

Universidade Federal do Rio Grande do Sul

Faculdade de Medicina

Programa de Pós-Graduação em Medicina: Ciências Cirúrgicas

Expressão da proteína *Biglicano* no câncer de mama

Ana Paula Thiesen Leindecker

Porto Alegre, 2022

Universidade Federal do Rio Grande do Sul

Faculdade de Medicina

Programa de Pós-Graduação em Medicina: Ciências Cirúrgicas

Expressão da proteína *Biglicano* no câncer de mama

Ana Paula Thiesen Leindecker

Orientador: Prof. Dr. Ricardo Francalacci Savaris

Dissertação apresentada como requisito parcial para
obtenção do título de Mestre no Programa de Pós-
Graduação em Medicina: Ciências Cirúrgicas,
Faculdade de Medicina, Universidade Federal do Rio
Grande do Sul.

Porto Alegre, 2022

Ficha catalográfica

CIP - Catalogação na Publicação

Thiesen Leindecker, Ana Paula
Expressão da proteína Biglicano no câncer de mama /
Ana Paula Thiesen Leindecker. -- 2022.
76 f.
Orientador: Ricardo Francalacci Savaris.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Faculdade de Medicina, Programa de
Pós-Graduação em Medicina: Ciências Cirúrgicas, Porto
Alegre, BR-RS, 2022.

1. câncer de mama. 2. biglicano. 3.
imuno-histoquímica. 4. aprendizado profundo. I.
Francalacci Savaris, Ricardo, orient. II. Título.

Agradecimentos

Ao meu esposo e filhas, agradeço o incentivo e ajuda com as tarefas de mãe que terceirizei por inúmeras vezes, agradeço pelos ensinamentos que me passaram ao incentivo nas horas em que precisei. Obrigada.

Ao meu orientador, Prof.Doutor Ricardo Francalacci Savaris pelo ensino e paciência que teve ao longo deste caminho. Obrigada por responder inúmeras vezes todas as perguntas que fiz, pela dedicação e ajuda na construção da dissertação, do artigo, da submissão. Sempre te fizeste presente desde o delineamento até o resgate de cada amostra.

Ao meu pai, Aureo Jorge Thiesen, pelo exemplo de vida e resiliência. Meu eterno obrigado.

Às pacientes, sem as quais nada poderia ser feito.

SUMÁRIO

LISTA DE ABREVIATURAS.....	6
LISTA DE TABELAS.....	11
LISTA DE FIGURAS.....	12
RESUMO E PALAVRAS-CHAVE.....	13
ABSTRACT E KEYWORDS.....	15
1. INTRODUÇÃO	17
2. REVISÃO DA LITERATURA.....	19
2.1. INCIDÊNCIA.....	19
2.2 FATORES DE RISCO	19
2.3. CLASSIFICAÇÃO HISTOLÓGICA.....	20
2.3.1. <i>Carcinoma in Situ</i>	20
2.3.2. <i>Carcinoma Ductal Invasor</i>	21
2.4. CLASSIFICAÇÃO MOLECULAR.....	21
2.4.1.1 Luminal A	23
2.4.1.2 Luminal B.....	23
2.4.1.3 Triplo Negativo/Basaloides.....	24
2.4.1.4 Receptor do Fator de Crescimento Epitelial Humano 2 (HER2)	25
3. PROTEOGLICANOS.....	26
3.1 <i>Proteoglicanos - Forma e função</i>	28
3.1.1 Proteoglicanos Pequenos Ricos em Leucina (<i>Small Leucine Rich Proteoglycans</i> - SLRPs).....	29
3.2 <i>Biglicano (BGN)</i>	31
3.2.1 Localização do gene	31
3.2.2 Estrutura da biglicano glicosilada.....	32
3.2.3 Localização da Proteína Biglicano	33
3.2.4 Moduladores da proteína biglican.....	35
3.2.5 Proteína Biglicano e doenças.....	36
3.2.6 Biglicano e câncer	37
3. REDES NEURAI CONVOLUCIONAIS (CNN).....	42

3. 1. MÉTRICAS PARA VERIFICAR O DESEMPENHO DAS REDES NEURAS CONVOLUCIONAIS.....	47
4. JUSTIFICATIVA.....	49
5. HIPÓTESES	51
6. OBJETIVOS.....	52
6.1. OBJETIVO PRINCIPAL	52
6.2. OBJETIVO SECUNDÁRIO	52
7. REFERÊNCIAS	53
8. ARTIGO EM INGLÊS	59
9. CONCLUSÃO.....	79

Lista de abreviaturas

AGRN	Agrin
amp	amplificação
APOBEC	polipeptídeo catalítico 3
ASPN	Asporina
AUC	Área da Característica operacional do receptor
BGN	biglicano
Bgn KO	knockout para biglicano no estroma
BMP-4	bone morphogenetic protein 4
BS1	Basaloide 1
BS2	Basaloide 2
CDI/SOE	carcinoma ductal invasivo/ sem outras especificações
CDK4	cyclin dependent kinase inhibitor 2A
CDKN2C	cyclin dependent kinase inhibitor 2C
CET	Células tumorais endoteliais
CHAD	Condroadarina
CNN	Convolutional Neural Networks
COL1A1	Collagen alpha-1(I) chain
COL1A2	Collagen alpha-2(I) chain

COL2A1	Collagen alpha-1(II) chain
CS	sulfato de condroitina
D-HScore	pontuação histológica digital
DAB	Diaminobenzidina
DAG1	Dystroglycan 1
DCN	Decorina
DS	sulfato de dermatan
ECM2	Proteína 2 da matriz extracelular
EGF	Fator de crescimento epidérmico
ELN	Elastin
EPYC	Epifican
ER +	receptor de estrógeno positivo
ERK	extracellular regulated MAP kinase
F-Score	Coeficiente de similaridade de dados
FAK	Focal adhesion kinase
FGFR4	receptor 4 do fator de crescimento de fibroblastos
FMOD	Fibromodulina
FN	Falso Negativo
FP	Falso Positivo

GAG	glicosaminoglicano
HER2	receptor do Fator de Crescimento Epitelial Humano 2
HER2-	receptor de crescimento epidérmico 2 negativo
HER2+	receptor de crescimento epidérmico positivo
HIF1A	hypoxia inducible factor 1 subunit alpha
IM	Imunomodulatório
INPP4B	inositol polyphosphate-4-phosphatase type II B
kDa	kilo Dalton
KERA	Keratocana
KS	sulfato de queratan
LUM	Lumican
M	Mesenquimal
MDM2	MDM2 proto-oncogene
MEC	Matriz Extracelular
MFAP2	Microfibrillar-associated protein 2
MLP	Multilayer Perceptron
mRNA	mensageiro do ácido ribonucleico
MSL	Mesenquimal, células tronco-símile
mut	mutação

MYC	MYC proto-oncogene bHLH transcription factor
NF- κ B	fator nuclear kappa-light-chain-enhancer de células B ativadas
NYX	Nictalopina
OGTC	Osteoglicina
OMD	Osteomodulina
OPTC	Opticina
PCOLCE	Procollagen C-endopeptidase enhancer 1
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PODN	Podocan
PODNL1	Proteína 1 tipo Podocan
PRELP	Proteína de repetição rica em leucina e extremidade rica em prolina e arginina
PRPF40B	Pre-mRNA-processing factor 40 homolog B
PTEN	phosphatase and tensin homolog
RB1	RB transcriptional corepressor 1
RBM25	RNA-binding protein 25
ROC	Característica operacional do receptor
ROI	Região de Interesse
SGCA	Alpha-sarcoglycan
SLRPs	Proteoglicanos Pequenos Ricos em Leucina (Small Leucine Rich Proteoglycans - SLRPs)

SNP	polimorfismo de nucleotídeo único.
T	Total
TFP	Taxa de Falso Positivo
TGF- β	fator transformador de crescimento- β ;
TGFB1	Transforming growth factor beta-1.
TLR2	Toll-like receptor 2
TLR4	Toll-like receptor 4
TNBC	Câncer de mama triplo negativo
TNF	Tumor necrosis factor
TP53	tumor protein 53
TSKU	Tsukushi
TVP	Taxa de verdadeiro Positivo
VEGF	vascular endothelial growth factor
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
VPP	Valor Preditivo Positivo

Lista de Tabelas

Tabela 1. Fatores que predis põe o c �ncer de mama	20
Tabela 2. Caracter�sticas gen �micas, cl�nicas e prote �micas dos subtipos	22
Tabela 3. Diversidade entre os principais proteoglicanos secretados - pequenos proteoglicanos ricos em leucina (SLRPs).	29
Tabela 4. Classifica �o dos Proteoglicanos Pequenos Ricos em Leucina (SLRP) com o respectivo s�mbolo do gene, de acordo com o comit � de nomenclatura de genes HUGO (HGNC). 30	
Tabela 5. Express �o do biglicano e sua relev �ncia cl�nica observada em diferentes estudos com c �ncer.	39
Tabela 6. Matriz de confus �o.....	47
Tabela 7. M �tricas de avalia �o do desempenho para o processamento de imagem.	47

Lista de Figuras

Figura 1. Os proteoglicanos	27
Figura 2. Estrutura básica da decorina, biglicano e lumicana.	32
Figura 3. Localização da proteína biglicano na célula	33
Figura 4. Estimativa da expressão(\log_{10}) da proteína biglicano nos diferentes tecidos humanos. ...	34
Figura 5. Rede interativa da proteína biglicano (BGN).....	36
Figura 6. Influência do biglicano no microambiente tumoral	38
Figura 7. Modelo de neurônio não linear.....	43
Figura 8. Modelo de rede neural Multilayer Perceptron.....	44
Figura 9. Arquitetura de uma rede neural convolucional, apresentando as camadas.	45
Figura 10. Representação gráfica da fórmula da função linear retificada.	46
Figura 11. Expressão da proteína biglicano por imuno-histoquímica usando diferentes anticorpos no câncer de mama.	50

Resumo e palavras-chave

Introdução: O câncer de mama é o principal câncer em mulheres. A busca por novos biomarcadores para marcadores prognósticos tem sido buscada por diversos autores. O biglicano (BGN) é um membro da pequena família de proteínas proteoglicanas ricas em leucina e é um importante componente da matriz extracelular. Altos níveis de expressão de mRNA de BGN foram encontrados em vários tipos de câncer sólido, em comparação com tecidos normais. No entanto, há poucos dados sobre a expressão da proteína BGN no câncer de mama. O objetivo deste estudo é comparar a expressão da proteína BGN no tecido mamário normal e no câncer de mama por meio de técnicas de imuno-histoquímica. A análise de imagem de sua expressão foi realizada usando pontuação histológica digital (D-HScore) e redes neurais de aprendizagem profunda supervisionadas.

Material e métodos: Neste estudo de caso-controle, 24 tecidos fixados em formalina e embebidos em parafina foram obtidos de arquivos patológicos para análise. Seções de tecido de mama normal (n=9) e câncer de mama (n=15) foram analisadas por imuno-histoquímica usando anticorpo monoclonal BGN (M01 - Abnova), clone 4E1-1G7 na diluição 1:300 em pH 6 e 3,3'-Diaminobenzidina (DAB) como cromogênio. As fotomicrografias das lâminas foram analisadas com o software ImageJ com "color deconvolution". Depois de selecionar as regiões de interesse (ROI), os painéis deconvoluídos apenas com DAB foram quantificados usando unidades DAB arbitrárias. Outro conjunto, com maior ampliação sem seleção de ROI, foi submetido ao modelo de reconhecimento de incorporação de imagem de rede neural profunda inceptionV3. Em seguida, a análise de rede neural supervisionada, usando validação cruzada estratificada de 20 vezes, com 200 camadas ocultas, ativação ReLu e regularização em $\alpha = 0,0001$ foram aplicadas para redes neurais de aprendizado supervisionado. O tamanho da amostra foi calculado para um mínimo de 7 casos e 7 controles, tendo um poder = 90%, um erro $\alpha = 5\%$ e um desvio padrão de 20, para identificar uma diminuição da média de 40 unidades DAB (controle) a 4 unidades DAB em câncer.

Resultados: A expressão de BGN (média \pm DP) foi de $6,1 \pm 3,9$ em tecido de câncer de mama, enquanto em tecido de mama normal foi de $39,6 \pm 21,9$, usando D-HScore ($p=0,0017$, teste t de Student, correção de Welch). O SDLNN foi capaz de classificar corretamente 110 das 129 fotomicrografias do conjunto de dados usando apenas painéis DAB, com uma precisão de classificação de 85,3% (95% CI = 78,1% a 90,3%) e a área sob a curva = 94,3%

Conclusões: D-HScore e redes neurais de aprendizado supervisionado foram capazes de revelar que a expressão da proteína BGN é reduzida no tecido com câncer de mama, em comparação com o tecido normal.

Palavras-chaves: Câncer de mama, biglicano, imuno-histoquímica, aprendizado profundo

Abstract e Keywords

Background: Breast cancer is the leading cancer in women. The search for new biomarkers for prognostic markers have been sought by many authors. Biglycan (BGN) is a member of the small leucine-rich proteoglycan family of proteins, and is an important component of the extracellular matrix. High mRNA expression levels of BGN were found in various types of solid cancer, compared to normal tissues. However, there is scant data on the protein expression of BGN in breast cancer. The objective of this study is to compare the protein expression of BGN in normal breast tissue and in breast cancer using immunohistochemical techniques. Image analysis of its expression was performed using digital histological score (D-HScore) and supervised deep learning neural networks.

Material and methods: In this case-control study, 24 formalin-fixed, paraffin-embedded tissues were obtained from pathological archives for analysis. Normal breast (n=9) and Breast cancer (n=15) tissue sections were analyzed by immunohistochemistry using BGN monoclonal antibody (M01 - Abnova), clone 4E1-1G7 at dilution 1: 300 at pH 6, and 3,3'-Diaminobenzidine (DAB) as the chromogen. Photomicrographs of the slides were analyzed with ImageJ software with "color deconvolution". After selecting the regions of interest (ROI), deconvoluted panels with DAB only were quantified using arbitrary DAB units. Another set, with higher magnification without ROI selection, was submitted to the inceptionV3 deep neural network image embedding recognition model. Next, supervised neural network analysis, using stratified 20 fold cross validation, with 200 hidden layers, ReLu activation, and regularization at $\alpha=0.0001$ were applied for supervised learning neural networks. The sample size was calculated for a minimum of 7 cases and 7 controls, having a power = 90%, an α error = 5%, and a standard deviation of 20, to identify a decrease from the average of 40 DAB units (control) to 4 DAB units in cancer.

Results: BGN expression (mean \pm SD) was 6.1 ± 3.9 in breast cancer tissue, while in normal breast tissue, it was 39.6 ± 21.9 , using D-HScore (p=0.0017, Student t-test, Welch corrected). SDLNN was

able to correctly classify 110 out of 129 photomicrographs of the dataset using DAB panels only, with a classification accuracy of 85.3% (95%CI=78.1% to 90.3%) and the area under the curve=94.3%

Conclusions: D-HScore and supervised learning neural networks were able to reveal that BGN protein expression is reduced in breast cancer tissue, compared to normal tissue.

Keywords: Breast cancer; Biglycan; immunohistochemistry; deep learning

1. INTRODUÇÃO

O câncer de mama é o mais frequentemente diagnosticado e a principal causa de morte entre as mulheres (SIEGEL *et al.*, 2021). Segundo a Organização Mundial da Saúde, 2,3 milhões de mulheres com diagnóstico de câncer de mama e 685.000 mortes globalmente, em 2020 (WHO, 2021). Dentre os cânceres, o câncer de mama feminino superou o câncer de pulmão como o câncer mais comumente diagnosticado (11,7%), sendo seguido de pulmão (11,4%), colorretal (10,0%), próstata (7,3%) e estômago (5,6%) câncer (SUNG *et al.*, 2021). A extensão de um câncer no momento do diagnóstico é o fator chave utilizado para definir o tratamento e para determinar a possibilidade de resultados bem-sucedidos. Sistemas de estadiamento do câncer foram desenvolvidos com a finalidade de codificar a extensão do câncer para proporcionar aos médicos e pacientes os meios para quantificar o prognóstico para o indivíduo e comparar grupos de pacientes em ensaios clínicos de tratamento, de forma padrão, em todo o mundo (EDGE; COMPTON, 2010).

O câncer de mama é uma doença heterogênea, compreendendo vários subgrupos de acordo com aspectos do seu comportamento clínico, aspectos histológicos, moleculares; a classificação destes subgrupos de câncer tem sido realizada de acordo com a sua expressão gênica (RUSSNES *et al.*, 2017). A expressão imuno-histoquímica do receptor de estrogênio, receptor de progesterona e o receptor de fator de crescimento epitelial humano 2 tem sido amplamente utilizada no prognóstico do câncer de mama e para fornecer estratégias terapêuticas (KIM *et al.*, 2017).

Na tentativa de compreender a fisiopatologia do câncer de mama, vários esforços têm sido realizados para a descoberta de novos biomarcadores para estratificação e prognósticos do câncer de mama, associando algoritmos de inteligência artificial (COCCO *et al.*, 2020; FITZGERALD *et al.*, 2021; MUELLER *et al.*, 2018; ROBERTSON *et al.*, 2018).

A seguir serão revisados os conceitos epidemiológicos do câncer de mama (incidência e fatores de risco), os aspectos histológicos, a classificação dos tumores de mama pela expressão de marcadores moleculares. A presente dissertação focará na expressão de um potencial marcador para o câncer de

mama, a proteína biglican, a qual faz parte dos proteoglicanos e uma breve revisão sobre as redes neurais para a análise de imagens.

2. REVISÃO DA LITERATURA

O câncer de mama é uma doença causada pela multiplicação desordenada de células da mama. Este processo gera células anormais que se multiplicam, formando um tumor. Há vários tipos de câncer de mama. Por isso, a doença pode evoluir de diferentes formas. Alguns tipos têm desenvolvimento rápido, enquanto outros crescem mais lentamente. Estes comportamentos distintos se devem à característica própria de cada tumor (CÂNCER DE MAMA, 2018).

2.1. Incidência

Cerca de 2,3 milhões de casos novos de câncer de mama ocorreram em 2020 globalmente, fazendo com que esta doença seja a causa mais frequente de morte entre as mulheres (WHO, 2021). A taxa ajustada de mortalidade por câncer de mama, na população mundial em 2020, foi estimada em 11,84 óbitos/100.000 mulheres (INCA, 2022) .

A incidência de câncer de mama tem uma estimativa de crescimento progressiva a partir dos 40 anos, assim como a mortalidade por essa neoplasia. A estimativa de novos casos no Brasil para o ano de 2021 foi de 66.280, sendo que as mortes contabilizaram 17.825 mulheres. A taxa ajustada de mortalidade por câncer de mama no Brasil, na mesma época, foi de 8,82/100.000 mulheres. Assim, este câncer continua sendo a primeira causa de morte na população feminina em todas as regiões do Brasil, exceto na região Norte, onde o câncer do colo do útero ainda ocupa o primeiro lugar (INCA, 2022).

2.2 Fatores de risco

Os fatores de risco para o câncer de mama podem ser evidenciados a partir do perfil socioeconômico e demográfico. A tabela 1 descreve a relação dos principais fatores de risco para o câncer de mama (CÂNCER DE MAMA, 2018).

Tabela 1. Fatores que predispõe o câncer de mama

Fatores ambientais e comportamentais	Fatores da história reprodutiva e hormonal	Fatores genéticos e hereditários
Obesidade e sobrepeso após a menopausa	Primeira menstruação antes dos 12 anos	História familiar de câncer de ovário
Sedentarismo e inatividade física	Não ter filhos	Casos de câncer de mama na família, principalmente antes dos 50 anos
Consumo de bebidas alcólicas	primeira gravidez após os 30 anos	História familiar de câncer de mama em homens
Exposição frequente a radiação ionizantes (Raio X)	Parar de menstruar(menopausa) após os 55 anos	Alteração genética especialmente nos genes BRCA 1 e BRCA 2
	Uso de contraceptivos hormonais (estrogênio-progesterona). Ter feito reposição hormonal pós-menopausa, principalmente por mais de cinco anos	

Fonte: Fatores de risco para câncer de mama (CÂNCER DE MAMA, 2018).

2.3. Classificação histológica

Os carcinomas mamários podem ser estratificados em diferentes entidades com base em comportamento clínico, características histológicas e / ou propriedades biológicas. A classificação do câncer de mama deve ser baseada pelo padrão genômico somático das mutações (RUSSNES *et al.*, 2017).

2.3.1. Carcinoma *in Situ*

O carcinoma *in situ* ou câncer não invasivo, refere-se a neoplasia de células limitadas aos ductos e lóbulos pela membrana basal, é o primeiro estágio em que o câncer, não originário das células do sangue, pode ser classificado. Nesse estágio, as células cancerosas estão somente na camada da qual elas se desenvolveram e ainda não se espalharam para outras camadas do órgão de origem. A maioria dos cânceres *in situ* é curável, se for tratada antes que progrida para a fase de câncer invasor (WHO, 2021).

2.3.2. Carcinoma Ductal Invasor

Os carcinomas mamários invasivos são os mais frequentes, com 75% deles subclassificados em carcinoma ductal invasivo/ sem outras especificações (CDI/SOE), 15% como lobular e 10% denominados subtipos especiais (LOPEZ-GARCIA *et al.*, 2010).

2.4. Classificação molecular

Os carcinomas de mama apresentam diferentes comportamentos. Isso se deve pela existência de diversos subtipos moleculares que exigem uma terapêutica específica (PEROU *et al.*, 2000). Essa classificação tem como base os perfis de expressão gênica de cada subgrupo, sendo avaliado principalmente a expressão dos receptores de estrogênio, receptores de progesterona, superexpressão de receptor de fator de crescimento epitelial humano 2, assim como câncer de mama triplo negativo. Estes marcadores podem ser identificados através de técnicas convencionais de imuno-histoquímicas (OLIVER; JAMUR, 2010).

Os sistemas atuais de classificação histopatológica para câncer de mama são baseados em fatores prognósticos. As tecnologias moleculares de alto rendimento estão reformulando nossa compreensão do câncer de mama assim como sua taxonomia molecular que tem um poder preditivo forte e está lentamente emergindo (GEYER; MARCHIO; REIS-FILHO, 2009).

A análise genômica dos tumores de mama revelou que existem, para cada um desses tipos histológicos, subtipos moleculares, conforme a Tabela 2.

Tabela 2. Características genômicas, clínicas e proteômicas dos subtipos

Subtipo	Luminal A	Luminal B	Basal-like	HER2
ER+/HER2- (%)	87%	82%	10%	20%
HER2+(%)	7%	15%	2%	68%
TNBCs via	2	1	80%	95
TP53 via	TP53 mut (12%); ganho do MDM2 (14%)	TP53 mut (32%); ganho do MDM2 (31%)	TP53 mut (84%); ganho do MDM2 (14%)	TP53 mut (75%); ganho do MDM2 (30%)
PIK3CA/PTEN via	PIK3CA mut (49%) PTEN mut/perda (13%); INPP4B perda (9%)	PIK3CA mut (32%)PTENmut/perda (24%); INPP4B perda (16%)	PIK3CA mut (7%)PTENmut/perda (35%); INPP4B perda (30%)	PIK3CA mut (42%)PTENmut/perda (19%); INPP4B perda (30%)
RB1 via	Cyclin D1 amp (29%); CDK4 ganho (14%); baixa expressão do CDKN2C; Alta expression of RB1	Cyclin D1 amp (58%); CDK4 ganho (25%)	RB1 mut/perda (20%); cyclin E1 amp (9%); Alta expressão of CDKN2A; baixa expressão of RB1	Cyclin D1 amp (38%); CDK4 ganho (24%)
mRNA expressão	ER alta cluster; baixa proliferação	ER Baixa cluster; Alta proliferação	Basal assinatura; Alta proliferação	HER2 assinatura do amplicon; Alta proliferação
Número de cópias	Diploide (maioria); muitos com genomas quiescentes; 1q,8q,8p11 ganho;8p, 16q perda; 11q13.3 amp (24%)	Aneuploidia (maioria); muitos com amplificação focal; 1q,8q,8p11 ganho;8p, 16q perda; 11q23 amp (28%)	Aneuploidia (maioria); Alta instabilidade genômica; 1q, 10p ganho; 8p, 5q perda; MYC focal ganho (40%)	Aneuploidia (maioria); Alta instabilidade genômica; 1q, 8p perda; 17q12 focal ERBB2 amp (71%)

amp: amplificação; CDK4: cyclin dependent kinase inhibitor 2A; CDKN2C: cyclin dependent kinase inhibitor 2C; ER +: receptor de estrogênio positivo; ER +/HER2-: receptor de estrogênio positivo/receptor de crescimento epidérmico 2 negativo; HER2 +:receptor de crescimento epidérmico positivo; INPP4B: inositol polyphosphate-4-phosphatase type II B; MDM2: MDM2 proto-oncogene; MYC: MYC proto-oncogene, bHLH transcription factor; mRNA: mensageiro do ácido ribonucleico; mut: mutação; PTEN: phosphatase and tensin homolog; TP53: tumor protein 53; RB1: RB transcriptional corepressor 1;

Fonte: Cancer Genome Atlas Network (CANCER GENOME ATLAS NETWORK, 2012)

2.4.1.1 Luminal A

O subtipo luminal A representa a maioria dos tumores luminais, tendo o melhor prognóstico de todos os cânceres de mama (FRAGOMENI; SCIALIS; JERUSS, 2018). O perfil imunohistoquímico destes tumores é receptor de estrogênio e progesterona positivos, fator de crescimento epitelial humano 2 negativo, e baixos índices proliferativos. O tumor luminal A é caracterizado pela elevada expressão de genes representados pelas células epiteliais luminais, como, por exemplo, as citoqueratinas 7, 8, 18 e 19. A resposta à terapêutica antiestrogênica traz um prognóstico melhor (SØRLIE *et al.*, 2001).

2.4.1.2 Luminal B

O luminal B é o grupo mais heterogêneo, tanto clínica, quanto molecularmente. Os cânceres de mama luminais B compõem cerca de 20% dos tumores de mama (FRAGOMENI; SCIALIS; JERUSS, 2018) e são caracterizados pela expressão do receptor de estrogênio, uma expressão baixa ou ausente do receptor de progesterona e um alto grau histológico (LI *et al.*, 2016). O fator de crescimento epitelial humano 2 pode ser negativo ou positivo. O subtipo luminal B tem maior índice de proliferação celular (a expressão aumentada do Ki-67 elevado), o que leva a um pior prognóstico em relação aos tumores luminais A (VODUC *et al.*, 2010). Ele também está associado ao maior risco de recorrência e à menor sobrevida livre da doença, mostrando beneficiar-se mais do que o subtipo luminal A de quimioterapia associada a antiestrogênicos (BARRETO-NETO *et al.*, 2014; SØRLIE *et al.*, 2001)

A variabilidade da expressão do tumor luminal B (receptor de estrogênio positivo, receptor de fator de crescimento epitelial humano 2 positivo ou ausente, altos índices proliferativos) faz com que as pacientes recebem um regime de terapia diferenciada (que incorpora terapia anti-HER2 direcionada) em comparação com outros subtipos luminais (BEDIAGA *et al.*, 2016).

2.4.1.3 Triplo Negativo/Basaloides

O câncer de mama triplo negativo (TNBC) é um subtipo de câncer de mama caracterizado por um comportamento agressivo; os receptores de estrogênio, progesterona e do crescimento epidérmico humano receptor de fator 2 têm expressão negativa na imuno-histoquímica (YANG *et al.*, 2015), (SANTUARIO-FACIO *et al.*, 2017). A prevalência deste subtipo é de 15 a 20% do total de casos com câncer de mama (BROWN *et al.*, 2008).

Clinicamente, os tumores triplo negativos ocorrem em mulheres mais jovens, que podem ser descobertos durante a mamografia de rastreamento; o risco de recorrência é maior nos primeiros 3 anos. Recorrências distantes no cérebro e metástases viscerais são mais comuns do que em tumores positivos para receptores hormonais. Terapeuticamente, apesar de serem altamente quimiossensíveis, seu tempo livre de progressão é geralmente curto (CHACÓN; COSTANZO, 2010).

Os subtipos mais comuns e mais bem caracterizados dos subtipos moleculares triplo-negativos são (LEHMANN *et al.*, 2011):

- Basaloide 1 (BS1)
- Basaloide 2 (BS2)
- Imunomodulatório (IM), mesenquimal (M)
- Mesenquimal, células tronco-símile (MSL)
- Luminal receptor de andrógeno.

Os carcinomas basaloides tem uma expressão forte das citoqueratina 5/6, 14 e 17, têm receptores hormonais negativos, baixa expressão do fator de crescimento epitelial humano 2, podem expressar receptores do fator de crescimento epidérmico e c-kit, tem alto grau histológico, pior prognóstico

do que tumores triplo-negativos não basalóides e estão associados com mutação BRCA1 (CHACÓN; COSTANZO, 2010).

2.4.1.4 Receptor do Fator de Crescimento Epitelial Humano 2 (HER2)

O receptor HER2 (receptor 2 do fator de crescimento epidérmico humano) é uma tirosina quinase de membrana e, quando ativado, afeta a proliferação e a sobrevivência celular (KRISHNAMURTI; SILVERMAN, 2014). Os tumores HER-2 enriquecidos constituem 12 a 20% dos carcinomas mamários infiltrantes (FRAGOMENI; SCIALLI; JERUSS, 2018).

Histologicamente são representados por tumores grau 2 e grau 3 (moderadamente e pouco diferenciados) sendo a sua grande maioria carcinomas infiltrantes do tipo não especial, além de carcinomas apócrinos e lobulares pleomórficos.¹⁵

Seu perfil genético é caracterizado pela superexpressão do HER-2 e genes relacionados, mutações de TP53, mutações de *phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha* (PIK3CA), superexpressão do receptor 4 do fator de crescimento de fibroblastos (FGFR4), superexpressão do receptor do fator de crescimento epidérmico (EGFR), mutações da enzima de edição de mRNA de apolipoproteína B, polipeptídeo catalítico 3 (APOBEC), amplificação de ciclina D1 e alta instabilidade genômica (CANCER GENOME ATLAS NETWORK, 2012; PEROU *et al.*, 2000; PRAT *et al.*, 2014; SØRLIE *et al.*, 2001; SORLIE *et al.*, 2003; STAAF *et al.*, 2010).

Seu perfil imuno-histoquímico é caracterizado por (KRISHNAMURTI; SILVERMAN, 2014):

- Receptor de Estrogênio negativo
- Receptor de progesterona negativo
- Expressão alta do Ki-67
- superexpressão do HER-2.

Os tumores HER-2 enriquecidos tendem a ter crescimento mais rápido que os luminais e têm pior prognóstico, porém, com o advento de terapias-alvo, têm excelente resposta ao tratamento e melhora expressiva de prognóstico (HAMILTON *et al.*, 2021).

3. Proteoglicanos

Entre 1930 e 1960, as investigações de “condromucoide” da cartilagem e preparações anticoagulantes do fígado (heparina) deram início ao estudo das proteoglicanos. A partir da análise química dos polissacarídeos dessas preparações, também conhecidos como “mucopolissacarídeos”, obtendo-se a estrutura do hialuronano, sulfato de dermatano, sulfato de queratano, diferentes formas isoméricas de sulfato de condroitina, heparina e sulfato de heparano). Juntos, esses polissacarídeos passaram a ser conhecidos como glicosaminoglicanos para indicar a presença de aminoácidos e outros açúcares em uma forma polimérica. Estudos subsequentes forneceram informações sobre a ligação das cadeias às proteínas do núcleo de proteoglicanos. Esses estudos estruturais abriram caminho para estudos biossintéticos com o desenvolvimento da purificação e análise de dos tecidos com proteoglicanos e glicosaminoglicanos; a ultracentrifugação com gradiente de densidade permitiu a separação dos grandes proteoglicanos agregadores da cartilagem (MERRY *et al.*, 2022).

A produção de proteoglicanos era uma propriedade geral das células animais; proteoglicanos e glicosaminoglicanos estavam presentes na superfície celular, no interior da célula e na matriz extracelular; verificou-se, então, que o proteoglicano tinha função na adesão celular, sinalização e outras atividades biológicas (XU *et al.*, 2022).

Os proteoglicanos são moléculas que consistem em um núcleo protéico ao qual estão ligadas a uma ou mais cadeias glicosaminoglicanas (Figura 1).

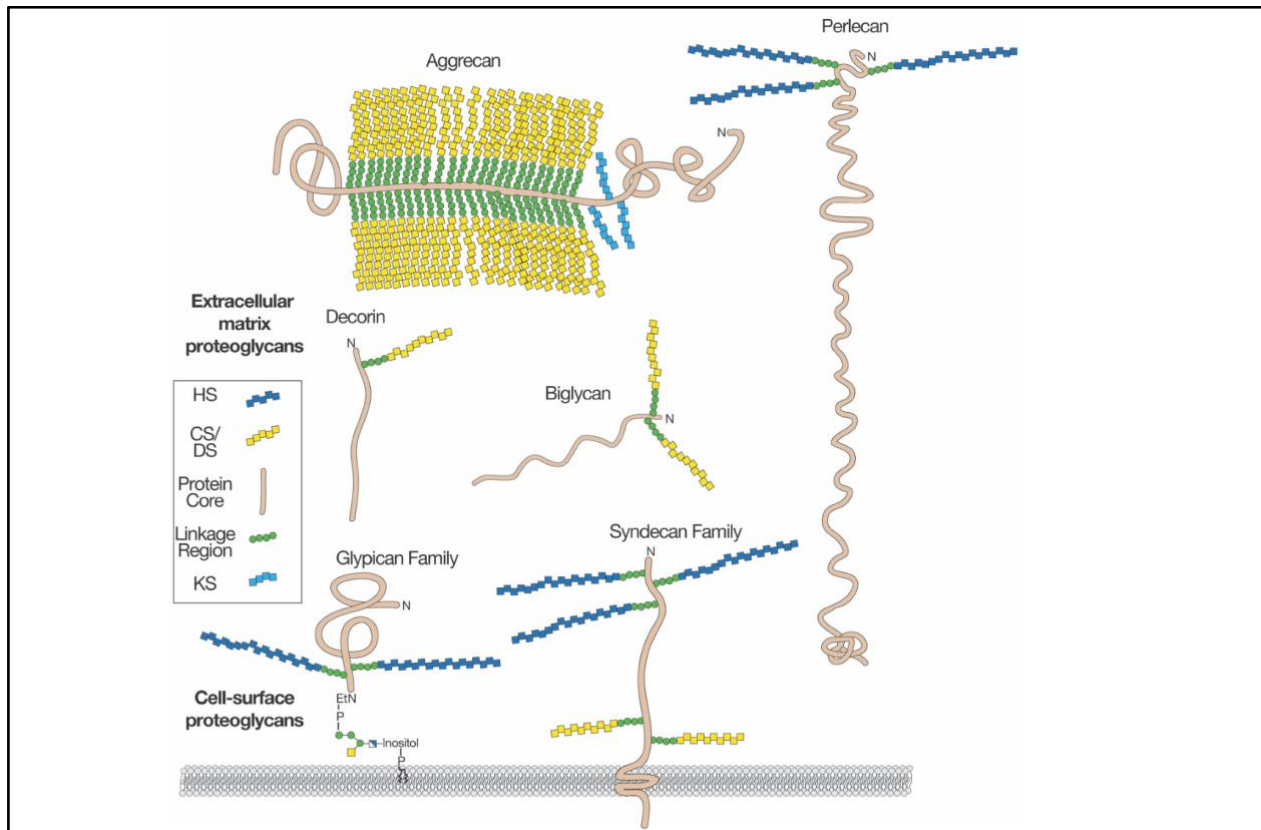


Figura 1. Os proteoglicanos. Os proteoglicanos consistem em um núcleo proteico (marrom) e uma ou mais cadeias de glicosaminoglicanos ligadas covalentemente (azul escuro, HS, sulfato de heparano; amarelo, CS/DS, sulfato de condroitina/sulfato de dermatano; azul claro, KS, sulfato de queratano). Os proteoglicanos de membrana atravessam a membrana plasmática (proteínas de membrana tipo I) ou estão ligados por uma âncora de glicosilfosfatidilinositol (GPI). Os proteoglicanos da matriz extracelular geralmente são secretados, mas alguns proteoglicanos podem ser clivados proteoliticamente e se desprender da superfície celular (não mostrado).

Fonte: Merry et al. (MERRY et al., 2022)

Praticamente todas as células de mamíferos produzem proteoglicanos, os quais são secretados na matriz extracelular (MEC), inseridos na membrana plasmática, ou os armazenam em grânulos de secreção. A MEC é um componente essencial na determinação das características físicas dos tecidos e muitas das propriedades biológicas das células que estão neste tecido. Os principais componentes da MEC são proteínas fibrilares que fornecem resistência à tração e elasticidade (por exemplo, vários colágenos e elastinas), glicoproteínas adesivas (por exemplo, fibronectina, lamininas e tenascinas) e os proteoglicanos que interagem com outros componentes da MEC para promover a

montagem da MEC, orientam suas propriedades físicas e servem como um reservatório de pequenas proteínas biologicamente ativas, como fatores de crescimento. Um único tipo de célula pode expressar vários proteoglicanos. Assim, um determinado proteoglicano (definido por sua proteína central) representa uma população diversificada de moléculas, cada uma potencialmente representando uma entidade estrutural única. Estas características, típicas de todos os proteoglicanos, criam uma enorme diversidade e potencial variação biológica na atividade (ESKO; KIMATA; LINDAHL, 2009).

3.1 Proteoglicanos - Forma e função

As principais classes de proteoglicanos podem ser classificadas por sua distribuição, homologies e função. Os proteoglicanos podem ser secretados, ligados à membrana ou localizados em grânulos intracelulares. Os proteoglicanos secretados podem ser divididos entre os da família de agrecanos (lecticanos), nos pequenos proteoglicanos ricos em leucina e outros. A Tabela 3 fornece uma visão geral dos principais pequenos proteoglicanos ricos em leucina.

Entre outras funções, os proteoglicanos estão envolvidos na resistência à tração da pele e dos tendões, afetam o crescimento e a invasão de células tumorais e influenciam a progressão das neurites (HALPER, 2014; IOZZO; SCHAEFER, 2015).

Tabela 3. Diversidade entre os principais proteoglicanos secretados - pequenos proteoglicanos ricos em leucina (SLRPs).

Proteoglicano	Proteína central (kDa)	Número e tipo de cadeias de glicosaminoglicanos	Distribuição nos tecidos	Função proposta	Doenças associadas em humanos
Proteoglicanos secretados - pequenos proteoglicanos ricos em leucina (SLRPs)					
Decorina	36	1 CS/DS	células do tecido conjuntivo	regula a fibrilogênese do colágeno intersticial, inibição da sinalização do TGF- β	superexpressão na esclerose sistêmica, mutações no estroma congênito distrofia da córnea
Biglicano	38	0–2 CS/DS	células do tecido conjuntivo, macrófagos	montagem da matriz de colágeno, formas solúveis ativam o sistema imunológico inato	
Lumicano	37	3–4 KS I	amplamente distribuído	montagem da matriz de colágeno	
Keratocano	37	3–4 KS I	amplamente distribuído, mas sulfatado apenas na córnea	montagem da matriz de colágeno, função na transparência da córnea	mutações associadas à córnea plana
Fibromodulina	59	2–4 KS I	amplamente distribuído	montagem da matriz de colágeno	variações intrônicas e SNPs associados com miopia de alto grau
Osteoglicina/ Mimecano	25	2–3 KS I	amplamente distribuído, mas sulfatado apenas na córnea	montagem da matriz de colágeno, função na formação óssea e na transparência da córnea.	potencial biomarcador de insuficiência cardíaca isquêmica

CS, sulfato de condroitina; DS, sulfato de dermatan, KS, sulfato de queratan; ECM, matriz extracelular; TGF- β , fator transformador de crescimento- β ; SNP, polimorfismo de nucleotídeo único.

Fonte: De Merry et al (MERRY *et al.*, 2022)

3.1.1 Proteoglicanos Pequenos Ricos em Leucina (*Small Leucine Rich Proteoglycans* - SLRPs)

Os SLRPs compõem um subgrupo diversificado de proteoglicanos. Eles estão envolvidos na organização da matriz extracelular e na sua regulação do crescimento, bem como na sinalização celular. Os SLRPs são distribuídos de forma onipresente na matriz extracelular. Além dos proteoglicanos apresentados na Tabela 3, existem outros 11 membros menos caracterizados, os

quais estão organizados em cinco classes distintas com base em sua conservação evolutiva de proteínas, repetições ricas em leucina (LRR), aglomerados ricos em cisteína N-terminal e organização cromossômica (NIKITOVIC *et al.*, 2012; SCHAEFER; IOZZO, 2008). Os membros das duas primeiras classes estão envolvidos na fibrogênese do colágeno e nos processos relacionados. Os proteoglicanos pequenos ricos em leucina da Classe I contêm 10 regiões ricas em leucina e possuem condroitina, com ou sem sulfato de dermatan, ligados ao núcleo da proteína (HALPER, 2014; IOZZO; SCHAEFER, 2015).

A Tabela 4 apresenta a classificação dos SLRPs com os seus respectivos símbolos genéticos.

Tabela 4. Classificação dos Proteoglicanos Pequenos Ricos em Leucina (SLRP) com o respectivo símbolo do gene, de acordo com o comitê de nomenclatura de genes HUGO (HGNC).

Proteoglicanos Pequenos Ricos em Leucina	
Classe I	Decorina (DCN) Biglicano (BGN) Asporina (ASPN) Proteína 2 da matriz extracelular (ECM2)
Classe II	Fibromodulina (FMOD) Lumican (LUM) Proteína de repetição rica em leucina e extremidade rica em prolina e arginina (PRELP) Keratocana (KERA) Osteomodulina (OMD)
Classe III	Epifican (EPYC) Opticina (OPTC) Osteoglicina (OGTC)
Classe IV	Chondroaderina (CHAD) Nictalopina (NYX) Tsukushi (TSKU)
Classe V	Podocan (PODN) Proteína 1 tipo Podocan (PODNL1)

Classificação adaptada de Schaefer et al. (SCHAEFER; IOZZO, 2008)

Uma característica dos proteoglicanos pequenos ricos em leucina é a sua capacidade de interagir com receptores extracelulares (como receptores de tirosina quinase e receptores do tipo *Toll-like*).

Esta interação inicia a sinalização que regula a função da matriz celular, especialmente a proliferação e a inflamação (KOBE; KAJAVA, 2001; SCHAEFER; IOZZO, 2008).

Os SLRPs da Classe I têm uma sequência Cys N-terminal única. Nesta classe, estão incluídas a decorina, biglicano e asporin, que são codificados por genes compostos por oito éxons com junções de íntrons em posições altamente conservadas (WEI; ZHANG, 2018). Dentes os SLRP da classe I, iremos aprofundar sobre a proteína biglicano.

3.2 Biglicano (BGN)

O gene do Biglicano está localizado no cromossomo Xq13-qter, na posição 153.494.980 - 153.509.546, com 14,567 bases (MCBRIDE; FISHER; YOUNG, 1990). O gene do biglicano codifica um núcleo de proteína de 42 kDa. No entanto, quando totalmente glicosilado, o peso molecular do biglicano aumenta até 100-250 kDa, devido à presença de duas cadeias de condroitina/dermatan sulfato (CS/DS) glicosaminoglicano (GAG) ligadas covalentemente à proteína (BIANCO *et al.*, 1990). O núcleo proteico contém 331 aminoácidos e há duas cadeias de condroitina/sulfato de dermatan ligadas à sua região N-terminal (TUFVESSON *et al.*, 2002).

O biglicano está intimamente relacionado na estrutura primária com a decorina, sendo ambos codificados por genes que consistem em oito éxons e sete íntrons e que provavelmente surgiram de um ancestral comum. Os núcleos maduros de proteínas biglicanas são altamente conservados entre as espécies (Figura 2).

3.2.1 Localização do gene

O gene do biglicano está localizado no cromossomo Xq13-qter, na posição 153.494.980 - 153.509.546 (FISHER *et al.*, 1991), com 14,567 bases (MCBRIDE; FISHER; YOUNG, 1990). Este gene codifica um núcleo de proteína de 42 kDa contendo repetições ricas em leucina, às quais uma ou duas cadeias laterais de GAG estão ligadas covalentemente (BIANCO *et al.*, 1990). O biglicano está intimamente relacionado na estrutura primária com a decorina, sendo ambos codificados por

genes que consistem em oito éxons e sete íntrons e que provavelmente surgiram de um ancestral comum. Os núcleos maduros das proteínas biglicanos são altamente conservados entre as espécies.

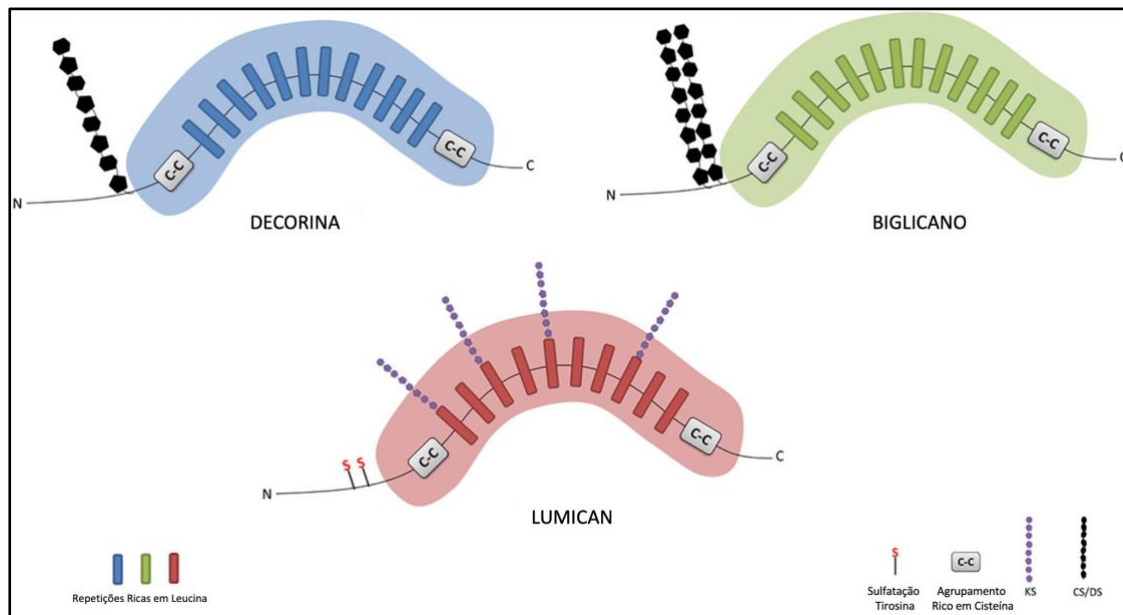


Figura 2. Estrutura básica da decorina, biglicano e lumican. Estes Proteoglicanos Pequenos Ricos em Leucina possuem uma proteína central (core) em forma de solenóide curvo compreendendo 10-12 motivos de repetição rica em leucina que são flanqueados em ambos os lados por agrupamentos ricos em cisteína. A decorina consiste em um único CS/DS e o biglicano tem dois CS/DS; o lumican tem quatro KS e resíduos de sulfatação de tirosina ao longo dos domínios N-terminais. (CS- Sulfato de condroitina; DS- Sulfato de Dermatan; KS- Sulfato de queratano; N: terminal Amino; C: terminal Carboxil).

Fonte: De Appunni et al. (APPUNNI *et al.*, 2019)

3.2.2 Estrutura da biglicano glicosilada

As cadeias GAG de condroitina ou dermatan-sulfato específicas do tecido do biglicano estão ligadas a resíduos de aminoácidos no terminal N da proteína central (CHOI *et al.*, 1989; ROUGHLEY *et al.*, 1993). Existem evidências de que o biglicano pode ser um proteoglicano de “tempo parcial”, pois sua forma não glicada foi encontrada na cartilagem articular envelhecida e nos discos intervertebrais (JOHNSTONE *et al.*, 1993; ROUGHLEY *et al.*, 1993).

A proteína tem uma porção principal de 42 kDa e duas cadeias de condroitina / sulfato de dermatana; é composto por um núcleo proteico de 331 aminoácidos e duas cadeias de condroitina / sulfato de dermatan ligadas à sua região N-terminal. Contudo, quando totalmente glicosilado, o peso molecular

do biglicano aumenta até 100-250 kDa, devido à presença de duas cadeias de condroitina/dermatan sulfato (CS/DS) glicosaminoglicano (GAG) ligadas covalentemente à proteína (TUFVESSON *et al.*, 2002).

3.2.3 Localização da Proteína Biglicano

O biglicano é ubíquo na matriz extracelular, onde serve como um componente essencial para os componentes da matriz extracelular e como uma molécula essencial para a sinalização (HU *et al.*, 2016). A sua expressão também é vista no espaço pericelular rico em leucina da matriz extracelular (SUBBARAYAN *et al.*, 2018). A Figura 3 apresenta a localização do biglicano

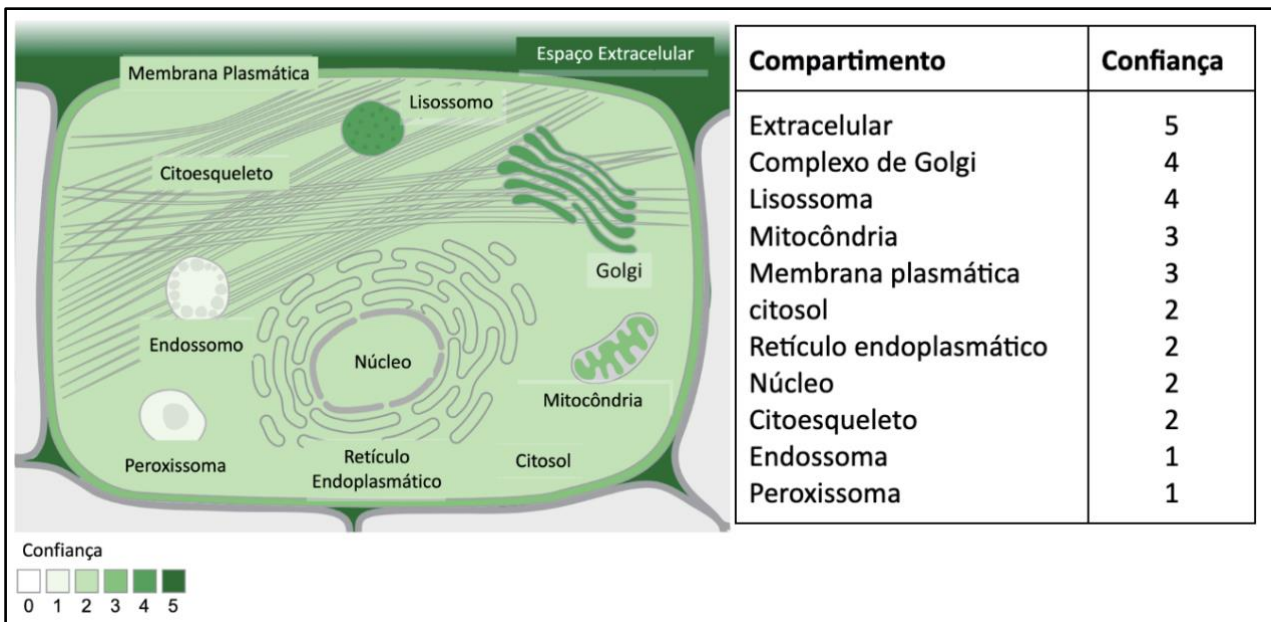


Figura 3. Localização da proteína biglicano na célula

Fonte: De Genecards.org (GENECARDS HUMAN GENE DATABASE, 2022)

A expressão da proteína biglicano ocorre em diferentes tecidos normais no ser humano, principalmente nas células mesenquimais da medula óssea, medula espinhal, mama, vesícula e pele, conforme demonstrado na Figura 4.

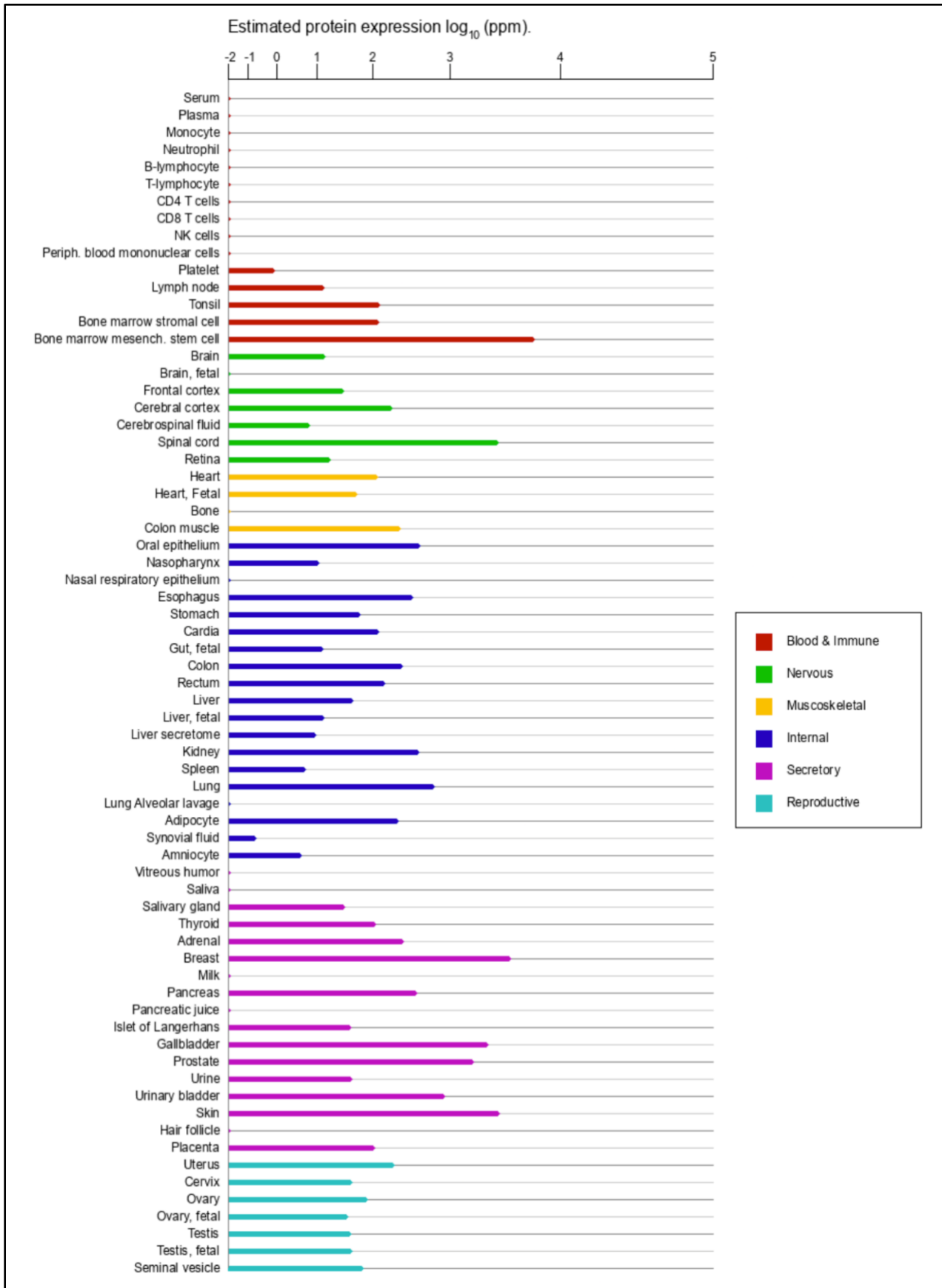


Figura 4. Estimativa da expressão(\log_{10}) da proteína biglicano nos diferentes tecidos humanos.

Fonte: Genecards.org (GENECARDS HUMAN GENE DATABASE, 2022)

3.2.4 Função da proteína biglicano

Cada função proposta depende do tecido ou órgão em questão. O biglicano serve como um componente chave da matriz e é uma molécula essencial de sinalização. Considera-se que o biglicano tenha um papel organizador na montagem da matriz extracelular e está estruturalmente relacionado à decorina, fibromodulina e lumican (YANG; ROTHRAUFF; TUAN, 2013).

As principais funções do biglicano incluem modulação da montagem da matriz, migração celular, adesão, mineralização óssea, inflamação, crescimento celular, regulação da autofagia e apoptose. Consequentemente, o biglicano está envolvido em vários processos fisiológicos e fisiopatológicos, como a tumorigênese (SUBBARAYAN *et al.*, 2018).

O biglicano interage com os *Toll Like Receptors 2* e *4* nas células imunes inatas para iniciar a inflamação através da via (HU *et al.*, 2016; NASTASE; YOUNG; SCHAEFER, 2012).

3.2.4 Moduladores da proteína biglican

A proteína biglicano interage diretamente com 15 proteínas conforme demonstrado na Figura 5. As interações aqui demonstradas são aquelas onde a interação apresentou um alto grau de confiança ($\geq 90\%$) e foi verificada em estudos *in vitro*, em modelos animais ou em humanos.

O biglicano é estimulado pelo TGF- β (KÄHÄRI; LARJAVA; UITTO, 1991), enquanto que o uso de dexametasona reduz a expressão do biglicano (KIMOTO *et al.*, 1994). A região flanqueadora 5' do biglicano humano possui atividade promotora que pode ser modulada pela interleucina 6 e pelo Fator Tumoral de Necrose- α (UNGEFROREN; KRULL, 1996).

O biglicano estimula o fator de crescimento vascular epitelial através da via dos *Toll Like Receptors* (HU *et al.*, 2016).

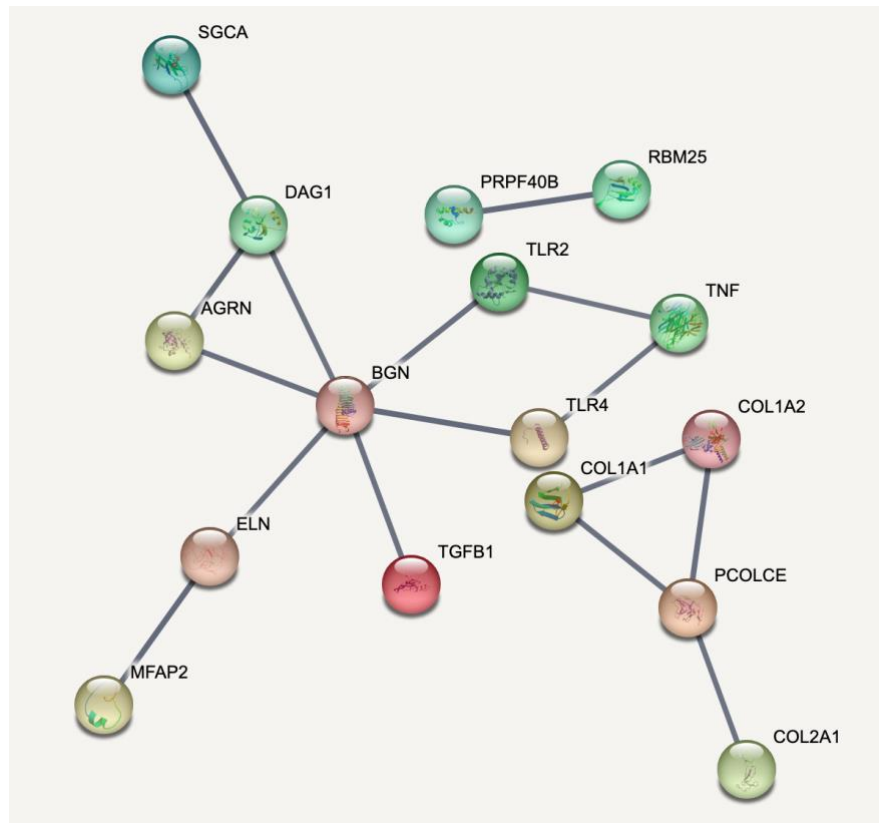


Figura 5. Rede interativa da proteína biglicano (BGN). **TGFB1**: Transforming growth factor beta-1. **COL1A1**: Collagen alpha-1(I) chain; **TLR2**: Toll-like receptor 2; **RBM25**:RNA-binding protein 25; **SGCA**:Alpha-sarcoglycan; **COL1A2**:Collagen alpha-2(I) chain; **ELN**:Elastin; **TLR4**: Toll-like receptor 4; **MFAP2**: Microfibrillar-associated protein 2; **AGRN**: Agrin; **PCOLCE**: Procollagen C-endopeptidase enhancer 1; **COL2A1**:Collagen alpha-1(II) chain; **TNF**:Tumor necrosis factor; **DAG1**:Dystroglycan 1; **PRPF40B**: Pre-mRNA-processing factor 40 homolog B.

Fonte: STRING Interaction Network (STRING INTERACTION NETWORK, 2022)

3.2.5 Proteína Biglicano e doenças

O biglicano tem sido associado a diferentes doenças que estão relacionadas ao colágeno. A displasia Espondilopimetafisária ligada ao X e a Síndrome de Meester-Loeysforam associadas a defeitos do biglicano (CAMERA; STELLA; CAMERA, 1994; CHO *et al.*, 2016; MEESTER *et al.*, 2017)

A lesão renal isquêmica aguda está associada com a ativação do biglicano através dos *Toll Like Receptors 2 e 4* (MORETH *et al.*, 2014).

3.2.6 Biglicano e câncer

Estudos iniciais sobre o papel do biglicano no microambiente do tumor produziram resultados conflitantes em que alguns acharam que era promotor de tumor (GU *et al.*, 2012; NABA *et al.*, 2012; WANG *et al.*, 2011), enquanto os resultados de outros estudos sugerem que o biglicano suprime o crescimento tumoral (NASH; LOERCHER; FREEDMAN, 1999; WEBER *et al.*, 2001). Esses diferentes efeitos podem depender do tipo do tecido e do tipo específico de glicosaminoglicanos ligados. O biglicano também pode influenciar o crescimento e a diferenciação celular diretamente pela modulação do armazenamento local e/ou disponibilidade de diferentes fatores de crescimento, incluindo *bone morphogenetic protein 4* (BMP-4). (CHEN *et al.*, 2004)

A regulação positiva do *biglicano* no estroma do câncer tem sido associada a proliferação celular, migração celular, metástase e angiogênese (HU *et al.*, 2016). Através de um modelo murino, usando células endoteliais tumorais de duas linhas de melanoma (alto e baixo poder metastático), foi verificado que as células endoteliais tumorais secretam biglicano; a secreção do biglicano ativou a migração das células tumorais via ativação das vias Nuclear Factor- κ B e *extracellular regulated MAP kinase* (ERK), através da sinalização dos *Toll Like Receptors* (MAISHI *et al.*, 2016), conforme demonstrado na Figura 6.

O biglicano é superexpresso em adenocarcinoma pancreático, câncer de cólon, câncer de ovário, câncer endometrial, colangiocarcinoma intra-hepático, câncer gástrico, carcinoma de células escamosas do esôfago e melanoma (GU *et al.*, 2012; LIU *et al.*, 2014; SCHAEFER *et al.*, 2017; WANG *et al.*, 2011). A tabela 5 apresenta um resumo da expressão do biglicano em diferentes estudos com câncer.

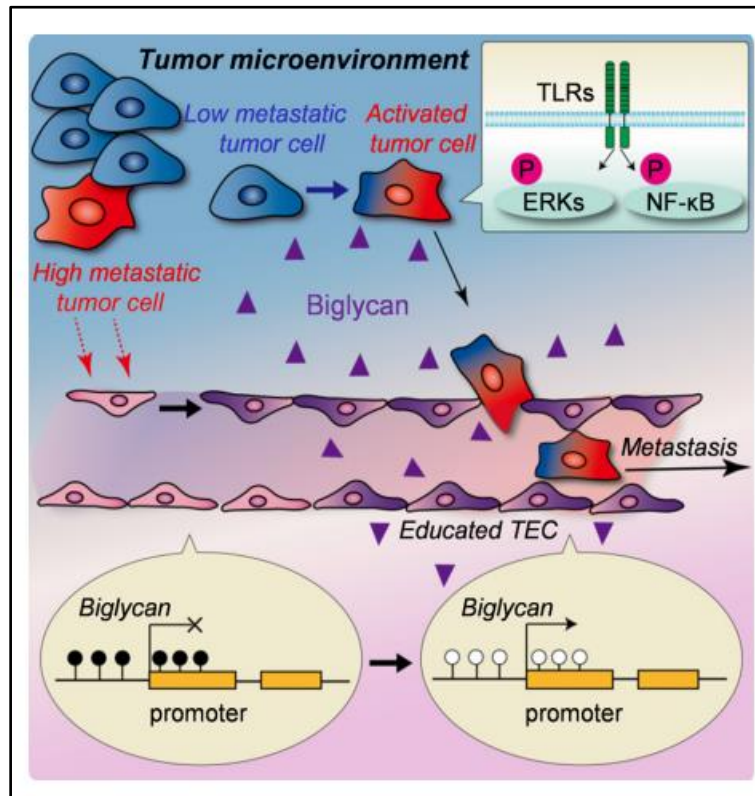


Figura 6. Influência do biglicano no microambiente tumoral. Os vasos sanguíneos do tumor atuam como um “portão” para metástase (durante a invasão das células tumorais), guardado-as como “porteiros” das células epiteliais tumorais (CET). As CET em tumores altamente metastáticos têm uma molécula “chave”, o biglicano, por meio da hipometilação do DNA para permitir que as células tumorais rompam esse portão e prossigam para a corrente sanguínea, o que resulta em metástase hematogênica. O complexo CET- biglicano estimula o receptor *Toll-like* expressando células tumorais metastáticas pela ativação das vias NF-κB e ERK

Fonte: Maishi et al. (MAISHI *et al.*, 2016).

Poucos estudos com modelos de câncer de mama e de adenocarcinoma ductal pancreático sugeriram um papel antiproliferativo para o biglicano (THEOCHARIS *et al.*, 2015; WEBER *et al.*, 2001). Há achados contraditórios no carcinoma urotelial entre linhagens celulares e tecidos, o primeiro mostrando uma correlação negativa entre expressão e proliferação do biglicano (APPUNNI *et al.*, 2019; SCHULZ *et al.*, 2019)

Tabela 5. Expressão do biglicano e sua relevância clínica observada em diferentes estudos com câncer.

Autores	Cancer/linha celular/xenoinxerto	Expressão	Desfecho/significado
Wang et al.(WANG <i>et al.</i> , 2011)	Câncer gástrico	Alta	Alta expressão de mRNA e proteína associada a mau prognóstico do paciente
Gu et al. (GU <i>et al.</i> , 2012; LIU <i>et al.</i> , 2014; SCHAEFER <i>et al.</i> , 2017; WANG <i>et al.</i> , 2011)	Câncer colorretal	Alta	Alta expressão de mRNA do biglicano associado a agressividade do tumor e metástases
Niedworok et al. (NIEDWOROK <i>et al.</i> , 2013)	Câncer de bexiga	Alta	Alta expressão de mRNA do biglicano associado a invasão muscular do câncer
	Linha celular J82		A expressão do biglicano está reduzida na proliferação celular <i>in vitro</i>
Liu et al. (APPUNNI <i>et al.</i> , 2017; LIU <i>et al.</i> , 2014)	Câncer Endometrial	Alta	Expressão elevada de soro, mRNA e proteína associada a mau prognóstico
Appunni et al.(APPUNNI <i>et al.</i> , 2017; LIU <i>et al.</i> , 2014)	Carcinoma Urotelial	Alta	Alta expressão circulatória (soro), mRNA e proteína no câncer de bexiga. Biomarcador de câncer plausível
Andrlová et al.(ANDRLOVÁ <i>et al.</i> , 2017)	Melanoma	Alta	Alta expressão gênica do biglicano associada a mau prognóstico clínico
	camundongo Bgn ^{-/-}		Rigidez tecidual reduzida e invasão reduzida
Jacobsen et al. (JACOBSEN <i>et al.</i> , 2017)	Câncer Próstata	Alta	Imunocoloração positiva de biglicano em 78,7% das amostras associada a desfecho clínico ruim

Fonte: De Appunni et al. (APPUNNI *et al.*, 2019)

Em modelos animais de câncer de cólon, os níveis de biglicano foram 6 vezes maiores em tumores em comparação com o normal. *In vitro*, o câncer de cólon e as células iniciadoras de câncer de cólon não expressaram biglicano, indicando a origem extracelular no nível do tecido (SUHOVSKIH *et al.*, 2015).

O biglicano tem potencial oncogênico ou supressor tumoral dependendo da origem celular. A superexpressão de HER-2/neu em fibroblastos murinos e sistemas de modelo humano está inversamente correlacionada com a expressão de biglicano. Os fibroblastos BGN^{high} HER-2/neu+ eram menos tumorigênicos em camundongos imunocompetentes, em comparação às células BGN^{low/neg} HER-2/neu+. O aumento da imunogenicidade dos fibroblastos BGN^{high} HER-2/neu+ parece ser devido a antígenos de superfície MHC de classe I regulados positivamente e níveis de expressão reduzidos das isoformas do fator de crescimento transformador (TGF)- β e do receptor 1 de TGF- β , sugerindo uma ligação entre biglicano, TGF - β via e regulação negativa mediada por HER-2/neu de antígenos MHC classe I (SUBBARAYAN *et al.*, 2018).

O biglicano superexpresso em tecido tumoral resultaria em um aumento metastático, angiogênese e pior prognóstico. Estes mecanismos ocorrem através dos *Toll Like Receptor 2 e 4*, e pela ativação das vias de sinalização *NF- κ B*, *Focal adhesion kinase (FAK)*, *ERK*, *hypoxia inducible factor 1 subunit alpha (HIF1A)*, *vascular endothelial growth factor (VEGF)* e pela inibição da TGF- β (APPUNNI *et al.*, 2019)

3.2.8 Biglicano e câncer de mama

O estudo do biglicano em caso de câncer de mama tem se restringido a modelos *in vitro*, *in silico* e modelos animais.

Leygue et al descreveram a expressão do biglicano em tecidos mamários normais e neoplásicos de mulheres. O estudo utilizou a técnica de Western blot de biópsias congeladas e a extração completa das proteínas. para tanto, eles utilizaram o anticorpo não comercial para identificação do biglicano (ROUGHLEY *et al.*, 1993). Eles verificaram que a expressão proteica do biglicano estava reduzida tanto nos tecidos tumorais de mama, como na mama normal. O biglicano foi detectado apenas em níveis baixos em 2 dos 6 tecidos normais e em 3 dos 9 tumores mamários (LEYGUE *et al.*, 2000).

Bischof et al. utilizaram um modelo tridimensional de matriz extracelular descelularizada para determinar se o mesênquima embrionário poderia induzir a normalização do câncer num modelo de co-cultura. Os autores verificaram que o biglicano estava implicado na regulação do câncer. A matriz extracelular deficiente de biglicano, a qual era produzida pelas células mamárias mesenquimais, falhou em normalizar o tumor mamário para um tecido normal. Estes resultados confirmaram que o mesênquima embrionário retém a capacidade de induzir a reversão parcial do câncer de mama; esta capacidade reside, pelo menos em parte, na proteína biglicano produzida na matriz extracelular (BISCHOF *et al.*, 2013).

Zhao et al. utilizaram o banco de dados da Oncomine. Oncomine é um banco de dados de microarray de câncer e uma plataforma de mineração de dados baseada na web destinada a facilitar a descoberta de análises de expressão em todo o genoma, bem como comparar os dados do transcriptoma em vários tipos de câncer, respectivos a tecidos normais. eles utilizaram 13 banco de dados com mais de 3000 amostras, com diferentes tipos de carcinoma mamário. Os autores não conseguiram identificar uma associação significativa da regulação positiva do biglicano com a sobrevida do paciente com câncer de mama (ZHAO *et al.*, 2020).

Cong et al. injetaram células de câncer de mama E0771 em camundongos selvagens e *knockout* para biglicano. Eles verificaram que os camundongos com câncer de mama e com alta expressão de biglicano tiveram pior sobrevida livre de metástase à distância. A expressão do biglicano foi maior no compartimento estromal do tumor em comparação com o compartimento epitelial. Os camundongos com *knockout* para biglicano no estroma (Bgn KO) tiveram uma inibição para das metástases das células E071 para o pulmão. Os camundongos Bgn KO também tiveram uma redução da angiogênese tumoral; uma redução da vasculatura tumoral, através da supressão da via de sinalização do fator de necrose tumoral- α /angiopoietina 2. A fibrose foi suprimida e a infiltração de células T CD8⁺ foi aumentada em camundongos Bgn KO portadores de tumor. A distribuição e a

eficácia da droga quimioterápica foram melhoradas *in vivo* em camundongos Bgn KO (CONG *et al.*, 2021)

Manupati *et al.* utilizaram células tronco de câncer de mama e verificaram que uma pequena subpopulação destas células tumorais tinha um alto potencial metastático. O RNAm biglicano superexpresso estava associado com os fenótipos agressivos dessas células. A supressão do RNAm do biglicano reduzia o metabolismo celular, a diminuição da ativação do fator de transcrição NFκB, da proteína p65 e dos níveis de fosfo-IκB após a supressão do biglicano, indicando regulação do fenótipo agressivo do biglicano era pela via NFκB (MANUPATI *et al.*, 2022).

3. REDES NEURAIAS CONVOLUCIONAIS (CNN)

Em 1968, após estudos sobre o córtex visual de primatas, foi postulada a ideia de uma estrutura hierárquica nos neurônios do sistema visual. No sistema visual, certas células formariam “campo receptivo” e enviaria esta informação para células em um nível superior. Desta forma, campos receptivos simples podem ser combinados para formar campos receptivos grandes e complexos (HUBEL; WIESEL, 1968). As redes neurais artificiais, também conhecidas como redes neurais, tiveram a sua analogia à forma pela qual o cérebro humano calcula: complexo, não linear e com funcionamento paralelo no processamento das informações (HAYKIN, 2008). A Figura 7 exemplifica o modelo de neurônio não linear.

A rede neural ainda conta com bias (b_k), cujo papel é aumentar ou diminuir a entrada líquida da função de ativação φ (a letra grega fi), de modo que a saída y do neurônio k é obtida através da equação abaixo:

$$y_k = \varphi(b_k + net_k)$$

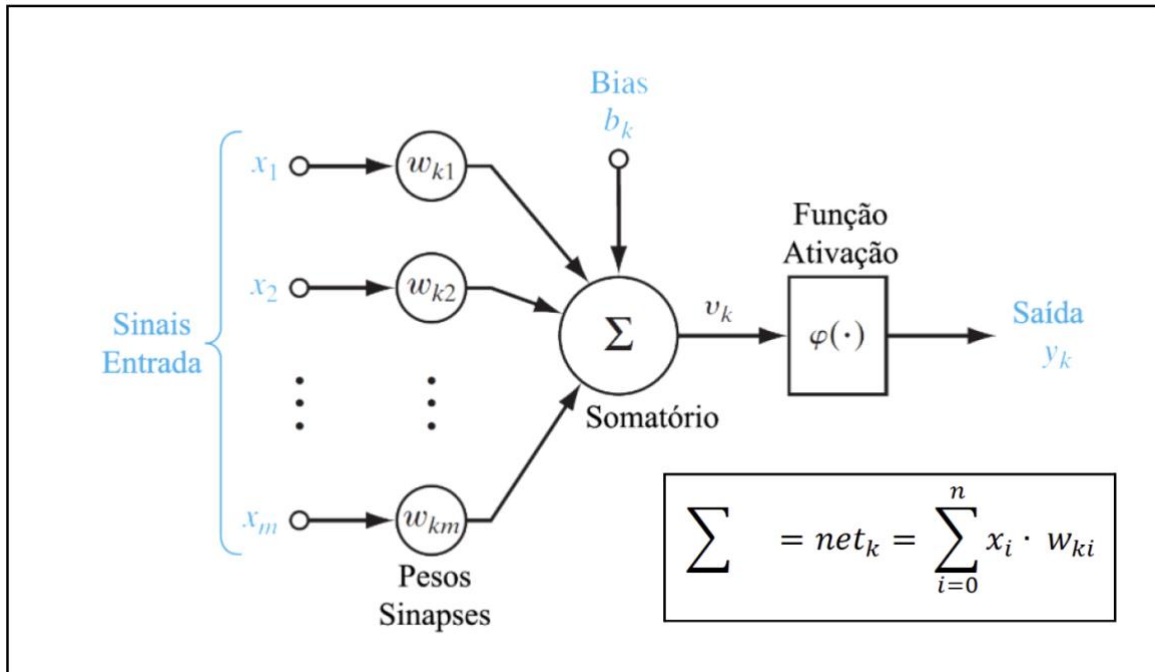


Figura 7. Modelo de neurônio não linear. Os sinais de entrada são x_1, x_2, \dots, x_m , bem como $w_{k1}, w_{k2}, \dots, w_{km}$ que recebem o peso das sinapses do neurônio k . O somatório da equação está representado no quadro dentro da figura.

Fonte: Adaptado de Haykin (HAYKIN, 2008)

O modelo de maior sucesso no contexto de reconhecimento de padrões são as redes neurais *Multilayer Perceptron* (MLP) feedforward (NASRABADI, 2007). Neste modelo, a capacidade de neurônios ocultos para extrair estatísticas de ordem mais elevada é particularmente valiosa quando o tamanho da camada de entrada é grande (HAYKIN, 2008). As três principais características das redes MLP são 1) o alto grau de conectividade utilizado, 2) a existência de uma ou mais camadas que estão escondidas dos nós de entrada e saída e 3) cada neurônio na rede inclui uma função de ativação não linear que é diferenciável (HAYKIN, 2008). Assim, uma rede neural com *Multilayer Perceptron* apresenta a configuração apresentada na Figura 8.

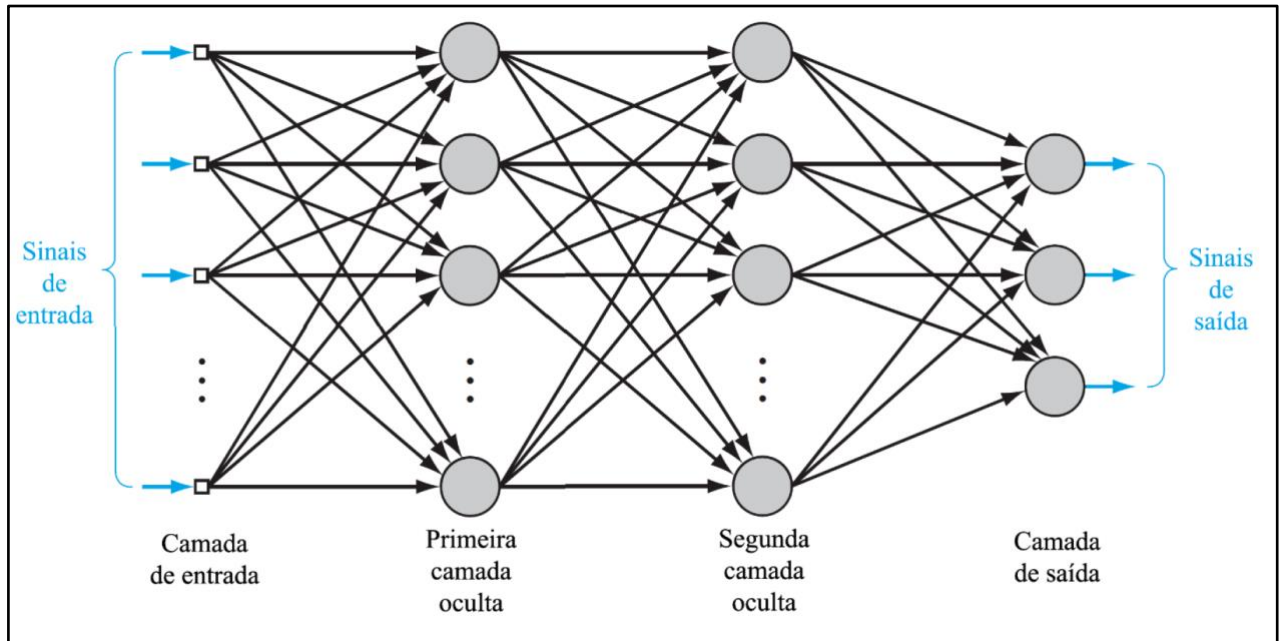


Figura 8. Modelo de rede neural Multilayer Perceptron. As camadas ocultas podem ser incluídas em cada camada, podendo possuir mais de um neurônio. aprendido. Um fator decisivo para o desempenho é que os neurônios em uma rede MLP estão conectados em camadas, com unidades de uma camada n passando as suas ativações apenas para neurônios da camada $n+1$. Logo, o processamento dos sinais de entrada nas múltiplas camadas significa que a análise da fonte de erro na camada de saída é extremamente complexa. Isto ocorre porque é o erro entre o valor esperado e o predito pela rede é facilmente calculado nos neurônios de saída, porém é bem mais difícil determinar o erro e qual é o neurônio responsável nas camadas ocultas.

Fonte: Adaptado de Haykin (HAYKIN, 2008)

As redes neurais convolucionais (do inglês, Convolutional Neural Networks - CNN) profundas foram desenvolvidas para o processamento e análise de imagens, vídeos e áudio (DONG; WANG; ABBAS, 2021). Este tipo de aprendizado de máquinas é considerado como o melhor algoritmo para o reconhecimento de imagens e o mais utilizado para análise de imagens médicas (SARVAMANGALA; KULKARNI, 2022).

O nome rede neural convolucional decorre do emprego de uma operação matemática chamada convolução, a qual é um tipo especializado de operação linear. A arquitetura das CNN apresenta 3 tipos de camadas: camada convolucional, camada de *pooling* e a camada totalmente conectada, conforme apresentado na Figura 9.

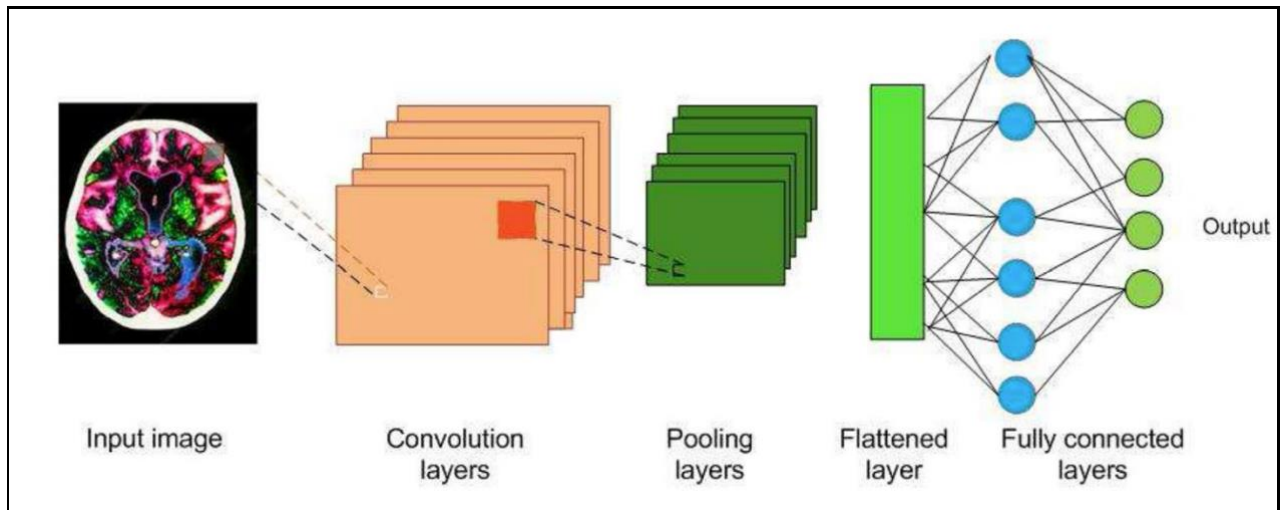


Figura 9. Arquitetura de uma rede neural convolucional, apresentando as camadas.
 Fonte: Sarvamangala et al. (SARVAMANGALA; KULKARNI, 2022)

A camada convolucional é modelada baseada nos neurônios. O objetivo desta camada é extrair características, como bordas, cores, textura e orientação de gradiente. As camadas convolucionais são feitas de filtros, ou *kernels*, de tamanho $n \times m \times d$, onde d é a profundidade da imagem. Durante a passagem direta, os kernels são convoluídos ao longo da largura e altura do volume de entrada e o produto escalar é calculado entre as entradas do filtro e a entrada. Intuitivamente, a CNN aprende filtros que são ativados quando se deparam com arestas, cores, textura e outras características. A saída da camada convolucional é alimentada em uma camada de função de ativação (SARVAMANGALA; KULKARNI, 2022).

A função de ativação serve para garantir a representação no espaço de entrada. Como os dados no mundo real são não lineares, as funções de ativação são usadas para transformá-los em dados lineares. Esta transformação, utilizada nas redes neurais, determina quais neurônios devem ser ativados ou não para determinada entrada. A função unidade linear retificada (ReLU) pega um número de valor real x e converte x em 0, se x for negativo. ReLU é a função não linear mais usada na CNN, leva menos tempo de computação e, portanto, é mais rápida e é expressa pela fórmula e Figura 10.

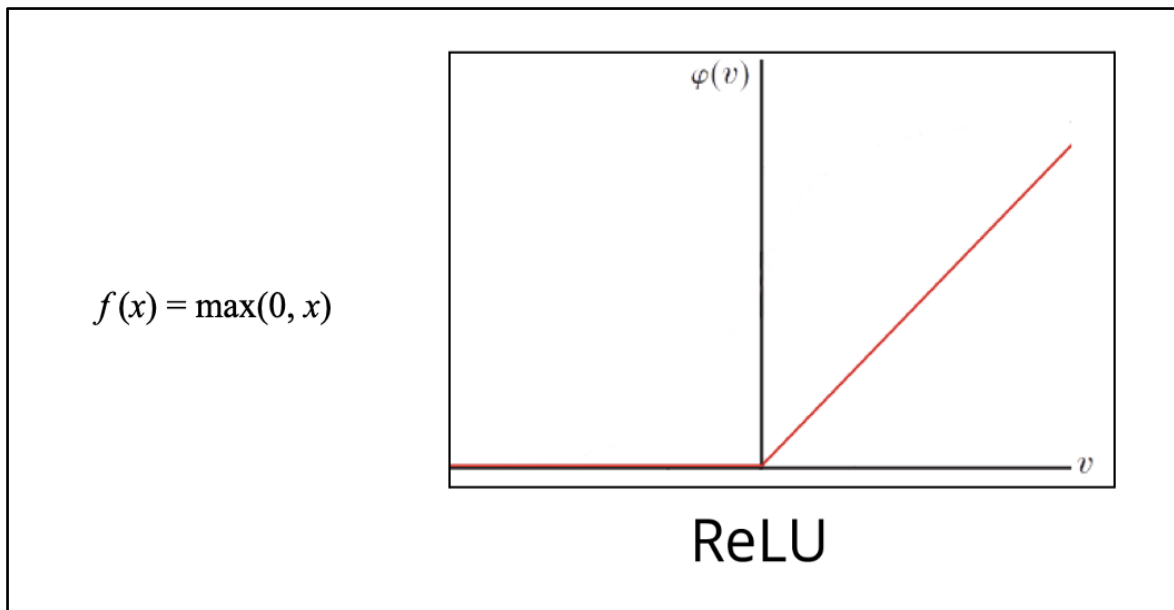


Figura 10. Representação gráfica da fórmula da função linear retificada.

Fonte: Sarvamangala et al. (SARVAMANGALA; KULKARNI, 2022)

A camada de *pooling* executa uma amostragem descendente não linear do recurso convoluto. Diminui o poder computacional necessário para processar os dados através da redução de dimensionalidade. O tamanho espacial agregando dados sobre o espaço ou tipo de recurso é reduzido, controla o *overfitting* e supera a variação de conversão e rotação de imagens. A operação de agrupamento resulta no particionamento de sua entrada em um conjunto de *patches* retangulares. Cada *patch* é substituído por um único valor, dependendo do tipo de agrupamento selecionado (SARVAMANGALA; KULKARNI, 2022).

A camada totalmente conectada é semelhante à rede neural artificial, onde cada nó possui conexões de entrada de todas as entradas, e todas as conexões possuem pesos associados a elas. A saída é a soma de todas as entradas multiplicadas pelos pesos correspondentes, conforme Figura 6. A camada FC é seguida por uma função de ativação e executa o trabalho do classificador para determinar como a imagem será classificada.

3. 1. Métricas para verificar o desempenho das redes neurais convolucionais.

Existem muitas métricas utilizadas para a avaliação do desempenho de algoritmos de compreensão de imagens médicas. A matriz de confusão, também conhecida como matriz de erro, é a tabela utilizada para visualizar o desempenho de um algoritmo e para o cálculo de diversas métricas de avaliação. Ela fornece uma visão sobre os tipos de erros cometidos pelo classificador. É uma matriz 2x2, onde as linhas representam os resultados reais e as colunas representam as instâncias dos resultados previstos do algoritmo. As Tabelas 6 e 7 explicam as métricas do desempenho.

Tabela 6. Matriz de confusão

Total das amostras	Resultado do algoritmo	
	Predição positiva	Predição negativa
caso verdadeiramente positivo	Verdadeiro Positivo (VP)	Falso negativo (FN)
caso verdadeiramente negativo	Falso Positivo (FP)	Verdadeiro Negativo (VN)

A fórmula dos casos é determinada pela equação $\Rightarrow \text{Total (T)} = \text{VP} + \text{VN} + \text{FP} + \text{FN}$.

Tabela 7. Métricas de avaliação do desempenho para o processamento de imagem.

Métrica	Interpretação	Fórmula	Valor desejado
Acurácia	No geral, com que frequência o resultado é verdadeiro	$\frac{VP + VN}{T}$	Alta (≈ 1)
Resultado de classificação incorreta (taxa de erro)	No geral, com que frequência o resultado é falso	$\frac{FP + FN}{T}$	Baixa (≈ 0)
Taxa de verdadeiro Positivo (TVP) ou (recall ou sensibilidade)	Quando é realmente verdadeiro, com que frequência ele prevê o verdadeiro	$\frac{VP}{FN + VP}$	Alta (≈ 1)
Taxa de Falso Positivo (TFP)	Quando é realmente falso, com que frequência ele prevê o verdadeiro	$\frac{FP}{FP + VN}$	Baixa (≈ 0)
Taxa de Falso Negativo (TFN)	Quando é realmente falso, com que frequência ele prevê falso	$\frac{FN}{FN + VP}$	Baixa (≈ 0)
Especificidade		$\frac{VN}{FP + VP}$	Alta (≈ 1)

Precisão (Valor Preditivo Positivo - VPP)	Quando prediz verdadeiro, com que frequência está correto?	$\frac{VP}{FP + VP}$	Alta (≈ 1)
F-Score (Coeficiente de similaridade de dados)	Média harmônica de recall e precisão.	$2 \times \frac{VPP \times T}{VPP + T}$	Alta (≈ 1)
Característica operacional do receptor (ROC)	Gráfico comumente usado para resumir o desempenho de um classificador em todos os limites possíveis.	Gráfico de recall por TVP	Alta (≈ 1)
Área da ROC (AUC)	Área da ROC (AUC)	Gráfico de TVP por TFP	Alta (≈ 1)

Fonte: Sarvamangala et al. (SARVAMANGALA; KULKARNI, 2022)

4. JUSTIFICATIVA

Há uma contradição entre os achados da expressão da proteína biglicano entre diferentes tipos de câncer. Por exemplo, no carcinoma esofágico epidermóide, o biglicano é 10x mais expresso no câncer, do que no tecido normal adjacente do mesmo caso (ZHU *et al.*, 2013). No carcinoma urotelial entre linhagens celulares, há uma correlação negativa entre expressão e proliferação do biglicano (APPUNNI *et al.*, 2019; SCHULZ *et al.*, 2019). A forma pela qual a análise é realizada também pode ser um motivo destas discrepâncias.

O Human Protein Atlas apresenta dois anticorpos validados para a investigação da expressão proteica do biglicano em amostras com imuno-histoquímica: HPA 003157 (Sigma-Aldrich) e o M01 (Abnova-CAB003678) (THE HUMAN PROTEIN ATLAS, 2022). O anticorpo da Abnova é monoclonal, enquanto o da Sigma-Aldrich é policlonal. A expressão da proteína é diferente de acordo com o anticorpo utilizado, conforme a Figura 11.

O único dado sobre a expressão do biglicano por imuno-histoquímica em mulheres com câncer foi identificado no trabalho de Leygue *et al.* neste estudo, o biglicano foi detectado apenas em níveis baixos em 2 dos 6 tecidos normais e em 3 dos 9 tumores mamários. O anticorpo utilizado para a análise não era comercial, e a análise da expressão era subjetiva, o que dificulta a reprodutibilidade dos achados (LEYGUE *et al.*, 2000).

Portanto, o presente estudo é justificado ao se propor a verificar a expressão da proteína biglicano por técnica adequada de imuno-histoquímica, utilizando um anticorpo monoclonal comercialmente disponível, e utilizando uma metodologia de avaliação da expressão com reduzida subjetividade

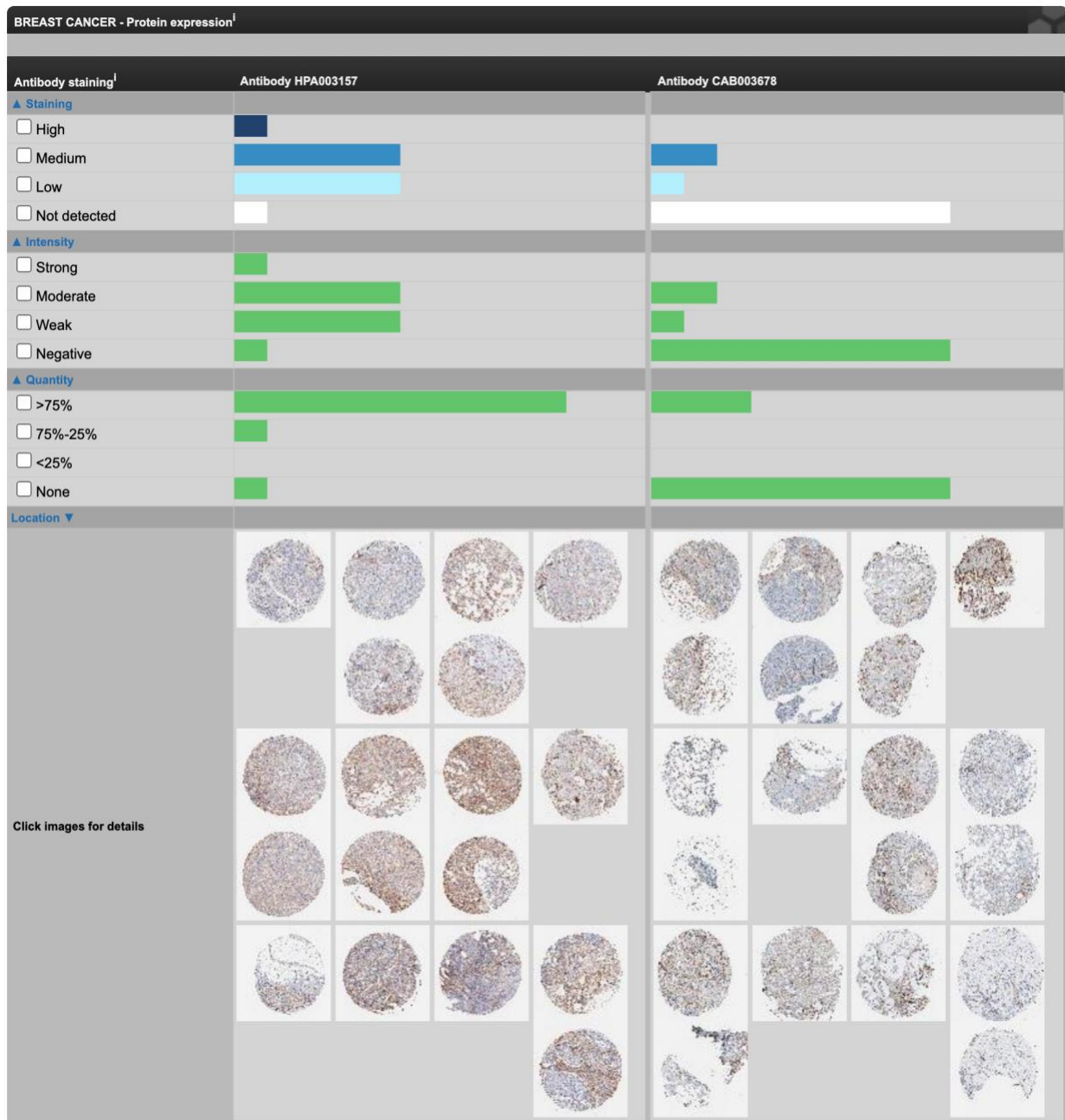


Figura 11. Expressão da proteína biglicano por imuno-histoquímica usando diferentes anticorpos no câncer de mama.

Antibody: anticorpo, High: Alto, Medium: média, Low: baixa, Not detected: não detectada, Staining: coloração, Quantity: quantidade, Intensity: intensidade, Location: localização.

Fonte: Human Protein Atlas (EXPRESSION OF BGN IN BREAST CANCER - THE HUMAN PROTEIN ATLAS, [s. d.]).

5. HIPÓTESES

H_0 : não existe diferença na média de expressão da proteína Biglicano entre tecido mamário com carcinoma ductal e tecido normal.

H_a : existe diferença na média de expressão da proteína Biglicano entre tecidos mamário com carcinoma ductal e tecido normal

6. OBJETIVOS

6.1. Objetivo principal

Comparar a expressão imuno-histoquímica da proteína biglicano, através da análise pelo programa ImageJ, entre tecido mamário normal e com carcinoma ductal de mama.

6.2. Objetivo secundário

Verificar a acurácia na identificação do de casos e controles, através da análise de redes neurais, utilizando somente a colocação obtida do DAB oriunda do Image J.

7. REFERÊNCIAS

- ANDRLOVÁ, Hana *et al.* Biglycan expression in the melanoma microenvironment promotes invasiveness via increased tissue stiffness inducing integrin- β 1 expression. **Oncotarget**, [s. l.], v. 8, n. 26, p. 42901–42916, 2017.
- APPUNNI, Sandeep *et al.* Altered expression of small leucine-rich proteoglycans (Decorin, Biglycan and Lumican): Plausible diagnostic marker in urothelial carcinoma of bladder. **Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine**, [s. l.], v. 39, n. 5, p. 1010428317699112, 2017.
- APPUNNI, Sandeep *et al.* Small Leucine Rich Proteoglycans (decorin, biglycan and lumican) in cancer. **Clinica chimica acta; international journal of clinical chemistry**, [s. l.], v. 491, p. 1–7, 2019.
- BARRETO-NETO, Nestor José de Souza *et al.* Perfil epidemiológico dos subtipos moleculares de carcinoma ductal da mama em população de pacientes em Salvador, Bahia. **Revista Brasileira de Mastologia**, [s. l.], v. 24, n. 4, p. 98–102, 2014.
- BEDIAGA, Naiara G. *et al.* Luminal B breast cancer subtype displays a dicotomic epigenetic pattern. **SpringerPlus**, [s. l.], v. 5, p. 623, 2016.
- BIANCO, P. *et al.* Expression and localization of the two small proteoglycans biglycan and decorin in developing human skeletal and non-skeletal tissues. **The journal of histochemistry and cytochemistry: official journal of the Histochemistry Society**, [s. l.], v. 38, n. 11, p. 1549–1563, 1990.
- BISCHOF, Ashley G. *et al.* Breast cancer normalization induced by embryonic mesenchyme is mediated by extracellular matrix biglycan. **Integrative biology: quantitative biosciences from nano to macro**, [s. l.], v. 5, n. 8, p. 1045–1056, 2013.
- BROWN, Monica *et al.* The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: the California Cancer Registry, 1999-2004. **Cancer**, [s. l.], v. 112, n. 4, p. 737–747, 2008.
- CAMERA, G.; STELLA, G.; CAMERA, A. New X linked spondyloepimetaphyseal dysplasia: report on eight affected males in the same family. **Journal of medical genetics**, [s. l.], v. 31, n. 5, p. 371–376, 1994.
- CÂNCER DE MAMA. [s. l.], 2018. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>. Acesso em: 18 out. 2019.
- CANCER GENOME ATLAS NETWORK. Comprehensive molecular portraits of human breast tumours. **Nature**, [s. l.], v. 490, n. 7418, p. 61–70, 2012.
- CHACÓN, Reinaldo D.; COSTANZO, María V. Triple-negative breast cancer. **Breast cancer research: BCR**, [s. l.], v. 12 Suppl 2, n. Suppl 2, p. S3, 2010.
- CHEN, Xiao-Dong *et al.* The small leucine-rich proteoglycan biglycan modulates BMP-4-induced osteoblast differentiation. **FASEB journal: official publication of the Federation of American Societies for Experimental Biology**, [s. l.], v. 18, n. 9, p. 948–958, 2004.
- CHO, Sung Yoon *et al.* BGN Mutations in X-Linked Spondyloepimetaphyseal Dysplasia. **American journal of human genetics**, [s. l.], v. 98, n. 6, p. 1243–1248, 2016.
- CHOI, H. U. *et al.* Characterization of the dermatan sulfate proteoglycans, DS-PGI and DS-PGII, from bovine articular cartilage and skin isolated by octyl-sepharose chromatography. **The Journal of biological chemistry**, [s. l.], v. 264, n. 5, p. 2876–2884, 1989.
- COCCO, Stefania *et al.* Biomarkers in Triple-Negative Breast Cancer: State-of-the-Art and Future Perspectives. **International journal of molecular sciences**, [s. l.], v. 21, n. 13, 2020. Disponível em:

<http://dx.doi.org/10.3390/ijms21134579>.

CONG, Li *et al.* Inhibition of stromal biglycan promotes normalization of the tumor microenvironment and enhances chemotherapeutic efficacy. **Breast cancer research: BCR**, [s. l.], v. 23, n. 1, p. 51, 2021.

DONG, S.; WANG, P.; ABBAS, K. A survey on deep learning and its applications. **Computer Science Review**, [s. l.], v. 40, p. 100379, 2021. Disponível em: Acesso em: 11 dez. 2022.

EDGE, Stephen B.; COMPTON, Carolyn C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. **Annals of surgical oncology**, [s. l.], v. 17, n. 6, p. 1471–1474, 2010.

ESKO, Jeffrey D.; KIMATA, Koji; LINDAHL, Ulf. Proteoglycans and Sulfated Glycosaminoglycans. *Em: ESSENTIALS OF GLYCOBIOLOGY. 2ND EDITION.* [S. l.]: Cold Spring Harbor Laboratory Press, 2009. *E-book*. Disponível em: Acesso em: 11 dez. 2022.

EXPRESSION OF BGN IN BREAST CANCER - THE HUMAN PROTEIN ATLAS. [S. l.], [s. d.]. Disponível em: <https://www.proteinatlas.org/ENSG00000182492-BGN/pathology/breast+cancer#Quantity>. Acesso em: 11 dez. 2022.

FISHER, L. W. *et al.* Human biglycan gene. Putative promoter, intron-exon junctions, and chromosomal localization. **The Journal of biological chemistry**, [s. l.], v. 266, n. 22, p. 14371–14377, 1991.

FITZGERALD, Jenny *et al.* Future of biomarker evaluation in the realm of artificial intelligence algorithms: application in improved therapeutic stratification of patients with breast and prostate cancer. **Journal of clinical pathology**, [s. l.], v. 74, n. 7, p. 429–434, 2021.

FRAGOMENI, Simona Maria; SCIALLIS, Andrew; JERUSS, Jacqueline S. Molecular Subtypes and Local-Regional Control of Breast Cancer. **Surgical oncology clinics of North America**, [s. l.], v. 27, n. 1, p. 95–120, 2018.

GENECARDS HUMAN GENE DATABASE. **BGN Gene - GeneCards**. [S. l.], 2022. Disponível em: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=BGN>. Acesso em: 10 dez. 2022.

GEYER, Felipe C.; MARCHIO, Caterina; REIS-FILHO, Jorge S. The role of molecular analysis in breast cancer. **Pathology**, [s. l.], v. 41, n. 1, p. 77–88, 2009.

GU, Xiaohu *et al.* Up-regulated biglycan expression correlates with the malignancy in human colorectal cancers. **Clinical and experimental medicine**, [s. l.], v. 12, n. 3, p. 195–199, 2012.

HALPER, Jaroslava. Proteoglycans and diseases of soft tissues. **Advances in experimental medicine and biology**, [s. l.], v. 802, p. 49–58, 2014.

HAMILTON, Erika *et al.* Targeting HER2 heterogeneity in breast cancer. **Cancer treatment reviews**, [s. l.], v. 100, p. 102286, 2021.

HAYKIN, Simon S. What is Neural Network?. *Em: HAYKIN, Simon S. (org.). Neural Networks and Learning Machines.* 3. ed. New York: Prentice Hall, 2008. p. 1–6.

HU, Lei *et al.* Biglycan stimulates VEGF expression in endothelial cells by activating the TLR signaling pathway. **Molecular oncology**, [s. l.], v. 10, n. 9, p. 1473–1484, 2016.

HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. **The Journal of physiology**, [s. l.], v. 195, n. 1, p. 215–243, 1968.

INCA. **Atlas da mortalidade**. [S. l.], 2022. Disponível em: <https://www.inca.gov.br/app/mortalidade>. Acesso em: 10 dez. 2022.

IOZZO, Renato V.; SCHAEFER, Liliana. Proteoglycan form and function: A comprehensive nomenclature

- of proteoglycans. **Matrix biology: journal of the International Society for Matrix Biology**, [s. l.], v. 42, p. 11–55, 2015.
- JACOBSEN, Frank *et al.* Up-regulation of Biglycan is Associated with Poor Prognosis and PTEN Deletion in Patients with Prostate Cancer. **Neoplasia**, [s. l.], v. 19, n. 9, p. 707–715, 2017.
- JOHNSTONE, B. *et al.* Identification and characterization of glycanated and non-glycanated forms of biglycan and decorin in the human intervertebral disc. **Biochemical Journal**, [s. l.], v. 292 (Pt 3), n. Pt 3, p. 661–666, 1993.
- KÄHÄRI, V. M.; LARJAVA, H.; UITTO, J. Differential regulation of extracellular matrix proteoglycan (PG) gene expression. Transforming growth factor-beta 1 up-regulates biglycan (PGI), and versican (large fibroblast PG) but down-regulates decorin (PGII) mRNA levels in human fibroblasts in culture. **The Journal of biological chemistry**, [s. l.], v. 266, n. 16, p. 10608–10615, 1991. Disponível em: Acesso em: 10 dez. 2022.
- KIM, Dong Hyun *et al.* High excision repair cross-complementation group 1 expression is associated with favorable prognostic factors in breast cancer. **Oncology letters**, [s. l.], v. 14, n. 4, p. 4995–5003, 2017.
- KIMOTO, S. *et al.* The effect of glucocorticoid on the synthesis of biglycan and decorin in human osteoblasts and bone marrow stromal cells. **Endocrinology**, [s. l.], v. 135, n. 6, p. 2423–2431, 1994.
- KOBE, B.; KAJAVA, A. V. The leucine-rich repeat as a protein recognition motif. **Current opinion in structural biology**, [s. l.], v. 11, n. 6, p. 725–732, 2001.
- KRISHNAMURTI, Uma; SILVERMAN, Jan F. HER2 in breast cancer: a review and update. **Advances in anatomic pathology**, [s. l.], v. 21, n. 2, p. 100–107, 2014.
- LEHMANN, Brian D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. **The Journal of clinical investigation**, [s. l.], v. 121, n. 7, p. 2750–2767, 2011.
- LEYGUE, E. *et al.* Lumican and decorin are differentially expressed in human breast carcinoma. **The Journal of pathology**, [s. l.], v. 192, n. 3, p. 313–320, 2000.
- LI, Zhi-Hua *et al.* Luminal B breast cancer: patterns of recurrence and clinical outcome. **Oncotarget**, [s. l.], v. 7, n. 40, p. 65024–65033, 2016.
- LIU, Yang *et al.* Expression and significance of biglycan in endometrial cancer. **Archives of gynecology and obstetrics**, [s. l.], v. 289, n. 3, p. 649–655, 2014.
- LOPEZ-GARCIA, Maria A. *et al.* Breast cancer precursors revisited: molecular features and progression pathways. **Histopathology**, [s. l.], v. 57, n. 2, p. 171–192, 2010.
- MAISHI, Nako *et al.* Tumour endothelial cells in high metastatic tumours promote metastasis via epigenetic dysregulation of biglycan. **Scientific reports**, [s. l.], v. 6, p. 28039, 2016.
- MANUPATI, Kanakaraju *et al.* Biglycan Promotes Cancer Stem Cell Properties, NFκB Signaling and Metastatic Potential in Breast Cancer Cells. **Cancers**, [s. l.], v. 14, n. 2, 2022. Disponível em: <http://dx.doi.org/10.3390/cancers14020455>.
- MCBRIDE, O. W.; FISHER, L. W.; YOUNG, M. F. Localization of PGI (biglycan, BGN) and PGII (decorin, DCN, PG-40) genes on human chromosomes Xq13-qter and 12q, respectively. **Genomics**, [s. l.], v. 6, n. 2, p. 219–225, 1990.
- MEESTER, Josephina A. N. *et al.* Loss-of-function mutations in the X-linked biglycan gene cause a severe syndromic form of thoracic aortic aneurysms and dissections. **Genetics in medicine: official journal of the American College of Medical Genetics**, [s. l.], v. 19, n. 4, p. 386–395, 2017.

- MERRY, Catherine L. R. *et al.* Proteoglycans and Sulfated Glycosaminoglycans. *Em: ESSENTIALS OF GLYCOBIOLOGY [INTERNET]. 4TH EDITION.* [s. l.]: Cold Spring Harbor Laboratory Press, 2022. *E-book*. Disponível em: Acesso em: 11 dez. 2022.
- MORETH, Kristin *et al.* Biglycan-triggered TLR-2- and TLR-4-signaling exacerbates the pathophysiology of ischemic acute kidney injury. **Matrix biology: journal of the International Society for Matrix Biology**, [s. l.], v. 35, p. 143–151, 2014.
- MUELLER, Claudius *et al.* Protein biomarkers for subtyping breast cancer and implications for future research. **Expert review of proteomics**, [s. l.], v. 15, n. 2, p. 131–152, 2018.
- NABA, Alexandra *et al.* The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. **Molecular & cellular proteomics: MCP**, [s. l.], v. 11, n. 4, p. M111.014647, 2012.
- NASH, M. A.; LOERCHER, A. E.; FREEDMAN, R. S. In vitro growth inhibition of ovarian cancer cells by decorin: synergism of action between decorin and carboplatin. **Cancer research**, [s. l.], v. 59, n. 24, p. 6192–6196, 1999.
- NASRABADI, Nasser M. Book Review: Pattern Recognition and Machine Learning. **Journal of Electronic Imaging**, [s. l.], v. 16, n. 4, p. 049901, 2007. Disponível em: Acesso em: 11 dez. 2022.
- NASTASE, Madalina V.; YOUNG, Marian F.; SCHAEFER, Liliana. Biglycan: a multivalent proteoglycan providing structure and signals. **The journal of histochemistry and cytochemistry: official journal of the Histochemistry Society**, [s. l.], v. 60, n. 12, p. 963–975, 2012.
- NIEDWOROK, Christian *et al.* Inhibitory role of the small leucine-rich proteoglycan biglycan in bladder cancer. **PLoS one**, [s. l.], v. 8, n. 11, p. e80084, 2013.
- NIKITOVIC, Dragana *et al.* The biology of small leucine-rich proteoglycans in bone pathophysiology. **The Journal of biological chemistry**, [s. l.], v. 287, n. 41, p. 33926–33933, 2012.
- OLIVER, Constance; JAMUR, Maria Célia. Immunocytochemical methods and protocols. **Methods in molecular biology**, [s. l.], v. 588, p. iv – v, 2010.
- PEROU, C. M. *et al.* Molecular portraits of human breast tumours. **Nature**, [s. l.], v. 406, n. 6797, p. 747–752, 2000.
- PRAT, Aleix *et al.* Molecular features and survival outcomes of the intrinsic subtypes within HER2-positive breast cancer. **Journal of the National Cancer Institute**, [s. l.], v. 106, n. 8, 2014. Disponível em: <http://dx.doi.org/10.1093/jnci/dju152>.
- ROBERTSON, Stephanie *et al.* Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. **Translational research: the journal of laboratory and clinical medicine**, [s. l.], v. 194, p. 19–35, 2018.
- ROUGHLEY, P. J. *et al.* Non-proteoglycan forms of biglycan increase with age in human articular cartilage. **Biochemical Journal**, [s. l.], v. 295 (Pt 2), n. Pt 2, p. 421–426, 1993.
- RUSSNES, Hege G. *et al.* Breast Cancer Molecular Stratification. **The American journal of pathology**, [s. l.], v. 187, n. 10, p. 2152–2162, 2017.
- SANTUARIO-FACIO, Sandra Karina *et al.* A New Gene Expression Signature for Triple Negative Breast Cancer Using Frozen Fresh Tissue before Neoadjuvant Chemotherapy. **Molecular medicine**, [s. l.], v. 23, p. 101–111, 2017.
- SARVAMANGALA, D. R.; KULKARNI, Raghavendra V. Convolutional neural networks in medical image understanding: a survey. **Evolutionary intelligence**, [s. l.], v. 15, n. 1, p. 1–22, 2022.

SCHAEFER, Liliana *et al.* Proteoglycan neofunctions: regulation of inflammation and autophagy in cancer biology. **The FEBS journal**, [s. l.], v. 284, n. 1, p. 10–26, 2017.

SCHAEFER, Liliana; IOZZO, Renato V. **Biological Functions of the Small Leucine-rich Proteoglycans: From Genetics to Signal Transduction**. [S. l.: s. n.], 2008. Disponível em: <http://dx.doi.org/10.1074/jbc.r800020200>.

SCHULZ, Gerald B. *et al.* Prognostic value and association with epithelial-mesenchymal transition and molecular subtypes of the proteoglycan biglycan in advanced bladder cancer. **Urologic oncology**, [s. l.], v. 37, n. 8, p. 530.e9–e530.e18, 2019.

SIEGEL, Rebecca L. *et al.* Cancer Statistics, 2021. **CA: a cancer journal for clinicians**, [s. l.], v. 71, n. 1, p. 7–33, 2021.

SØRLIE, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 98, n. 19, p. 10869–10874, 2001.

SORLIE, Therese *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 100, n. 14, p. 8418–8423, 2003.

STAAF, Johan *et al.* Identification of subtypes in human epidermal growth factor receptor 2--positive breast cancer reveals a gene signature prognostic of outcome. **Journal of clinical oncology: official journal of the American Society of Clinical Oncology**, [s. l.], v. 28, n. 11, p. 1813–1820, 2010.

STRING INTERACTION NETWORK. [S. l.], 2022. Disponível em: <https://string-db.org/cgi/network?taskId=btSNhzNKHCCXX&sessionId=b4vKuIornPpq>. Acesso em: 10 dez. 2022.

SUBBARAYAN, Karthikeyan *et al.* Biglycan-mediated upregulation of MHC class I expression in HER-2/neu-transformed cells. **Oncoimmunology**, [s. l.], v. 7, n. 4, p. e1373233, 2018.

SUHOVSKIH, Anastasia V. *et al.* Proteoglycans as potential microenvironmental biomarkers for colon cancer. **Cell and tissue research**, [s. l.], v. 361, n. 3, p. 833–844, 2015.

SUNG, Hyuna *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA: a cancer journal for clinicians**, [s. l.], v. 71, n. 3, p. 209–249, 2021.

THE HUMAN PROTEIN ATLAS. **BGN - Antibodies - The Human Protein Atlas**. [S. l.], 2022. Disponível em: <https://www.proteinatlas.org/ENSG00000182492-BGN/summary/antibody>. Acesso em: 11 dez. 2022.

THEOCHARIS, Achilleas D. *et al.* Insights into the key roles of proteoglycans in breast cancer biology and translational medicine. **Biochimica et biophysica acta**, [s. l.], v. 1855, n. 2, p. 276–300, 2015.

TUFVESSON, Ellen *et al.* Biglycan isoforms with differences in polysaccharide substitution and core protein in human lung fibroblasts. **European journal of biochemistry / FEBS**, [s. l.], v. 269, n. 15, p. 3688–3696, 2002.

UNGEFROREN, H.; KRULL, N. B. Transcriptional regulation of the human biglycan gene. **The Journal of biological chemistry**, [s. l.], v. 271, n. 26, p. 15787–15795, 1996.

VODUC, K. David *et al.* Breast cancer subtypes and the risk of local and regional relapse. **Journal of clinical oncology: official journal of the American Society of Clinical Oncology**, [s. l.], v. 28, n. 10, p. 1684–1691, 2010.

WANG, Bin *et al.* Biglycan expression correlates with aggressiveness and poor prognosis of gastric cancer. **Experimental biology and medicine**, [s. l.], v. 236, n. 11, p. 1247–1253, 2011.

WEBER, C. K. *et al.* Biglycan is overexpressed in pancreatic cancer and induces G1-arrest in pancreatic cancer cell lines. **Gastroenterology**, [s. l.], v. 121, n. 3, p. 657–667, 2001.

WEI, Qingshen; ZHANG, Yuntao. Characteristics of Small Leucine-rich Proteoglycans in the Intervertebral Disc Degeneration. **Anatomy Physiology & biochemistry international journal**, [s. l.], v. 4, n. 3, p. 1–6, 2018. Disponível em: Acesso em: 10 dez. 2022.

WHO. **Breast cancer**. [S. l.], 2021. Disponível em: <https://www.who.int/news-room/factsheets/detail/breast-cancer>. Acesso em: 30 jun. 2022.

XU, Ding *et al.* Proteins That Bind Sulfated Glycosaminoglycans. *Em: ESSENTIALS OF GLYCOBIOLOGY [INTERNET]. 4TH EDITION.* [S. l.]: Cold Spring Harbor Laboratory Press, 2022. *E-book*. Disponível em: Acesso em: 10 dez. 2022.

YANG, Qi *et al.* Ultrasonographic features of triple-negative breast cancer: a comparison with other breast cancer subtypes. **Asian Pacific journal of cancer prevention: APJCP**, [s. l.], v. 16, n. 8, p. 3229–3232, 2015.

YANG, Guang; ROTHRAUFF, Benjamin B.; TUAN, Rocky S. Tendon and Ligament Regeneration and Repair: Clinical Relevance and Developmental Paradigm. **Birth defects research. Part C, Embryo today: reviews**, [s. l.], v. 99, n. 3, p. 203, 2013. Disponível em: Acesso em: 23 set. 2022.

ZHAO, Shou-Feng *et al.* Biglycan as a potential diagnostic and prognostic biomarker in multiple human cancers. **Oncology letters**, [s. l.], v. 19, n. 3, p. 1673–1682, 2020.

ZHU, Y. H. *et al.* High expression of biglycan is associated with poor prognosis in patients with esophageal squamous cell carcinoma. **International journal of clinical and experimental pathology**, [s. l.], v. 6, n. 11, 2013. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/24228112/>. Acesso em: 21 set. 2022.

8. ARTIGO EM INGLÊS

Artigo submetido para a revista PLOS ONE

Deep Learning Neural Network Image Analysis Of Immunohistochemical Protein Expression Reveals A Significantly Reduced Expression of Biglycan in Breast Cancer

Ana Paula Thiesen¹, Bruna Mielczarski², Ricardo Francalacci Savaris^{1,2}

¹Universidade Federal do Rio Grande do Sul, Postgraduate Program in Health Science: Surgical Sciences, Porto Alegre, Rio Grande do Sul, Brazil, 90035-003

² Department of Obstetrics and Gynecology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, 90035-003

*Corresponding author:

Prof. Ricardo Francalacci Savaris, MD, MSc, PhD

Department of Gynecology and Obstetrics, Department of Medicine

Federal University of Rio Grande do Sul

Rua Ramiro Barcelos 2350/1125

Porto Alegre, RS Brazil

90035-903

Phone: +55 51 33598117

e-mail: rsavaris@hcpa.edu.br

Abstract

New breast cancer biomarkers have been sought for better tumor characterization and treatment. Among these putative markers, there is Biglycan (BGN). BGN is a class I small leucine-rich proteoglycan family of proteins characterized by a protein core with leucine-rich repeats. The objective of this study is to compare the protein expression of BGN in breast tissue with and without cancer, using immunohistochemical technique associated with digital histological score (D-HScore) and supervised deep learning neural networks (SDLNN).

In this case-control study, 24 formalin-fixed, paraffin-embedded tissues were obtained for analysis. Normal (n=9) and cancerous (n=15) tissue sections were analyzed by immunohistochemistry using BGN monoclonal antibody (M01-Abnova) and 3,3'-Diaminobenzidine (DAB) as the chromogen. Photomicrographs of the slides were analysed with D-HScore, using arbitrary DAB units. Another set (n=129) with higher magnification without ROI selection, was submitted to the inceptionV3 deep neural network image embedding recognition model. Next, supervised neural network analysis, using stratified 20 fold cross validation, with 200 hidden layers, ReLu activation, and regularization at $\alpha=0.0001$ were applied for SDLNN. The sample size was calculated for a minimum of 7 cases and 7 controls, having a power = 90%, an α error = 5%, and a standard deviation of 20, to identify a decrease from the average of 40 DAB units (control) to 4 DAB units in cancer.

BGN expression in DAB units [median (range)] was 6.2 (0.8 to 12.4) and 27.31 (5.3 to 81.7) in cancer and normal breast tissue, respectively, using D-HScore ($p=0.0017$, Mann-Whitney test). SDLNN classification accuracy was 85.3% (110 out of 129; 95%CI=78.1% to 90.3%).

BGN protein expression is reduced in breast cancer tissue, compared to normal tissue.

Keywords: breast cancer; Biglycan ; immunohistochemistry, deep learning

Introduction

In women, breast cancer is the most commonly diagnosed cancer in the world and the leading cause of death from cancer [1]. Breast cancers are heterogeneous in nature, both at the histological and molecular levels, and the molecular profiling of breast cancer guides diagnostic and therapeutic strategies for the disease [2]. New breast cancer biomarkers have been sought in order to better characterize tumors and to select the best possible treatment [3]. Small leucine-rich proteoglycans (SLRPs), a diverse sub-group of proteoglycans, are involved in matrix organization and the regulation of cell growth and signaling [4]. Biglycan (BGN) is one SLRP whose gene has been mapped to the Xq28 chromosome [5]. BGN is a class I SLRP characterized by a protein core with leucine-rich repeats and is composed of 331 amino acids and a molecular weight of 42 kDa [6]. The molecular weight of BGN increases up to 100-250 kDa, when fully glycosylated. This glycosylation is due to the presence of two chondroitin/dermatan sulfate and glycosaminoglycan chains covalently attached to the N-terminal region [7]. The glycosaminoglycan chains consist of repeating disaccharide units of either chondroitin sulfate or dermatan sulfate and are attached to the core protein via an O-linked glycosidic bond. This proteoglycan is ubiquitously expressed; it can be incorporated into the extracellular matrix (ECM) or exist in the blood in its soluble form under certain disease conditions [8,9].

BGN could alter tumor proliferation by modulating the receptors and cellular expression molecules within the tumor microenvironment [9]. Zhao *et al.*, using a cancer microarray database and a web-based data-mining platform (Oncomine), have reported that BGN gene expression was upregulated in breast and other cancers [10]. However, the clinical impact of BGN on cancer is still poorly understood and sometimes contradictory. For instance, in bladder cancer, silencing of BGN resulted in enhanced tumor cell proliferation, indicating that BGN acts as a growth suppressor in this disease [11], while another study using animal model found that the inhibition of stromal BGN promoted normalization of the tumor microenvironment and enhanced chemotherapeutic efficacy in mice

injected with breast cancer cells [12]. Bischof *et al.* demonstrated that the injection of normal, early stage, embryonic mesenchyme cells was sufficient to induce differentiation and suppress growth of mouse mammary tumor epithelial cells both *in vitro* and *in vivo*; they reported that BGN was required for tumor normalization [13].

These apparent contradictions may be explained by the different types of models used, i.e., *in vitro*, animal model, and detection method (mRNA, or immunohistochemistry). The type of antibodies used in immunohistochemistry and the specific site for BGN identification may also explain some of these discrepancies. For instance, mature and functionally active BGN protein was detected using a polyclonal antibody after glycosaminoglycan removal by enzymatic digestion with chondroitinase ABC [14]. This selective removal was performed, because the presence of a complex of two chondroitin/dermatan sulfates with glycosaminoglycan chains could hinder antibody binding, leading to the misinterpretation of results [14]. Nevertheless, this polyclonal antibody was discontinued, making new studies with this antibody unsuitable. These differences are more evident when comparisons between different antibodies are applied in the same tissue. For instance, the [Human Protein Atlas database](#) [15] has two antibodies validated against BGN to be used in immunohistochemistry: HPA003157 (Sigma, Aldrich), a polyclonal antibody that targets 140 amino acids, and H00000633-M01 (Abnova, Taipei, Taiwan), and a monoclonal antibody against the full sequence of BGN (368 amino acids); the protein expression of BGN in breast cancer is different between these two antibodies. While the former had 8.3% negative expression (1 out of 12 cases), the latter had a negative expression in 75% (9 out of 12 cases) [16]. In addition, manual evaluation of immunohistochemically stained specimens is a subjective and highly individual task, which, as has been reported by others, depends on intra- and inter-observer variability [17].

To reduce the highly subjective and easily biased nature of these tasks, the digital histological scoring method (D-HSCORE) has been reported [18]. Another emerging area in image analysis is

deep learning. Deep learning (DL) is a form of machine learning that relies on both supervised and unsupervised learning; DL applied to digital pathology uses artificial neural networks (ANN) to determine if the output or interpretation of a digital image is correct [19]. ANN uses multiple layers of calculations imitating the complex network of neurons in the human brain to analyze this complex data [19].

Data on BGN expression in human breast cancer is scant and contradictory [13,20], and little information is available for human BGN protein expression *in vivo* using an antibody against the full length of the protein. Therefore, the objective of this study is to verify the immunohistochemical expression of BGN in breast cancer biopsies compared to normal breast tissue using a validated monoclonal antibody and two digital imaging methods of analysis: D-HSCORE and deep learning neural network image analysis.

Material and Methods

Material and Methods

Ethics statement

This study was submitted and approved by Hospital de Clínicas de Porto Alegre Ethical Review Board, under the approval number 2019/0337 and registered at *Plataforma Brasil* under the certificate of submission for ethical analysis (CAAE 15329119.9.0000.5327).

Study design and setting

In this case-control study, paraffin blocks were obtained from the pathological archive of Hospital de Clínicas in Porto Alegre, Brazil. Slides were dated between January 1st, 2012, and December 30th, 2015. The original pathological report was reviewed by a certified board pathologist to confirm the diagnosis of benign and cancerous breast tissue. The study was conducted between May 20, 2019, and July 30, 2020.

Patients and methods

Women with diagnoses of invasive ductal carcinoma and those who underwent breast surgery for benign conditions (e.g., mammoplasty, benign mammary cyst) were included in the sample. Patients with lobular carcinoma or intraductal papilloma who had undergone chemotherapy or radiotherapy and aged below 20- and over 79-years-old were excluded. These cases were excluded since chemo and radiotherapy may change protein expression of the tumor.

Variables

BGN protein expression was the primary continuous variable, i.e., DAB units, varying between 0 and 255 units.

Healthy breast tissue (benign - control group) and breast cancer tissue (cancer) were categorical data. Other variables were age and ethnicity. Estrogen and progesterone receptor status were described in cancerous tissues, along with human epidermal growth factor receptor 2 (HER2), Nottingham Grading and tumor staging.

Data sources / measurement

Immunohistochemistry

Immunohistochemistry methodology was performed according to standard technique [21] and as previously reported by our group with minor modifications [22]. Modifications included deparaffinized at 75°C for two hours, followed by xylol rinse, rehydrated in successive steps of ethanol, water, and phosphate-buffered saline solution (PBS); slide incubation was done in sodium citrate solution, pH 6 at 90°C for 45 min. for antigen retrieval; primary antibody against the full length of recombinant BGN was BGN monoclonal antibody (M01), clone 4E1-1G7, IgG2a kappa (Abnova, Taipei, Taiwan). It was used at dilution 1:1000 at pH 6 and incubated overnight. After 2 x 5 minutes in a PBS rinse, secondary antibody anti-mouse IgG (whole molecule), namely

peroxidase antibody produced in rabbit (A9044, Sigma-Aldrich, Darmstadt, Germany), was incubated for 90 minutes at 22°C in the same chamber using a dilution of 1:1000. The primary antibody visualization and counterstained were performed as previously reported [22]. Negative controls were obtained by replacing the primary antibody with mouse IgG2a, kappa monoclonal [18C8BC7AD10] - Isotype Control (ab170191) - Abcam, Cambridge, UK). Human lung cancer samples were used as external positive control. These procedures followed the REporting recommendations for tumor MARKer prognostic studies (REMARK) guidelines [23].

Images from stained sections were obtained using an optical microscope (Olympus BX51 microscope; Olympus Optical Co., Tokyo, Japan) with a 40x objective U Plan Fluorite dry objective (numerical aperture 0.65 mm, Olympus). A digital color camera (Olympus DP73; OM Digital Solutions Co., Tokyo, Japan) captured digital images, at a size of 4800 x 3600 pixels (resolution: 1 mm = 6000 pixels), under standard conditions for ImageJ analysis. Another set of images (n=129 photomicrographs) with a 100x objective (UPLFL 100x; Oil Immersion, Olympus) were taken of entire slides for supervised deep learning neural network analysis.

Image analysis with ImageJ

Photomicrographs were coded and blindly analyzed using Digital HSCORE (D-HSCORE) as previously reported [18,24,25]. Briefly, only the glandular and tumor sites of the tissue sections were selected as regions of interest (ROI). After selecting the ROI, images were submitted for “color deconvolution” analysis. The image with DAB staining was used for analysis.

Supervised deep learning neural network

The 129 photomicrographs of DAB-only images, with 100x magnification and without ROI selection, were submitted to the inceptionV3 deep neural network image embedding recognition model using [Orange](#) 3.31.0 software (University of Ljubljana, Slovenia). Next, supervised neural

network analysis (SDLNN), using stratified 20-fold cross validation with 200 hidden layers, ReLu activation, and regularization at $\alpha = 0.0001$, were submitted to SDLNN, in Orange software.

Bias

Bias was reduced by using D-HSCORE and SDLNN.

Study size

The sample size for ImageJ analysis was calculated according to the literature [26] in order to have a power = 90%, an α error = 5%, and a standard deviation of 20, to identify a decrease from the average of 40 DAB units (control) to 4 DAB units in cancer. With these figures, at least 7 samples in each group were necessary.

Sample size for supervised neural network analysis was chosen for convenience after obtaining the maximum number of photomicrographs from the slides.

Quantitative variables

The average DAB units intensity, derived from up to three images obtained from color deconvolution, was calculated according to the formula: $f=255-i$, where f =final DAB intensity, and i =mean DAB intensity obtained from the software, as previously described [18].

Statistical analysis

Groups, with categorical data, were compared using Fisher's exact test. Continuous data of BGN expression, in arbitrary DAB units, between groups, were compared using unpaired Student t -test with Welch's correction if data had a Gaussian distribution and different SDs, otherwise Mann-Whitney test was used. D'Agostino & Pearson omnibus normality test was used to verify Gaussian distribution. These analyses were performed using [GraphPad Prism](#) version 9.3.1 for Macintosh (GraphPad Software Inc. San Diego, CA).

Deep neural network image embedding recognition model analysis was performed using [Orange](#) 3.31.0 software (University of Ljubljana, Slovenia). Supervised neural network was set with 200 hidden layers, the rectified linear activation function (ReLU) was used for activation, Adam was used as the optimization algorithm [27], regularization was set at $\alpha = 0.0001$, and maximum number of iterations was 300, with training replication. Test and training were set using a stratified 20-fold cross validation.

Results

Participants and descriptive data

A total of 24 samples were obtained for the study: benign = 9; cancerous = 15. The mean age between groups was not significant ($p = 0.1$; unpaired Student t -test). Details of the sampled population are depicted in Table 1.

Table 1. Characteristics of the studied population

Characteristics	Breast Cancer (n=15)	Controls (n=9)	<i>p</i>
Age (years) mean (SD)	52(11.2)	43(11.6)	0.1 ^a
Ethnicity			
caucasian/non-caucasian	14/1	8/1	1 ^b
Pathology Report (n)			
Invasive ductal carcinoma	15		
Normal breast tissue		2	
normal with adenosis		2	
stromal fibrosis		4	
Fibro-microcystic changes		1	
Estrogen Receptor (positive/negative)	6/9		
Progesterone Receptor (positive/negative)	10/5		
HER2 (n)%			
0, +1	11 (73.4)		
+2	2 (13.3)		
+3	2 (13.3)		
Nottingham Grading (n)%			
I	1 (6.7)		
II	8 (53.3)		
III	6 (40)		
Staging - n (%)			
I	3 (20)		
IIA	5 (33.3)		
IIB	1 (6.7)		
IIIA	4 (26.7)		
IIIB	1 (6.7)		
IV	1 (6.7)		

^a Unpaired Student *t*-test with Welch's correction ^b Fisher's exact test

Outcome data

BGN protein expression in DAB units [median (range)] was significantly lower in breast tissue with cancer 6.2 (0.8 to 12.4) compared to benign breast tissue 27.31(5.3 to 81.7) as shown in Figure 1 (p=0.0017, Mann-Whitney test).

The area [median(range) in mm²] analyzed between benign tissue [0.67(0.39 to 4.6)] and cancer tissue [1.1(0.08 to 2.8)] was not significantly different (p=0.48; Mann-Whitney U test).

Supervised neural network analysis

A total of 129 high-power magnification (100x) photomicrographs with DAB staining only and derived from ImageJ (benign, n=69; cancer, n=60) were submitted for supervised neural network analysis (Figure 2).

The performance of the supervised neural network analysis yielded an area under the curve of 94.3%, further details are depicted in Table 2.

Table 2. Performance of supervised neural network analysis for classifying samples as benign or cancerous based on DAB expression only.

confusion matrix		Parameter	Results (95% confidence interval)										
	predicted	Accuracy	85.3	(78.1 - 90.3)									
	<table border="1"> <tr> <td></td> <td>benign</td> <td>cancer</td> </tr> <tr> <td>benign</td> <td>59</td> <td>10</td> </tr> <tr> <td>cancer</td> <td>9</td> <td>51</td> </tr> </table>		benign	cancer	benign	59	10	cancer	9	51	F1	86.1	
	benign	cancer											
benign	59	10											
cancer	9	51											
	benign	Precision	86.8	(76.7 - 92.8)									
actual	cancer	Recall	85.5	(75.3 - 91.9)									

F1: weighted harmonic mean of precision and recall

Precision: the proportion of true positives among instances classified as positive

Recall: the proportion of true positives among all positive instances in the data)

Discussion

BCG expression, using the antibody M01-4E1-1G7 under the methodological conditions

described here, was mainly located in the cytoplasm and the extracellular matrix. The location of

the protein is in accordance with others. It has been reported by different authors that BGN protein expression in gastric cancer was mainly located in the cytoplasm of epithelial cells [28] [29].

In breast cancer tissues, the expression of BGN was significantly lower when compared to normal breast tissue. These results are in accordance with those published by Bischof *et al.* who reported that adding BGN to an *in vitro* model has the ability to reverse neoplastic progression and ‘reboot’ breast cancer cells [13]. Nevertheless, our results are different from those published by others [10,12]. Possible explanations for these discrepancies can be related to a) number of cases, b) type of model used, c) type of antibody, and d) its biological effect. Apparently, in the human protein atlas, only two cases were accessed [16], while our sample had nine cases. The standard deviation of the BGN protein expression in normal tissue was high: two cases had low levels of expression (9.37 and 13.8 DAB units), while one case reached 80 DAB units (Figure 1). The use of different species may also explain these differences. Cong *et al.* noted that suppressing stromal BGN may yield a potent and superior anticancer effect in breast cancer induced in BGN knockout mice, compared to wild type [12]. While Cong *et al.* used a knockout mice model, we used human breast cancer tissue. In addition, data obtained from Zhao *et al.* was based on an Oncomine database using mRNA [10]. Another explanation for the lower levels of BGN protein expression in breast cancer tissue may be related to the full-length monoclonal antibody that was used here. Identifying various segments of the BGN protein may yield different outcomes due to the presence of chondroitin/dermatan sulfate + glycosaminoglycan side chains, which would then lead to the misinterpretation of results [14]. Finally, the amount of BGN in tissue sections does not necessarily reflect its biological effect as it mainly indicates BGN that has been sequestered in the extracellular matrix, for example, as part of the fibrotic scar [30]. Despite these disagreements, the difference in the protein expression in benign and malignant breast tissue reveals that the local

microenvironment, including the ECM, may have an important role in controlling cell growth, survival, and fate determination [20].

This study has some limitations. We neither analyzed subgroups, nor did any mechanistic experiment. In addition, we are unaware if there is a difference in the BGN expression across the menstrual cycle, which may explain the differences in BGN expression in normal breast tissue. Finally, we did not make a side-by-side comparison of different antibodies in order to identify putative differences.

The results here are strengthened by several aspects. The use of ImageJ software reduced the subjective bias of DAB quantification. The artificial intelligence was able to classify, with a high degree of accuracy, 129 photomicrographs based only on DAB expression, confirming the results obtained with ImageJ analysis. The use of both techniques together is not widely used yet, but it is promising. The criteria used for classification of the slides by the artificial intelligence was not completely understood; it is likely that the artificial intelligence used other factors beyond DAB expression. However, it is unlikely that the shape of the cells seen in DAB pictures had a major influence in the classification of the slides. The use of monoclonal antibody and non-specific primary antibody, as negative control, are evidence of the special care taken in the quality control of the immunohistochemical methodology. Immunohistochemical staining was performed using a non-specific primary antibody; this procedure has been considered a better methodology, compared to the omission of the primary antibody. Lung cancer was used as external positive and negative controls. External validity is expected once the same methodology is applied.

With our results, further studies may investigate the use of BGN as a biomarker or as a prognostic factor in breast cancer. The functional significance and the role of BGN alterations in breast tumorigenesis and progression remain to be determined.

Acknowledgement

This work was supported by FIPE (Fundo de Incentivo à Pesquisa e a Eventos) grant 19-0337 at Hospital de Clínicas de Porto Alegre. Support was given to buy laboratory reagents, and by Conselho Nacional de Pesquisa (CNPq) Grant 302448/2018-7 (RFS).

Competing interests

No competing interests were disclosed.

Data availability

Photomicrograph data

Harvard Dataverse: Deep learning neural network image analysis of immunohistochemical protein expression reveals a significantly reduced expression of Biglycan in breast cancer.

<https://doi.org/10.7910/DVN/NJGKWV>.

References

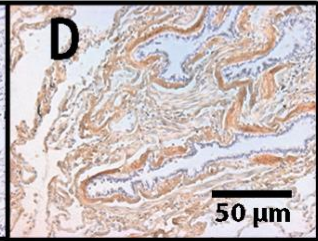
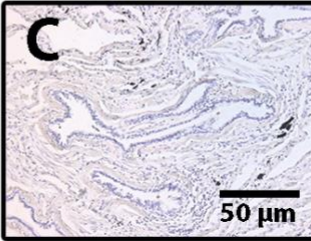
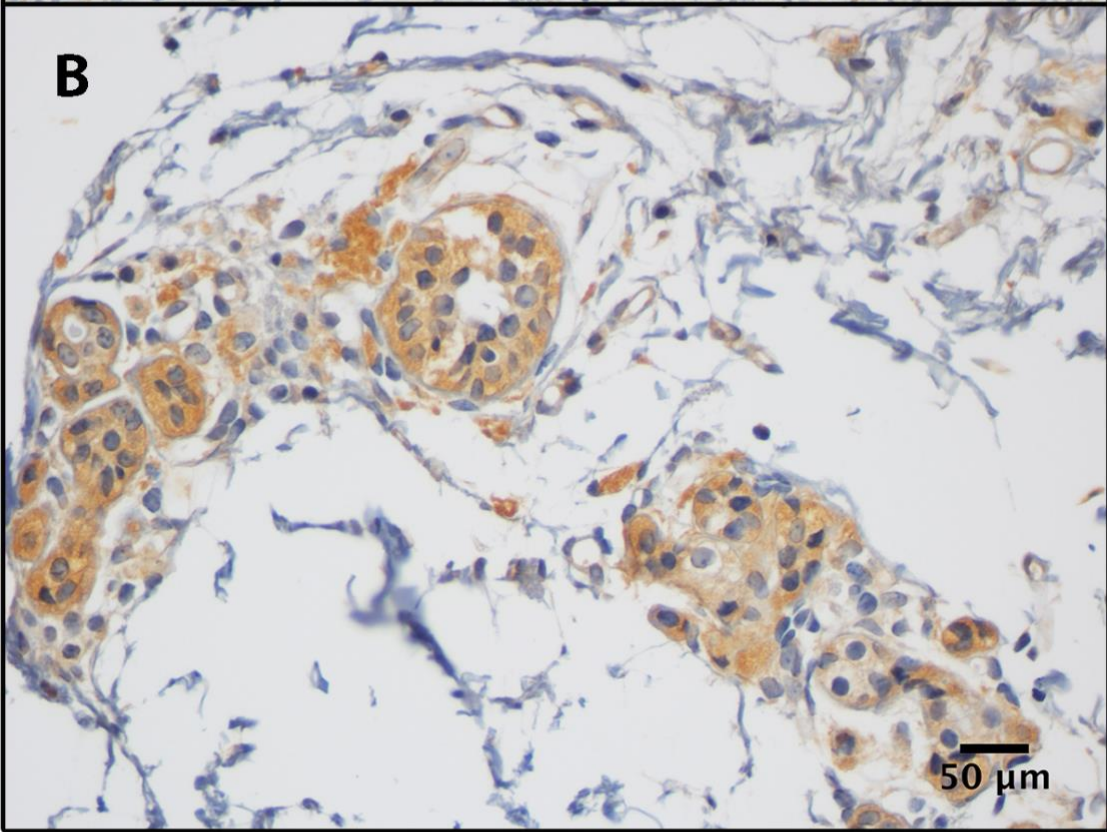
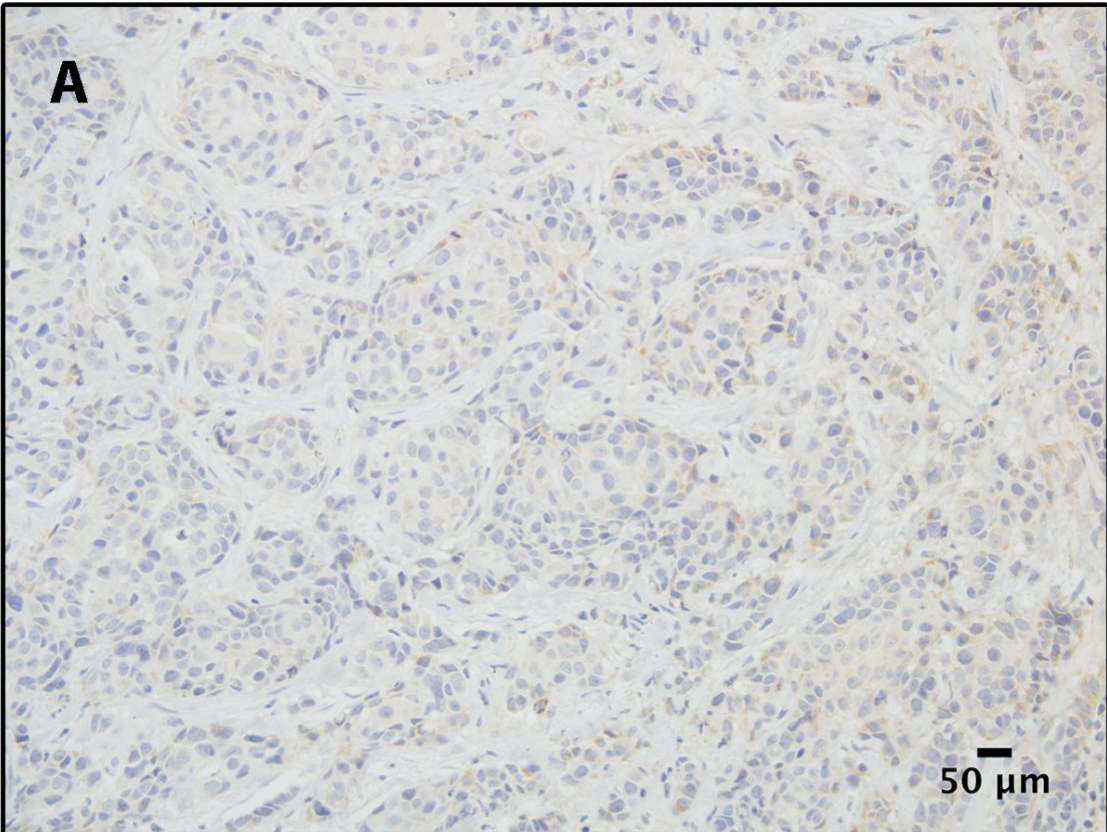
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71: 209–249.
2. Zubair M, Wang S, Ali N. Advanced Approaches to Breast Cancer Classification and Diagnosis. *Front Pharmacol.* 2020;11: 632079.
3. Ramirez-Valles EG, Rodríguez-Pulido A, Barraza-Salas M, Martínez-Velis I, Meneses-Morales I, Ayala-García VM, et al. A Quest for New Cancer Diagnosis, Prognosis and Prediction Biomarkers and Their Use in Biosensors Development. *Technol Cancer Res Treat.* 2020;19: 1533033820957033.
4. Appunni S, Anand V, Khandelwal M, Gupta N, Rubens M, Sharma A. Small Leucine Rich Proteoglycans (decorin, biglycan and lumican) in cancer. *Clin Chim Acta.* 2019;491: 1–7.
5. McBride OW, Fisher LW, Young MF. Localization of PGI (biglycan, BGN) and PGII (decorin, DCN, PG-40) genes on human chromosomes Xq13-qter and 12q, respectively. *Genomics.* 1990;6: 219–225.
6. Bianco P, Fisher LW, Young MF, Termine JD, Robey PG. Expression and localization of the two small proteoglycans biglycan and decorin in developing human skeletal and non-skeletal tissues. *J Histochem Cytochem.* 1990;38: 1549–1563.
7. Fisher LW, Termine JD, Young MF. Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species. *J Biol Chem.* 1989;264: 4571–4576.
8. Tufvesson E, Malmström J, Marko-Varga G, Westergren-Thorsson G. Biglycan isoforms with differences in polysaccharide substitution and core protein in human lung fibroblasts. *Eur J Biochem.* 2002;269: 3688–3696.

9. Nastase MV, Young MF, Schaefer L. Biglycan: a multivalent proteoglycan providing structure and signals. *J Histochem Cytochem*. 2012;60: 963–975.
10. Zhao S-F, Yin X-J, Zhao W-J, Liu L-C, Wang Z-P. Biglycan as a potential diagnostic and prognostic biomarker in multiple human cancers. *Oncol Lett*. 2020;19: 1673–1682.
11. Niedworok C, Röck K, Kretschmer I, Freudenberger T, Nagy N, Szarvas T, et al. Inhibitory role of the small leucine-rich proteoglycan biglycan in bladder cancer. *PLoS One*. 2013;8: e80084.
12. Cong L, Maishi N, Annan DA, Young MF, Morimoto H, Morimoto M, et al. Inhibition of stromal biglycan promotes normalization of the tumor microenvironment and enhances chemotherapeutic efficacy. *Breast Cancer Res*. 2021;23: 51.
13. Bischof AG, Yüksel D, Mammoto T, Mammoto A, Krause S, Ingber DE. Breast cancer normalization induced by embryonic mesenchyme is mediated by extracellular matrix biglycan. *Integr Biol* . 2013;5: 1045–1056.
14. Hara T, Yoshida E, Shinkai Y, Yamamoto C, Fujiwara Y, Kumagai Y, et al. Biglycan Intensifies ALK5-Smad2/3 Signaling by TGF- β and Downregulates Syndecan-4 in Cultured Vascular Endothelial Cells. *J Cell Biochem*. 2017;118: 1087–1096.
15. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347: 1260419.
16. Expression of BGN in cancer - Summary - The Human Protein Atlas. [cited 6 Jan 2022]. Available: <https://www.proteinatlas.org/ENSG00000182492-BGN/pathology>
17. Rüdiger T, Höfler H, Kreipe H-H, Nizze H, Pfeifer U, Stein H, et al. Quality assurance in immunohistochemistry: results of an interlaboratory trial involving 172 pathologists. *Am J Surg Pathol*. 2002;26: 873–882.
18. Fuhrich DG, Lessey BA, Savaris RF. Comparison of HSCORE assessment of endometrial beta3 integrin subunit expression with digital HSCORE using computerized image analysis (ImageJ). *Anal Quant Cytopathol Histopathol*. 2013;35: 210–216.
19. Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association. *J Pathol Inform*. 2019;10: 9.
20. Bruno RD, Smith GH. A potential mechanism for extracellular matrix induction of breast cancer cell normality. *Breast cancer research: BCR*. 2014. p. 302.
21. Rasmussen OF, Rudbeck L. Immunohistochemistry: A Dako Perspective. *Handbook of Practical Immunohistochemistry*. 2015. pp. 57–67.
22. Lopes Q dos A, dos Anjos Lopes Q, de Paula Guedes neto E, Gross LA, Pedrini JL, Savaris RF. Heart and Neural Crest Derivatives Expressed Transcript 2 (HAND2) is Reduced in Women with Breast Cancer. A Case-Control Study. *Clinical Oncology and Research*. 2019. pp. 1–6. doi:10.31487/j.cor.2019.5.12
23. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med*. 2012;9: e1001216.

24. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol.* 2001;23: 291–299.
25. Grudzinski M, Fuhrich DG, Savaris RF. Expression of elafin in fallopian tubes of ectopic pregnancies is reduced. *Appl Immunohistochem Mol Morphol.* 2015;23: 349–354.
26. Julious SA. *Sample Sizes for Clinical Trials.* CRC Press; 2009.
27. Kingma DP, Ba JL. ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. In: Bengio Y LY, editor. 3rd International Conference on Learning Representations. 7-9 May 2015. pp. 1–15.
28. Hu L, Duan Y-T, Li J-F, Su L-P, Yan M, Zhu Z-G, et al. Biglycan enhances gastric cancer invasion by activating FAK signaling pathway. *Oncotarget.* 2014;5: 1885–1896.
29. Pinto F, Santos-Ferreira L, Pinto MT, Gomes C, Reis CA. The Extracellular Small Leucine-Rich Proteoglycan Biglycan Is a Key Player in Gastric Cancer Aggressiveness. *Cancers.* 2021;13. doi:10.3390/cancers13061330
30. Schaefer L. Small Leucine-Rich Proteoglycans in Kidney Disease. *Journal of the American Society of Nephrology.* 2011. pp. 1200–1207. doi:10.1681/asn.2010050570

Figure 1. Photomicrographs of BGN protein expression in (A) invasive ductal carcinoma, and in (B) normal breast tissue. Lung tissue was used as an external negative (C) and positive (D) control. Bars represent 50 μm .

Scatter dot plot (E) demonstrates a significant reduction of BGN protein expression in breast cancer tissue, compared to benign breast tissue ($p=0.0017$; Mann-Whitney test). Each dot represents a sample, bars represent median values.



E

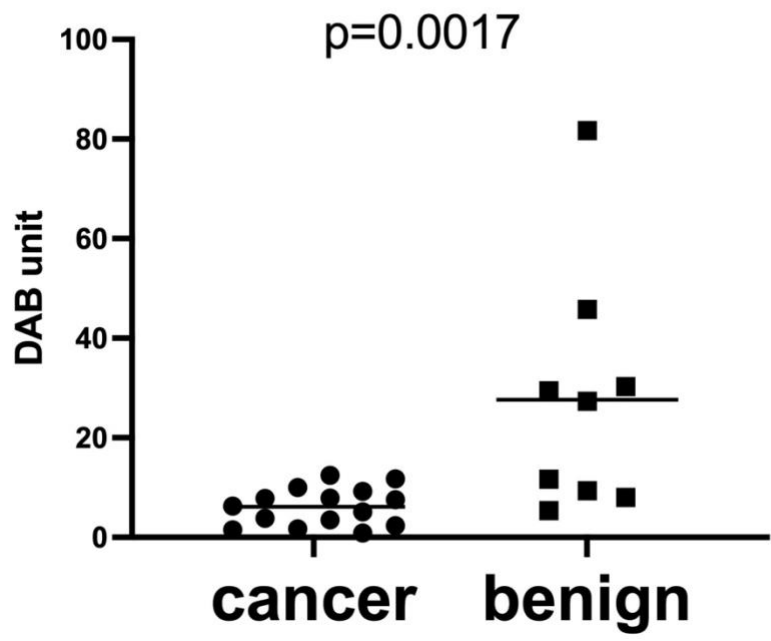
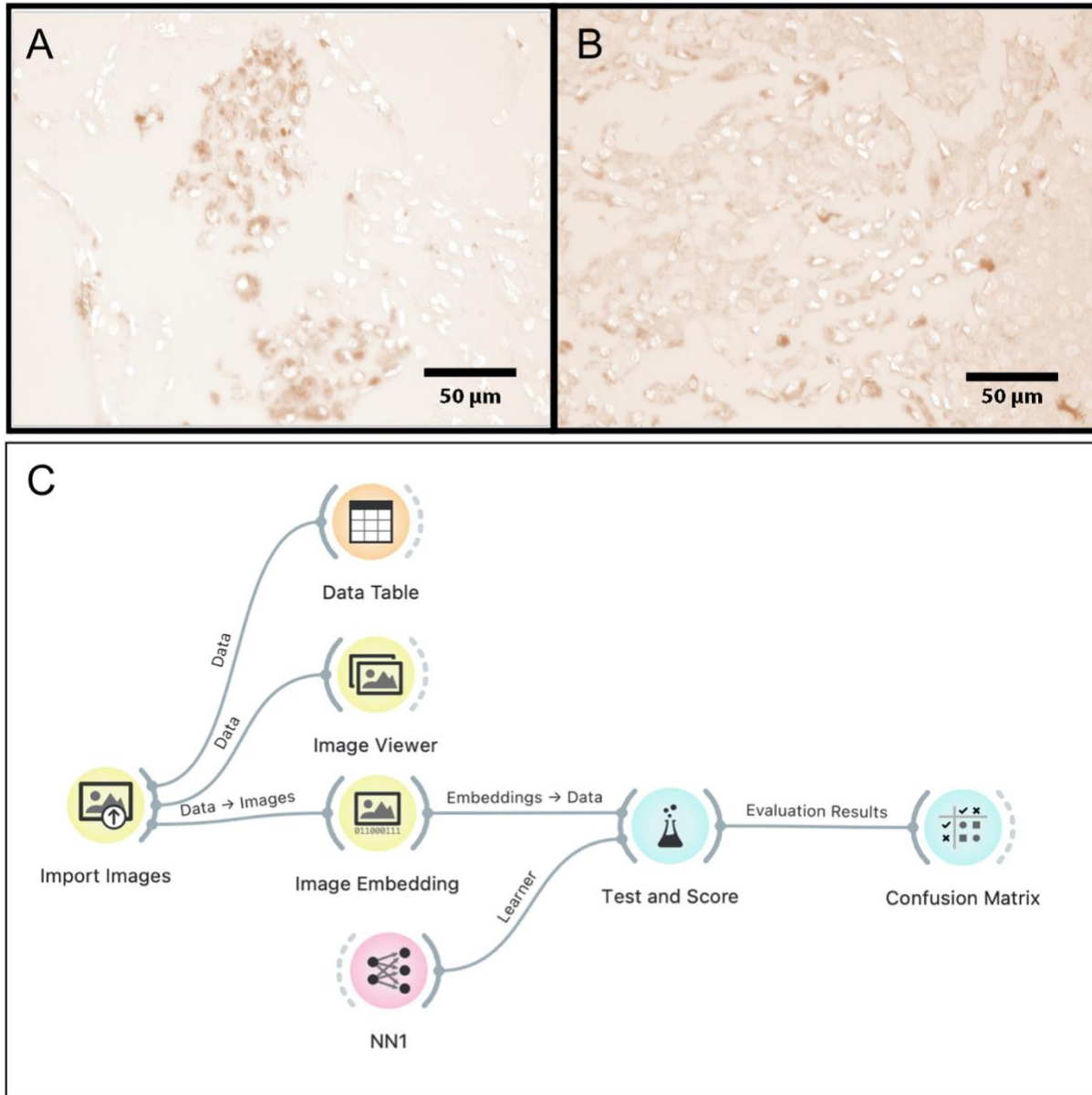


Figure 2. Example of BGN expression in DAB only in photomicrographs, after color deconvolution in normal breast tissue (A) and with cancer (B), bars represent 50 μm . Flowchart of the supervised neural network analysis using Orange 3 software (C). NN1:supervised neural network analysis.



9. CONCLUSÃO

A expressão proteica por imuno-histoquímica, utilizando o anticorpo monoclonal da Abnova (M01) e a análise pelo D-HScore e redes neurais convolucionais com aprendizado supervisionado, é reduzida no tecido com câncer de mama, em comparação com o tecido normal.