hormonais (HER-2+/ER+), esse subtipo difere da classificação clínica de câncer de mama HER-2+ por apresentar amplificação específica do cromossomo 17q22, além de expressão da assinatura gênica luminal. Por isso, é também chamado de subtipo HER-2 superexpresso (Asif, Hafiz Muhammad *et al.* 2016).

O receptor HER-2, também conhecido como erbB-2, CD340 ou receptor do fator de crescimento epidérmico humano tipo 2, é uma tirosina-quinase transmembrana que participa de importantes vias de transdução de sinal relacionadas ao crescimento e à proliferação celular (Sørlie, T *et al.* 2001).

Destaca-se como um proto-oncogene que tem papel fundamental na progressão de tumores agressivos e sua superexpressão está associada a um mau prognóstico e ao aumento da agressividade tumoral (Nolan, Emma *et al.* 2023).

Esse receptor forma dímeros com outros membros da família HER, como HER-1 e HER-3. A dimerização induz a ativação de vias oncogênicas, podendo resultar no desenvolvimento do câncer.

Além disso, a superexpressão de HER-2 tem impacto negativo quando associada ao tratamento com agentes quimioterápicos, como a doxorrubicina, aumentando a

cardiotoxicidade, o que leva a um resultado clínico desfavorável (Asif, Hafiz Muhammad et al. 2016; Mitri, Z., Constantine, T., O'Regan, R. 2012).

Estudos mostram que aproximadamente 72% dos tumores HER-2-enriquecidos apresentam mutações no gene *TP53*, fator que contribui ainda mais para a instabilidade genômica desse subtipo (Perou et. al 2000) pois leva a ativação do microRNA 21, levando à ativação da via de PI3K de forma indireta.

Em contrapartida, a identificação do HER-2 como biomarcador foi um avanço importante na terapia do câncer de mama, sendo utilizado em terapias direcionadas que visam bloquear a atividade do receptor ou impedir sua dimerização. Estratégias como a remoção do HER-2 da superfície celular ou a inibição de suas atividades enzimáticas demonstram potencial na diminuição da oncogenicidade (Perou *et. al.* 2000).

Apesar de seu envolvimento patológico, o HER-2 também desempenha funções fisiológicas no desenvolvimento e crescimento normal da mama. Isso reforça a complexidade de seu papel biológico e a importância de abordagens terapêuticas específicas e bem direcionadas.

#### **1.2.1.4. Basal-Like**

Os tumores de subtipo tipo basal, do inglês *basal-like*, são altamente proliferativos e apresentam expressão de genes relacionados às citoqueratinas basais. Expressam genes como CK5 e CK17, EGFR, laminina e outras proteínas ligadas ao fenótipo basal/mioepitelial, também chamados de marcadores basais, além da baixa expressão dos genes associados ao subtipo Luminal A (Rakha *et. al.* 2009).

Apresentam alta instabilidade genômica, com destaque para a elevada taxa de mutações no gene *TP53*, presente em aproximadamente 82% dos casos (Perou *et al.* 2000), e BRCA1 (Rakha *et. al.* 2009), gene supressor tumoral responsável pela manutenção da integridade genômica por meio da recombinação homóloga de reparo do DNA, de forma que sua perda favorece o acúmulo de quebras em fita dupla e alterações cromossômicas.

Os tumores triplo-negativos representam um subtipo clínico agressivo, aparentemente de maior frequência em mulheres jovens com ascendência africana ou hispânica (Derakhshan, F. e Reis-Filho, J. S. 2022). Seu tratamento é considerado o mais desafiador dentre os demais, uma vez que não é suscetível às terapias hormonais ou anti-HER-2 (Rakha *et al.* 2009).

Apresentam grau histológico elevado — parâmetro de avaliação do quão diferencial se encontra a arquitetura do tecido tumoral em relação ao tecido mamário normal (Carriaga M.T., Henson. D.E. 1995) — especialmente os associados a mutações nos genes *TP53* e *BRCA1* (Rakha *et al.* 2009). A ausência da expressão dos receptores limita a utilização de terapias direcionadas, o que leva a um mau prognóstico para esse grupo (Rakha *et al.* 2009).

Ainda dentro do espectro triplo-negativo, destaca-se o subtipo baixo em claudinas, do inglês *claudin-low*, que se caracteriza pela baixa expressão de proteínas claudinas, componentes fundamentais das junções célula-célula de células epiteliais (Nolan E. 2023). Esses tumores são pouco diferenciados e seu perfil de expressão gênica se assemelha às populações de células-tronco mamárias normais (Nolan E. 2023).

Seu perfil é de baixo nível de marcadores de proliferação, proteínas de adesão celular e genes de diferenciação luminal. Demonstram enriquecimento de características mesenquimais e infiltrado de leucócitos, apresentando baixa susceptibilidade à quimioterapia e representando a maior parte dos carcinomas metaplásicos e medulares (Nolan E. 2023).

O subtipo *Basal-like* se diferencia do triplo negativo, ainda que cerca de 80% dos tumores basal-like sejam triplo-negativos, ou seja, não expressam receptores de estrogênio (ER), progesterona (PR) e HER-2. Essa distinção decorre da expressão de citoqueratinas basais e genes característicos *basal-like* de células da mama (Rakha, E. A. *et al.* 2008).

Desse modo, nem todos os tumores triplo-negativos possuem o perfil molecular *basal-like*, já que somente 15–45% deles podem expressar algum desses marcadores (Rakha E. A. *et al.* 2009).

O termo *basal-like* foi atribuído a esse subtipo por sua semelhança genética com as células basais e mioepiteliais da mama (Rakha, E. A *et al.* 2009), incluindo expressão de marcadores como *EGFR*, laminina e *FABP7* (Sorlie *et al.* 2001).

### 1.3. O microambiente tumoral do câncer de mama

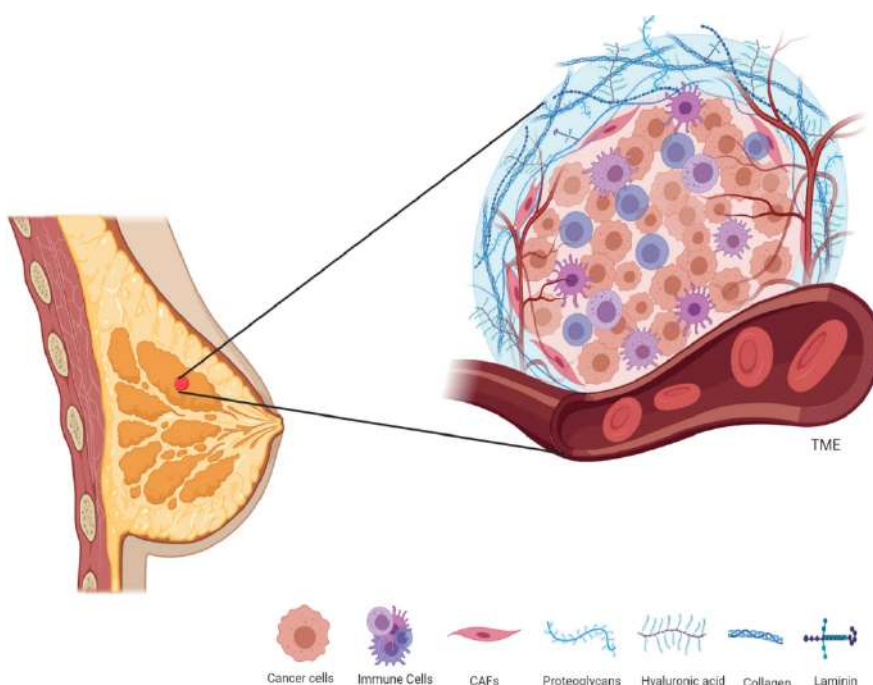
O microambiente tumoral engloba a presença de células neoplásicas em conjunto a células não tumorais, chamadas frequentemente de estroma. As células estromais podem ser fibroblastos associados ao tumor (CAFs-cancer associated fibroblasts), células do sistema imunológico infiltradas, endoteliais.

Os componentes da matriz extracelular, entre outros (**Figura 1**), e as células estromais influenciam o comportamento das células tumorais (Mittal, S., Brown, N. J., e Holen, I.

2018). Esses constituintes são essenciais para o estabelecimento do tumor, de seu nicho pré-metastático, da metástase e contribuem para a privação de nutrientes e oxigênio pela secreção de fatores e citocinas que levam a esse quadro (Mehraj, Umar *et al.* 2021).

A interação entre estroma e tumor, levando a alterações epigenéticas, perpetua um ambiente favorável para o crescimento, a invasividade e alteração morfológica das células, sendo fatores que afetam diretamente a resposta terapêutica (Mittal, S., Brown, N. J., Holen, I. 2018; Mehraj, Umar *et al.* 2021)

Assim, estudos recentes revelaram que o microambiente tumoral tem se mostrado um fator imprescindível como potencial alvo terapêutico, com tratamentos desenvolvidos contra células estromais (Soysal, S. D. *et al.* 2015).



**Figura 1. Microambiente tumoral do câncer de mama.** O microambiente tumoral é do câncer de mama é heterogêneo, sendo composto de células cancerosas, células imune infiltradas, CAFs, proteoglicanos, ácido hialurônico, colágeno e laminina. Essa constituição contribui para o crescimento e proliferação do tumor. Adaptado de Mehraj, Umar *et al.* 2021.

A comunicação bidirecional entre células epiteliais e estromais é primordial para o desenvolvimento e diferenciação da glândula mamária. O papel fisiológico do estroma abrange a manutenção da polaridade das células epiteliais, em que a perda das células mioepiteliais promove a transição do carcinoma ductal *in situ* para invasivo (Hu, M. *et al.* 2005).

Isso ocorre por esse tipo celular normal ser capaz de suprimir o crescimento, a invasão e a angiogênese de células tumorais mamárias, evitando o crescimento descontrolado e a possível formação de neoplasias (Hu, M. *et al.* 2005).

Por sua vez, a interação desse microambiente com células do sistema imunológico se demonstra importante na evasão imune das células tumorais (Burugu, S. *et al.* 2017). Ao serem recrutados para o microambiente tumoral, as células imunológicas sofrem alterações em suas funções, compondo um infiltrado no estroma tumoral (Burugu, S. *et al.* 2017).

No câncer de mama, a importância dos infiltrados imunes é percebido nas alterações genômicas dos tumores em que, no subtipo Triplo negativo, a presença de leucócitos infiltrados leva a percepção de uma menor instabilidade genômica (Miglietta, F. *et al.* 2021).

Enquanto que para subtipos positivos, em relação a presença de receptores hormonais e HER-2, a imunogenicidade do tumor é associada aos receptores e células do tumor, mediadores inflamatórios, níveis de estrogênio, terapias endócrinas e inclusive a menopausa, que pode levar a um melhor prognóstico (Miglietta, F. *et al.* 2021).

### **1.3.1. Os eosinófilos no microambiente tumoral do câncer de mama**

Os eosinófilos são leucócitos com núcleo bilobado, derivados de células mieloides precursoras CD34<sup>+</sup> na medula óssea (Rothenberg, M. E., e Hogan, S. P. 2006). Eles atuam em diferentes processos inflamatórios e patogênicos, principalmente em alergias e infecções helmínticas, e são recrutados a partir da circulação sanguínea para o local de inflamação (Rothenberg, M. E., e Hogan, S. P. 2006).

Descritos por Paul Ehrlich a partir da coloração com corantes acídicos, os eosinófilos derivam de células tronco hematopoiéticas pluripotentes, dependendo de fatores transcricionais específicos, com destaque para o GATA-1 (Klion, Amy D *et al.* 2020). Essa proteína controla a expressão de proteínas catiônicas específicas dos eosinófilos que ficam estocadas em seus grânulos citoplasmáticos, como a *major basic protein* 1 (MBP-1) e a peroxidase de eosinófilo, além da Galectina 10, e receptores como os de eotaxina (CCR3) e IL-5 (IL-5R $\alpha$ ) (Klion, Amy D *et al.* 2020).

Os eosinófilos são produzidos na medula óssea a partir de células-tronco hematopoiéticas multipotentes, em um processo de diferenciação que envolve o comprometimento progressivo da linhagem e a maturação das células comprometidas

(Blanchard, C. e Rothenberg M. E. 2009). A regulação crucial desse desenvolvimento depende de diversos fatores de crescimento e citocinas, em que a Interleucina-5 (IL-5) é a citocina mais importante, desempenhando um papel fundamental no desenvolvimento da linhagem eosinofílica e na diferenciação final, em conjunto com a IL-3 e o GM-CSF (Fator estimulador de colônias de granulócitos e macrófagos, do inglês *Granulocyte-macrophage colony-stimulating factor*) (Blanchard, C. e Rothenberg M. E. 2009). Essa diferenciação também é guiada por fatores de transcrição, como o GATA-1, que são essenciais para o desenvolvimento desde a célula-tronco até o eosinófilo maduro (Blanchard, C. e Rothenberg M. E. 2009).

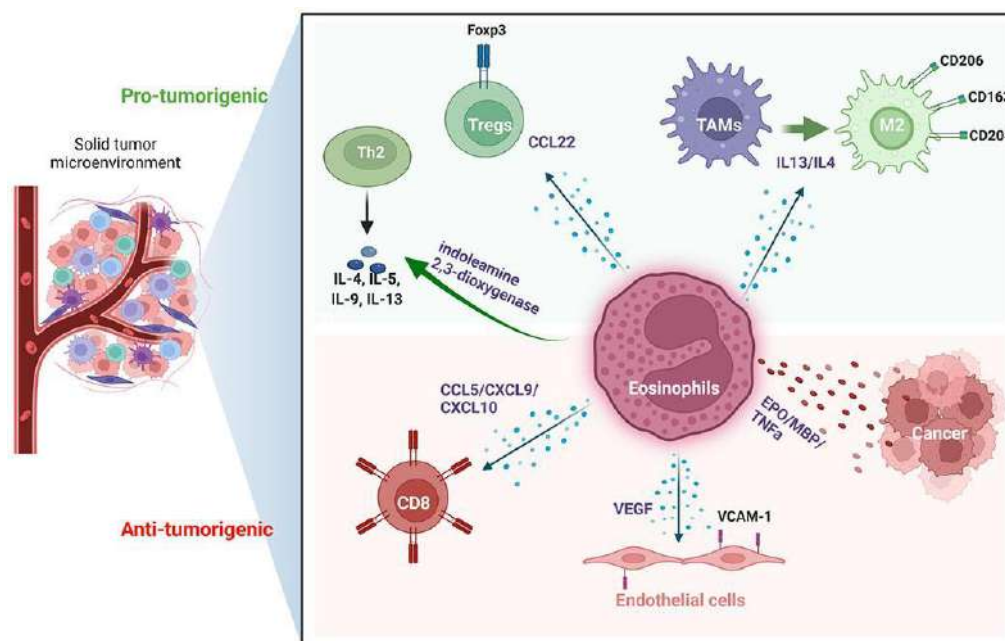
Esses granulócitos são pré-ativados e mantidos vivo no microambiente para o qual são recrutados por meio da ação da citocina IL-5, secretada pelos linfócitos T CD4<sup>+</sup> de perfil do tipo 2 e por células linfóides inatas do tipo 2 (ILC2). Logo após serem produzidos e mobilizados da medula óssea para o sangue, os eosinófilos são recrutados para os tecidos inflamados a partir da circulação por meio da ação de quimiocinas, como a CCL11 (eotaxina-1) e PGD<sub>2</sub>, e são ativados a secretar mediadores lipídicos e o conteúdo de seus grânulos (Kita H. 2011). Em microambientes submetidos a respostas imune do perfil do tipo 1, modula o ambiente, por exemplo, secretando MBP e EPO, responsáveis por gerar espécies reativas de oxigênio (ROS) e haletos tóxicos que causam dano celular (Kita H. 2011).

A ativação desses leucócitos decorre da ativação de seus receptores para citocinas, quimiocinas e mediadores lipídicos bioativos, levando à secreção de citocinas com um amplo espectro de funções, desde anti-inflamatórias a pró-inflamatórias, incluindo IL-10, TGF (fator de crescimento transformador, do inglês *transforming growth factor*), quimiocinas (RANTES e eotaxina 1) e mediadores lipídicos (fator ativador de plaquetas, do inglês *platelet activating factor* (PAF)), leucotrieno C<sub>4</sub> (LTC<sub>4</sub>) e prostaglandina D<sub>2</sub> (PGD<sub>2</sub>) (Rothenberg, M. E., e Hogan, S. P. 2006).

Apesar de tradicionalmente serem associadas às respostas imunes do tipo 2, hoje se sabe que estas células têm outras funções na manutenção da homeostasia tecidual em diferentes tecidos onde residem fisiologicamente. Por exemplo, sabe-se que essas células são residentes de tecido adiposo saudável, contribuindo para o processo de termogênese pela secreção de IL-4. A IL-4 deriva de eosinófilos via polarização de macrófagos M2 e participa do controle das funções do tecido adiposo, impedindo a instalação do quadro inflamatório característico da síndrome metabólica da obesidade (Zhang, Yi *et al.* 2015).

A eosinofilia associada ao tecido tumoral foi descrita por Lowe e colaboradores, sendo observada em diversos pacientes diagnosticados com câncer (Lowe D. *et al.* 1981). De forma geral e no tecido tumoral, o recrutamento de eosinófilos para os tecidos depende da sinalização da quimiocina eotaxina 1 (quimiocina 11 de motivo C-C, do inglês *CC motif chemokine ligand* (CCL11)), que se liga ao receptor CCR3 (Artham, Sandeep *et al.* 2023). Além disso, a IL-13 promove o aumento da expressão de eotaxina 1, enquanto a IL-5 intensifica a sensibilidade das células a esse sinal (Lowe D. *et al.* 1981).

No microambiente tumoral, a infiltração e acúmulo desses leucócitos sugere sua capacidade de eliminar células tumorais por meio da liberação de grânulos citotóxicos ou secreção de citocinas que regulam células do sistema imunológico inato e adaptativo, como CCL5 (quimiocina 5 de motivo C-C, do inglês *CC motif chemokine ligand*), CXCL9 (ligante 9 da quimiocina motivo C-X-C, do inglês *Chemokine C-X-C motif ligand 9*) e CXCL10 (ligante 10 da quimiocina motivo C-X-C, do inglês *Chemokine C-X-C motif ligand 10*), recrutando e ativando principalmente linfócitos TCD4<sup>+</sup> e TCD8<sup>+</sup>, levando à expressão de perfis anti-tumorais, sendo geralmente associado a um bom prognóstico (**Figura 2**) (Artham, Sandeep *et al.* 2023).



**Figura 2. Eosinófilos infiltrados no tumor podem influenciar células do microambiente tumoral, levando a expressão de perfis anti-tumorais ou pró-tumorais.** No microambiente tumoral, os eosinófilos podem assumir o perfil pró-tumorigênico ou anti-tumorigênico a depender das citocinas e células que estimula. Para a promoção do crescimento e proliferação do tumor, esses leucócitos ativam uma resposta do tipo 2, em que ativa células T auxiliares (T *helper*) do tipo 2, que secretam as citocinas anti-inflamatórias IL-4, IL-5, IL-9 e IL-13, células T

reguladoras pela secreção de CCL22, e macrófagos, que se diferenciam no perfil anti-inflamatório M2, o qual expressa CD206, CD163 e CD204, pela secreção de IL-13 e IL-4. Em contraste, podem estimular células T citotóxicas, TCD8<sup>+</sup>, pela expressão de CCL5, CXCL9 e CXCL10, as células tumorais, pela secreção de EPO, MBP, e as células endoteliais, por VEGF que leva à expressão de VCAM nessas células. Adaptado de Artham, Sandeep *et al.* 2023.

Estudos demonstraram que esses leucócitos são residentes do tecido mamário saudável e, com isso, estão diretamente associados ao seu desenvolvimento durante a puberdade e à maturação das estruturas na gravidez (Vickers, R., Porter, W. 2024). Sua participação se destaca no desenvolvimento dos botões terminais da mama por volta dos 9 anos de idade, se situando no tecido adjacente à cabeça dessas estruturas, que posteriormente serão diferenciadas e alongadas em ramos ductais (Sferruzzi-Perri, Amanda N *et al.* 2003).

Durante a puberdade, se observa a ocorrência de indução de eotaxinas, principalmente a CCL11, levando ao recrutamento de eosinófilos na glândula mamária, que secretam citocinas importantes na diferenciação e remodelação tecidual, como IL-4, IL-13 e TGF- $\beta$ , sugerindo que o recrutamento é um mecanismo regulado por estrogênio (Coussens, L. M, e Pollard. J. W. 2011). Após a diferenciação das estruturas, os eosinófilos não são mais encontrados nesse tecido até a gravidez (Sferruzzi-Perri, Amanda N *et al.* 2003).

Entretanto, foi notado que o super recrutamento dos eosinófilos leva à inibição da formação de botões terminais e, consecutivamente, de ramificações ductais, devido a inibição da proliferação das células epiteliais (Coussens, L. M, e Pollard. J. W. 2011). Assim, a preferência de localização e a deleção desses leucócitos após a diferenciação da estrutura sugere que eles possam estar envolvidos na regulação da ramificação e na complexidade dos ductos mamários (Coussens, L. M, e Pollard. J. W. 2011)

Em um modelo murino de carcinoma mamário, utilizando a linhagem celular metastática de câncer de mama do subtipo triplo negativo, 4T1, a administração exógena de Interleucina-33 (IL-33) no tumor implantado em camundongo demonstrou ser um fator imunomodulador fundamental (Qi, Lu, *et al.* 2020). Esta citocina promoveu o aumento do recrutamento de células Natural Killer (NK) para o microambiente pulmonar pela produção de CCL5 (RANTES), fator quimiotático, secretado por populações de eosinófilos e linfócitos TCD8<sup>+</sup> (Poncin, Aurélie *et al.* 2021).

Adicionalmente, a IL-33 conferiu um estado de ativação às células NK *in situ* que foi associado ao aumento da expressão do receptor ST2 (também conhecido como IL-1R1,

receptor da IL-33) nas membranas da célula imunológica, sendo importante para potencializar a vigilância imune antitumoral contra a progressão metastática (Qi, Lu, *et al.* 2020; Poncin, Aurélie *et al.* 2021).

No contexto do câncer de mama, o papel dos eosinófilos ainda é pouco compreendido, com escassos registros de infiltração nesse tipo de carcinoma (Poncin, Aurélie *et al.* 2021). Em outros tumores, sua presença é controversa, variando conforme o tipo tumoral e o estágio da doença (Blomberg, Olga S *et al.* 2023). Observou-se que em pacientes com uma contagem relativa de eosinófilos basal  $\geq 1,5\%$  exibiram uma taxa de resposta patológica completa (do inglês *pathological Complete Response - pCR*) significativamente maior em comparação com aquelas com REC inferior (Onesti, Concetta Elisa *et al.* 2018).

A pCR é definida como a ausência de câncer invasivo e *in situ* após o paciente ser submetido à terapia neoadjuvante, geralmente quimioterapia, sendo associado a um melhor desfecho clínico e aumento de sobrevida (Cortazar, Patricia *et al.* 2014). Esses dados sugerem que o aumento dos eosinófilos circulantes contribui para uma resposta imune antitumoral mais robusta e eficaz contra os subtipos agressivos (Onesti, Concetta Elisa *et al.* 2018).

A eosinofilia tecidual associada ao tumor (TATE do inglês *tumor-associated tissue eosinophilia*) tem sido frequentemente observada em lesões malignas, principalmente após o início do tratamento com inibidores de *checkpoint* imunológico (Blomberg, Olga S *et al.* 2023).

Os inibidores de *checkpoint* imunológico são imunoterapias utilizadas no câncer que tem como alvo receptores imunológicos na superfície de células T, visando o melhoramento da resposta anti-tumoral (Shiravand, Yavar *et al.* 2022). Estudos recentes demonstraram que o tratamento com inibidores de *checkpoint* imunológico aumenta a produção de IL-5 por células T CD4+, o que estimula a medula óssea a produzir mais eosinófilos (Blomberg, Olga S *et al.* 2023).

O tratamento com inibidores de *checkpoint* imunológico, associado à IL-33, favorece a infiltração intratumoral desses granulócitos, o que tem sido relacionado a um melhor prognóstico (Blomberg, Olga S *et al.* 2023). Essa infiltração já foi reportada em cânceres do colo do útero, colorretal, de mama, na doença de Hodgkin, além de cânceres de próstata e ovário (Poncin, Aurélie *et al.* 2021; Itan, Michal *et al.* 2024).

Em estudos de análise transcriptômica, observou-se que tumores positivos para receptor de estrogênio que apresentavam infiltração eosinofílica intratumoral estavam associados a um melhor prognóstico (Poncin, Aurélie *et al.* 2021). Por outro lado, a presença estromal de eosinófilos foi menor nos casos com alta expressão do receptor hormonal (Poncin, Aurélie *et al.* 2021).

Evidências indicam que os eosinófilos podem atuar de forma efetora ou moduladora no microambiente tumoral, exercendo citotoxicidade, promovendo angiogênese, e participando da regeneração tecidual (Reichman, H., Karo-Atar, D., e Munitz, A. 2016). Eosinófilos presentes no tecido também favorecem a polarização de macrófagos para o perfil M1, o que estimula a fagocitose e a produção do fator indutor de hipóxia 1- $\alpha$  (HIF-1 $\alpha$ ), associado à normalização da vasculatura e a uma resposta antitumoral (Hasan, Farah Falah *et al.* 2024).

Entretanto, quando encontrados tanto no estroma, como como infiltrados intratumorais, esses leucócitos parecem também desempenhar funções pró-tumorigênicas, como a secreção de MMP9, que contribui para a angiogênese, metástase e proliferação de fibroblastos, sugerindo um papel contraditório (Hasan, Farah Falah *et al.* 2024).

#### **1.4. Análise transcriptômica e bioinformática no câncer de mama**

O avanço na compreensão do câncer de mama exige o estudo aprofundado dos mecanismos moleculares relacionados à sua progressão e heterogeneidade. Nesse contexto, a transcriptômica surge como uma ferramenta fundamental, sendo definida como o estudo do transcriptoma, o conjunto completo de moléculas de RNA (incluindo mRNA, rRNA, tRNA e ncRNAs) expressas por uma célula, tecido ou organismo em um determinado momento e condição (Tzec-Interián J.A., González-Padilla D., Góngora-Castillo E.B. 2025).

A análise do transcriptoma permite quantificar os níveis de expressão gênica, revelando quais genes estão ativos e em que intensidade, o que reflete diretamente no estado funcional da célula tumoral. Isso é crucial, visto que a alteração na expressão gênica é uma característica importante na carcinogênese (Mohini *et al.* 2023).

A principal técnica utilizada para a análise de todo o transcriptoma é o RNA-*sequencing* (RNA-seq), que emprega o sequenciamento de nova geração (NGS) para gerar milhões a bilhões de leituras curtas de sequências de RNA (Mohini *et al.* 2023). Por meio de *pipelines* computacionais e ferramentas estatísticas, a Bioinformática torna possível

realizar o alinhamento das sequências, a quantificação da expressão gênica e a análise de expressão diferencial entre amostras de câncer e amostras controle, considerando os grandes volumes de sequências que são mapeados e quantificados (Mohini *et al.* 2023).

A Bioinformática, portanto, transforma os dados brutos de sequenciamento em informações biológicas interpretáveis, permitindo a identificação de assinaturas gênicas e de potenciais biomarcadores associados ao câncer de mama (Tzec-Interián J.A., González-Padilla D., Góngora-Castillo E.B. 2025).

A análise do transcriptoma pela Bioinformática possibilita a identificação de novos genes *driver* (mutações, alterações ou genes específicos que têm um papel fundamental na manutenção e progressão do câncer) e vias de sinalização que estão desreguladas, contribuindo para uma maior compreensão da biologia tumoral (Curtis C. *et al.* 2012).

Dessa forma, o potencial da análise de RNA-Seq no contexto oncológico reside em identificar biomarcadores prognósticos e preditivos, abrindo caminho para a medicina de precisão, onde os tratamentos podem ser direcionados para as alterações específicas do transcriptoma (Tsimberidou, A. M. *et al.* 2022).

## 2. JUSTIFICATIVA

O câncer de mama representa um grave problema de saúde pública entre as mulheres, no Brasil e no mundo. Essa patologia é classificada com base em diferentes tipos moleculares, que desempenham um papel crucial na determinação da agressividade e invasividade do tumor, além de determinar o prognóstico dos pacientes.

A investigação desse câncer exige o aprofundamento do entendimento sobre sua biologia e mecanismos de progressão, de forma a possibilitar o desenvolvimento de novas metodologias para testar terapias e fármacos que possam não apenas auxiliar no tratamento, mas também na identificação da oncogênese. Nesse contexto, a detecção de células imunológicas no microambiente tumoral têm se demonstrado imprescindível no prognóstico e na resposta ao tratamento.

Estudos comprovaram a relevância da participação de leucócitos no microambiente tumoral. Dentre elas destacam-se os eosinófilos, granulócitos presentes em processos imunopatológicos e inflamatórios, que têm sido associados a um papel anti-tumoral no câncer de mama. A presença de TATE (*tumor-associated tissue eosinophilia*) e eosinofilia periférica

têm sido associadas a maior sobrevivência e melhor resposta à imunoterapia, predizendo uma resposta mais favorável ao tratamento com inibidores de *checkpoint* imunológico.

Em modelos murinos de câncer de mama, foi evidenciado que a inibição do eixo estrogênio-receptor de estrogênio (ER) resulta em aumento significativo de TATE e maior eficácia dos ICIs (Artham, Sandeep et al. 2024), ressaltando o potencial terapêutico de moduladores do receptor de estrogênio em diferentes subtipos tumorais.

Uma das abordagens altamente empregadas para a identificação de novos alvos terapêuticos é a observação de alterações genéticas relevantes que podem ser encontradas em dados de amostras sequenciadas, disponibilizados em repositórios públicos. Plataformas como GEO (*Gene Expression Omnibus*), TCGA (*The Cancer Genome Atlas Program*) e GDC (*Genomic Data Commons*) são essenciais e disponibilizam informações genômicas processadas e normalizadas, prontas para serem analisadas.

A automatização desse processo, por meio do uso de linguagens de programação como Python e R, representa um diferencial metodológico, caracterizando a importância da bioinformática. Essa integração com ferramentas computacionais otimiza a análise e permite o entendimento do funcionamento de organismos *in vivo*, como o microambiente tumoral e sua interação com células do sistema imunológico, o que tem um impacto direto e positivo na compreensão do desenvolvimento do câncer.

Assim, a busca por assinaturas gênicas que modulam a presença ou a função desses granulócitos no tumor triplo negativo pode revelar alvos terapêuticos inovadores, personalizando e melhorando, potencialmente, o tratamento de pacientes com câncer de mama mais agressivo.

Os resultados obtidos a partir dessa pesquisa têm o potencial de enriquecer significativamente o conhecimento sobre a interação célula imune-tumor, contribuindo para o aprimoramento de estratégias terapêuticas atuais pela observação de como os eosinófilos podem exercer um papel anti-tumoral no tecido. Dessa forma, este estudo representa um passo importante na busca por novas abordagens no combate ao câncer de mama, visando não apenas melhorar os resultados clínicos, mas também a sobrevida das pacientes acometidas por essa patologia.

### **3. OBJETIVOS**

#### **3.1. Objetivo geral**

Analisar a expressão e interação de genes associados a presença de eosinófilos no microambiente tumoral do câncer de mama por assinaturas gênicas que modulam a presença ou função desses granulócitos no tumor.

#### **3.2. Objetivos específicos**

**3.2.1.** Analisar a influência da expressão gênica da presença de eosinófilos, por meio de amostras de RNAseq, nos subtipos moleculares de câncer de mama e no prognóstico da doença;

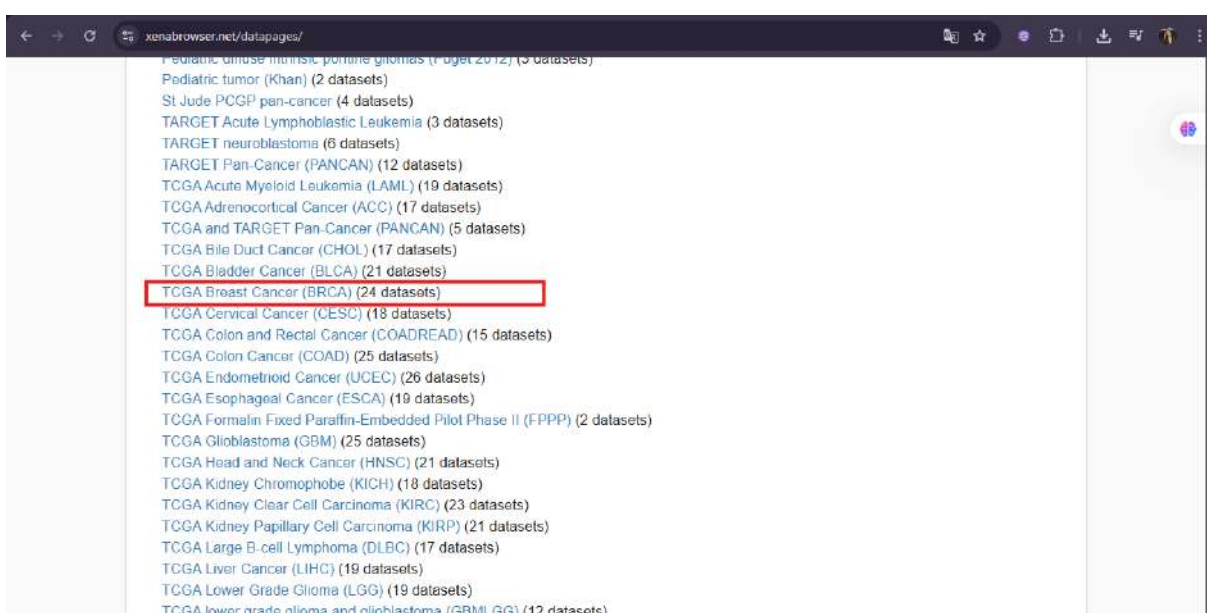
**3.2.2.** Entender a interação dos genes de recrutamento de eosinófilos com o microambiente dos diferentes subtipos moleculares de câncer de mama e sua relação com o prognóstico dos pacientes;

**3.3.3.** Entender o impacto da expressão gênica da manutenção da eosinofilia no microambiente tumoral e no prognóstico dos pacientes;

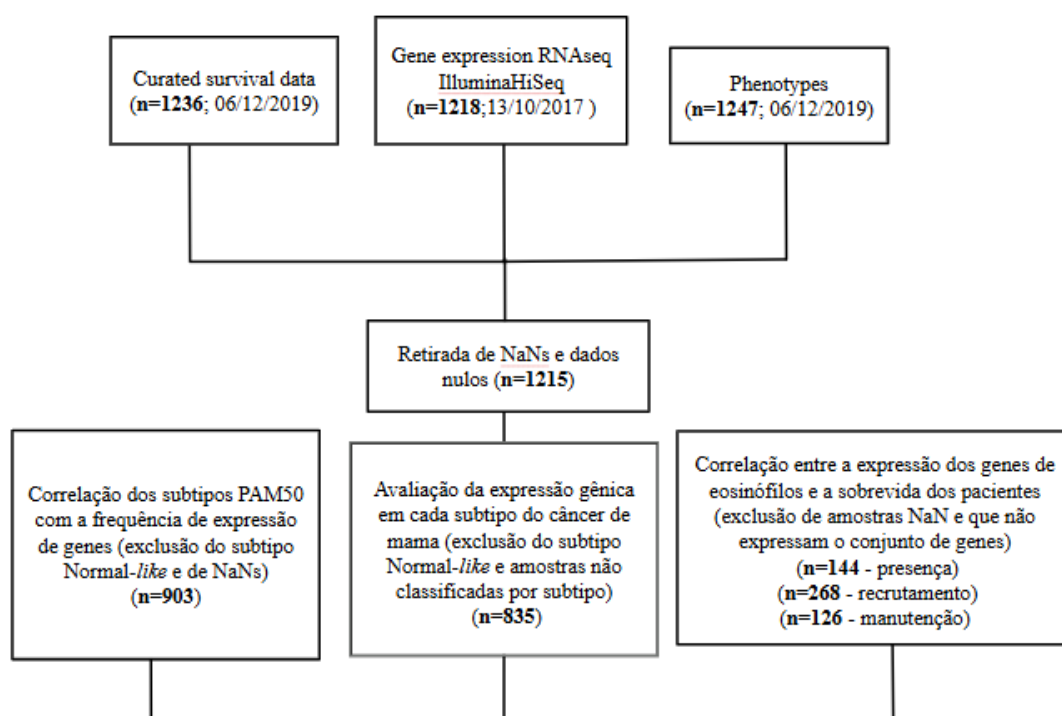
## 4. METODOLOGIA

### 4.1. Obtenção de amostras por bancos de dados públicos

Os dados foram obtidos a partir do banco de dados público “TCGA” (*The Cancer Genome Atlas Program*) (**Figura 3**), do qual, do total de 3.701 amostras, foram usadas 1215 que incluem dados de RNAseq, fenotípicos e clínicos. A diminuição do número de amostras se deu por conta da filtragem em relação a dados vazios (NaNs) ou de origem desconhecida (**Figura 4**). Os *datasets* usados foram: expressão gênica RNAseq (do inglês “*gene expression RNAseq*”) e fenótipos (do inglês “*phenotypes*”).



**Figura 3.** Site UCSC (*University of California Santa Cruz*) Xena Browser. O site armazena os dados de bancos de dados públicos direcionados a câncer, como o TCGA e GDC, dentro de Cohorts que abrangem diversificados datasets. O retângulo vermelho indica o coorte selecionado para este trabalho. Imagem obtida no dia 19/11/2025, pelo endereço eletrônico: <https://xenabrowser.net/datapages/>



**Figura 4. Fluxograma de amostras utilizadas de acordo com a filtragem para cada análise.** O fluxograma indica nos 3 primeiros quadros o n amostral dos dados obtidos de cada *dataset*. Em seguida, o quadro central mostra o n de amostras utilizado após a retirada de dados nulos e/ou vazios. Por fim, os 3 últimos quadros mostram o n utilizado em cada análise, com os grupos retirados sendo especificados entre parênteses.

As amostras representam os pacientes que são identificados pelos códigos TCGA-XX-YYYY-ZZ-A, em que TCGA se refere ao banco de dados da onde as amostras são provenientes, o código XX identifica a instituição responsável pela coleta, sendo chamado de local de origem do tecido (*Tissue Source Site* - TSS), YYYY é o código atribuído a identificação do indivíduo, ZZ indica o tipo de tecido ou preparação da amostra e o código

A se refere à porção da amostra do paciente, podendo ser A1, para diferenciar os frascos das amostras, A2 que indica a porção física do tecido e A3 se referencia ao analito, que é o tipo de material molecular extraído, identificado pelas letras D para DNA, R para RNA, T para DNA tumoral amplificado e X para quando o material não é especificado (National Cancer Institute. 2025).

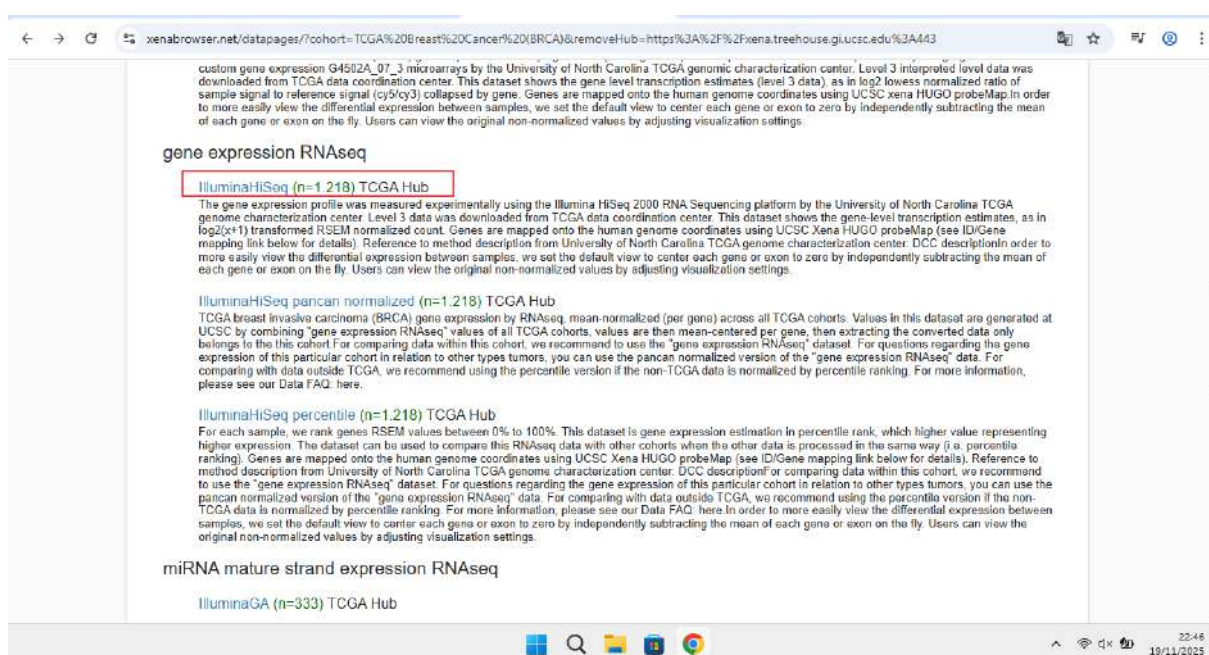
Para o código A das amostras há 19 tipos, sendo eles 01 para tumor sólido primário, que geralmente se origina de ressecção cirúrgica; 02 para tumor recorrente; 03 para câncer sanguíneo primário, 04 para câncer recorrente medular, 05 para adicional - novo primário, 06 para metástase que frequentemente provém de Biópsia, 07 para metástase adicional, 08 para

células originais de tumor humano, 09 para câncer primário medular, 10 para normal derivado de sangue, 11 para tecido sólido normal, adjacente à ressecção, 12 para célula bucal normal, 13 para normal imortalizado por EBV, 14 para medula óssea normal, 20 para analito de controle, 40 para câncer recorrente sanguíneo, 50 para linhagens celulares, 60 para tecido xenográfico primário, 61 para tecido xenográfico derivado de linhagem.

Foi selecionado o *dataset* IlluminaHiSeq por ser um método de alta precisão ao permitir sequenciar milhões a bilhões de *reads* (leituras curtas) em uma única corrida e sequenciar várias amostras na mesma *flow cell* usando *barcodes* ou índices.

O conjunto IlluminatiHiSeq pancan *normalized* é recomendado para a comparação de um coorte em que o valor de um gene é centrado na média em todos os tipos de tumor. Já o IlluminaHiSeq percentile, o valor da expressão do gene é transformado em percentil, onde se observa o quão altamente expresso está o gene em relação a outros genes de uma mesma amostra, de 0 a 100.

O *dataset* de expressão gênica (**Figura 5**) contém 1.215 amostras, atualizados em 13 de Outubro de 2017. São de nível 3 — dados processados, agregados e normalizados que passaram pela pipeline padronizada do TCGA, próprio para uso em bioinformática — processados e normalizados por RSEM significativamente, transformada em  $\log_2(x+1)$ .



**Figura 5.** *Dataset* expressão gênica RNAseq do cohort TCGA Breast Cancer (BRCA) para análise de dados. O retângulo vermelho indica o dataset escolhido para análise de expressão gênica como IlluminaHiSeq

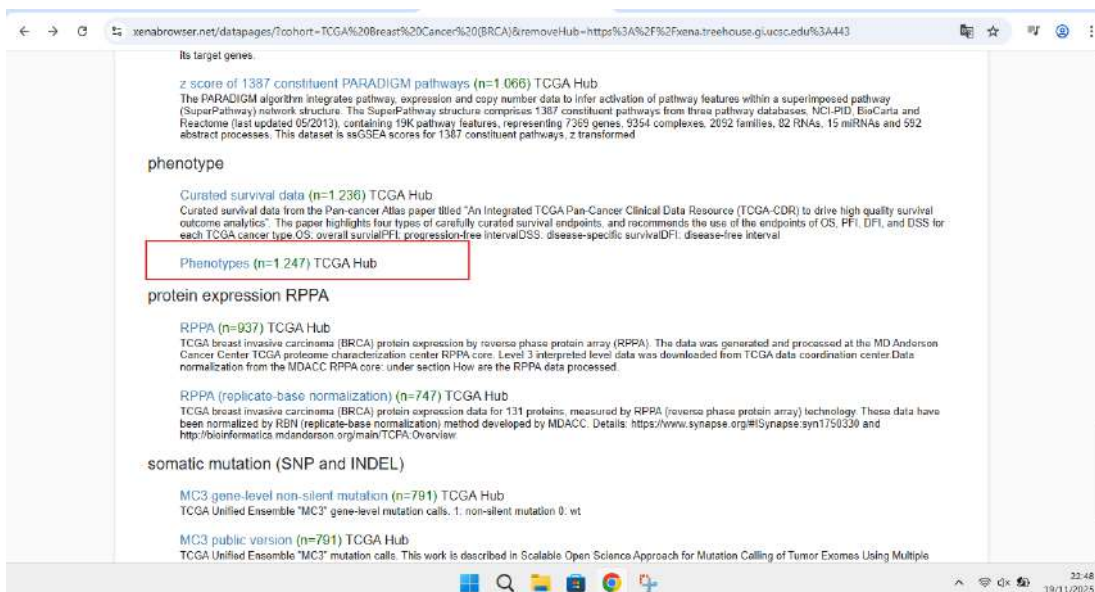
(n=1218) TCGA Hub. Imagem obtida no dia 19/11/2025, pelo endereço eletrônico: [https://xenabrowser.net/datapages/?cohort=TCGA%20Breast%20Cancer%20\(BRCA\)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=TCGA%20Breast%20Cancer%20(BRCA)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443)

A transformação é padrão para esse tipo de análise, por comprimir a escala de variação para que fique mais próxima da distribuição normal, estabilizando a variância. O RSEM é um *software* para estimar com precisão os níveis de expressão de genes e de suas isoformas, calculando a probabilidade de uma *read* se alinhar a um transcrito específico. Essa contagem gera um valor normalizado, que é o TPM, transcritos por milhão, do inglês “*transcripts per million*”.

A normalização do valor leva à correção da compensação de algumas amostras terem sido sequenciadas com mais *reads* totais do que outras, além de equiparar que genes longos terão mais *reads* mapeados do que genes curtos. Esses dados são descritos a partir do método do centro de caracterização do genoma TCGA da Universidade da Carolina do Norte (UCSC), que consiste na medida do perfil de expressão gênica de forma experimental usando a plataforma de Sequenciamento de RNA Illumina HiSeq 2000 pelo centro de caracterização do genoma TCGA da Universidade da Carolina do Norte (UCSC).

Os genes são mapeados nas coordenadas do genoma humano usando o mapa de sondas HUGO do UCSC Xena — que é um dicionário padronizado, usado pelo portal UCSC Xena, visando garantir a padronização de nomes de genes mapeados em diferentes experimentos.

Em relação ao *dataset* “*phenotype*”, usamos “*phenotype - Phenotypes*” (**Figura 6**) para análise de Subtipos PAM50 do câncer de mama e suas correlações com os genes e “*phenotype - Curated survival data*” para análise de sobrevida global (**Figura 7**).

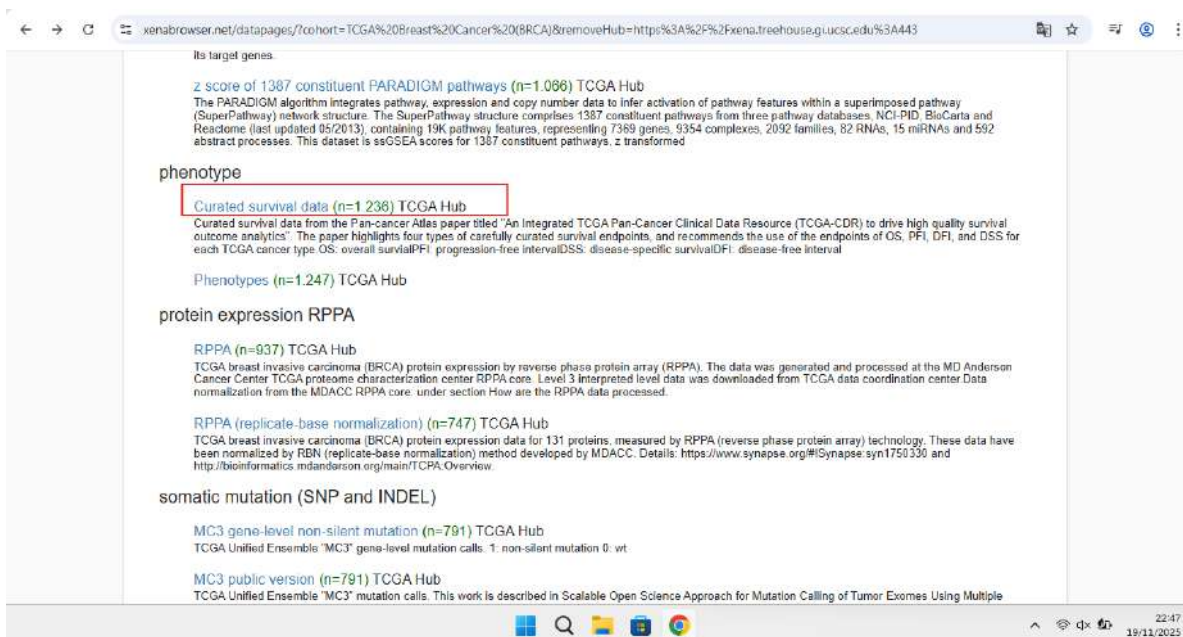


**Figura 6. Dataset de expressão fenotípica de genes.** O dataset inclui dados como idade, gênero, raça, dados de diagnóstico e patologia do tumor, informações de tratamento, status de sobrevida, tempo de sobrevida e tempo até o evento de recorrência e/ou progressão da doença. Imagem obtida no dia 19/11/2025, pelo endereço eletrônico:

[https://xenabrowser.net/datapages/?cohort=TCGA%20Breast%20Cancer%20\(BRCA\)&removeHub=https%3A%2F%2Fxfena.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=TCGA%20Breast%20Cancer%20(BRCA)&removeHub=https%3A%2F%2Fxfena.treehouse.gi.ucsc.edu%3A443)

Os subtipos PAM50 são uma forma de classificação com base na assinatura gênica de 50 genes, subdividindo o câncer de mama em cinco subtipos moleculares: Luminal A, Luminal B, Her-2, Basal-like e Normal-like. Essa expressão gênica da subtipagem pode ser identificada por meio de microarrays, RNA e/ou qRT-PCR (Kensler, Kevin H et al. 2019). Neste trabalho, realizamos as análises focadas nos 4 subtipos descritos por Perou et. al (2000), excluindo o Normal-like.

O grupo de “*phenotypes-Phenotypes*” contém 1.247 amostras relacionadas a informações demográficas, de diagnóstico e patologia, além de desfechos clínicos e sobrevida e foi atualizado pela última vez em 06 de Dezembro de 2019. Os dados clínicos de sobrevida, do dataset “*phenotype - Curated survival data*” inclui dados curados de sobrevida das amostras, incluindo o tempo e o status de sobrevida (**Figura 7**).



**Figura 7. Dataset de expressão fenotípica de genes.** Apresenta de amostras igual a 1236, indicado que provém do TCGA Hub. O retângulo vermelho indica o dataset selecionado para a análise e correlação gênica de sobrevida nesse trabalho. Imagem obtida no dia 19/11/2025, pelo endereço eletrônico: [https://xenabrowser.net/datapages/?cohort=TCGA%20Breast%20Cancer%20\(BRCA\)&removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=TCGA%20Breast%20Cancer%20(BRCA)&removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443)

Esse grupo contém 1.236 amostras, atualizado em 13 de Setembro de 2018, sendo interpretado com base nos eventos ocorridos, em que 0 representa um dado censurado, pois o paciente ainda está vivo ou parou o acompanhamento, e 1 indica o evento, que é a morte do indivíduo pela doença.

Esses dados foram obtidos publicamente a partir do recurso TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR), conforme a curadoria e as recomendações publicadas por Li et al. (2018). Destaca-se o uso dos termos técnicos: OS (Sobrevida global, do inglês “*Overall Survival*”), PFI (Intervalo livre de progressão, do inglês *Progression Free-Interval*), DSS (Sobrevivência Específica da Doença do inglês *Disease Specific-Survival*) e DFI (Intervalo livre de Doença, do inglês *Disease Free Interval*).

A sobrevida global infere quanto tempo desde o diagnóstico do paciente até o tempo da morte, seja por qualquer causa, enquanto que o PFI avalia o tempo desde o diagnóstico ou primeiro tratamento até o primeiro evento de progressão, recorrência, nova malignidade ou morte pelo câncer. O DSS avalia o tempo desde o diagnóstico clínico até a morte causada pelo

câncer, especificamente. Em contraste, o DFI é definido pelo tempo desde a conclusão do tratamento, com eliminação da detecção até a recorrência da doença.

Os dados dos 3 *datasets* são provenientes de distintas instituições localizadas nos Estados Unidos da América e identificadas pelos códigos de local de origem do tecido, do inglês *Tissue Source Site* (TSS) *code*: *Fox Chase Cancer Center* (A1), *MD Anderson Cancer Center* (A2), *University of North California* (A7), *Indivumed* (A8), *Roswell Park* (AQ), *Hospital of the University of Pennsylvania* (B6), *Beth Israel Deaconess Medical Center* (BH), *International Genomics Consortium (IGC)* (AO), *Massachusetts General Hospital* (AR), *MD Anderson Cancer Center* (Coleta Secundária - AN), *Memorial Sloan Kettering Cancer Center* (C8), *Cleveland Clinic Foundation* (D8), *Roswell Park* (Coleta Secundária - E2), *University of Pittsburgh* (E9), *University of Miami* (EW), *Cedars-Sinai Medical Center* (GI), *Washington University in St. Louis* (GM) e *Duke University* (HN).

Os dados foram obtidos de 1095 (90,12%) amostras de tumor sólido primário, retirados por ressecção cirúrgica; 7 (0,58%) amostras de biópsia de tecido metastático e 113 (9,3%) amostras de tecido sólido normal adjacentes à ressecção de códigos. A análise foi realizada com dados das amostras de tumor sólido primário devido a homogeneidade da amostra pelo maior subconjunto de dados, sendo assim mais representativa.

#### 4.2. Processamento de dados

O processamento de dados foi realizado a partir do uso de bibliotecas da linguagem de programação *Python*, versão 3.13.9, com o interpretador — IDLE, ambiente de desenvolvimento, do inglês “*integrated development environment*”) *VSCode* versão 1.106.2 (user setup) — e os scripts que deram origem às representações gráficas e tabelas estão disponíveis publicamente no repositório gratuito *Github*, disponíveis pela URL: [https://github.com/ClementeTamires/TCC\\_analise\\_dados](https://github.com/ClementeTamires/TCC_analise_dados).

#### 4.3. Análise dos dados

A análise dos dados foi realizada a partir da abordagem de 3 parâmetros de avaliação: a presença de eosinófilos, seu recrutamento e sua manutenção. Analisamos a frequência da expressão gênica por subtipo PAM50 e se a presença dos genes se correlaciona com a sobrevida global (OS) dos pacientes.

Visando avaliar a presença de eosinófilos no câncer de mama, consideramos como eosinófilos a expressão simultânea do grupo de genes *Galectina 10*, *Peroxidase de eosinófilos*,

receptor alfa de IL-5 e proteína principal de eosinófilos, representados respectivamente pelas siglas padronizadas por HUGO Gene Nomenclature Committee: *CLC*, *EPX*, *IL5RA*, *PRG2*.

Para analisar recrutamento desses leucócitos no microambiente tumoral, consideramos a expressão simultânea dos genes Eotaxina 1, Eotaxina 2 e Eotaxina 3, representados respectivamente pelas siglas padronizadas por HUGO Gene Nomenclature Committee: *CCL11*, *CCL24* e *CCL26*.

Por sua vez, para avaliar a manutenção de eosinófilos no contexto do microambiente tumoral, consideramos a expressão simultânea dos genes das citocinas Interleucina-5, Interleucina-33, Interleucina-25 e Linfopoetina estomal tímica, representados respectivamente pelas siglas padronizadas por HUGO Gene Nomenclature Committee: *IL5*, *IL33*, *IL25* e *TSLP* (thymic stromal lymphopoietin).

#### **4.3.1. Análise dos grupos de sobrevida global**

A fim de analisar a influência das assinaturas gênicas na sobrevida do paciente, optamos por avaliar a alta contra baixa expressão dos genes para presença, recrutamento e manutenção de eosinófilos. Para tal, foi realizada a mediana da expressão dos genes analisados, em que acima da mediana define o grupo de alta expressão e, abaixo ou igual a mediana, com baixa expressão.

#### **4.3.2. Análise da frequência de expressão gênica de acordo com os subtipos**

Para essa análise comparamos o grupo que expressa simultaneamente os genes constituintes da assinatura gênica de cada parâmetro com o grupo demais. O grupo “demais combinações” se refere às combinações da expressão simultânea dos genes avaliados, em que analisamos os grupos com 3 genes, 2 genes, 1 gene e nenhum gene.

#### **4.4 Análise estatística**

A análise estatística foi realizada para o tempo de sobrevida e para a relação de frequência de genes por subtipo usando Python e *Graphpad Prism* 8. Para a descrever a curva de sobrevida dos pacientes, foi usado o método de Kaplan-Meier, foi utilizado para comparar os grupos que os indivíduos expressam determinado gene contra o grupos “demais combinações” o teste Log-Rank, em que “demais combinações” se refere às combinações da expressão simultânea de 3 genes, 2 genes, 1 gene e nenhum gene.

O teste Log-Rank é não paramétrico e utilizado para testar a hipótese nula de que as funções de sobrevida não diferem ao longo do período de observação, considerando como

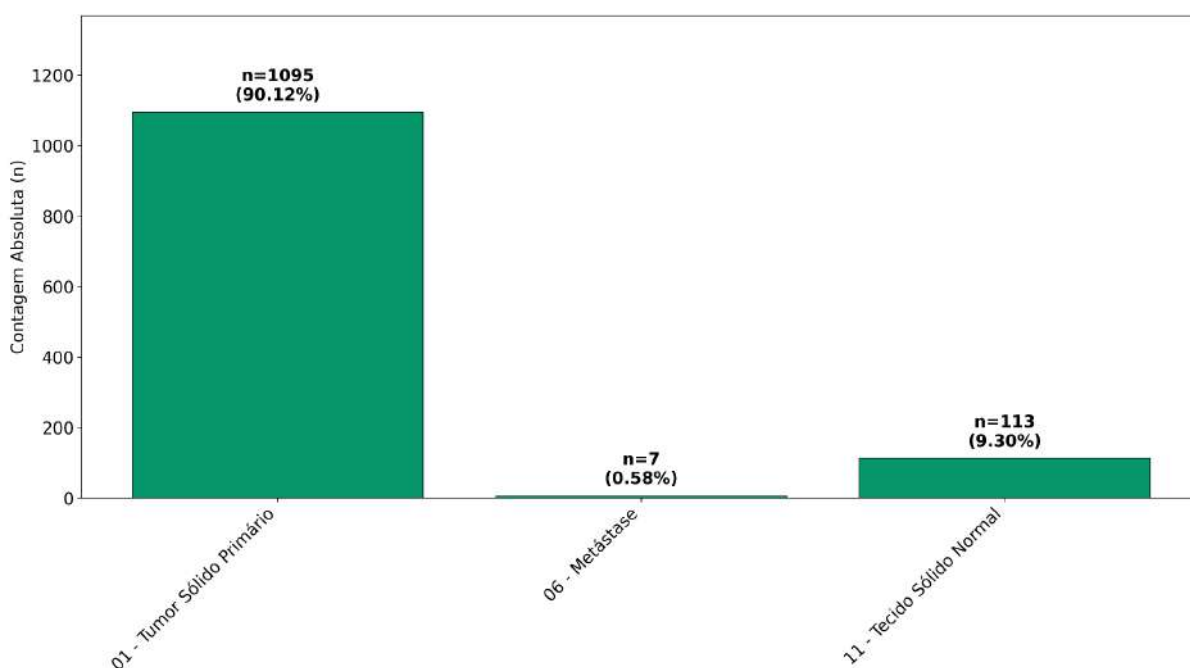
estatisticamente significativo quando o valor de  $p < 0,05$ , rejeitando a hipótese e afirmando que há uma diferença entre os grupos.

Por sua vez, o teste Qui-Quadrado foi usado para analisar associação entre variáveis qualitativas e verificar a independência entre elas, comparando as frequências observadas com as frequências esperadas sob a hipótese de independência (ou homogeneidade), utilizando uma Tabela de Contingência. Um valor de  $p < 0,05$  foi adotado para rejeição da hipótese nula de que não há associação entre as variáveis.

## 5. RESULTADOS

### 5.1. Classificação de amostras analisadas

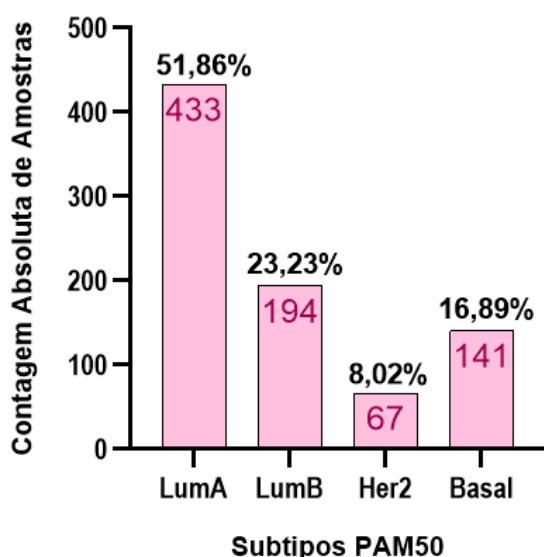
A partir do dataset TCGA Breast Cancer (BRCA)- RNAseq - IlluminaHiseq, foram classificadas 1215 amostras que contém a expressão dos genes analisados de acordo com os tipos de amostras (**Figura 8**). Nesta análise, as amostras se distribuíram com maior frequência em relação ao tumor sólido primário (de código 01), com 1095 amostras (92,5%) em seguida em tecido sólido normal (código 11) com 113 amostras (9,3%) e baixa em tecido de metástase (código 06) com 7 amostras (0,58%).



**Figura 8. Frequência dos tipos de amostras no dataset.** Gráfico em barras que exprime a distribuição dos tipos de amostras que foram obtidas a partir dos datasets. O eixo y representa a contagem absoluta das amostras,

enquanto o eixo x apresenta os tipos de amostras. Acima das barras está a contagem absoluta seguida pela porcentagem respectiva de cada tipo de amostra logo abaixo.

Em seguida, foi verificada a distribuição de amostras por subtipos PAM50 (**Figura 9**). Observa-se que dentre as 1215 amostras, de apenas 935 havia classificação molecular conhecida. Visando focar as próximas análises nos 4 subtipos mais comuns — Luminal A, Luminal B, Her-2 e Basal-like — excluindo o subtipo Normal-like, foram analisadas 835 amostras no total. Dentre estas, foi do subtipo Luminal A (n=433, 51,86%) a maior frequência, observada seguido do Luminal B (n=194, 23,23%), Basal-like (n=141, 16,89%) e o menos frequente Her-2 (n=67, 8,02%).



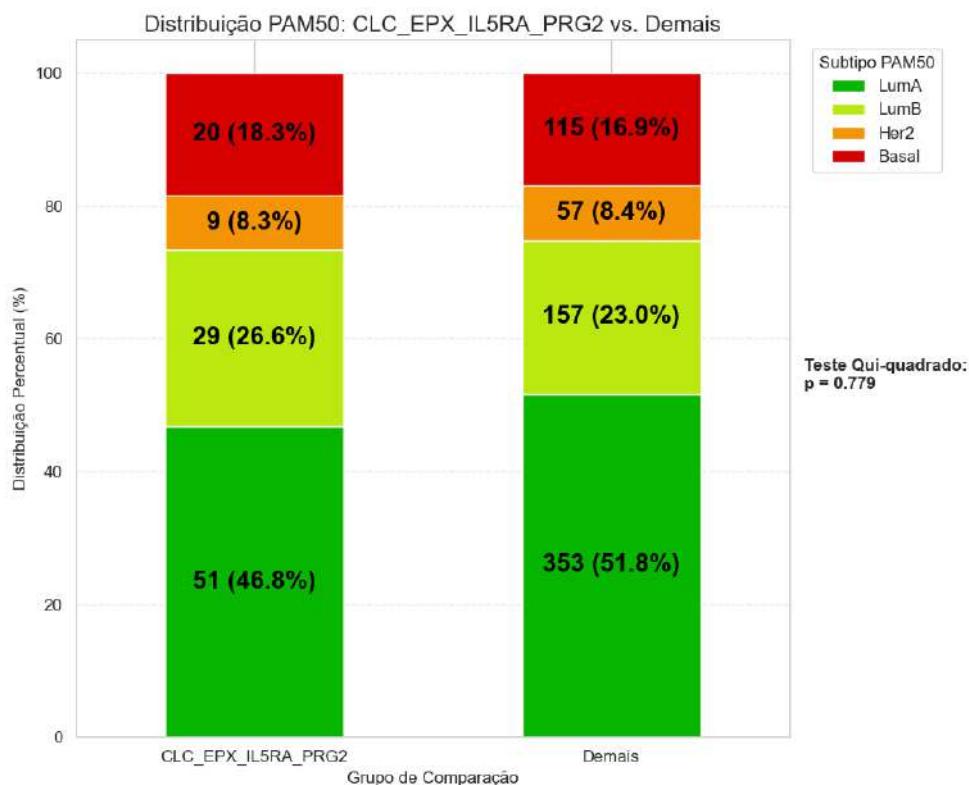
**Figura 9. Distribuição de amostras entre os subtipos PAM50 de câncer de mama.** Gráfico de barras contendo a distribuição de pacientes por subtipos PAM50, onde o eixo x aponta os subtipos moleculares de câncer de mama: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal), e o eixo y apresenta a contagem absoluta das amostras. Em cada barra há a respectiva contagem absoluta referente .

## 5.2. Avaliação da expressão gênica da presença de eosinófilos no tumor de mama

### 5.2.1. Correlação dos subtipos PAM50 com a frequência de expressão de genes relacionados à presença de eosinófilos

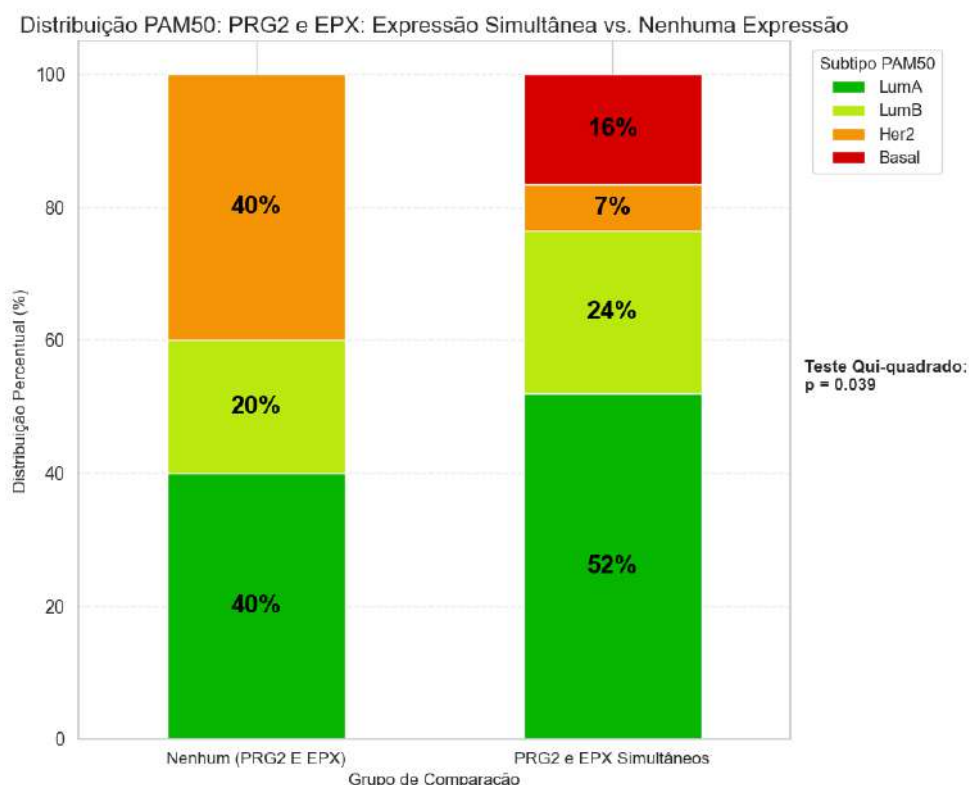
Ao comparar a expressão da assinatura gênica com a expressão dos demais grupos de combinação de expressão simultânea de genes, percebemos que não houve uma diferença significativa (Teste Qui-Quadrado com valor de  $p = 0,779$ ) da expressão entre os subtipos moleculares (**Figura 10**). Notou-se que o Luminal A (n=51, 46,8%) apresentou maior expressão do grupo que representa a presença de eosinófilos (CLC\_EPX\_IL5RA\_PRG2),

seguido de Luminal B (n=29, 26,6%), Basal-like (n=20, 18,3%) e Her-2 (n=9, 8,3%) com baixa expressão. Para essa análise, utilizamos o total de 903 amostras.



**Figura 10. Comparação da expressão simultânea dos genes CLC, EPX, IL5RA, PRG2 contra as demais combinações de genes por subtipo.** O eixo y exprime a distribuição percentual enquanto o eixo x apresenta os grupos de comparação: CLC\_EPX\_IL5RA\_PRG2 e os Demais, que se refere às combinações de genes expressos de forma simultânea: 3 genes, 2 genes, 1 gene e nenhum gene. A expressão de cada subtipo é representada por uma coloração, em que Luminal A (LumA) apresenta a cor verde, Luminal B (LumB) a cor amarela, Her-2 (Her2) a cor laranja e Basal-like (Basal) em vermelho. Ao lado direito se encontra a descrição do valor de p = 0,779 pelo Teste Qui-Quadrado.

Em seguida, avaliamos a ativação desses leucócitos por meio da expressão simultânea dos genes PRG2 (MBP) e EPX (EPO) (**Figura 11**), que são genes essenciais para definir um eosinófilo. Observou-se que são ativados em maior recorrência no subtipo Luminal A (52%), seguido por Luminal B (24%), Basal-like (16%) e Her-2 (7%), enquanto que a ausência dessas proteínas leva a ausência de eosinófilos no Basal-like e maior expressão em Her-2. Para tal análise, utilizamos o total de 903 amostras.

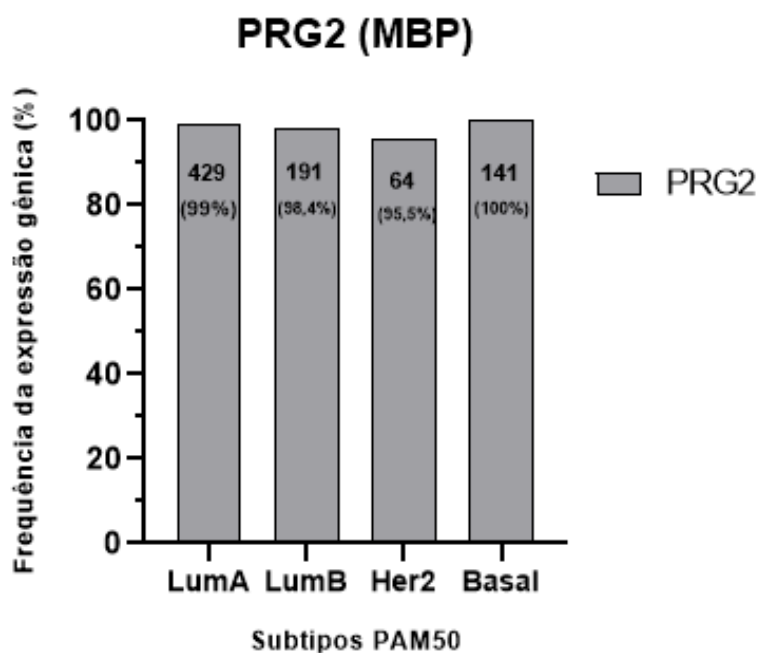


**Figura 11. Comparação da expressão simultânea dos genes EPX e PRG2 contra a expressão de nenhum dos dois.** O eixo y exprime a distribuição percentual enquanto o eixo x apresenta os grupos de comparação: EPX\_PRG2 e o Nenhum (não expressão de forma simultânea de PRG2 e EPX), que. A expressão de cada subtipo é representada por uma coloração, em que Luminal A (LumA) apresenta a cor verde, Luminal B (LumB) a cor amarela, Her-2 (Her2) a cor laranja e Basal-like (Basal) em vermelho. Ao lado direito se encontra a descrição do valor de  $p = 0,039$  pelo Teste Qui-Quadrado.

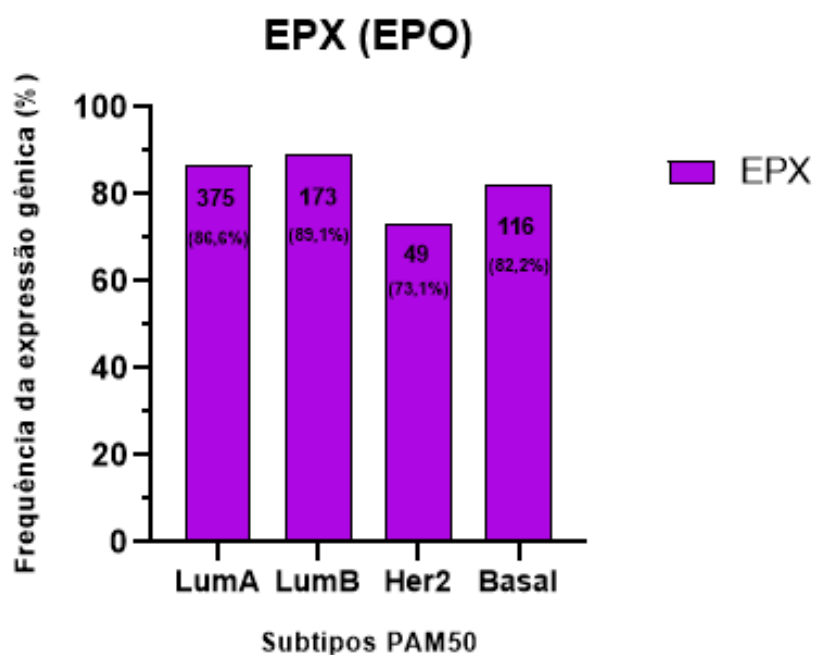
### 5.2.2. Avaliação da expressão gênica em cada subtipo do câncer de mama

Analizamos os genes da presença de eosinófilos, de forma individual, expressos nos subtipos de câncer de mama. Para isso, avaliamos os genes relacionados à ativação de eosinófilos, PRG2 (MBP) e EPX (EPO), com a contagem absoluta e a respectiva porcentagem referente aos pacientes que expressam os genes dentro do subtipo. Para essa análise, utilizamos o total de 835 amostras.

Observamos que tanto PRG2 quanto EPX estão expressos em todos os subtipos, independente da agressividade, em que PRG2 apresenta para Luminal A o  $n=429$  (99%), Luminal B com  $n=191$  (98,4%), Her-2 com  $n=64$  (95,5%) e Basal-like  $n=141$  (100%). Para EPX, notou-se que Luminal A apresenta  $n=375$  (86,6%), Luminal B com  $n=173$  (89,1%), Her-2 com  $n=49$  (73,1%) e Basal-like com  $n=116$  (82,2%) (**Figuras 12 e 13**).

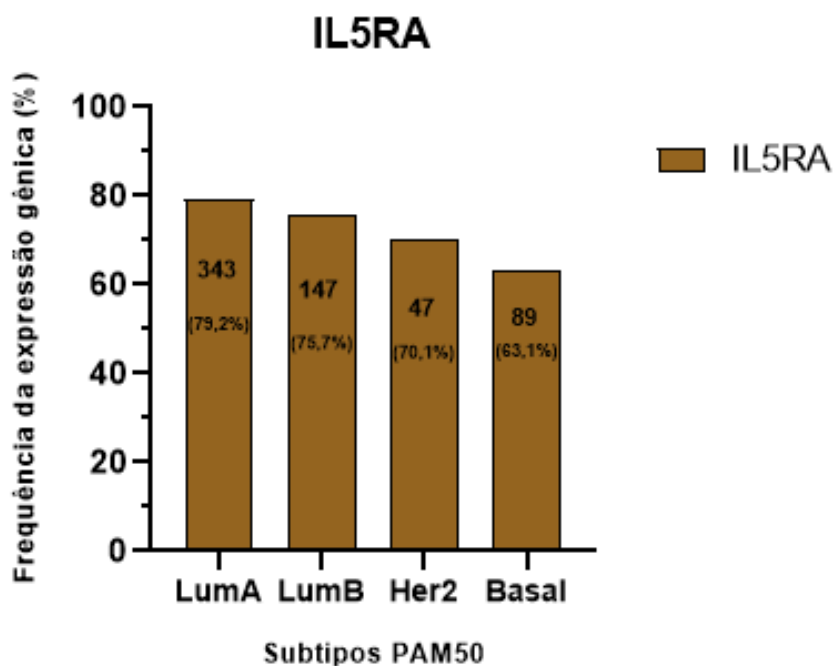


**Figura 12.** Frequência da expressão gênica de *PRG2* por subtipos PAM50. Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.

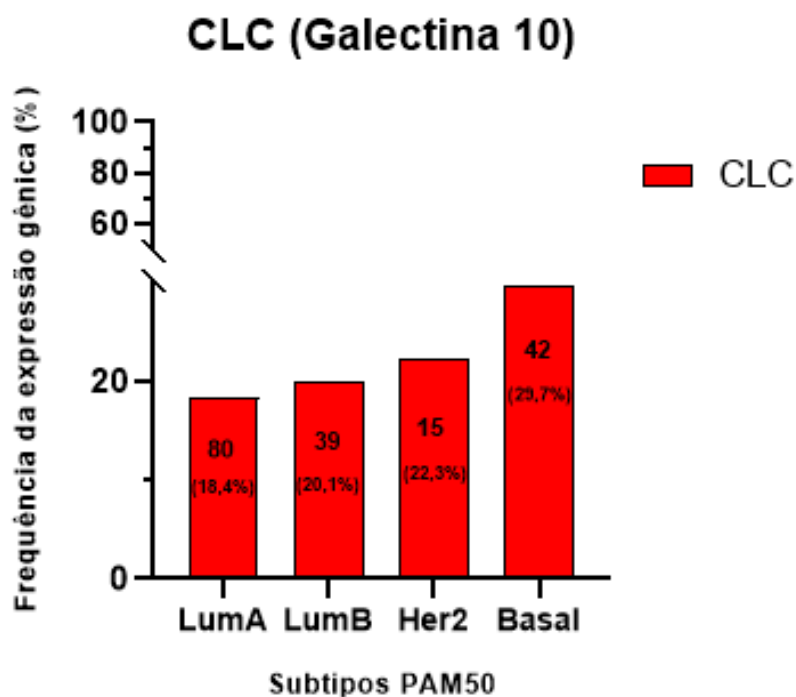


**Figura 13. Frequência da expressão gênica de *EPX* por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.

Além disso, analisamos os genes *IL5RA* (receptor alfa de IL5) e *CLC* (Galectina 10), que também indicam a presença dos granulócitos (**Figuras 14 e 15**). A expressão de *IL5RA* se distribuiu em Luminal A com n=343 (79,2%), Luminal B com n=147 (75,7%), Her-2 com n=47 (70,1%) e Basal-like com n=89 (63,1%). Para *CLC*, observou-se que Luminal A apresenta n=80 (18,4%), Luminal B com n=39 (20,1%), Her-2 com n=15 (22,3%) e Basal-like com n=42 (82,2%).



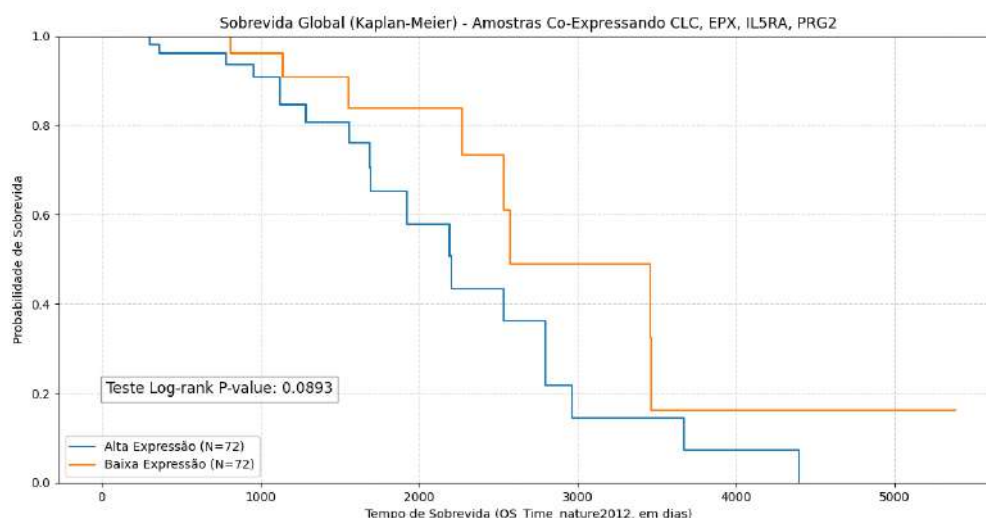
**Figura 14. Frequência da expressão gênica de *IL5RA* por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.



**Figura 15. Frequência da expressão gênica de CLC por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.

### 5.2.3. Correlação entre a expressão dos genes de eosinófilos e a sobrevida dos pacientes

Notou-se que a expressão de genes associados a presença de eosinófilos não se relaciona com a menor sobrevida dos pacientes (Log-Rank valor de  $p=0,0893$ ), onde observamos que a alta expressão leva a uma menor sobrevida (**Figura 16**). Utilizamos um total de 144 amostras que co-expressam os 4 genes para essa análise.

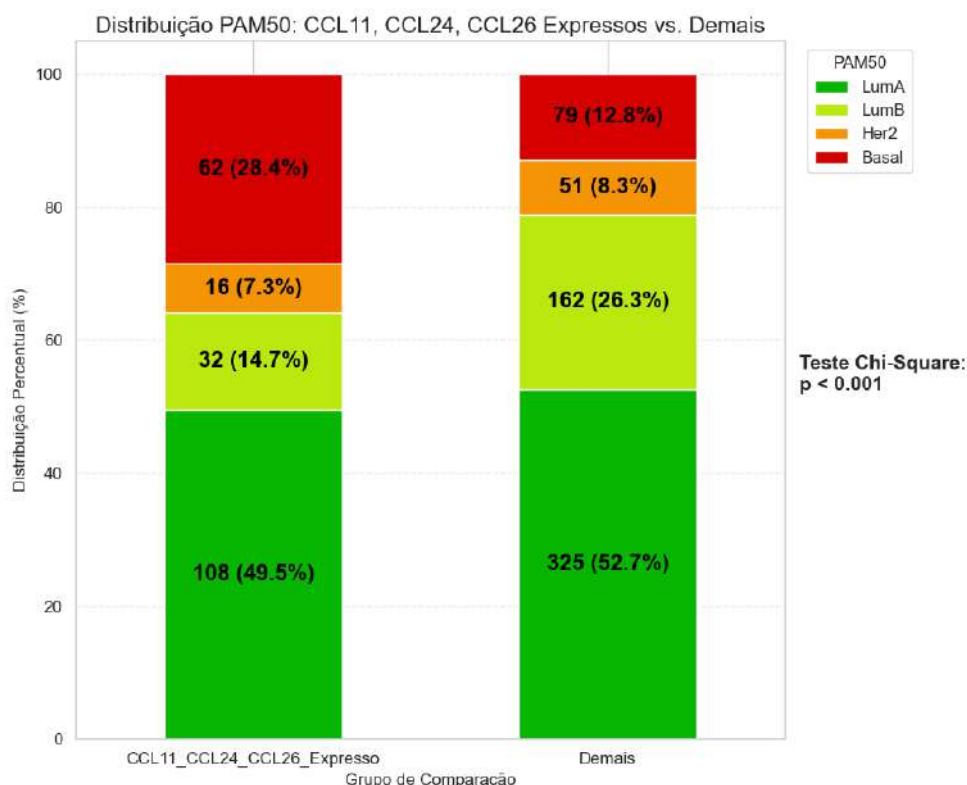


**Figura 16. Gráfico Kaplan-Meier de sobrevida de alta vs baixa expressão de genes de presença de eosinófilos.** Gráfico Kaplan-meier que representa a curva de sobrevida global dos pacientes, comparando a alta expressão contra a baixa expressão dos genes associados a presença do eosinófilo, *CLC*, *EPX*, *IL5RA*, *PRG2*. O eixo y indica a probabilidade de sobrevida global, enquanto que o eixo x expressa o tempo em dias. A curva amarela indica a baixa expressão, sinalizada por um n de amostras igual a 72, e a curva azul a alta expressão com um n de 72 amostras. A figura ainda indica o valor de  $p = 0,0893$  do teste estatístico de Log-rank.

### 5.3. Avaliação do recrutamento de eosinófilos no tumor de mama

#### 5.3.1. Correlação dos subtipos PAM50 com a frequência de expressão de quimiocinas de recrutamento de eosinófilos

Ao analisar a expressão da assinatura gênica com a expressão dos demais grupos de combinação de expressão simultânea de genes, percebemos uma diferença significativa (Teste Qui-Quadrado com valor de  $p < 0,001$ ) entre os subtipos moleculares (**Figura 17**). Observou-se que o Luminal A ( $n=108$ , 49,5%) apresentou maior expressão do grupo que representa a presença de eosinófilos (*CCL11\_CCL24\_CCL26*), seguido de Luminal B ( $n= 32$ , 14,7%), Basal-like ( $n=62$ , 28,4%), e Her-2 ( $n=16$ , 7,3%) com baixa expressão. Para essa análise, utilizamos o total de 903 amostras.



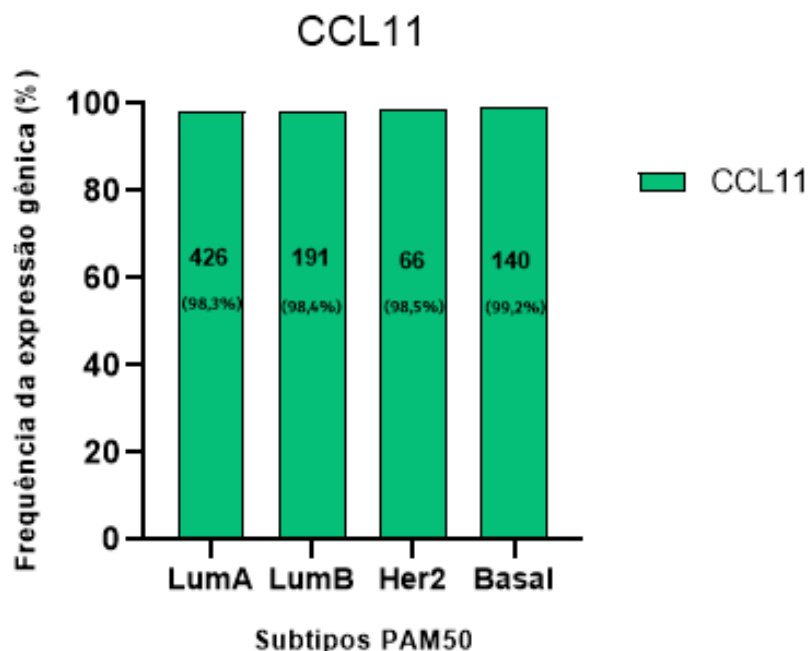
**Figura 17. Comparação da expressão dos genes *CCL11*, *CCL24*, *CCL26* contra as demais combinações de genes por subtipo.** O eixo y exprime a distribuição percentual enquanto o eixo x apresenta os grupos de comparação: *CLC\_EPX\_IL5RA\_PRG2* e os Demais, que se refere às combinações de genes expressos de forma simultânea: 3 genes, 2 genes, 1 gene e nenhum gene. A expressão de cada subtipo é representada por uma coloração, em que Luminal A (LumA) apresenta a cor verde, Luminal B (LumB) a cor amarela, Her-2 (Her2) a cor laranja e Basal-like (Basal) em vermelho. Ao lado direito se encontra a descrição do valor de  $p < 0,001$  pelo Teste Qui-Quadrado.

### 5.3.2. Avaliação da expressão de quimiocinas capazes de recrutar eosinófilos para o microambiente tumoral

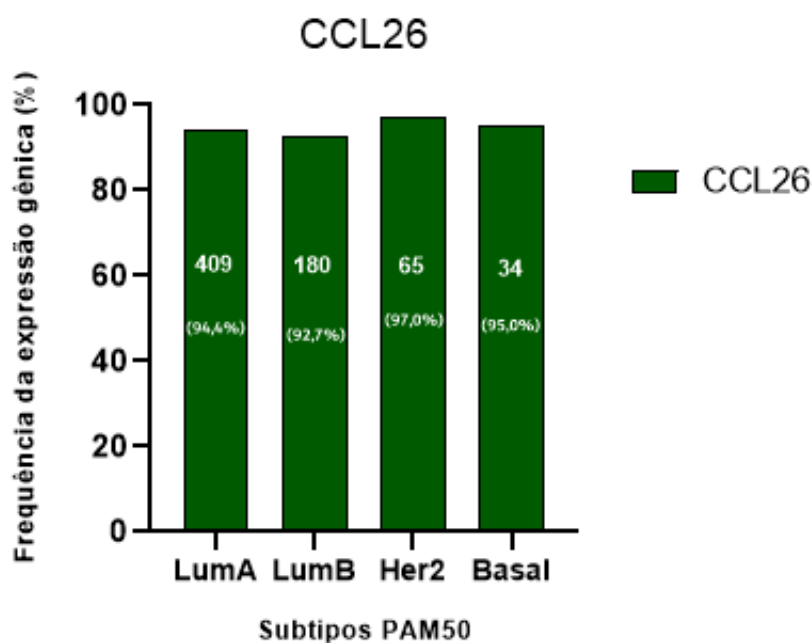
Analizamos os genes do recrutamento de eosinófilos expressos nos subtipos de câncer de mama. Para isso, avaliamos o conjunto de genes *CCL11*, *CCL24*, *CCL26*, com a contagem absoluta e a respectiva porcentagem referente aos pacientes que expressam os genes dentro do subtipo. Para essa análise, utilizamos o total de 835 amostras.

Observamos que *CCL11* e *CCL26* (**Figuras 18 e 19**) estão expressos em todos os subtipos, independente da agressividade, e *CCL24* está subexpresso nos subtipos de menor agressividade e com maior expressão no Basal-like. Para *CCL11*, Luminal A apresenta  $n=426$  (98,3%), Luminal B com  $n=191$  (98,4%), Her-2 com  $n=66$  (98,5%) e Basal-like  $n=140$

(99,2%). Em relação ao CCL26, Luminal A apresenta n=426 (98,3%), Luminal B com n=191 (98,4%), Her-2 com n=66 (98,5%) e Basal-like com n=140 (99,2%).

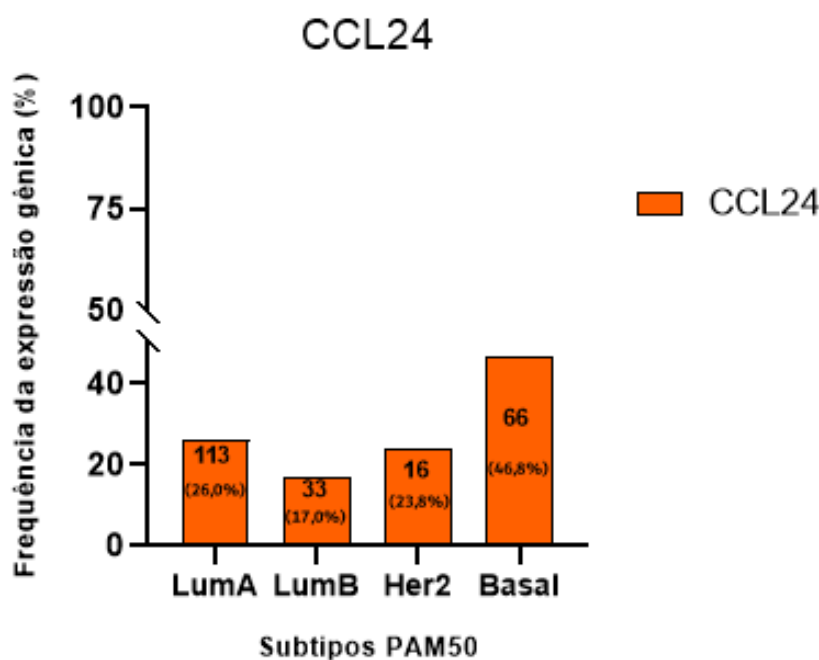


**Figura 18.** Frequência da expressão gênica de *CCL11* por subtipos PAM50. Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.



**Figura 19. Frequência da expressão gênica de *CCL26* por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.

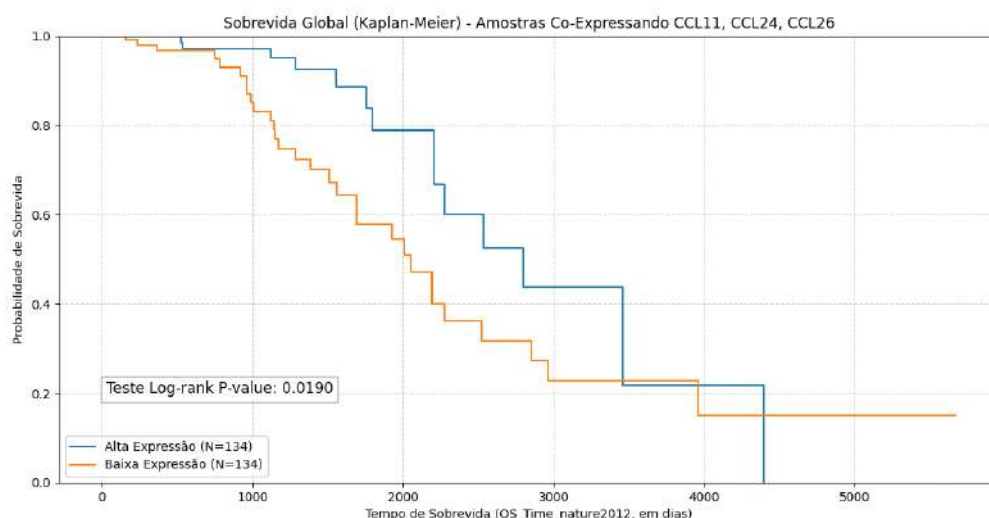
Já para a expressão de *CCL24*, observou-se em Luminal A um n=113(26,0%), para Luminal B n=33 (17,0%), que se apresenta com menor expressão, Her-2 com n=16 (23,8%) e Basal-like com n=66 (46,8%), destacando-se pela alta expressão do gene.



**Figura 20. Frequência da expressão gênica de *CCL24* por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.

### 5.3.3. Correlação da expressão de quimiocinas de recrutamento de eosinófilos com a sobrevida global dos pacientes

Observamos que a co-expressão das 3 eotaxinas, quando em altos níveis, está associada a um pior prognóstico (Log-Rank valor de  $p=0,0190$ ). Essa análise foi realizada com 268 amostras. (**Figura 21**).

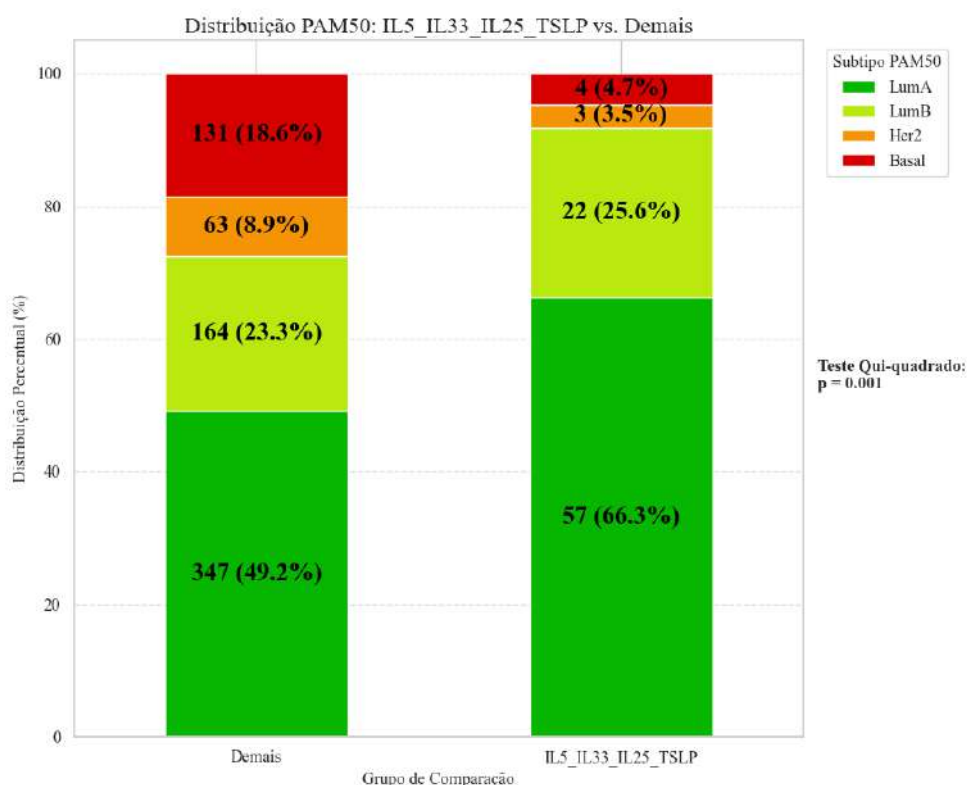


**Figura 21. Gráfico Kaplan-Meier de sobrevida de alta vs baixa expressão de genes de recrutamento de eosinófilos.** Gráfico Kaplan-meier que representa a curva de sobrevida global dos pacientes, comparando a alta expressão contra a baixa expressão dos genes associados a presença do eosinófilo, CCL11, CCL24, CCL26. O eixo y indica a probabilidade de sobrevida global, enquanto que o eixo x expressa o tempo em dias. A curva amarela indica a baixa expressão, sinalizada por um n de amostras igual a 134, e a curva azul a alta expressão com um n de 134 amostras. A figura ainda indica o valor de  $p = 0,0190$  do teste estatístico de Log-rank.

## 5.4. Avaliação da manutenção de eosinófilos no microambiente tumoral do câncer de mama

### 5.4.1. Correlação dos subtipos PAM50 com a frequência de expressão de citocinas de manutenção de eosinófilos

Nesta análise se observou uma diferença significativa (Teste Qui-Quadrado com valor de  $p = 0,001$ ) entre os subtipos moleculares. Notamos que o Luminal A ( $n=57$ , 63,3%) apresentou maior expressão do grupo avaliado (*IL5\_IL33\_IL25\_TSLP*), seguido de Luminal B ( $n= 22$ , 25,6%), Basal-like ( $n=4$ , 4,7%) e Her-2 ( $n=3$ , 3,5%) com baixa expressão. Para essa análise, utilizamos o total de 903 amostras. (**Figura 22**).



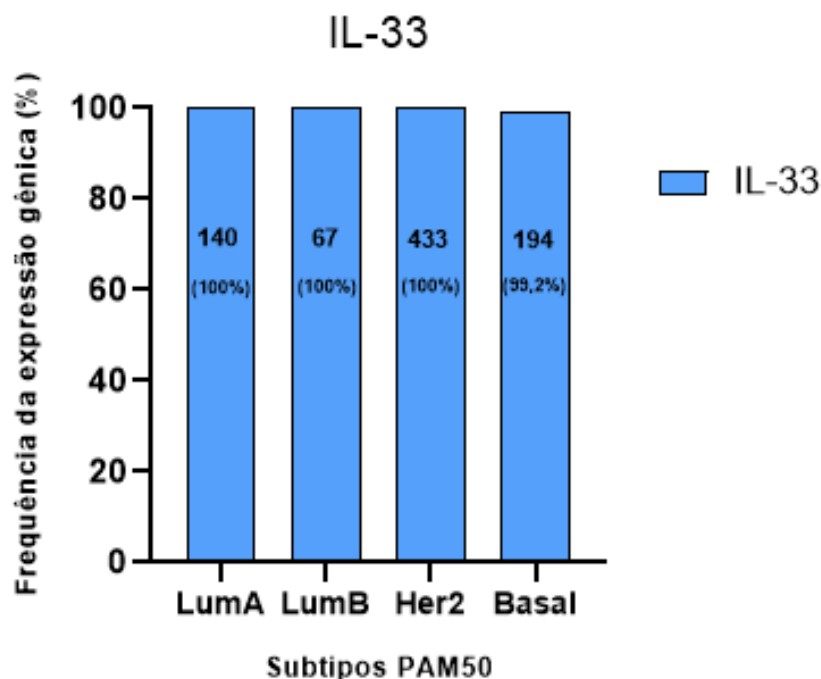
**Figura 22. Comparação da expressão dos genes *IL5*, *IL33*, *IL25*, *TSLP* contra as demais combinações de genes por subtipo.** O eixo y exprime a distribuição percentual enquanto o eixo x apresenta os grupos de comparação: *IL5\_IL33\_IL25\_TSLP* e os Demais, que se refere às combinações de genes expressos de forma simultânea: 3 genes, 2 genes, 1 gene e nenhum gene. A expressão de cada subtipo é representada por uma coloração, em que Luminal A (LumA) apresenta a cor verde, Luminal B (LumB) a cor amarela, Her-2 (Her2) a cor laranja e Basal-like (Basal) em vermelho. Ao lado direito se encontra a descrição do valor de  $p = 0,001$  pelo Teste Qui-Quadrado.

#### 5.4.2. Avaliação da expressão de citocinas capazes de realizar a manutenção de sobrevivência dos eosinófilos

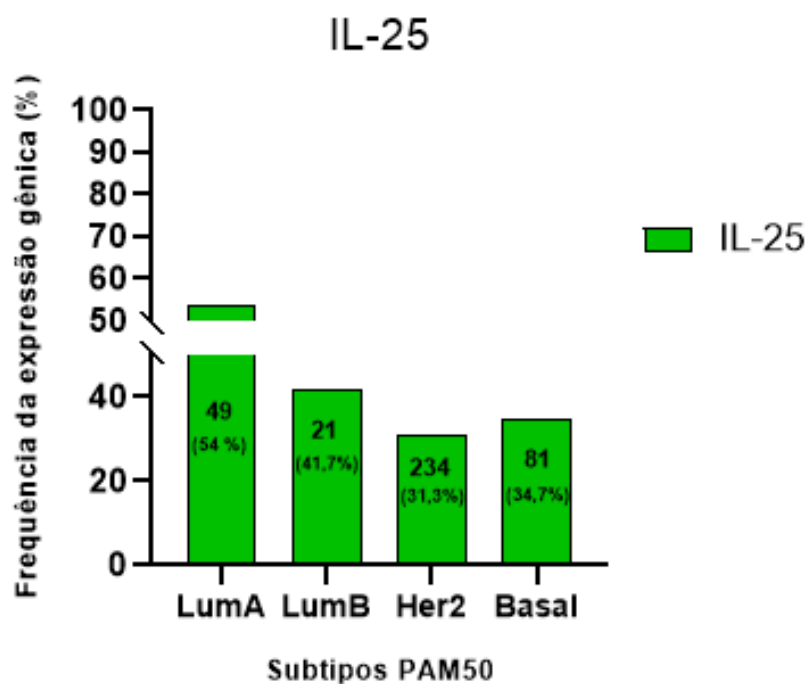
A fim de analisar a manutenção da sobrevivência dos eosinófilos correlacionada a sua expressão nos diferentes subtipos de câncer de mama, avaliamos individualmente os genes *IL33* (Figura 23), *IL25* (Figura 24), *TSLP* (Figura 25) e *IL5* (Figura 26, com a contagem absoluta e a respectiva porcentagem referente aos pacientes que expressam os genes dentro do subtipo. Para essa análise, utilizamos o total de 835 amostras.

Para *IL33* (Figura 23), Luminal A apresenta  $n=140$  (100%), Luminal B com  $n=67$  (100%), Her-2 com  $n=433$  (100%) e Basal-like  $n=194$  (99,2%). Em relação ao *IL25* (Figura 24), Luminal A apresenta  $n=49$  (54%), Luminal B com  $n=21$  (41,7%), Her-2 com

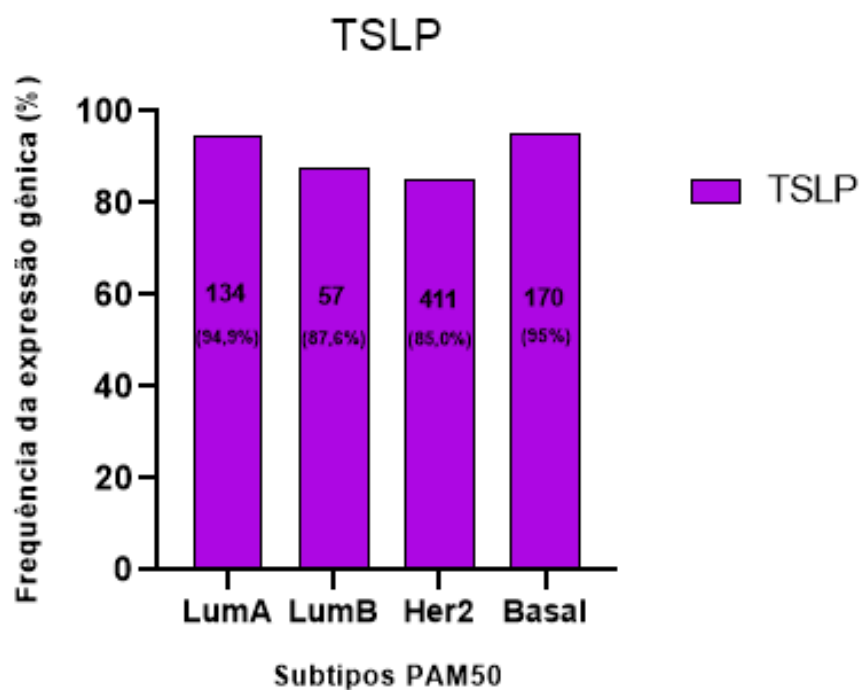
n=234 (31,3%) e Basal-like com n=81 (34,7%). Por sua vez, *TSLP* se destacou pela expressão mais alta nos subtipos Luminal A, com n=134 (94,9%), e Basal-like com n=170 (95%). Os demais subtipos obtiveram para Luminal B n=57 (87,6%) e para Her-2 n=411 (85,0%) (Figura 25).



**Figura 23. Frequência da expressão gênica de *IL33* por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.



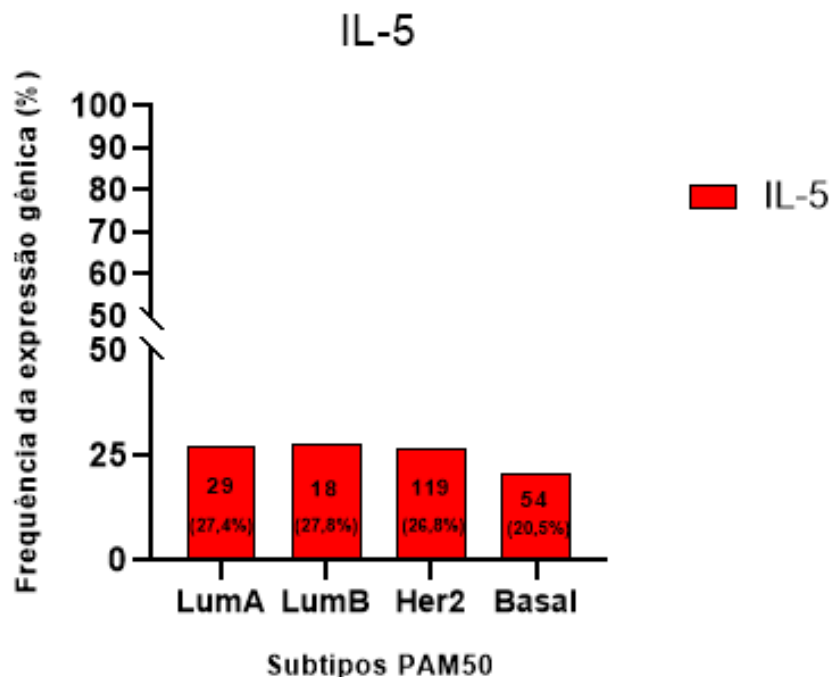
**Figura 24. Frequência da expressão gênica de *IL25* por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.



**Figura 25. Frequência da expressão gênica de *IL25* por subtipos PAM50.** Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.

Já para a expressão de *IL5*, observamos que está subexpresso nos subtipos, principalmente no de maior agressividade. Em Luminal A, notou-se um  $n=113(26,0\%)$ , para Luminal B  $n=33$

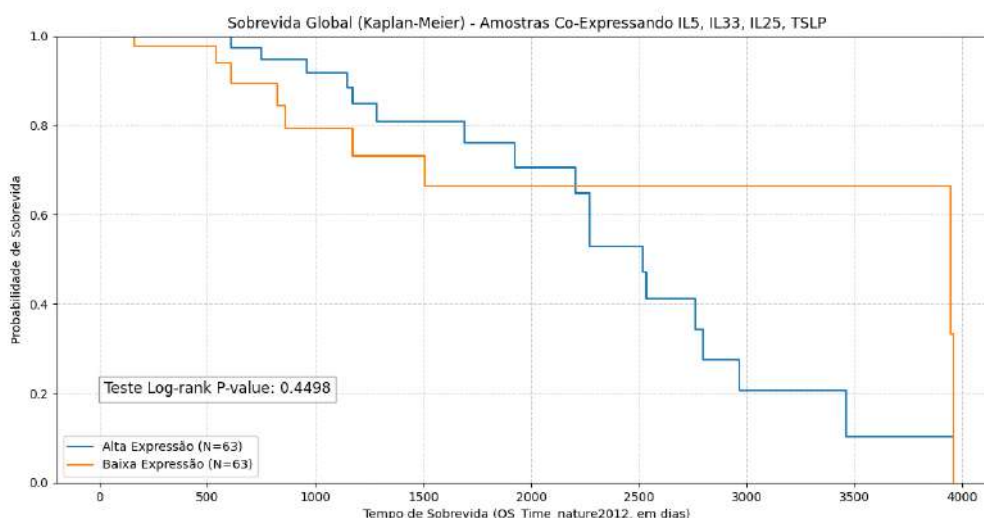
(17,0%), Her-2 com  $n=16$  (23,8%) e Basal-like com  $n=66$  (46,8%).



**Figura 26.** Frequência da expressão gênica de *IL5* por subtipos PAM50. Gráfico de barras em que o eixo y exprime a distribuição percentual da frequência de expressão gênica, enquanto o eixo x apresenta os subtipos PAM50: Luminal A (LumA), Luminal B (LumB), Her-2 (Her2) e Basal-like (Basal). Em cada coluna há a contagem absoluta dos pacientes e a sua respectiva porcentagem entre parênteses.

#### 5.4.3. Correlação da expressão de citocinas de manutenção de eosinófilos com a sobrevida global dos pacientes

Observamos que a co-expressão das 3 eotaxinas, quando em altos níveis, está associada a um pior prognóstico (Log-Rank valor de  $p=0,0190$ ). Essa análise foi realizada com 268 amostras. (**Figura 21**).



**Figura 16. Gráfico Kaplan-Meier de sobrevida do grupo de expressão de genes de manutenção .** Gráfico Kaplan-meier que representa a curva de sobrevida global dos pacientes, comparando a alta expressão contra a baixa expressão simultânea dos genes associados ao recrutamento de eosinófilos, *IL5*, *IL33*, *IL25* e *TSLP*. O eixo y indica a probabilidade de sobrevida global, enquanto que o eixo x expressa o tempo em dias. A curva vermelha indica a alta expressão, sinalizada por um n de amostras igual a 63 amostras, e a curva azul a baixa expressão com um n de 63 amostras. A figura ainda indica o valor de  $p = 0,4498$  do teste estatístico de Log-rank.

## 6. DISCUSSÃO

A análise *in silico* tem se mostrado uma ferramenta crucial na pesquisa oncológica atual, impactando a forma de estudo de novos alvos terapêuticos e de compreensão do microambiente tumoral e de suas interações. Essa metodologia possibilita a interpretação de dados vastos disponíveis publicamente, ainda que pouco usados como estratégia.

A partir da escolha de métodos de bioinformática como metodologia, nosso trabalho usou 3 *datasets* para avaliação da modulação da atividade eosinofílica no tumor. Inicialmente, obtivemos 3.701 amostras, dentre as quais, ao aplicar filtros que removem dados nulos ou não acessíveis (NaNs), restaram 1.215 amostras. Dentre estas, 1095 (90,12%) se apresentaram como originárias de tumor primário, como era o esperado, seguido pelo tecido normal com 113 amostras (9,3%) e tecido metastático com 7 (0,58%) amostras somente.

A etapa seguinte revelou que o subtipo Luminal A foi o que apresentou a maior frequência no conjunto de dados analisado, o que é consistente por ser o subtipo mais recorrente. Em seguida, a frequência seguiu-se de Luminal B, Basal-like e, por último, o HER-2.

O microambiente tumoral é um ecossistema complexo que abrange a presença de infiltrados imunológicos, como os eosinófilos, células com a capacidade secretória de grânulos e proteínas que resultam em danos teciduais, participando ativamente da resposta inflamatória, principalmente da relacionada ao tipo II e à hipersensibilidade do tipo I (Artham, S., Chang, C. Y., McDonnell, D. P. 2023). A relação desses leucócitos com o câncer de mama é contraditório pela literatura, podendo depender do contexto tumoral (Segovia-Mendoza, M., Morales-Montor, J. 2019). Com isso, visamos avaliar a influência da presença de eosinófilos no tumor de mama.

A análise de expressão gênica nos subtipos PAM50 demonstrou uma maior expressão do conjunto de genes *CLC* (Galectina 10), *EPX* (EPO), *IL5RA* (receptor alfa de IL5) e *PRGA* (MBP) nos subtipos Luminais. Entretanto, foi identificado que o subtipo Basal-*like* teve um nível de expressão considerável entre os pacientes, sendo maior que o Her-2. Em comparação ao grupo que contém a expressão dos demais grupos de genes, o dado indicou uma expressão aumentada no subtipo de maior agressividade.

Em seguida, ao analisar a frequência de expressão gênica individual por subtipos PAM50, observamos uma subexpressão da Galectina 10 (*CLC*) e baixa da *IL5RA*. Entretanto, no Basal-*like*, destacou-se a alta expressão desse gene em relação aos demais subtipos, o que pode sugerir que a Galectina 10 está relacionada a um pior prognóstico. Ainda foram observadas variações dentre os subtipos de maior agressividade, em que no subtipo HER-2, a expressão dos genes tende a ser menor em relação aos demais.

Ainda que a expressão desses genes não se associem a sobrevida, devido aos dados de sobrevida não apresentarem significância estatística, os resultados remanescentes demonstram uma correlação da presença de eosinófilos com o mau prognóstico do paciente, pela correlação dos genes com o subtipo mais agressivo, Basal-*like*.

No que concerne à migração, a liberação das quimiocinas eotaxina-1 (*CCL11*), eotaxina-2 (*CCL24*) e eotaxina-3 (*CCL26*) resulta no recrutamento de eosinófilos ao sítio inflamado. A secreção de *CCL11* é comum a células como CAFs (Huang, Wen-Yen et al., 2022), células tumorais do câncer de mama e também ocorre de forma autócrina pelos próprios eosinófilos, apresentando um estado superexpresso no tumor mamário (Thomas, J.K. et al., 2019).

De modo semelhante, *CCL24* contribui para um influxo sustentado de eosinófilos por meio de sua alta afinidade pelo receptor CCR3, intensificando a migração e promovendo a ativação dessas células após sua chegada ao microambiente tumoral (White, J.R. et al., 1997). Já *CCL-26*, cuja expressão é induzida predominantemente por estímulos associados ao eixo IL-4/IL-13, é relevante em tumores com perfil Th2, favorecendo não apenas a quimiotaxia eosinofílica, mas também a interação destas células com o estroma tumoral, fibroblastos e células imunes infiltradas (Shinkai, A. et al., 1999).

Com base nisso, observamos a presença de *CCL11* em todos os subtipos, independente da agressividade, o que pode ser justificado por ser uma quimiocina importante no desenvolvimento da glândula mamária. Esse padrão se alinha ao descrito na literatura: a *CCL11*, produzida por CAFs, células tumorais e pelos próprios eosinófilos, é frequentemente encontrada em estado de superexpressão no câncer de mama (Huang, Wen-Yen et al., 2022; Thomas, J.K. et al., 2019), favorecendo um recrutamento contínuo dessas células ao microambiente tumoral.

Além disso, notou-se baixa expressão de *CCL24* nas amostras dos subtipos de menor agressividade, enquanto que no *Basal-like* apresentou uma maior expressão comparado aos demais. Essa dado é um resultado importante pois, tendo em vista que é o subtipo de maior agressividade, essa elevação da Eotaxina-2 sugere que essa quimiocina pode ter a capacidade de exercer um papel pró-tumoral e de recrutamento eosinofílico específico nesse microambiente agressivo, podendo influenciar a progressão da doença.

Ao correlacionar a expressão desses genes com a sobrevida dos pacientes, observamos que os dados conversam entre si. A sobrevida global dos pacientes tende a diminuir com a alta expressão desses genes, inferindo que o recrutamento eosinofílico ao microambiente tumoral pode levar a um pior prognóstico.

No contexto da manutenção da sobrevida dos eosinófilos no microambiente tumoral, diversas citocinas desempenham papéis essenciais para garantir sua sobrevivência e funcionalidade. Além da IL-5, a IL-33, membro da família IL-1, atua como alarmina liberada por células estromais e epiteliais sob dano tecidual, potencializando respostas do tipo 2, ativando ILC2 e promovendo a expansão de eosinófilos, além de modular o perfil inflamatório do estroma tumoral (Liew, F.Y., Girard, J.P. & Turnquist, H.R., 2016). A IL-25 amplifica a resposta Th2 e estimula ILC2, resultando em produção secundária de IL-5 e IL-13

e, consequentemente, em maior recrutamento e manutenção dos eosinófilos em tumores associados a respostas alérgicas ou inflamação crônica (Fort, M.M. et al., 2001). Já a TSLP (do inglês “*Thymic Stromal Lymphopoietin*”), produzida por células epiteliais, fibroblastos e células tumorais, atua promovendo a ativação de células dendríticas e o condicionamento de respostas Th2, criando um ambiente favorável à sobrevivência eosinofílica e à persistência de mediadores pró-inflamatórios no tumor (Ziegler, S.F., 2010).

A partir da análise da correlação desses genes com o microambiente tumoral do câncer de mama. Por um lado, observou-se que a expressão desses genes estava elevada no subtipo Luminal A, associado ao melhor prognóstico. Em adição, a curva de sobrevida demonstrou que a alta expressão dessa mesma assinatura gênica está correlacionada com uma maior sobrevida do paciente.

Esses dados revelam que, ainda que a presença e o recrutamento de eosinófilos demonstre uma relação com um microambiente mais inflamatório, a sua manutenção nesse ecossistema leva a liberação de citocinas que podem estar associadas a um melhor prognóstico da doença.

## 7. CONCLUSÃO

O papel dos eosinófilos no câncer de mama permanece amplamente contraditório na literatura, uma vez que sua atuação depende do tipo de resposta imune predominante e das características do microambiente tumoral. Diante dessa complexidade, foram empregadas análises *in silico* para investigar como o microambiente pode modular a atividade desses leucócitos, explorando padrões de expressão associados a assinaturas moleculares de câncer de mama.

Dessa forma, nossos resultados indicam que a expressão simultânea dos genes de presença de eosinófilos está associada a uma menor sobrevida global, ainda que em maior recorrência no Luminal A. Isso pode ser justificado também pela presença do aumento de Galectina 10 apenas no subtipo de maior agressividade, que em conjunto aos dados de frequência de genes, se mostrou como um potencial de piora de prognóstico.

Foi observado que o padrão de expressão das quimiocinas associadas aos eosinófilos sugere um possível vínculo com desfechos desfavoráveis, especialmente diante dos níveis de baixa expressão de CCL24. Embora o conjunto dos dados revele que o conjunto de quimiocinas é mais superexpresso no subtipo Luminal A, a elevada expressão de CCL24 no

Basal-like, somada a menor sobrevida diante da alta expressão dos genes, inclina-se à piora do quadro clínico.

Em relação à manutenção eosinofílica, observamos que a expressão das citocinas apresentou um padrão marcado pela maior expressão de IL33, seguida de TSLP, IL25 enquanto IL5 permaneceu subexpressa. Quando esses dados são combinados às análises de sobrevida, verificamos que a coexpressão dos quatro genes não se associa a sobrevida global dos pacientes. Ainda assim, esses genes mostraram-se mais frequentemente expressos nos subtipos Luminal A e Luminal B, sugerindo que o impacto funcional dessas citocinas depende do contexto biológico e imunológico de cada subtipo tumoral.

Os dados obtidos reforçam a complexidade do papel eosinofílico. Porém, apontam para um predomínio de fatores de agressividade que, quando avaliados em conjunto com os marcadores fenotípicos, a partir da correlação de fatores do microambiente celular, sugerem um papel desfavorável desse leucócito.

Assim, este estudo permitiu esclarecer como diferentes componentes da assinatura eosinofílica e das vias associadas modulam o microambiente tumoral no câncer de mama. A análise simultânea dos genes sugeriu um papel pró-tumoral, avaliada pela presença, pelo recrutamento e sobrevida desses leucócitos, onde a atuação dos eosinófilos pode ser associado a relações contraditórias que dependem do contexto tumoral em que se encontram.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ADES, Felipe et al. Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. *Journal of Clinical Oncology*, v. 32, n. 25, p. 2794–2803, 2014. DOI: 10.1200/JCO.2013.54.1870.

ASIF, Hafiz Muhammad et al. Triple negative breast cancer: current status and future directions. *Asian Pacific Journal of Cancer Prevention*, v. 17, n. 4, p. 1883–1890, 2016.

BADVE, Sunil et al. Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Modern Pathology*, v. 24, p. 157–167, 2011. DOI: 10.1038/modpathol.2010.200.

BLANCHARD, C.; ROTHENBERG, M. E. Biology of the eosinophil. *Advances in Immunology*, v. 101, p. 81–121, 2009. DOI: 10.1016/S0065-2776(08)01003-1.

BLOMBERG, Olga S. et al. Spatial immune profiling of the tumor microenvironment in breast cancer. *Cancer Cell*, v. 41, n. 3, p. 485–501, 2023. DOI: 10.1016/j.ccell.2023.01.006.

BURUGU, S.; ASLEH-ABURAYA, K.; NIELSEN, T. O. Immune infiltrates in the breast cancer microenvironment: detection, characterization and clinical implication. *Breast Cancer*, v. 24, p. 3–15, 2017. DOI: 10.1007/s12282-016-0698-x.

CARRIAGA, M. T.; HENSON, D. E. The prognostic significance of tumor infiltrating lymphocytes in solid human tumors. *Cancer*, v. 76, n. 10, p. 1843–1851, 1995. DOI: 10.1002/1097-0142(19951115)76:10<1843::AID-CNCR2820761021>3.0.CO;2-9.

CORTAZAR, Patricia et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *The Lancet*, v. 384, n. 9938, p. 164–172, 2014. DOI: 10.1016/S0140-6736(13)62422-8.

COUSSENS, Lisa M.; POLLARD, Jeffrey W. Leukocytes in mammary development and cancer. *Cold Spring Harbor Perspectives in Biology*, v. 3, n. 3, a003285, 2011. DOI: 10.1101/cshperspect.a003285.

DERAKHSHAN, F.; REIS-FILHO, J. S. Pathogenesis of basal-like breast cancer. *Journal of Pathology*, v. 258, n. 1, p. 9–20, 2022. DOI: 10.1002/path.5961.

EROLES, P. et al. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treatment Reviews*, v. 38, n. 6, p. 698–707, 2012. DOI: 10.1016/j.ctrv.2011.11.005.

FORT, M. M. et al. IL-25 induces IL-4, IL-5, and IL-13 and Th2-associated pathologies in vivo. *Immunity*, v. 15, n. 6, p. 985–995, 2001. DOI: 10.1016/S1074-7613(01)00243-6.

HASAN, Farah Falah et al. Eosinophils in cancer: double-edged sword of the tumor microenvironment. *Frontiers in Immunology*, v. 15, 2024. DOI: 10.3389/fimmu.2024.1342210.

HU, M. et al. Regulation of in situ to invasive breast carcinoma transition. *Cancer Cell*, v. 8, n. 2, p. 113–127, 2005. DOI: 10.1016/j.ccr.2005.07.007.

HUANG, Wen-Yen et al. Tumor-infiltrating immune cells and breast cancer prognosis. *Cancers*, v. 14, n. 3, 2022. DOI: 10.3390/cancers14030664.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2023: Incidência de câncer no Brasil. Rio de Janeiro: INCA, 2023.

ITAN, Michal et al. Eosinophils and cancer: emerging roles in tumor immunity. *Nature Reviews Cancer*, v. 24, p. 1–17, 2024. DOI: 10.1038/s41568-024-00676-5.

KITA, H. Eosinophils: multifunctional and distinctive properties. *Immunological Reviews*, v. 242, n. 1, p. 161–177, 2011. DOI: 10.1111/j.1600-065X.2011.01026.x.

KLION, Amy D. et al. Eosinophils and eosinophil-associated disorders. *Journal of Allergy and Clinical Immunology*, v. 145, n. 6, p. 1511–1523, 2020. DOI: 10.1016/j.jaci.2020.03.036.

KUMAR, V.; ABBAS, A. K.; ASTER, J. C. Robbins and Cotran pathologic basis of disease. 9. ed. Philadelphia: Elsevier, 2017.

LI, Lian Tao et al. Molecular classification of breast cancer. *Oncology Letters*, v. 8, n. 6, p. 2560–2566, 2014. DOI: 10.3892/ol.2014.2580.

- LI, Zhi-Hua et al. Basal-like breast cancer: molecular features and clinical implications. *Oncology Letters*, v. 11, n. 5, p. 3129–3134, 2016. DOI: 10.3892/ol.2016.4382.
- LIEW, F. Y.; GIRARD, J. P.; TURNQUIST, H. R. Interleukin-33 in health and disease. *Nature Reviews Immunology*, v. 16, p. 676–689, 2016. DOI: 10.1038/nri.2016.95.
- LOWE, D. et al. Eosinophilia in malignant tumors. *British Journal of Cancer*, v. 43, n. 3, p. 361–366, 1981. DOI: 10.1038/bjc.1981.57.
- MEHRAJ, Umar et al. Tumor microenvironment and immune evasion in breast cancer. *Cancer Cell International*, v. 21, 2021. DOI: 10.1186/s12935-021-01895-7.
- MIGLIETTA, F. et al. Immune microenvironment in breast cancer subtypes. *Frontiers in Oncology*, v. 11, 2021. DOI: 10.3389/fonc.2021.678093.
- MITRI, Z.; CONSTANTINE, T.; O'REGAN, R. The HER2 receptor in breast cancer: pathophysiology, clinical use, and new advances in therapy. *Chemotherapy Research and Practice*, 2012. DOI: 10.1155/2012/743193.
- NOLAN, Emma et al. The immune landscape of breast cancer. *Nature Reviews Cancer*, v. 23, p. 1–18, 2023. DOI: 10.1038/s41568-023-00579-9.
- ONESTI, Concetta Elisa et al. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *European Journal of Surgical Oncology*, v. 44, n. 4, p. 429–436, 2018. DOI: 10.1016/j.ejso.2018.01.025.
- PONCIN, Aurélie et al. Eosinophils contribute to breast cancer progression by promoting tumor cell proliferation and invasion. *Cancer Research*, v. 81, n. 4, p. 990–1002, 2021. DOI: 10.1158/0008-5472.CAN-20-2315.
- QI, Lu et al. Tumor-associated eosinophils and prognosis of solid tumors. *OncoImmunology*, v. 9, n. 1, 2020. DOI: 10.1080/2162402X.2020.1727119.
- RAKHA, Emad A. et al. Breast cancer prognostic classification in the molecular era. *Pathology*, v. 41, n. 7, p. 615–629, 2009. DOI: 10.3109/00313020903257385.
- RAKHA, Emad A. et al. Prognostic markers in triple-negative breast cancer. *Cancer*, v. 113, n. 10, p. 2645–2653, 2008. DOI: 10.1002/cncr.23861.

REICHMAN, H.; KARO-ATAR, D.; MUNITZ, A. Emerging roles for eosinophils in the tumor microenvironment. *Trends in Cancer*, v. 2, n. 11, p. 664–675, 2016. DOI: 10.1016/j.trecan.2016.10.002.

ROTHENBERG, M. E.; HOGAN, S. P. The eosinophil. *Annual Review of Immunology*, v. 24, p. 147–174, 2006. DOI: 10.1146/annurev.immunol.24.021605.090720.

SFERRUZZI-PERRI, Amanda N. et al. Interleukin-5 transgene expression and eosinophilia. *Journal of Immunology*, v. 170, n. 1, p. 169–176, 2003. DOI: 10.4049/jimmunol.170.1.169.

SØRLIE, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses. *Proceedings of the National Academy of Sciences*, v. 98, n. 19, p. 10869–10874, 2001. DOI: 10.1073/pnas.191367098.

SOYSAL, S. D. et al. Prognostic significance of tumor-associated eosinophils in breast cancer. *Breast Cancer Research and Treatment*, v. 152, p. 487–495, 2015. DOI: 10.1007/s10549-015-3494-6.

VICKERS, R.; PORTER, W. Immune cell contribution to mammary gland development. *Journal of Mammary Gland Biology and Neoplasia*, v. 29, n. 1, p. 1–15, 2024. DOI: 10.1007/s10911-024-09568-y.

WHO. Breast cancer. Geneva: World Health Organization, 2025. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. Acesso em: 28 nov. 2025.

WILKINSON, Louise; GATHANI, Toral. Understanding breast cancer as a global health concern. *British Medical Bulletin*, v. 141, n. 1, p. 5–17, 2022. DOI: 10.1093/bmb/ldac002.

YERSAL, O.; BARUTCA, S. Biological subtypes of breast cancer: prognostic and therapeutic implications. *World Journal of Clinical Oncology*, v. 5, n. 3, p. 412–424, 2014. DOI: 10.5306/wjco.v5.i3.412.

ZHANG, Xinmin. Molecular subtypes of breast cancer and their clinical relevance. *Cancer Biology & Medicine*, v. 20, n. 1, p. 1–12, 2023. DOI: 10.20892/j.issn.2095-3941.2022.0453.

ZHANG, Yi et al. Eosinophils in cancer immunology. *Journal of Hematology & Oncology*, v. 8, 2015. DOI: 10.1186/s13045-015-0204-1.

ZIEGLER, S. F. Interleukin-33 in disease. *Journal of Immunology*, v. 185, n. 9, p. 5079–5085, 2010. DOI: 10.4049/jimmunol.1000911.